

Third Edition

MODERN DIFFERENTIAL GEOMETRY

of Curves and Surfaces with MATHEMATICA®

Alfred Gray, Elsa Abbena, and Simon Salamon

STUDIES IN ADVANCED MATHEMATICS

Chapman & Hall/CRC
Taylor & Francis Group

Third Edition

MODERN DIFFERENTIAL GEOMETRY

of Curves and Surfaces with MATHEMATICA®

Studies in Advanced Mathematics

Titles Included in the Series

John P. D'Angelo, Several Complex Variables and the Geometry of Real Hypersurfaces

Steven R. Bell, The Cauchy Transform, Potential Theory, and Conformal Mapping

John J. Benedetto, Harmonic Analysis and Applications

John J. Benedetto and Michael W. Frazier, Wavelets: Mathematics and Applications

Albert Boggess, CR Manifolds and the Tangential Cauchy–Riemann Complex

Keith Burns and Marian Gidea, Differential Geometry and Topology: With a View to Dynamical Systems

George Cain and Gunter H. Meyer, Separation of Variables for Partial Differential Equations: An Eigenfunction Approach

Goong Chen and Jianxin Zhou, Vibration and Damping in Distributed Systems
Vol. 1: Analysis, Estimation, Attenuation, and Design
Vol. 2: WKB and Wave Methods, Visualization, and Experimentation

Carl C. Cowen and Barbara D. MacCluer, Composition Operators on Spaces of Analytic Functions

Jewgeni H. Dshalalow, Real Analysis: An Introduction to the Theory of Real Functions and Integration

Dean G. Duffy, Advanced Engineering Mathematics with MATLAB®, 2nd Edition

Dean G. Duffy, Green's Functions with Applications

Lawrence C. Evans and Ronald F. Gariepy, Measure Theory and Fine Properties of Functions

Gerald B. Folland, A Course in Abstract Harmonic Analysis

José García-Cuerva, Eugenio Hernández, Fernando Soria, and José-Luis Torrea, Fourier Analysis and Partial Differential Equations

Peter B. Gilkey, Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem, 2nd Edition

Peter B. Gilkey, John V. Leahy, and Jeonghueong Park, Spectral Geometry, Riemannian Submersions, and the Gromov-Lawson Conjecture

Alfred Gray, Elsa Abbena, and Simon Salamon Modern Differential Geometry of Curves and Surfaces with Mathematica, Third Edition

Eugenio Hernández and Guido Weiss, A First Course on Wavelets

Kenneth B. Howell, Principles of Fourier Analysis

Steven G. Krantz, The Elements of Advanced Mathematics, Second Edition

Steven G. Krantz, Partial Differential Equations and Complex Analysis

Steven G. Krantz, Real Analysis and Foundations, Second Edition

Kenneth L. Kuttler, Modern Analysis

Michael Pedersen, Functional Analysis in Applied Mathematics and Engineering

Clark Robinson, Dynamical Systems: Stability, Symbolic Dynamics, and Chaos, 2nd Edition

John Ryan, Clifford Algebras in Analysis and Related Topics

John Scherk, Algebra: A Computational Introduction

Pavel Šolín, Karel Segeth, and Ivo Doležel, High-Order Finite Element Method

André Unterberger and Harald Upmeier, Pseudodifferential Analysis on Symmetric Cones

James S. Walker, Fast Fourier Transforms, 2nd Edition

James S. Walker, A Primer on Wavelets and Their Scientific Applications

Gilbert G. Walter and Xiaoping Shen, Wavelets and Other Orthogonal Systems, Second Edition

Nik Weaver, Mathematical Quantization

Kehe Zhu, An Introduction to Operator Algebras

Third Edition

MODERN DIFFERENTIAL GEOMETRY

of Curves and Surfaces with MATHEMATICA®

Alfred Gray, Elsa Abbena, and Simon Salamon

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Published in 2006 by
Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2006 by Taylor & Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-58488-448-7 (Hardcover)
International Standard Book Number-13: 978-1-58488-448-4 (Hardcover)
Library of Congress Card Number 2006003764

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Gray, Alfred, 1939-1998.
Modern differential geometry of curves and surfaces with Mathematica.-- 3rd ed. / Alfred Gray, Simon Salamon, Elsa Abbena.
p. cm. -- (Studies in advanced mathematics)
Includes bibliographical references and indexes.
ISBN-13: 978-0-58488-448-4 (acid-free paper)
ISBN-10: 1-58488-448-7 (acid-free paper)
1. Geometry, Differential--Data processing--Textbooks. 2. Curves on surfaces--Textbooks. 3. Mathematica (Computer file)--Textbooks. I. Salamon, Simon. II. Abbena, Elsa. III. Title. IV. Series.

QA641.G72 2006
516.3'6--dc22 2006003764

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface to the Second Edition[1]

Modern Differential Geometry of Curves and Surfaces is a traditional text, but it uses the symbolic manipulation program *Mathematica*. This important computer program, available on PCs, Macs, NeXTs, Suns, Silicon Graphics Workstations and many other computers, can be used very effectively for plotting and computing. The book presents standard material about curves and surfaces, together with accurate interesting pictures, *Mathematica* instructions for making the pictures and *Mathematica* programs for computing functions such as curvature and torsion.

Although Curves and Surfaces makes use of *Mathematica*, the book should also be useful for those with no access to *Mathematica*. All calculations mentioned in the book can in theory be done by hand, but some of the longer calculations might be just as tedious as they were for differential geometers in the 19th century. Furthermore, the pictures (most of which were done with *Mathematica*) elucidate concepts, whether or not *Mathematica* is used by the reader.

The main prerequisite for the book is a course in calculus, both single variable and multi-variable. In addition, some knowledge of linear algebra and a few basic concepts of point set topology are needed. These can easily be obtained from standard sources. *No computer knowledge is presumed.* In fact, the book provides a good introduction to *Mathematica*; the book is compatible with both versions 2.2 and 3.0. For those who want to use Curves and Surfaces to learn *Mathematica*, it is advisable to have access to Wolfram's book Mathematica for reference. (In version 3.0 of *Mathematica*, Wolfram's book is available through the help menus.)

Curves and Surfaces is designed for a traditional course in differential geometry. At an American university such a course would probably be taught at the junior-senior level. When I taught a one-year course based on Curves and Surfaces at the University of Maryland, some of my students had computer experience, others had not. All of them had acquired sufficient knowledge of *Mathematica* after one week. I chose not to have computers in my classroom because I needed the classroom time to explain concepts. I assigned all of the problems at the end of each chapter. The students used workstations, PCs

[1]This is a faithful reproduction apart from the updating of chapter references. It already incorporated the Preface to the First Edition dating from 1993.

and Macs to do those problems that required *Mathematica*. They either gave me a printed version of each assignment, or they sent the assignment to me by electronic mail.

Symbolic manipulation programs such as *Mathematica* are very useful tools for differential geometry. Computations that are very complicated to do by hand can frequently be performed with ease in *Mathematica*. However, they are no substitute for the theoretical aspects of differential geometry. So Curves and Surfaces presents theory and uses *Mathematica* programs in a complementary way.

Some of the aims of the book are the following.

- To show how to use *Mathematica* to plot many interesting curves and surfaces, more than in the standard texts. Using the techniques described in Curves and Surfaces, students can understand concepts geometrically by plotting curves and surfaces on a monitor and then printing them. The effect of changes in parameters can be strikingly portrayed.

- The presentation of pictures of curves and surfaces that are informative, interesting, and accurate. The book contains over 400 illustrations.

- The inclusion of as many topics of the classical differential geometry and surfaces as possible. In particular, the book contains many examples to illustrate important theorems.

- Alleviation of the drudgery of computing things such as the curvature and torsion of a curve in space. When the curvature and torsion become too complicated to compute, they can be graphed instead. There are more than 175 miniprograms for computing various geometric objects and plotting them.

- The introduction of techniques from numerical analysis into differential geometry. *Mathematica* programs for numerical computation and drawing of geodesics on an arbitrary surface are given. Curves can be found numerically when their torsion and curvature are specified.

- To place the material in perspective through informative historical notes. There are capsule biographies with portraits of over 75 mathematicians and scientists.

- To introduce interesting topics that, in spite of their simplicity, deserve to be better known. I mention triply orthogonal systems of surfaces (Chapter 19), Björling's formula for constructing a minimal surface containing a given plane curve as a geodesic (Chapter 22) and canal surfaces and cyclides of Dupin as Maxwell discussed them (Chapter 20).

- To develop a dialect of *Mathematica* for handling functions that facilitates the construction of new curves and surfaces from old. For example, there is a simple program to generate a surface of revolution from a plane curve.

- To provide explicit definitions of curves and surfaces. Over 300 *Mathematica* definitions of curves and surfaces can be used for further study.

The approach of Curves and Surfaces is admittedly more computational than is usual for a book on the subject. For example, Brioschi's formula for the Gaussian curvature in terms of the first fundamental form can be too complicated for use in hand calculations, but *Mathematica* handles it easily, either through computations or through graphing the curvature. Another part of *Mathematica* that can be used effectively in differential geometry is its special function library. For example, nonstandard spaces of constant curvature can be defined in terms of elliptic functions and then plotted.

Frequently, I have been asked if new mathematical results can be obtained by means of computers. Although the answer is generally no, it is certainly the case that computers can be an effective supplement to pure thought, because they allow experimentation and the graphs provide insights into complex relationships. I hope that many research mathematicians will find Curves and Surfaces useful for that purpose. Two results that I found with the aid of *Mathematica* are the interpretation of torsion in terms of tube twisting in Chapter 7 and the construction of a conjugate minimal surface without integration in Chapter 22. I have not seen these results in the literature, but they may not be new.

The programs in the book, as well as some descriptive *Mathematica* notebooks, will eventually be available on the web.

Sample Course Outlines

There is ample time to cover the whole book in three semesters at the undergraduate level or two semesters at the graduate level. Here are suggestions for courses of shorter length.

- One semester undergraduate differential geometry course: Chapters 1, 2, 7, 9–13, parts of 14–17, 27.

- Two semester undergraduate differential geometry course: Chapters 1–3, 9–19, 27.

- One semester graduate differential geometry course: Chapters 1, 2, 7–13, 15–19, parts of 22–27.

- One semester course on curves and their history: Chapters 1–8.

- One semester course on *Mathematica* and graphics: Chapters 1–6, 7–11, parts of 14–19, 23, and their notebooks.

I have tried to include more details than are usually found in mathematics books. This, plus the fact that the *Mathematica* programs can be used to elucidate theoretical concepts, makes the book easy to use for independent study.

Curves and Surfaces is an ongoing project. In the course of writing this book, I have become aware of the vast amount of material that was well known over a hundred years ago, but now is not as popular as it should be. So I plan to develop a web site, and to write a problem book to accompany the present text. Spanish, German, Japanese and Italian versions of Curves and Surfaces are already available.

Graphics

Although *Mathematica* graphics are very good and can be used to create *QuickTime* movies, the reader may also want to consider the following additional display methods:

- *Acrospin* is an inexpensive easy-to-use program that works on even the humblest PC.
- *Geomview* is a program for interactive display of curves and surfaces. It works on most unix-type systems and can be freely downloaded from http://www.geomview.org
- *Dynamic Visualizer* is an add-on program to *Mathematica* that allows interactive display. Details are available from http://www.wolfram.com
- *AVS* programs (see the commercial site http://www.avs.com) have been developed by David McNabb at the University of Maryland (http://www.umd.edu) for spectacular stereo three-dimensional images of the surfaces described in this book.

A Perspective

Mathematical trends come and go. R. Osserman (in his article "The Geometry Renaissance in America: 1938–1988" [Os2]) makes the point that in the 1950s when he was a student at Harvard, algebra dominated mathematics, the attention given to analysis was small, and the interest in differential geometry was converging to zero.

It was not always that way. In the last half of the 19th century surface theory was a very important area of mathematics, both in research and teaching. Brill, then Schilling, made an extensive number of plaster models available to the

mathematical public. Darboux's Leçons sur la Théorie Générale des Surfaces and Bianchi's Lezioni di Geometria Differenziale were studied intensely. I attribute the decline of differential geometry, especially in the United States, to the rise of tensor analysis. Instead of drawing pictures it became fashionable to raise and lower indices.

I strongly feel that pictures need to be much more stressed in differential geometry than is presently the case. It is unfortunate that the great differential geometers of the past did not share their extraordinary intuitions with others by means of pictures. I hope that the present book contributes in some way to returning the differential geometry of curves and surfaces to its proper place in the mathematics curriculum.

I wish to thank Elsa Abbena, James Anderson, Thomas Banchoff, Marcel Berger, Michel Berry, Nancy Blachman, William Bruce, Renzo Caddeo, Eugenio Calabi, Thomas Cecil, Luis A. Cordero, Al Currier, Luis C. de Andrés, Mirjana Djorić, Franco Fava, Helaman Fergason, Marisa Fernández, Frank Flaherty, Anatoly Fomenko, V.E. Fomin, David Fowler, George Francis, Ben Friedman, Thomas Friedrich, Pedro M. Gadea, Sergio Garbiero, Laura Geatti, Peter Giblin, Vladislav Goldberg, William M. Goldman, Hubert Gollek, Mary Gray, Joe Grohens, Garry Helzer, A.O. Ivanov, Raj Jakkumpudi, Gary Jensen, Alfredo Jiménez, David Johannsen, Joe Kaiping, Ben Kedem, Robert Kragler, Steve Krantz, Henning Leidecker, Stuart Levy, Mats Liljedahl, Sanchez Santiago Lopez de Medrano, Lee Lorch, Armando Machado, Roman Maeder, Mikhail A. Malakhaltsev, Steen Markvorsen, David McNabb, José J. Mencía, Michael Mezzino, Vicente Miquel Molina, Deanne Montgomery, Tamara Munzner, Emilio Musso, John Novak, Barrett O'Neill, Richard Palais, Mark Phillips, Lori Pickert, David Pierce, Mark Pinsky, Paola Piu, Valeri Pouchnia, Rob Pratt, Emma Previato, Andreas Iglesias Prieto, Lilia del Riego, Patrick Ryan, Giacomo Saban, George Sadler, Simon Salamon, Isabel Salavessa, Jason P. Schultz, Walter Seaman, B.N. Shapukov, V.V. Shurygin, E.P. Shustova, Sonya Šimek, Cameron Smith, Dirk Struik, Rolf Sulanke, John Sullivan, Daniel Tanré, C. Terng, A.A. Tuzhilin, Lieven Vanhecke, Gus Vlahacos, Tom Wickam-Jones and Stephen Wolfram for valuable suggestions.

Alfred Gray
July 1998


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface to the Third Edition

Most of the material of this book can be found, in one form or another, in the Second Edition. The exceptions to this can be divided into three categories.

First, a number of modifications and new items had already been prepared by Alfred Gray following publication of the Second Edition, and we have been able to incorporate some of these in the Third Edition. The most obvious is Chapter 21. In addition, we have liberally expanded a number of sections by means of additional text or graphics, where we felt that this was warranted.

The second is Chapter 23, added by the editors to present the popular theory of quaternions. This brings together many of the techniques in the rest of the book, combining as it does the theory of space curves and surfaces.

The third concerns the *Mathematica* code presented in the notebooks. While this is closely based on that written by the author and displayed in previous editions, many programs have been enhanced and sibling ones added. This is to take account of the progressive presentation that *Mathematica* notebooks offer, and a desire to publish instructions to generate the book's figures, *all* of which were produced with *Mathematica* in the Third Edition.

The new edition does differ notably from the previous one in the *manner* in which the material is organized. All *Mathematica* code has been separated from the body of the text and organized into notebooks, to give readers interactive access to the material. There is one notebook to accompany each chapter, and it contains relevant programs in parallel with the text, section by section. An abridged version is printed at the end of the chapter, for close reference and to present a fair idea of the programs that ride in tandem with the mathematics. The distillation of computer code into notebooks makes it easier to conceive of writing it in a different language, and a project is underway to do this in *Maple*.

The full notebooks can be downloaded from the publisher's site

http://www.crcpress.com

Their organization and layout is discussed in more detail in Notebook 0 below. They contain no output, as this can be generated at will. All the figures in the book were compiled automatically by merely evaluating the notebooks chapter by chapter, and this served to "validate" the notebooks using Version 5.2 of *Mathematica*. It is the editors' intention to build up an online database of solutions to the exercises at the end of each chapter. Those marked M are designed to be solved with the help of a suitable *Mathematica* program.

The division of the material into chapters and the arrangement of later chapters has also been affected by the presence of the notebooks. We have chosen to shift the exposition of differentiable manifolds and abstract surfaces towards the end of the book. In addition, there are a few topics in the Second Edition that have been relegated to electronic form in an attempt to streamline the volume. This applies to the fundamental theorems of surfaces, and some more advanced material on minimal surfaces. (It is also a recognition of other valuable sources, such as [Oprea2] to mention one.) The diagram after Notebook 0 provides an idea of the relationship between the various topics; touching blocks represent a group of chapters that are probably best read in numerical order, and 21, 22, 27 are independent peaks to climb.

In producing the Third Edition, the editors were fortunate in having ready access to electronic versions of the author's files. For this, they are grateful to Mike Mezzino, as well as the editors of the Spanish and Italian versions, Marisa Fernández and Renzo Caddeo. Above all, they are grateful to Mary Gray and Bob Stern for entrusting them with the editing task.

As regards the detailed text, the editors acknowledge the work of Daniel Drucker, who diligently scanned the Second Edition for errors, and provided helpful comments on much of the Third Edition. Some of the material was used for the course "Geometrical Methods for Graphics" at the University of Turin in 2004 and 2005, and we thank students of that course for improving some of the figures and associated computer programs. We are also grateful to Simon Chiossi, Sergio Console, Antonio Di Scala, Anna Fino, and Sergio Garbiero for proofreading parts of the book, to John Sullivan and others for providing photo images, and to Steven Wilkinson for making his program ImplicitPlot3D freely available.

The years since publication of the First Edition have seen a proliferation of useful websites containing information on curves and surfaces that complements the material of this book. In particular, the editors acknowledge useful visits to the Geometry Center's site http://www.geom.uiuc.edu, and Richard Palais's pages http://vmm.math.uci.edu/3D-XplorMath. Finally, we are grateful to Wolfram Research for their ongoing support, and to Paul Wellin in particular.

The editors initially worked with Alfred Gray at the University of Maryland in 1979, and on various occasions subsequently. Although his more abstract research had an enormous influence on many branches of differential geometry (a hint of which can be found in the volume [FeWo]), we later witnessed the pleasure he experienced in preparing material for both editions of this book. We hope that our more modest effort for the Third Edition will help extend this pleasure to others.

Elsa Abbena and Simon Salamon
December 2005

Notebook 0

■ N0.0

The next three pages serve as an introduction to the notebooks, which form an integral part of the Third Edition. An overview is followed, in N0.1, by an explanation of cell styles and evaluation techniques.

It was the author's view, and one shared by the editors, that *Mathematica* is an excellent tool for investigating elementary differential geometry as represented by the many topics in this book. However, readers can choose to what extent they rely on the notebooks for an understanding of the mathematics. Whilst the notebooks were primarily envisaged as playing the role of "chapter notes" to back up the text and its formulas, one could conceivably *begin* with the notebooks and use the text as reference. In any case, each notebook is inexorably linked to the chapter it follows.

The online version of each notebook contains no output and (at the time the book went to press) averages 60KB in length. Readers in possession of *Mathematica* can use the notebook files to generate all the book's 316 figures in color, as well as other images and output designed to support many proofs.

The notebooks printed at the end of each chapter are abridged versions, and subsections with gray numerical headings (the first being N0.1.1) indicate the presence of further material in the *Mathematica* file, such as less enlightening graphics code. The advantage of this approach is that the electronic notebooks can be kept up to date, and more material relevant to the book's chapters added in time, for example in support of the exercises.

Each notebook is self-contained, and requires no specialized definitions to be loaded other than one or more standard packages. This enables us to focus on the requirements of each particular chapter, and customize settings for the tasks at hand. An alternative approach would be to load text files containing all the definitions (including those on pages 977-984) and such files will be available on the book's website for the many readers who will prefer to construct their own notebooks.

The notebook sections run (apart from initial ones labelled Nn.0) in parallel to those of the text, though not every chapter section has a notebook counterpart. Early notebooks contain more basic explanation and less sophisticated techniques, and their development reflects the editors' own learning process that we hope others can follow.

■ N0.1

Each *Mathematica* notebook incorporates an imported style file, similar to that of ArticleModern. There are two versions of this file available, one to display background color for cells in the electronic notebooks, and a black and white style for the book.

In addition to title and subheadings, seven different types of cells have been designated, including text (this one). The remaining six are assigned distinct colors, though only three shades of gray are distinguished in print. Here is a list of the cell types:

Input

Definition

Package

Miniprogram

Animation

NumberedFigure

Figure 0.1

Working backwards, NumberedFigure contains the code of the corresponding figure shown in the chapter text, and to some extent the entire notebook presentation is designed around the regular appearance of such cells.

Animation is used for "movies" in which multiple images are presented in quick succession, though to run the movie it is usually necessary to replace the command Show by the single character S, as explained below. The designation Miniprogram is used for an important definition incorporating the command Module; it may be code to compute curvature or something much simpler. A Package cell is used to load a *Mathematica* package or set of options for plotting. All remaining definitions worth

recording are set in the style Definition, while Input is used for "local" instructions that only have significance to the current section.

Cells designated Definition, Package, Miniprogram are *initialization cells,* and their contents should be automatically evaluated when the user is invited to do so at the opening stage of each notebook. This makes all the definitions of that notebook (many of which may be hidden in a section labelled Nn.0.1) instantly available wherever they appear. With this proviso, the notebooks are designed to evaluate correctly from the top downwards, either manually or with the command Evaluate Notebook.

The figures for each chapter were first saved as EPS files into a directory for that chapter, so that they could be picked up automatically when the LaTeX file was compiled. This imposed two constraints. First, the order of the NumberedFigure cells coincides with the order of the figures in the chapter. Second, no other graphics is produced when the notebook is evaluated automatically (this also helps to speed up the evaluation process). In practice, to keep early notebooks simple, this second feature is satisfied only from Notebook 6 onwards.

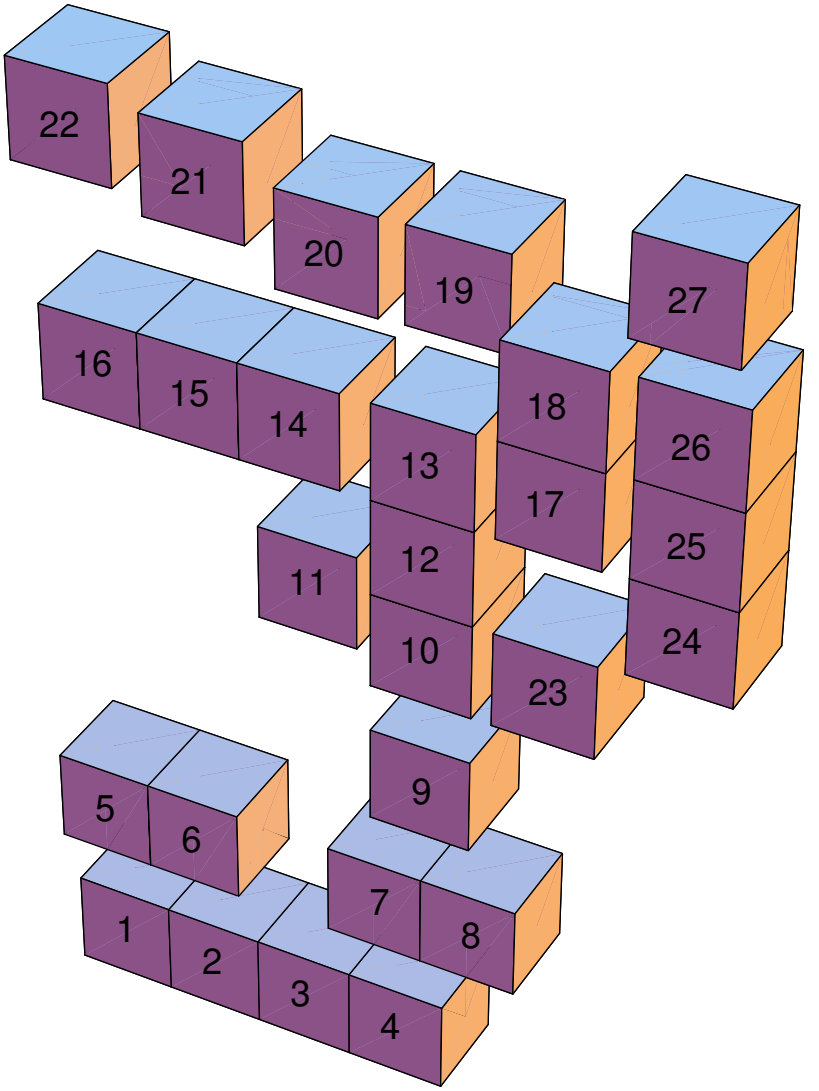
To prevent much of the graphics from being output from Notebook 6 onwards, we have introduced a variant of the Show command. The user is required to insert the command S to display a plot that is not a numbered figure, or change Show to S to activate a more complicated graphics program. Another reason for adopting this method is to suppress all the intermediate plots needed to produce a composite figure, without having to type DisplayFunction->Identity each time. More details are given in N6.0, N9.0, and N10.0. It is however likely that it will be easier to switch graphics displays on and off in future versions of *Mathematica.*

The course of production saw the successive use of *Mathematica* 4.2, 5.0, 5.1, 5.2, and the latter was used to test the notebooks on various operating systems. As new versions of *Mathematica* are released, so corresponding versions of the book's notebooks will be made readily available on the book's website.

Neither the author nor the editors were trained as computer programmers, and our approach is at times naïve. This is evident from an inspection of the varied programs in the Notebook Index on page 977. No use has been made of *Mathematica contexts*, though we have tried to increase the *modularity* of the programs found in the appendices of the Second Edition by using more abbreviations and shortcuts. Those with a certain level of expertise will prefer to achieve the same results in different ways.

The final subsection N0.1.1 was used to produce the diagram overleaf, and can be opened in the electronic version of this notebook.

■ N0.1.1


22
21
20
19
27
16
15
14
18
26
13
17
25
11
12
24
10
23
9
5
6
7
8
1
2
3
4

Chapter Scheme

1 Curves in the Plane 1
2 Famous Plane Curves 39
3 Alternative Ways of Plotting Curves 73
4 New Curves from Old 99
5 Determining a Plane Curve from Its Curvature 127
6 Global Properties of Plane Curves 153
7 Curves in Space 191
8 Construction of Space Curves 229
9 Calculus on Euclidean Space 263
10 Surfaces in Euclidean Space 287
11 Nonorientable Surfaces 331
12 Metrics on Surfaces 361
13 Shape and Curvature 385
14 Ruled Surfaces 431
15 Surfaces of Revolution and Constant Curvature 461
16 A Selection of Minimal Surfaces 501
17 Intrinsic Surface Geometry 531
18 Asymptotic Curves and Geodesics on Surfaces 557
19 Principal Curves and Umbilic Points 593
20 Canal Surfaces and Cyclides of Dupin 639
21 The Theory of Surfaces of Constant Negative Curvature 683
22 Minimal Surfaces via Complex Variables 719
23 Rotation and Animation Using Quaternions 767
24 Differentiable Manifolds 809
25 Riemannian Manifolds 847
26 Abstract Surfaces and Their Geodesics 871
27 The Gauss–Bonnet Theorem 901


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Full Contents

1. Curves in the Plane 1

1.1 Euclidean Space 2
1.2 Curves in Space 5
1.3 The Length of a Curve 9
1.4 Curvature of Plane Curves 14
1.5 Angle Functions 17
1.6 First Examples of Plane Curves 20
1.7 The Semicubical Parabola and Regularity 25
1.8 Exercises 27

Notebook 1 29

2. Famous Plane Curves 39

2.1 Cycloids 40
2.2 Lemniscates of Bernoulli 43
2.3 Cardioids 45
2.4 The Catenary 45
2.5 The Cissoid of Diocles 47
2.6 The Tractrix 50
2.7 Clothoids 53
2.8 Pursuit Curves 54
2.9 Exercises 57

Notebook 2 60

3. Alternative Ways of Plotting Curves 73

3.1 Implicitly Defined Plane Curves 74
3.2 The Folium of Descartes 76
3.3 Cassinian Ovals 79
3.4 Plane Curves in Polar Coordinates 81
3.5 A Selection of Spirals 82
3.6 Exercises 85

Notebook 3 88

4. New Curves from Old 99

4.1 Evolutes 99
4.2 Iterated Evolutes 102
4.3 Involutes 104
4.4 Osculating Circles to Plane Curves 107
4.5 Parallel Curves 110
4.6 Pedal Curves 113
4.7 Exercises 115

Notebook 4 118

5. Determining a Plane Curve from Its Curvature 127

5.1 Euclidean Motions 128
5.2 Isometries of the Plane 133
5.3 Intrinsic Equations for Plane Curves 136
5.4 Examples of Curves with Assigned Curvature 140
5.5 Exercises 142

Notebook 5 146

6. Global Properties of Plane Curves 153

6.1 Total Signed Curvature 154
6.2 Trochoid Curves 157
6.3 The Rotation Index of a Closed Curve 160
6.4 Convex Plane Curves 164
6.5 The Four Vertex Theorem 166
6.6 Curves of Constant Width 169
6.7 Reuleaux Polygons and Involutes 172
6.8 The Support Function of an Oval 174
6.9 Exercises 178

Notebook 6 181

7. Curves in Space 191

7.1 The Vector Cross Product 192
7.2 Curvature and Torsion of Unit-Speed Curves 195
7.3 The Helix and Twisted Cubic 200
7.4 Arbitrary-Speed Curves in $\mathbb{R}^3$ 203
7.5 More Constructions of Space Curves 206
7.6 Tubes and Tori 209
7.7 Torus Knots 211
7.8 Exercises 213

Notebook 7 217

8. Construction of Space Curves 229

8.1 The Fundamental Theorem of Space Curves 230
8.2 Assigned Curvature and Torsion 233
8.3 Contact 236
8.4 Space Curves That Lie on a Sphere 242
8.5 Curves of Constant Slope 246
8.6 Loxodromes on Spheres 250
8.7 Exercises 252

Notebook 8 254

9. Calculus on Euclidean Space 263

9.1 Tangent Vectors to $\mathbb{R}^n$ 263
9.2 Tangent Vectors as Directional Derivatives 265
9.3 Tangent Maps or Differentials 268
9.4 Vector Fields on $\mathbb{R}^n$ 272
9.5 Derivatives of Vector Fields 275
9.6 Curves Revisited 280
9.7 Exercises 281

Notebook 9 283

10. Surfaces in Euclidean Space 287

10.1 Patches in $\mathbb{R}^n$ 288
10.2 Patches in $\mathbb{R}^3$ and the Local Gauss Map 295
10.3 The Definition of a Regular Surface 297
10.4 Examples of Surfaces 302
10.5 Tangent Vectors and Surface Mappings 307
10.6 Level Surfaces in $\mathbb{R}^3$ 311
10.7 Exercises 316

Notebook 10 320

11. Nonorientable Surfaces 331

11.1 Orientability of Surfaces 331
11.2 Surfaces by Identification 336
11.3 The Möbius Strip 339
11.4 The Klein Bottle 341
11.5 Realizations of the Real Projective Plane 343
11.6 Twisted Surfaces 348
11.7 Exercises 350

Notebook 11 352

12. Metrics on Surfaces 361

12.1 The Intuitive Idea of Distance 361
12.2 Isometries between Surfaces 365
12.3 Distance and Conformal Maps 369
12.4 The Intuitive Idea of Area 372
12.5 Examples of Metrics 374
12.6 Exercises 377

Notebook 12 379

13. Shape and Curvature 385

13.1 The Shape Operator 386
13.2 Normal Curvature 389
13.3 Calculation of the Shape Operator 393
13.4 Gaussian and Mean Curvature 397
13.5 More Curvature Calculations 403
13.6 A Global Curvature Theorem 410
13.7 Nonparametrically Defined Surfaces 411
13.8 Exercises 415

Notebook 13 420

14. Ruled Surfaces 431

14.1 Definitions and Examples 432
14.2 Curvature of a Ruled Surface 437
14.3 Tangent Developables 440
14.4 Noncylindrical Ruled Surfaces 444
14.5 Exercises 449

Notebook 14 452

15. Surfaces of Revolution and Constant Curvature ... 461

15.1 Surfaces of Revolution 462
15.2 Principal Curves 465
15.3 Curvature of a Surface of Revolution 468
15.4 Generalized Helicoids 470
15.5 Surfaces of Constant Positive Curvature 473
15.6 Surfaces of Constant Negative Curvature 477
15.7 More Examples of Constant Curvature 480
15.8 Exercises 485

Notebook 15 488

16. A Selection of Minimal Surfaces 501

16.1 Normal Variation 502
16.2 Deformation from the Helicoid to the Catenoid 504
16.3 Minimal Surfaces of Revolution 507
16.4 More Examples of Minimal Surfaces 509
16.5 Monge Patches and Scherk's Minimal Surface 513
16.6 The Gauss Map of a Minimal Surface 516
16.7 Isothermal Coordinates 518
16.8 Exercises 521

Notebook 16 523

17. Intrinsic Surface Geometry 531

17.1 Intrinsic Formulas for the Gaussian Curvature 531
17.2 Gauss's Theorema Egregium 535
17.3 Christoffel Symbols 537
17.4 Geodesic Curvature of Curves on Surfaces 540
17.5 Geodesic Torsion and Frenet Formulas 545
17.6 Exercises 547

Notebook 17 548

18. Asymptotic Curves and Geodesics on Surfaces ... 557

18.1 Asymptotic Curves 558
18.2 Examples of Asymptotic Curves and Patches 562
18.3 The Geodesic Equations 565
18.4 First Examples of Geodesics 569
18.5 Clairaut Patches 572
18.6 Use of Clairaut Patches 576
18.7 Exercises 579

Notebook 18 582

19. Principal Curves and Umbilic Points 593

19.1 The Differential Equation for Principal Curves 594
19.2 Umbilic Points 597
19.3 The Peterson–Mainardi–Codazzi Equations 599
19.4 Hilbert's Lemma and Liebmann's Theorem 602
19.5 Triply Orthogonal Systems of Surfaces 604
19.6 Elliptic Coordinates 609
19.7 Parabolic Coordinates and a General Construction 615
19.8 Parallel Surfaces 620
19.9 The Shape Operator of a Parallel Surface 622

19.10 Exercises 625

Notebook 19 628

20. Canal Surfaces and Cyclides of Dupin 639

20.1 Surfaces Whose Focal Sets Are 2-Dimensional 641
20.2 Canal Surfaces 646
20.3 Cyclides of Dupin via Focal Sets 655
20.4 The Definition of Inversion 661
20.5 Inversion of Surfaces 665
20.6 Exercises 669

Notebook 20 673

21. The Theory of Surfaces of Constant Negative Curvature 683

21.1 Intrinsic Tchebyshef Patches 684
21.2 Patches on Surfaces of Constant Negative Curvature 687
21.3 The Sine–Gordon Equation 691
21.4 Tchebyshef Patches on Surfaces of Revolution 692
21.5 The Bianchi Transform 697
21.6 Moving Frames on Surfaces in $\mathbb{R}^3$ 701
21.7 Kuen's Surface as Bianchi Transform of the Pseudosphere . 704
21.8 The Bäcklund Transform 706
21.9 Exercises 710

Notebook 21 712

22. Minimal Surfaces via Complex Variables 719

22.1 Isometric Deformations of Minimal Surfaces 720
22.2 Complex Derivatives 725
22.3 Minimal Curves 728
22.4 Finding Conjugate Minimal Surfaces 732
22.5 The Weierstrass Representation 736
22.6 Minimal Surfaces via Björling's Formula 741
22.7 Costa's Minimal Surface 746
22.8 Exercises 752

Notebook 22 755

23. Rotation and Animation Using Quaternions 767

23.1 Orthogonal Matrices 769
23.2 Quaternion Algebra 775
23.3 Unit Quaternions and Rotations 779
23.4 Imaginary Quaternions and Rotations 783
23.5 Rotation Curves 785
23.6 Euler Angles 789
23.7 Further Topics 793
23.8 Exercises 797

Notebook 23 799

24. Differentiable Manifolds 809

24.1 The Definition of a Differentiable Manifold 810
24.2 Differentiable Functions on Manifolds 814
24.3 Tangent Vectors on Manifolds 819
24.4 Induced Maps 826
24.5 Vector Fields on Manifolds 831
24.6 Tensor Fields 836
24.7 Exercises 839

Notebook 24 842

25. Riemannian Manifolds 847

25.1 Covariant Derivatives 848
25.2 Pseudo-Riemannian Metrics 854
25.3 The Classical Treatment of Metrics 858
25.4 The Christoffel Symbols in Riemannian Geometry 861
25.5 The Riemann Curvature Tensor 863
25.6 Exercises 867

Notebook 25 868

26. Abstract Surfaces and Their Geodesics 871

26.1 Christoffel Symbols on Abstract Surfaces 872
26.2 Examples of Abstract Metrics 874
26.3 The Abstract Definition of Geodesic Curvature 878
26.4 Geodesics on Abstract Surfaces 880
26.5 The Exponential Map and the Gauss Lemma 884
26.6 Length Minimizing Properties of Geodesics 888
26.7 Exercises 892

Notebook 26 895

27. The Gauss–Bonnet Theorem 901
27.1 Turning Angles and Liouville's Theorem 902
27.2 The Local Gauss–Bonnet Theorem 906
27.3 An Area Bound 910
27.4 A Generalization to More Complicated Regions 911
27.5 The Topology of Surfaces 914
27.6 The Global Gauss–Bonnet Theorem 918
27.7 Applications of the Gauss–Bonnet Theorem 920
27.8 Exercises 921
Notebook 27 923

Bibliography 931

Name Index 953

Subject Index 957

Notebook Index 977

Chapter 1

Curves in the Plane

Geometry before calculus involves only the simplest of curves: straight lines, broken lines, arcs of circles, ellipses, hyperbolas and parabolas. These are the curves that form the basis of Greek geometry. Although other curves (such as the cissoid of Diocles and the spiral of Archimedes) were known in antiquity, the general theory of plane curves began to be developed only after the invention of Cartesian coordinates by Descartes[1] in the early 1600s.

Interesting plane curves arise in everyday life. Examples include the trajectory of a projectile (a parabola), the form of a suspension bridge (a catenary), the path of a point on the wheel of a car (a cycloid), and the orbit of a planet (an ellipse). In this chapter we get underway with the study of the local properties of curves that has developed over the last few centuries.

We begin in Section 1.1 by recalling some standard operations on Euclidean space $\mathbb{R}^n$, and in Section 1.2 define the notions of curves in $\mathbb{R}^n$ and vector fields along them. Arc length for curves in $\mathbb{R}^n$ is defined and discussed in Section 1.3. In Section 1.4, we define the important notion of *signed curvature* of a curve in the plane $\mathbb{R}^2$. Section 1.5 is devoted to the problem of defining an angle function between plane curves, and this allows us to show that the signed curvature is the derivative of the turning angle. The examples discussed in Section 1.6 include the logarithmic spiral, while the arc length of the semicubical parabola is computed in Section 1.7.

Our basic approach in Chapters 1 and 2 is to study curves by means of their parametric representations, whereas implicitly defined curves in $\mathbb{R}^2$ are discussed in Section 3.1. In this chapter, we must therefore show that any



[1] René du Perron Descartes (1596–1650). French mathematician and philosopher. Descartes developed algebraic techniques to solve geometric problems, thus establishing the foundations of analytic geometry. Although most widely known as a philosopher, he also made important contributions to physiology, optics and mechanics.

geometric invariant depending only on the point set traced out by the curve is independent of the parametrization, at least up to sign. The most important geometric quantities associated with a curve are of two types: (1) those that are totally independent of parametrization, and (2) those that do not change under a positive reparametrization, but change sign under a negative reparametrization. For example, we show in Section 1.3 that the length of a curve is independent of the parametrization chosen. The curvature of a plane curve is a more subtle invariant, because it does change sign if the curve is traversed in the opposite direction, but is otherwise independent of the parametrization. This fact we prove in Section 1.4.

In Section 1.6 we give some examples of plane curves that generalize circles, discuss their properties and graph associated curvature functions.

1.1 Euclidean Space

Since we shall be studying curves and surfaces in a Euclidean space, we summarize some of the algebraic properties of Euclidean space in this section.

Definition 1.1. ***Euclidean*** *n-**space*** $\mathbb{R}^n$ *consists of the set of all real n-tuples*

$$\mathbb{R}^n = \big\{\, (p_1, \ldots, p_n) \mid p_j \text{ is a real number for } j = 1, \ldots, n \,\big\}.$$

*We write $\mathbb{R}$ for $\mathbb{R}^1$; it is simply the set of all real numbers. $\mathbb{R}^2$ is frequently called the **plane**.*

Elements of $\mathbb{R}^n$ represent both points in n-dimensional space, and the *position vectors* of points. As a consequence, $\mathbb{R}^n$ is a vector space, so that the operations of *addition* and *scalar multiplication* are defined. Thus if

$$\mathbf{p} = (p_1, \ldots, p_n) \quad \text{and} \quad \mathbf{q} = (q_1, \ldots, q_n),$$

then $\mathbf{p} + \mathbf{q}$ is the element of $\mathbb{R}^n$ given by

$$\mathbf{p} + \mathbf{q} = (p_1 + q_1, \ldots, p_n + q_n).$$

Similarly, for $\lambda \in \mathbb{R}$ the vector $\lambda\mathbf{p}$ is defined by

$$\lambda\mathbf{p} = (\lambda p_1, \ldots, \lambda p_n).$$

Furthermore, we shall denote by $\cdot$ the ***dot product*** (or scalar product) of $\mathbb{R}^n$. It is an operation that assigns to each pair of vectors $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ the real number

$$\mathbf{p} \cdot \mathbf{q} = \sum_{j=1}^{n} p_j q_j.$$

The ***norm*** and ***distance*** functions of $\mathbb{R}^n$ can then be defined by

$$\|\mathbf{p}\| = \sqrt{\mathbf{p} \cdot \mathbf{p}}, \qquad \text{and} \qquad \mathsf{distance}(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|$$

for $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$.

These functions have the properties

$$\begin{aligned} \mathbf{p} \cdot \mathbf{q} &= \mathbf{q} \cdot \mathbf{p}, & (\mathbf{p} + \mathbf{r}) \cdot \mathbf{q} &= \mathbf{p} \cdot \mathbf{q} + \mathbf{r} \cdot \mathbf{q}, \\ (\lambda \mathbf{p}) \cdot \mathbf{q} &= \lambda(\mathbf{p} \cdot \mathbf{q}) = \mathbf{p} \cdot (\lambda \mathbf{q}), & \|\lambda \mathbf{p}\| &= |\lambda| \, \|\mathbf{p}\|, \end{aligned}$$

for $\lambda \in \mathbb{R}$ and $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbb{R}^n$. Furthermore, the ***Cauchy–Schwarz*** and ***triangle*** inequalities state that for all $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$ we have

$$|\mathbf{p} \cdot \mathbf{q}| \leqslant \|\mathbf{p}\| \, \|\mathbf{q}\| \qquad \text{and} \qquad \|\mathbf{p} + \mathbf{q}\| \leqslant \|\mathbf{p}\| + \|\mathbf{q}\|.$$

(See, for example, [MS, pages 22–24].)

We shall also need the notion of the ***angle*** between nonzero vectors $\mathbf{p}, \mathbf{q}$ of $\mathbb{R}^n$, which is a number θ defined by:

$$\cos\theta = \frac{\mathbf{p} \cdot \mathbf{q}}{\|\mathbf{p}\| \, \|\mathbf{q}\|}.$$

The Cauchy–Schwarz inequality implies that

$$-1 \leqslant \frac{\mathbf{p} \cdot \mathbf{q}}{\|\mathbf{p}\| \, \|\mathbf{q}\|} \leqslant 1$$

for nonzero $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$, so that the definition of the angle makes sense. At times, it is convenient to specify the range of θ using Lemma 1.3 below.

A ***linear map*** of $\mathbb{R}^n$ into $\mathbb{R}^m$ is a function $A\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$A(\lambda \mathbf{p} + \mu \mathbf{q}) = \lambda A\mathbf{p} + \mu A\mathbf{q}$$

for $\lambda, \mu \in \mathbb{R}$ and $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$. We may regard such a map A as a $m \times n$ matrix, though to make the above definition consistent, one needs to represent elements of $\mathbb{R}^n$ by columns rather than rows. This is because the usual convention dictates that matrices, like functions, should act on the left.

So far we have been dealing with $\mathbb{R}^n$ for general n. When n is 2 or 3, the vector space $\mathbb{R}^n$ has special structures, namely a complex structure and a vector cross product, which are useful for describing curves and surfaces respectively. In Chapters 1–6 we shall be studying curves in the plane, so we concentrate attention on $\mathbb{R}^2$ now. Algebraic properties of $\mathbb{R}^3$ will be considered in Section 7.1, and those of $\mathbb{R}^4$ in Chapter 23.

For the differential geometry of curves in the plane an essential tool is the ***complex structure*** of $\mathbb{R}^2$; it is the linear map $J\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$J(p_1, p_2) = (-p_2, p_1).$$

Geometrically, J is merely a rotation by $\pi/2$ in a counterclockwise direction. It is easy to show that the complex structure J has the properties

$$J^2 = -\mathbf{1},$$
$$(J\mathbf{p}) \cdot (J\mathbf{q}) = \mathbf{p} \cdot \mathbf{q},$$
$$(J\mathbf{p}) \cdot \mathbf{p} = 0,$$

for $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2$ (where $\mathbf{1}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the identity map).

A point in the plane $\mathbb{R}^2$ can be considered a complex number via the canonical isomorphism

$$\mathbf{p} = (p_1, p_2) \;\leftrightarrow\; p_1 + ip_2 = \mathfrak{Re}\,\mathbf{p} + i\,\mathfrak{Im}\,\mathbf{p}, \tag{1.1}$$

where $\mathfrak{Re}\,\mathbf{p}$ and $\mathfrak{Im}\,\mathbf{p}$ denote the real and imaginary parts of $\mathbf{p}$. In a moment, we shall need descriptions of the dot product $\cdot$ and the complex structure J in terms of complex numbers. Recall that the ***complex conjugate*** and ***absolute value*** of a complex number $\mathbf{p}$ are defined by

$$\overline{\mathbf{p}} = \mathfrak{Re}\,\mathbf{p} - i\,\mathfrak{Im}\,\mathbf{p} \qquad \text{and} \qquad |\mathbf{p}| = \sqrt{\mathbf{p}\overline{\mathbf{p}}}.$$

The proof of the following lemma is elementary.

Lemma 1.2. Identify the plane $\mathbb{R}^2$ with the set of complex numbers $\mathbb{C}$, and let $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2 = \mathbb{C}$. Then

$$J\mathbf{p} = i\mathbf{p}, \qquad |\mathbf{p}| = \|\mathbf{p}\| \qquad \textit{and} \qquad \mathbf{p}\overline{\mathbf{q}} = \mathbf{p} \cdot \mathbf{q} + i\,\mathbf{p} \cdot (J\mathbf{q}). \tag{1.2}$$

The angle between vectors in $\mathbb{R}^n$ does not distinguish between the order of the vectors, but there is a refined notion of angle between vectors in $\mathbb{R}^2$ that makes this distinction.

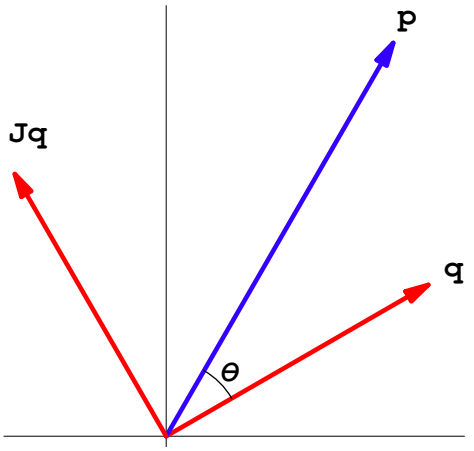


Figure 1.1: The oriented angle

Lemma 1.3. *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be nonzero vectors in* $\mathbb{R}^2$. *There exists a unique number* θ *with the properties*

$$\cos\theta = \frac{\mathbf{p}\cdot\mathbf{q}}{\|\mathbf{p}\|\,\|\mathbf{q}\|}, \qquad \sin\theta = \frac{\mathbf{p}\cdot J\mathbf{q}}{\|\mathbf{p}\|\,\|\mathbf{q}\|}, \qquad 0 \leqslant \theta < 2\pi. \tag{1.3}$$

We call θ *the* ***oriented angle*** *from* $\mathbf{q}$ *to* $\mathbf{p}$.

Proof. The oriented angle θ defines $\mathbf{p}$ in relation to the frame $(\mathbf{q},\, J\mathbf{q})$, as indicated in Figure 1.1. A formal proof proceeds as follows.

Since $\mathbf{p}\overline{\mathbf{q}}/(|\mathbf{p}|\,|\mathbf{q}|)$ is a complex number of absolute value 1, it lies on the unit circle in $\mathbb{C}$; thus there exists a unique θ with $0 \leqslant \theta < 2\pi$ such that

$$\frac{\mathbf{p}\,\overline{\mathbf{q}}}{|\mathbf{p}|\,|\mathbf{q}|} = e^{i\theta}. \tag{1.4}$$

Then using the expressions for $\mathbf{p}\overline{\mathbf{q}}$ in (1.2) and (1.4), we find that

$$\mathbf{p}\cdot\mathbf{q} + i\mathbf{p}\cdot(J\mathbf{q}) = e^{i\theta}|\mathbf{p}|\,|\mathbf{q}| = |\mathbf{p}|\,|\mathbf{q}|\cos\theta + i\,|\mathbf{p}|\,|\mathbf{q}|\sin\theta. \tag{1.5}$$

When we take the real and imaginary parts in (1.5), we get (1.3). ∎

1.2 Curves in Space

In the previous section we reviewed some of the algebraic properties of $\mathbb{R}^n$. We need to go one step further and study differentiation. In this chapter, and generally throughout the rest of the book, we shall use the word "differentiable" to mean "possessing derivatives or partial derivatives of all orders." We begin by studying $\mathbb{R}^n$-valued functions of one variable.

Definition 1.4. *Let* $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ *be a function, where* (a,b) *is an open interval in* $\mathbb{R}$. *We write*

$$\boldsymbol{\alpha}(t) = \big(a_1(t), \ldots, a_n(t)\big),$$

where each a_j *is an ordinary real-valued function of a real variable. We say that* $\boldsymbol{\alpha}$ *is* ***differentiable*** *provided* a_j *is differentiable for* $j = 1, \ldots, n$. *Similarly,* $\boldsymbol{\alpha}$ *is* ***piecewise-differentiable*** *provided* a_j *is piecewise-differentiable for* $j = 1, \ldots, n$.

Definition 1.5. *A* ***parametrized curve*** *in* $\mathbb{R}^n$ *is a piecewise-differentiable function*

$$\boldsymbol{\alpha}\colon (a,b) \longrightarrow \mathbb{R}^n,$$

where (a,b) *is an open interval in* $\mathbb{R}$. *We allow the interval to be finite, infinite or half-infinite. If* I *is any other subset of* $\mathbb{R}$, *we say that*

$$\boldsymbol{\alpha}\colon I \longrightarrow \mathbb{R}^n$$

is a ***curve*** *provided there is an open interval* (a,b) *containing* I *such that* $\boldsymbol{\alpha}$ *can be extended as a piecewise-differentiable function from* (a,b) *into* $\mathbb{R}^n$.

It is important to distinguish a curve $\boldsymbol{\alpha}$, which is a function, from the set of points traced out by $\boldsymbol{\alpha}$, which we call the ***trace*** of $\boldsymbol{\alpha}$. The trace of $\boldsymbol{\alpha}$ is just the image $\boldsymbol{\alpha}\big((a,b)\big)$, or more generally $\boldsymbol{\alpha}(I)$. We say that a subset $\mathscr{C}$ of $\mathbb{R}^n$ is ***parametrized by*** $\boldsymbol{\alpha}$ provided there is a subset $I \subseteq \mathbb{R}$ such that $\boldsymbol{\alpha}\colon I \to \mathbb{R}^n$ is a curve for which $\boldsymbol{\alpha}(I) = \mathscr{C}$.

Definition 1.6. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve with $\boldsymbol{\alpha}(t) = \big(a_1(t), \ldots, a_n(t)\big)$. Then the **velocity** of $\boldsymbol{\alpha}$ is the function $\boldsymbol{\alpha}'\colon (a,b) \to \mathbb{R}^n$ given by*

$$\boldsymbol{\alpha}'(t) = \big(a_1'(t), \ldots, a_n'(t)\big).$$

*The function v defined by $v(t) = \big\|\boldsymbol{\alpha}'(t)\big\|$ is called the **speed** of $\boldsymbol{\alpha}$. The **acceleration** of $\boldsymbol{\alpha}$ is given by $\boldsymbol{\alpha}''(t)$.*

Notice that $\boldsymbol{\alpha}'(t)$ is defined for those t for which $a_1'(t), \ldots, a_n'(t)$ are defined. When $n = 2$, the acceleration can be compared with the vector $J\boldsymbol{\alpha}'(t)$ as shown in Figure 1.2; this will be the basis of the definition of curvature on page 14.

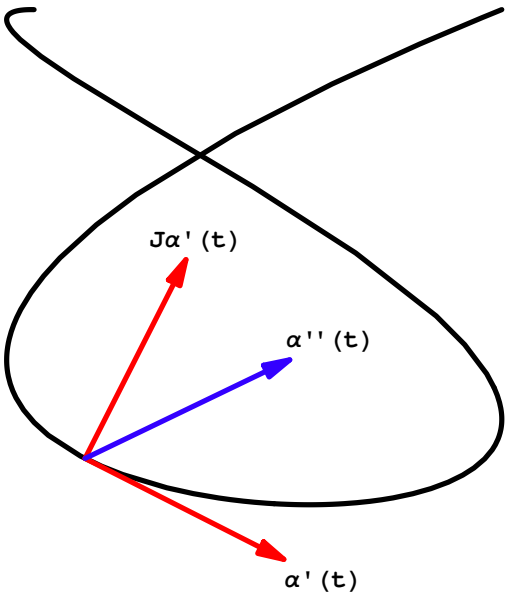


Figure 1.2: Velocity and acceleration of a plane curve

Some curves are "better" than others:

Definition 1.7. *A curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ is said to be **regular** if it is differentiable and its velocity is everywhere defined and nonzero. If $\big\|\boldsymbol{\alpha}'(t)\big\| = 1$ for $a < t < b$, then $\boldsymbol{\alpha}$ is said to have **unit speed**.*

The simplest example of a parametrized curve in $\mathbb{R}^n$ is a ***straight line***. If it contains distinct points $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$, it is most naturally parametrized by

$$\boldsymbol{\beta}(t) = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) = (1-t)\mathbf{p} + t\,\mathbf{q}, \tag{1.6}$$

with $t \in \mathbb{R}$. In the plane $\mathbb{R}^2$, the second simplest curve is a ***circle***. If it has radius r and center $(p_1, p_2) \in \mathbb{R}^2$, it can be parametrized by

$$\boldsymbol{\gamma}(t) = \big(p_1 + r\cos t,\ p_2 + r\sin t\big), \tag{1.7}$$

with $0 \leqslant t < 2\pi$.

Our definition of a curve in $\mathbb{R}^n$ is the *parametric form* of a curve. In contrast, the ***nonparametric form*** of a curve would typically consist of a system of $n-1$ equations

$$F_1(p_1, \ldots, p_n) = \ \cdots\ = F_{n-1}(p_1, \ldots, p_n) = 0,$$

where each $F_i \colon \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. We study the nonparametric form of curves in $\mathbb{R}^2$ in Section 3.1. It is usually easier to work with the parametric form of a curve, but there is one disadvantage: distinct parametrizations may trace out the same point set. Therefore, it is important to know when two curves are equivalent under a change of variables.

Definition 1.8. *Let $\boldsymbol{\alpha} \colon (a,b) \to \mathbb{R}^n$ and $\boldsymbol{\beta} \colon (c,d) \to \mathbb{R}^n$ be differentiable curves. Then $\boldsymbol{\beta}$ is said to be a* ***positive reparametrization*** *of $\boldsymbol{\alpha}$ provided there exists a differentiable function $h \colon (c,d) \to (a,b)$ such that $h'(u) > 0$ for all $c < u < d$ and $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$.*

Similarly, $\boldsymbol{\beta}$ is called a ***negative reparametrization*** *of $\boldsymbol{\alpha}$ provided there exists a differentiable function $h \colon (c,d) \to (a,b)$ such that $h'(u) < 0$ for all $c < u < d$ and $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$.*

We say that β is a ***reparametrization*** *of $\boldsymbol{\alpha}$ if it is either a positive or negative reparametrization of $\boldsymbol{\alpha}$.*

Next, we determine the relation between the velocity of a curve $\boldsymbol{\alpha}$ and the velocity of a reparametrization of $\boldsymbol{\alpha}$.

Lemma 1.9. (The chain rule for curves.) *Suppose that $\boldsymbol{\beta}$ is a reparametrization of $\boldsymbol{\alpha}$. Write $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$, where $h \colon (c,d) \to (a,b)$ is differentiable. Then*

$$\boldsymbol{\beta}'(u) = h'(u)\boldsymbol{\alpha}'\big(h(u)\big) \tag{1.8}$$

for $c < u < d$.

Proof. Write $\boldsymbol{\alpha}(t) = \big(a_1(t), \ldots, a_n(t)\big)$ and $\boldsymbol{\beta}(u) = \big(b_1(u), \ldots, b_n(u)\big)$. Then we have $b_j(u) = a_j\big(h(u)\big)$ for $j = 1, \ldots, n$. The ordinary chain rule implies that $b_j'(u) = a_j'\big(h(u)\big)h'(u)$ for $j = 1, \ldots, n$, and so we get (1.8). ∎

For example, the straight line (1.6) and circle (1.7) have slightly more complicated unit-speed parametrizations:

$$\begin{aligned} s &\mapsto \frac{\|\mathbf{p}-\mathbf{q}\|-s}{\|\mathbf{p}-\mathbf{q}\|}\,\mathbf{p} + \frac{s}{\|\mathbf{p}-\mathbf{q}\|}\,\mathbf{q} \\ s &\mapsto \Big(p_1 + r\cos\frac{s}{r},\; p_2 + r\sin\frac{s}{r}\Big). \end{aligned}$$

The quantities $\boldsymbol{\alpha}'$, $J\boldsymbol{\alpha}'$, $\boldsymbol{\alpha}''$ are examples of ***vector fields*** along a curve, because they determine a vector at *every* point of the curve, not just at the one point illustrated. We now give the general definition.

Definition 1.10. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve. A **vector field along $\boldsymbol{\alpha}$** is a function $\mathbf{Y}$ that assigns to each t with $a < t < b$ a vector $\mathbf{Y}(t)$ at the point $\boldsymbol{\alpha}(t)$.*

At this stage, we shall not distinguish between a vector at $\boldsymbol{\alpha}(t)$ and the vector parallel to it at the origin. This means a vector field $\mathbf{Y}$ along a curve $\boldsymbol{\alpha}$ is really an n-tuple of functions:

$$\mathbf{Y}(t) = \big(y_1(t), \ldots, y_n(t)\big).$$

(Later, for example in Section 9.6, we *will* need to distinguish between a vector at $\boldsymbol{\alpha}(t)$ and the vector parallel to it at the origin.) Differentiability of $\mathbf{Y}$ means that each of the functions $y_1, \ldots, y_n$ is differentiable, and in this case $\mathbf{Y}$ is effectively another curve and its derivative is defined in the obvious way:

$$\mathbf{Y}'(t) = \big(y_1'(t), \ldots, y_n'(t)\big).$$

Addition, scalar multiplication and dot product of vector fields along a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ are defined in the obvious way. Furthermore, we can multiply a vector field $\mathbf{Y}$ along $\boldsymbol{\alpha}$ by a function $f\colon (a,b) \to \mathbb{R}$: we define the vector field $f\mathbf{Y}$ by $(f\mathbf{Y})(t) = f(t)\mathbf{Y}(t)$. Finally, if $n = 2$ and $\mathbf{X}$ is a vector field along a curve with $\mathbf{X}(t) = (x(t), y(t))$, then another vector field $J\mathbf{X}$ can be defined by $J\mathbf{X}(t) = (-y(t), x(t))$. Taking the derivative is related to these operations in the obvious way:

Lemma 1.11. *Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are differentiable vector fields along a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$, and let $f\colon (a,b) \to \mathbb{R}$ be differentiable. Then*

(i) $(f\mathbf{Y})' = f'\mathbf{Y} + f\mathbf{Y}'$;

(ii) $(\mathbf{X}+\mathbf{Y})' = \mathbf{X}' + \mathbf{Y}'$;

(iii) $(\mathbf{X}\cdot\mathbf{Y})' = \mathbf{X}'\cdot\mathbf{Y} + \mathbf{X}\cdot\mathbf{Y}'$;

(iv) *for $n = 2$,* $(J\mathbf{Y})' = J\mathbf{Y}'$.

Proof. For example, we prove (i):

$$\begin{aligned}(f\mathbf{Y})' &= \big((fy_1)', \ldots, (fy_n)'\big) = (f'y_1 + fy_1', \ldots, f'y_n + fy_n') \\ &= f'(y_1, \ldots, y_n) + f(y_1', \ldots, y_n') = f'\mathbf{Y} + f\mathbf{Y}'. \ \blacksquare\end{aligned}$$

1.3 The Length of a Curve

One of the simplest and most important geometric quantities associated with a curve is its length. Everyone has a natural idea of what is meant by length, but this idea must be converted into an exact definition. For our purposes the simplest definition is as follows:

Definition 1.12. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve. Assume that $\boldsymbol{\alpha}$ is defined on a slightly larger interval containing (a,b), so that $\boldsymbol{\alpha}$ is defined and differentiable at a and b. Then the* ***length*** *of $\boldsymbol{\alpha}$ over the interval $[a,b]$ is given by*

$$\mathsf{length}[\boldsymbol{\alpha}] = \int_a^b \big\|\boldsymbol{\alpha}'(t)\big\|\, dt. \tag{1.9}$$

One can emphasize the role of the endpoints with the notation $\mathsf{length}[a,b][\boldsymbol{\alpha}]$ (compatible with Notebook 1); this is essential when we wish to compute the length of a curve restricted to an interval smaller than that on which it is defined. It is clear intuitively that length should not depend on the parametrization of the curve. We now prove this.

Theorem 1.13. *Let $\boldsymbol{\beta}$ be a reparametrization of $\boldsymbol{\alpha}$. Then*

$$\mathsf{length}[\boldsymbol{\alpha}] = \mathsf{length}[\boldsymbol{\beta}].$$

Proof. We do the positive reparametrization case first. Let $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$, where $h\colon (c,d) \to (a,b)$ and $h'(u) > 0$ for $c < u < d$. Then by Lemma 1.9 we have

$$\big\|\boldsymbol{\beta}'(u)\big\| = \big\|\boldsymbol{\alpha}'\big(h(u)\big)h'(u)\big\| = \big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|h'(u).$$

Using the change of variables formula for integrals, we compute

$$\begin{aligned}\mathsf{length}[\boldsymbol{\alpha}] \ = \int_a^b \big\|\boldsymbol{\alpha}'(t)\big\|\, dt \ &= \int_c^d \big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|h'(u)\, du \\ &= \int_c^d \big\|\boldsymbol{\beta}'(u)\big\|\, du \ = \ \mathsf{length}[\boldsymbol{\beta}].\end{aligned}$$

In the negative reparametrization case, we have

$$\lim_{u \downarrow c} h(u) = b, \qquad \lim_{u \uparrow d} h(u) = a,$$

where $u \downarrow c$ denotes $u \to c + 0$ and $u \uparrow d$ denotes $u \to d - 0$. Also

$$\big\|\boldsymbol{\beta}'(u)\big\| = \big\|\boldsymbol{\alpha}'\big(h(u)\big)h'(u)\big\| = -\big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|h'(u).$$

Again, using the change of variables formula for integrals,

$$\begin{aligned}\mathsf{length}[\boldsymbol{\alpha}] \;&= \int_a^b \big\|\boldsymbol{\alpha}'(t)\big\|\,dt \;= \int_d^c \big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|h'(u)\,du \\ &= \int_c^d \big\|\boldsymbol{\beta}'(u)\big\|\,du \;=\; \mathsf{length}[\boldsymbol{\beta}]. \;\blacksquare\end{aligned}$$

The trace of a curve $\boldsymbol{\alpha}$ can be approximated by a series of line segments connecting a sequence of points on the trace of $\boldsymbol{\alpha}$. Intuitively, the length of this approximation to $\boldsymbol{\alpha}$ should tend to the length of $\boldsymbol{\alpha}$ as the individual line segments become smaller and smaller. This is in fact true. To explain exactly what happens, let $\boldsymbol{\alpha}\colon (c,d) \to \mathbb{R}^n$ be a curve and $[a,b] \subset (c,d)$ a closed finite interval. For every partition

$$P = \{\, a = t_0 < t_1 < \cdots < t_N = b \,\} \tag{1.10}$$

of $[a,b]$, let

$$|P| = \max_{1 \leqslant j \leqslant N}(t_j - t_{j-1}) \qquad \text{and} \qquad \ell(\boldsymbol{\alpha}, P) = \sum_{j=1}^{N} \big\|\boldsymbol{\alpha}(t_j) - \boldsymbol{\alpha}(t_{j-1})\big\|.$$

Geometrically, $\ell(\boldsymbol{\alpha}, P)$ is the length of the polygonal path in $\mathbb{R}^n$ whose vertices are $\boldsymbol{\alpha}(t_j)$, $j = 0, \ldots, N$. Then $\mathsf{length}[\boldsymbol{\alpha}]$ is the limit of the lengths of inscribed polygonal paths in the sense of

Theorem 1.14. *Let $\boldsymbol{\alpha}\colon (c,d) \to \mathbb{R}^n$ be a curve, and let $\mathsf{length}[\boldsymbol{\alpha}]$ denote the length of the restriction of $\boldsymbol{\alpha}$ to a closed subinterval $[a,b]$. Given $\epsilon > 0$, there exists $\delta > 0$ such that*

$$|P| < \delta \qquad \textit{implies} \qquad \big|\,\mathsf{length}[\boldsymbol{\alpha}] - \ell(\boldsymbol{\alpha}, P)\big| < \epsilon. \tag{1.11}$$

Proof. Given a partition (1.10), write $\boldsymbol{\alpha} = (a_1, \ldots, a_n)$. For each i, the Mean Value Theorem implies that for $1 \leqslant j \leqslant N$ there exists $t_j^{(i)}$ with $t_{j-1} < t_j^{(i)} < t_j$ such that

$$a_i(t_j) - a_i(t_{j-1}) = a_i'(t_j^{(i)})(t_j - t_{j-1}).$$

Then

$$\begin{aligned}\big\|\boldsymbol{\alpha}(t_j) - \boldsymbol{\alpha}(t_{j-1})\big\|^2 \;&=\; \sum_{i=1}^{n} \big(a_i(t_j) - a_i(t_{j-1})\big)^2 \\ &=\; \sum_{i=1}^{n} a_i'(t_j^{(i)})^2 (t_j - t_{j-1})^2 \\ &=\; (t_j - t_{j-1})^2 \sum_{i=1}^{n} a_i'(t_j^{(i)})^2.\end{aligned}$$

Hence

$$\big\|\boldsymbol{\alpha}(t_j) - \boldsymbol{\alpha}(t_{j-1})\big\| \;=\; (t_j - t_{j-1})\sqrt{\sum_{i=1}^{n} a_i'(t_j^{(i)})^2} \;=\; (t_j - t_{j-1})(A_j + B_j),$$

where

$$A_j = \sqrt{\sum_{i=1}^{n} a_i'(t_j)^2} = \big\|\boldsymbol{\alpha}'(t_j)\big\| \quad \text{and} \quad B_j = \sqrt{\sum_{i=1}^{n} a_i'(t_j^{(i)})^2} - \sqrt{\sum_{i=1}^{n} a_i'(t_j)^2}.$$

An application of the triangle inequality allows us to deduce that

$$\begin{aligned} |B_j| \;&=\; \Big|\big\|\big(a_1'(t_j^{(1)}), \ldots, a_n'(t_j^{(n)})\big)\big\| - \big\|\boldsymbol{\alpha}'(t_j)\big\|\Big| \\ &\leqslant\; \big\|\big(a_1'(t_j^{(1)}), \ldots, a_n'(t_j^{(n)})\big) - \boldsymbol{\alpha}'(t_j)\big\| \;=\; \left(\sum_{i=1}^{n} \big(a_i'(t_j^{(i)}) - a_i'(t_j)\big)^2\right)^{1/2} \\ &\leqslant\; \sum_{i=1}^{n} \big|a_i'(t_j^{(i)}) - a_i'(t_j)\big|. \end{aligned}$$

Let $\epsilon > 0$. Since $\boldsymbol{\alpha}'$ is continuous on the closed interval $[a, b]$, there exists a positive number δ_1 such that

$$|u - v| < \delta_1 \qquad \text{implies} \qquad \big|a_i'(u) - a_i'(v)\big| < \frac{\epsilon}{2n(b-a)}$$

for $i = 1, \ldots, n$. Then $|P| < \delta_1$ implies that $|B_j| < \epsilon/(2(b-a))$, for each j, so that

$$\left|\sum_{j=1}^{N}(t_j - t_{j-1})B_j\right| \;\leqslant\; \sum_{j=1}^{N}(t_j - t_{j-1})\frac{\epsilon}{2(b-a)} \;=\; \frac{\epsilon}{2}.$$

Hence

$$\begin{aligned} \big|\,\mathsf{length}[\boldsymbol{\alpha}] - \ell(\boldsymbol{\alpha}, P)\,\big| \;&=\; \left|\mathsf{length}[\boldsymbol{\alpha}] - \sum_{j=1}^{N}\big\|\boldsymbol{\alpha}(t_j) - \boldsymbol{\alpha}(t_{j-1})\big\|\right| \\ &=\; \left|\mathsf{length}[\boldsymbol{\alpha}] - \sum_{j=1}^{N}(A_j + B_j)(t_j - t_{j-1})\right| \\ &\leqslant\; \left|\mathsf{length}[\boldsymbol{\alpha}] - \sum_{j=1}^{N} A_j(t_j - t_{j-1})\right| + \left|\sum_{j=1}^{N} B_j(t_j - t_{j-1})\right| \\ &\leqslant\; \left|\int_a^b \big\|\boldsymbol{\alpha}'(t)\big\|\,dt - \sum_{j=1}^{N}\big\|\boldsymbol{\alpha}'(t_j)(t_j - t_{j-1})\big\|\right| + \frac{\epsilon}{2}. \end{aligned}$$

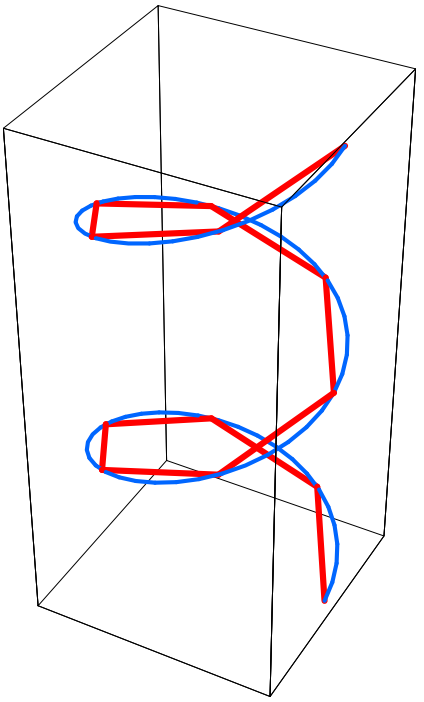


Figure 1.3: Polygonal approximation of the curve $t \mapsto (\cos t,\ \sin t,\ t/3)$

But

$$\sum_{j=1}^{N} \big\|\boldsymbol{\alpha}'(t_j)(t_j - t_{j-1})\big\|$$

is a Riemann sum approximation to the right-hand side of (1.9). Hence there exists $\delta_2 > 0$ such that $|P| < \delta_2$ implies that

$$\left| \sum_{j=1}^{N} \big\|\boldsymbol{\alpha}'(t_j)(t_j - t_{j-1})\big\| - \int_a^b \big\|\boldsymbol{\alpha}'(t)\big\| dt \right| < \frac{\epsilon}{2}.$$

Thus if we take $\delta = \min(\delta_1, \delta_2)$, we get (1.11). ∎

We need to know how the length of a curve varies when we change the limits of integration.

Definition 1.15. *Fix a number c with $a < c < b$. The* ***arc length function*** *s_α of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$, starting at c, is defined by*

$$s_\alpha(t) = \mathsf{length}[c,t][\boldsymbol{\alpha}] = \int_c^t \big\|\boldsymbol{\alpha}'(u)\big\|\, du,$$

for $a < t < b$. When there is no danger of confusion, we simplify s_α to s.

Note that if $a < d \leqslant f < b$, then $\mathsf{length}[d, f][\boldsymbol{\alpha}] = s_\alpha(f) - s_\alpha(d) \geqslant 0$.

Theorem 1.16. *Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ be a regular curve. Then there exists a unit-speed reparametrization $\boldsymbol{\beta}$ of $\boldsymbol{\alpha}$.*

Proof. By the fundamental theorem of calculus, any arc length function s of $\boldsymbol{\alpha}$ satisfies

$$\frac{ds}{dt}(t) = s'(t) = \big\|\boldsymbol{\alpha}'(t)\big\|.$$

Since $\boldsymbol{\alpha}$ is regular, $\boldsymbol{\alpha}'(t)$ is never zero; hence ds/dt is always positive. The Inverse Function Theorem implies that $t \mapsto s(t)$ has an inverse $s \mapsto t(s)$, and that

$$\frac{dt}{ds}\Big|_{s(t)} = \frac{1}{\dfrac{ds}{dt}\Big|_{t(s)}}.$$

Now define $\boldsymbol{\beta}$ by $\boldsymbol{\beta}(s) = \boldsymbol{\alpha}\big(t(s)\big)$. By Lemma 1.9, we have $\boldsymbol{\beta}'(s) = (dt/ds)\boldsymbol{\alpha}'\big(t(s)\big)$. Hence

$$\big\|\boldsymbol{\beta}'(s)\big\| = \left\|\frac{dt}{ds}\boldsymbol{\alpha}'\big(t(s)\big)\right\| = \frac{dt}{ds}\big\|\boldsymbol{\alpha}'\big(t(s)\big)\big\| = \frac{dt}{ds}\Big|_s \frac{ds}{dt}\Big|_{t(s)} = 1. \ \blacksquare$$

The arc length function of any unit-speed curve $\boldsymbol{\beta}\colon (c, d) \to \mathbb{R}^n$ starting at c satisfies

$$s_\beta(t) = \int_c^t \big\|\boldsymbol{\beta}'(u)\big\|\, du = t - c.$$

Thus the function s_β actually measures length along $\boldsymbol{\beta}$. This is the reason why unit-speed curves are said to be ***parametrized by arc length***.

We conclude this section by showing that unit-speed curves are unique up to starting point and direction.

Lemma 1.17. *Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ be a unit-speed curve, and let $\boldsymbol{\beta}\colon (c, d) \to \mathbb{R}^n$ be a reparametrization of $\boldsymbol{\alpha}$ such that $\boldsymbol{\beta}$ also has unit speed. Then*

$$\boldsymbol{\beta}(s) = \boldsymbol{\alpha}(\pm s + s_0)$$

for some constant real number s_0.

Proof. By hypothesis, there exists a differentiable function $h\colon (c, d) \to (a, b)$ such that $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$ and $h'(u) \neq 0$ for $c < u < d$. Then Lemma 1.9 implies that

$$1 = \big\|\boldsymbol{\beta}'(u)\big\| = \big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|\, |h'(u)| = |h'(u)|.$$

Therefore, $h'(u) = \pm 1$ for all t. Since h' is continuous on a connected open set, its sign is constant. Thus there is a constant s_0 such that $h(u) = \pm s + s_0$ for $c < u < d$, s being arc length for $\boldsymbol{\alpha}$. $\blacksquare$

1.4 Curvature of Plane Curves

Intuitively, the curvature of a curve measures the failure of a curve to be a straight line. The faster the velocity $\boldsymbol{\alpha}'$ turns along a curve $\boldsymbol{\alpha}$, the larger the curvature.[2]

In Chapter 7, we shall define the notion of curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ of a curve $\boldsymbol{\alpha}$ in $\mathbb{R}^n$ for arbitrary n. In the special case that $n = 2$ there is a refined version $\boldsymbol{\kappa 2}$ of $\boldsymbol{\kappa}$, which we now define. We first give the definition that is the easiest to use in computations. In the next section, we show that the curvature can be interpreted as the derivative of a "turning angle."

Definition 1.18. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular curve. The* ***curvature*** *$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ of $\boldsymbol{\alpha}$ is given by the formula*

$$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) = \frac{\boldsymbol{\alpha}''(t) \cdot J\boldsymbol{\alpha}'(t)}{\left\|\boldsymbol{\alpha}'(t)\right\|^3}. \tag{1.12}$$

The positive function $1/\left|\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\right|$ is called the ***radius of curvature*** *of $\boldsymbol{\alpha}$.*

The notation $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ is consistent with that of Notebook 1, though there is no danger in omitting $\boldsymbol{\alpha}$ if there is only one curve under consideration.

We can still speak of the curvature of a plane curve $\boldsymbol{\alpha}$ that is only piecewise-differentiable. At isolated points where the velocity vector $\boldsymbol{\alpha}'(t)$ vanishes or does not exist, the curvature $\boldsymbol{\kappa 2}(t)$ remains undefined. Similarly, the radius of curvature $1/|\boldsymbol{\kappa 2}(t)|$ is undefined at those t for which either $\boldsymbol{\alpha}'(t)$ or $\boldsymbol{\alpha}''(t)$ vanishes or is undefined.

Notice that the curvature is proportional (with a positive constant) to the orthogonal projection of $\boldsymbol{\alpha}''(t)$ in the direction of $J\boldsymbol{\alpha}'(t)$ (see Figure 1.2). If the acceleration $\boldsymbol{\alpha}''(t)$ vanishes, so does $\boldsymbol{\kappa 2}(t)$. In particular, the curvature of the straight line parametrized by (1.6) vanishes identically.

It is important to realize that the curvature can assume both positive and negative values. Often $\boldsymbol{\kappa 2}$ is called the ***signed curvature*** in order to distinguish it from a curvature function $\boldsymbol{\kappa}$ that we shall define in Chapter 7. It is useful to have some pictures to understand the meaning of positive and negative curvature for plane curves. There are four cases, each of which we illustrate by a parabola in Figures 1.4 and 1.5. These show that, briefly, positive curvature means "bend to the left" and negative curvature means "bend to the right."

---

[2]The notion of curvature of a plane curve first appears implicitly in the work of Apollonius of Perga (262–180 BC). Newton (1642–1727) was the first to study the curvature of plane curves explicitly, and found in particular formula (1.13).

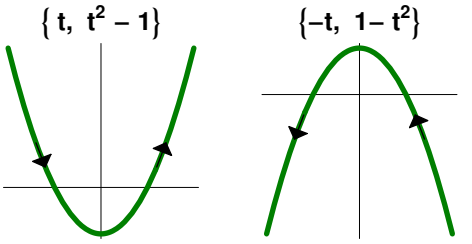


Figure 1.4: Parabolas with $\boldsymbol{\kappa 2} > 0$

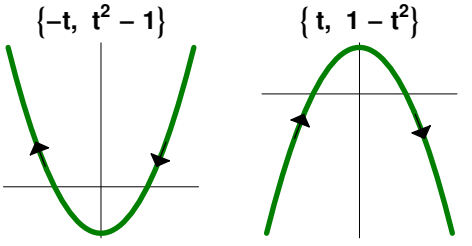


Figure 1.5: Parabolas with $\boldsymbol{\kappa 2} < 0$

Next, we derive the formula for curvature that can be found in most calculus books, and a version using complex numbers and the notation of Lemma 1.2.

Lemma 1.19. *If $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ is a regular curve with $\boldsymbol{\alpha}(t) = (x(t), y(t))$, then the curvature of $\boldsymbol{\alpha}$ is given in the equivalent ways*

$$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{\big(x'^2(t) + y'^2(t)\big)^{3/2}}, \tag{1.13}$$

$$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) = \mathfrak{Im}\, \frac{\boldsymbol{\alpha}''(t)\,\overline{\boldsymbol{\alpha}'(t)}}{\big\|\boldsymbol{\alpha}'(t)\big\|^3}. \tag{1.14}$$

Proof. We have $\boldsymbol{\alpha}''(t) = (x''(t), y''(t))$ and $J\boldsymbol{\alpha}'(t) = (-y'(t), x'(t))$, so that

$$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) = \frac{\big(x''(t), y''(t)\big) \cdot \big(-y'(t), x'(t)\big)}{\big(x'^2(t) + y'^2(t)\big)^{3/2}} = \frac{x'(t)y''(t) - x''(t)y'(t)}{\big(x'^2(t) + y'^2(t)\big)^{3/2}},$$

proving (1.13). Equation (1.14) follows from (1.12) and (1.2) (Exercise 5). ∎

Like the function $\mathsf{length}[\boldsymbol{\alpha}]$ the curvature $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ is a geometric quantity associated with a curve $\boldsymbol{\alpha}$. As such, it should be independent of the parametrization of the curve $\boldsymbol{\alpha}$. We now show that this is almost true. If λ is a nonzero real number, then $\lambda/|\lambda|$ is understandably denoted by "sign λ."

Theorem 1.20. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular curve, and let $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^2$ be a reparametrization of $\boldsymbol{\alpha}$. Write $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$, where $h\colon (c,d) \to (a,b)$ is differentiable. Then*

$$\boldsymbol{\kappa 2}[\beta](u) = \big(\operatorname{sign} h'(u)\big)\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big(h(u)\big), \tag{1.15}$$

wherever $h'(u) \neq 0$. Thus the curvature of a plane curve is independent of the parametrization up to sign.

Proof. We have $\boldsymbol{\beta}' = (\boldsymbol{\alpha}' \circ h)h'$, so that $J\boldsymbol{\beta}' = J(\boldsymbol{\alpha}' \circ h)\, h'$, and

$$\boldsymbol{\beta}'' = (\boldsymbol{\alpha}'' \circ h)h'^2 + (\boldsymbol{\alpha}' \circ h)h''. \tag{1.16}$$

Thus

$$\begin{aligned} \boldsymbol{\kappa 2}[\beta] &= \frac{\big((\boldsymbol{\alpha}'' \circ h)h'^2 + (\boldsymbol{\alpha}' \circ h)h''\big) \cdot J(\boldsymbol{\alpha}' \circ h)\, h'}{\big\|(\boldsymbol{\alpha}' \circ h)h'\big\|^3} \\ &= \left(\frac{h'^3}{|h'|^3}\right) \frac{(\boldsymbol{\alpha}'' \circ h) \cdot J(\boldsymbol{\alpha}' \circ h)}{\big\|(\boldsymbol{\alpha}' \circ h)\big\|^3}. \end{aligned}$$

Hence we get (1.15). ∎

The formula for $\boldsymbol{\kappa 2}$ simplifies dramatically for a unit-speed curve:

Lemma 1.21. *Let $\boldsymbol{\beta}$ be a unit-speed curve in the plane. Then*

$$\boldsymbol{\beta}'' = \boldsymbol{\kappa 2}[\boldsymbol{\beta}]J\boldsymbol{\beta}'. \tag{1.17}$$

Proof. Differentiating $\boldsymbol{\beta}' \cdot \boldsymbol{\beta}' = 1$, we obtain $\boldsymbol{\beta}'' \cdot \boldsymbol{\beta}' = 0$. Thus $\boldsymbol{\beta}''$ must be a multiple of $J\boldsymbol{\beta}'$. In fact, it follows easily from (1.12) that this multiple is $\boldsymbol{\kappa 2}[\boldsymbol{\beta}]$, and we obtain (1.17) . ∎

Finally, we give simple characterizations of straight lines and circles by means of curvature.

Theorem 1.22. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular curve.*

(i) *$\boldsymbol{\alpha}$ is part of a straight line if and only if $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) \equiv 0$.*

(ii) *$\boldsymbol{\alpha}$ is part of a circle of radius $r > 0$ if and only if $\big|\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big|(t) \equiv 1/r$.*

Proof. It is easy to show that the curvature of the straight line is zero, and that the curvature of a circle of radius r is $+1/r$ for a counterclockwise circle and $-1/r$ for a clockwise circle (see Exercise 3).

For the converse, we can assume without loss of generality that $\boldsymbol{\alpha}$ is a unit-speed curve. Suppose $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) \equiv 0$. Lemma 1.21 implies that $\boldsymbol{\alpha}''(t) = 0$ for $a < t < b$. Hence there exist constant vectors $\mathbf{p}$ and $\mathbf{q}$ such that

$$\boldsymbol{\alpha}(t) = \mathbf{p}t + \mathbf{q}$$

for $a < t < b$; thus $\boldsymbol{\alpha}$ is part of a straight line.

Next, assume that $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)$ is identically equal to a positive constant $1/r$. Without loss of generality, $\boldsymbol{\alpha}$ has unit speed. Define a curve $\boldsymbol{\gamma}\colon (b,c) \to \mathbb{R}^2$ by

$$\boldsymbol{\gamma}(t) = \boldsymbol{\alpha}(t) + rJ\boldsymbol{\alpha}'(t).$$

Lemma 1.21 implies that $\boldsymbol{\gamma}'(t) = 0$ for $a < t < b$. Hence there exists $\mathbf{q}$ such that $\boldsymbol{\gamma}(t) = \mathbf{q}$ for $a < t < b$. Then $\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\| = \big\|rJ\boldsymbol{\alpha}'(t)\big\| = r$. Hence $\boldsymbol{\alpha}(t)$ lies on a circle of radius r centered at $\mathbf{q}$.

Similarly, $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) \equiv -1/r$ also implies that $\boldsymbol{\alpha}$ is part of a circle of radius r (see Exercise 4). ∎

1.5 Angle Functions

In Section 1.1 we defined the notion of oriented angle between vectors in $\mathbb{R}^2$; now we wish to define a similar notion for curves. Clearly, we can compute the oriented angle between corresponding velocity vectors. This oriented angle θ satisfies $0 \leqslant \theta < 2\pi$, but such a restriction is not desirable for curves. The problem is that the two curves can twist and turn so that eventually the angle between them lies outside the interval $[0, 2\pi)$. To arrive at the proper definition, we need the following useful lemma of O'Neill [ON1].

Lemma 1.23. *Let $f, g\colon (a,b) \to \mathbb{R}$ be differentiable functions with $f^2 + g^2 = 1$. Fix t_0 with $a < t_0 < b$ and let θ_0 be such that $f(t_0) = \cos\theta_0$ and $g(t_0) = \sin\theta_0$. Then there exists a unique continuous function $\theta\colon (a,b) \to \mathbb{R}$ such that*

$$\theta(t_0) = \theta_0, \quad f(t) = \cos\theta(t) \quad \textit{and} \quad g(t) = \sin\theta(t) \tag{1.18}$$

for $a < t < b$.

Proof. We use complex numbers. Let $h = f + ig$, so $h\overline{h} = 1$. Define

$$\theta(t) = \theta_0 \; - \; i\int_{t_0}^{t} h'(s)\overline{h(s)}\,ds. \tag{1.19}$$

Then

$$\frac{d}{dt}(h\,e^{-i\theta}) = e^{-i\theta}(h' - i\,h\theta') = e^{-i\theta}(h' - h\,h'\,\overline{h}) = 0.$$

Thus $h\,e^{-i\theta} = c$ for some constant c. Since $h(t_0) = e^{i\theta_0}$, it follows that

$$c = h(t_0)e^{-i\theta(t_0)} = 1.$$

Hence $h(t) = e^{i\theta(t)}$, and so we get (1.18).

Let $\hat{\theta}$ be another continuous function such that

$$\hat{\theta}(t_0) = \theta_0, \quad f(t) = \cos\hat{\theta}(t) \quad \text{and} \quad g(t) = \sin\hat{\theta}(t)$$

for $a < t < b$. Then $e^{i\theta(t)} = e^{i\hat{\theta}(t)}$ for $a < t < b$. Since both θ and $\hat{\theta}$ are continuous, there is an integer n such that

$$\theta(t) - \hat{\theta}(t) = 2\pi n$$

for $a < t < b$. But $\theta(t_0) = \hat{\theta}(t_0)$, so that $n = 0$. Hence θ and $\hat{\theta}$ coincide. ∎

We can now apply this lemma to deduce the existence and uniqueness of a differentiable angle function between curves in $\mathbb{R}^2$.

Corollary 1.24. *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be regular curves in $\mathbb{R}^2$ defined on the same interval (a, b), and let $a < t_0 < b$. Choose θ_0 such that*

$$\frac{\boldsymbol{\alpha}'(t_0) \cdot \boldsymbol{\beta}'(t_0)}{\|\boldsymbol{\alpha}'(t_0)\| \, \|\boldsymbol{\beta}'(t_0)\|} = \cos\theta_0 \quad \textit{and} \quad \frac{\boldsymbol{\alpha}'(t_0) \cdot J\boldsymbol{\beta}'(t_0)}{\|\boldsymbol{\alpha}'(t_0)\| \, \|\boldsymbol{\beta}'(t_0)\|} = \sin\theta_0.$$

Then there exists a unique differentiable function $\theta\colon (a, b) \to \mathbb{R}$ such that

$$\theta(t_0) = \theta_0, \quad \frac{\boldsymbol{\alpha}'(t) \cdot \boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\beta}'(t)\|} = \cos\theta(t) \quad \textit{and} \quad \frac{\boldsymbol{\alpha}'(t) \cdot J\boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\beta}'(t)\|} = \sin\theta(t)$$

*for $a < t < b$. We call θ the **angle function** from $\boldsymbol{\beta}$ to $\boldsymbol{\alpha}$ determined by θ_0.*

Proof. We take

$$f(t) = \frac{\boldsymbol{\alpha}'(t) \cdot \boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\beta}'(t)\|} \quad \text{and} \quad g(t) = \frac{\boldsymbol{\alpha}'(t) \cdot J\boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\beta}'(t)\|}$$

in Lemma 1.23. ∎

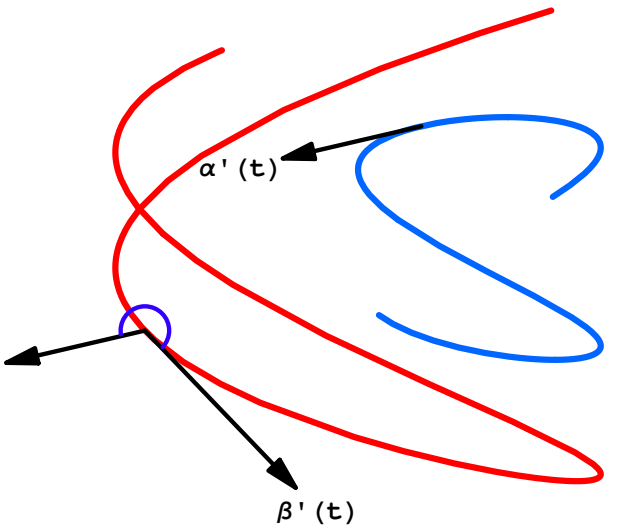


Figure 1.6: The angle function between curves

Intuitively, it is clear that the velocity of a highly curved plane curve changes rapidly. This idea can be made precise with the concept of ***turning angle***.

Lemma 1.25. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular curve and fix t_0 with $a < t_0 < b$. Let θ_0 be a number such that*

$$\frac{\boldsymbol{\alpha}'(t_0)}{\|\boldsymbol{\alpha}'(t_0)\|} = (\cos\theta_0,\ \sin\theta_0).$$

Then there exists a unique differentiable function $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\alpha}]\colon (a,b) \to \mathbb{R}$ such that $\boldsymbol{\theta}(t_0) = \theta_0$ and

$$\frac{\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|} = \big(\cos\boldsymbol{\theta}(t),\ \sin\boldsymbol{\theta}(t)\big) \tag{1.20}$$

*for $a < t < b$. We call $\boldsymbol{\theta}[\boldsymbol{\alpha}]$ the **turning angle** determined by θ_0.*

Proof. Let $\boldsymbol{\beta}(t) = (t,0)$; then $\boldsymbol{\beta}$ parametrizes a horizontal straight line and $\boldsymbol{\beta}'(t) = (1,0)$ for all t. Write $\boldsymbol{\alpha}(t) = (a_1(t), a_2(t))$; then

$$\boldsymbol{\alpha}'(t)\cdot\boldsymbol{\beta}'(t) = a_1'(t) \qquad \text{and} \qquad \boldsymbol{\alpha}'(t)\cdot J\boldsymbol{\beta}'(t) = a_2'(t).$$

Corollary 1.24 implies the existence of a unique function $\boldsymbol{\theta}\colon (a,b) \to \mathbb{R}$ such that $\boldsymbol{\theta}(t_0) = \theta_0$ and

$$\begin{cases} \cos\boldsymbol{\theta}(t) = \dfrac{\boldsymbol{\alpha}'(t)\cdot\boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\|\,\|\boldsymbol{\beta}'(t)\|} = \dfrac{a_1'(t)}{\|\boldsymbol{\alpha}'(t)\|}, \\[2ex] \sin\boldsymbol{\theta}(t) = \dfrac{\boldsymbol{\alpha}'(t)\cdot J\boldsymbol{\beta}'(t)}{\|\boldsymbol{\alpha}'(t)\|\,\|\boldsymbol{\beta}'(t)\|} = \dfrac{a_2'(t)}{\|\boldsymbol{\alpha}'(t)\|}. \end{cases} \tag{1.21}$$

Then (1.21) is equivalent to (1.20). ∎

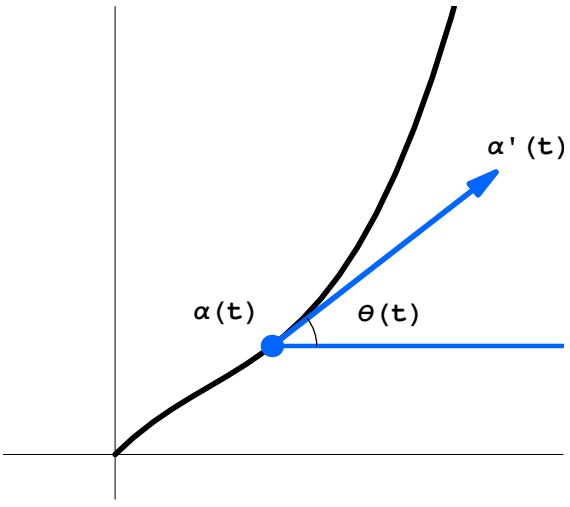


Figure 1.7: The turning angle of a plane curve

Geometrically, the turning angle $\boldsymbol{\theta}[\boldsymbol{\alpha}](t)$ is the angle between the horizontal and $\boldsymbol{\alpha}'(t)$. Next, we derive a relation (first observed by Kästner[3]) between the turning angle and the curvature of a plane curve.

Lemma 1.26. *The turning angle and curvature of a regular curve $\boldsymbol{\alpha}$ in the plane are related by*

$$\boldsymbol{\theta}[\boldsymbol{\alpha}]'(t) = \big\|\boldsymbol{\alpha}'(t)\big\|\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t). \tag{1.22}$$

Proof. The derivative of the left-hand side of (1.20) is

$$\frac{\boldsymbol{\alpha}''(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|} + \boldsymbol{\alpha}'(t)\frac{d}{dt}\Big(\frac{1}{\big\|\boldsymbol{\alpha}'(t)\big\|}\Big). \tag{1.23}$$

Furthermore, by the chain rule (Lemma 1.9) it follows that the derivative of the right-hand side of (1.20) is

$$\boldsymbol{\theta}'(t)\big(-\sin\boldsymbol{\theta}(t),\, \cos\boldsymbol{\theta}(t)\big) = \boldsymbol{\theta}'(t)\frac{J\boldsymbol{\alpha}'(t)}{\big\|J\boldsymbol{\alpha}'(t)\big\|}. \tag{1.24}$$

Setting (1.24) equal to (1.23) and taking the dot product with $J\boldsymbol{\alpha}'(t)$, we obtain

$$\boldsymbol{\theta}'(t)\big\|\boldsymbol{\alpha}'(t)\big\| = \frac{\boldsymbol{\alpha}''(t)\cdot J\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|} = \big\|\boldsymbol{\alpha}'(t)\big\|^2\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t).$$

Equation (1.22) now follows. ∎

Corollary 1.27. *The turning angle and curvature of a unit-speed curve $\boldsymbol{\beta}$ in the plane are related by*

$$\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s) = \frac{d\boldsymbol{\theta}[\boldsymbol{\beta}](s)}{ds}.$$

Curvature therefore measures rate of change of the turning angle with respect to arc length.

1.6 First Examples of Plane Curves

The equation of a circle of radius a with center $(0,0)$ is

$$x^2 + y^2 = a^2.$$

Its simplest parametrization is

$$\mathsf{circle}[a](t) = a\big(\cos t,\, \sin t\big), \qquad 0 \leqslant t < 2\pi, \tag{1.25}$$



[3] Abraham Gotthelf Kästner (1719–1800). German mathematician, professor at Leipzig and Göttingen. Although Kästner's German contemporaries ranked his mathematical and expository skills very high, Gauss found his lectures too elementary to attend. Gauss declared that Kästner was first mathematician among poets and first poet among mathematicians.

as in (1.7) but with a now representing the radius (not an initial parameter value). A more interesting parametrization is obtained by setting $\tau = \tan(t/2)$, so as to give

$$\boldsymbol{\beta}(\tau) = a\left(\frac{1-\tau^2}{1+\tau^2},\ \frac{2\tau}{1+\tau^2}\right), \qquad \tau \in \mathbb{R},$$

though the trace of $\boldsymbol{\beta}$ omits the point $(-a, 0)$.

We shall now study two classes of curves that include the circle as a special case. In both cases, we shall graph their curvature.

Ellipses

These are perhaps the next simplest curves after the circle. The name "ellipse" (which means "falling short") is due to Apollonius.[4] The nonparametric form of the ellipse centered at the origin with axes of lengths a and b is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

and the standard parametrization is

$$\mathsf{ellipse}[a, b](t) = \big(a\cos t,\ b\sin t\big), \qquad 0 \leqslant t < 2\pi. \tag{1.26}$$

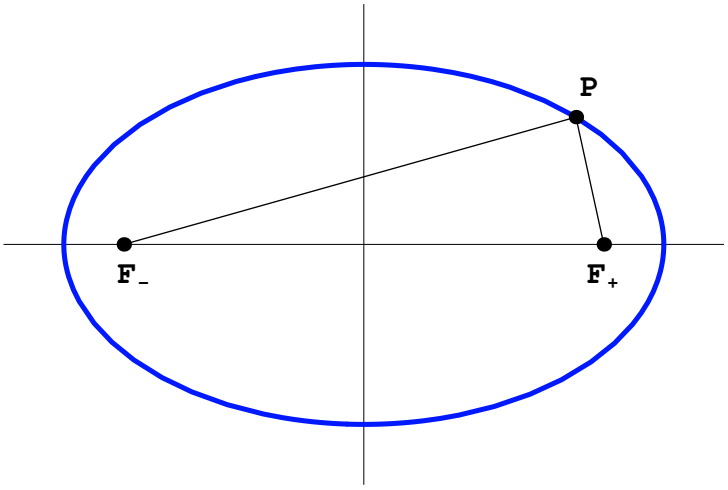


Figure 1.8: An ellipse and its foci

The curvature of this ellipse is

$$\boldsymbol{\kappa}\mathbf{2}[\mathsf{ellipse}[a, b]](t) = \frac{ab}{(b^2\cos^2 t + a^2\sin^2 t)^{3/2}}.$$

[4] Apollonius of Perga (262–180 BC). His eight volume treatise on conic sections is the standard ancient source of information about ellipses, hyperbolas, and parabolas.

Its graph for $a = 3,\ b = 4$ (the ellipse in Figure 1.8) is illustrated below with an exaggerated vertical scale. The fact that there are two maxima and two minima is related to the Four Vertex Theorem, which will be discussed in Section 6.5.

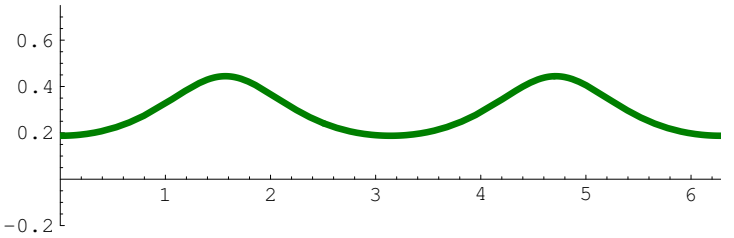


Figure 1.9: The curvature of an ellipse

An ellipse is traditionally defined as the locus of points such that the sum of the distances from two fixed points F_- and F_+ is constant. The points F_- and F_+ are called the ***foci*** of the ellipse. In the case of $\mathsf{ellipse}[a, b]$ with $a > b$ the foci are $\big(\pm\sqrt{a^2 - b^2},\ 0\big)$:

Theorem 1.28. *Let $\mathscr{E}$ be an ellipse with foci F_- and F_+. Then*

(i) *for any point P on $\mathscr{E}$, the sum*

$$\mathsf{distance}(F_-, P) + \mathsf{distance}(F_+, P)$$

is constant.

(ii) *the tangent line to $\mathscr{E}$ at P makes equal angles with the line segments connecting P to F_- and F_+.*

Proof. To establish these two properties of an ellipse, it is advantageous to use complex numbers. Clearly, the parametrization (1.26) is equivalent to

$$\mathbf{z}(t) = \tfrac{1}{2}(a + b)e^{it} + \tfrac{1}{2}(a - b)e^{-it} \qquad (0 \leqslant t < 2\pi). \tag{1.27}$$

The vector from $F_\mp$ to P is given by

$$\mathbf{z}_\pm(t) = \tfrac{1}{2}(a + b)e^{it} + \tfrac{1}{2}(a - b)e^{-it} \ \pm\ \sqrt{a^2 - b^2}.$$

It is remarkable that this can be written as a perfect square

$$\mathbf{z}_\pm(t) = \tfrac{1}{2}\left(\sqrt{a + b}\, e^{it/2} \pm \sqrt{a - b}\, e^{-it/2}\right)^2. \tag{1.28}$$

An easy calculation now shows that

$$|\mathbf{z}_\pm(t)| = a \pm \sqrt{a^2 - b^2}\cos t.$$

Consequently, $|\mathbf{z}_+(t)| + |\mathbf{z}_-(t)| = 2a$; this proves (i).

To prove (ii), we first use (1.27) to compute

$$\begin{aligned}\mathbf{z}'(t) &= \frac{i}{2}(a+b)\,e^{it} - \frac{i}{2}(a-b)\,e^{-it} \\ &= \frac{i}{2}\left(\sqrt{a+b}\,e^{it/2}+\sqrt{a+b}\,e^{-it/2}\right)\left(\sqrt{a+b}\,e^{it/2}-\sqrt{a+b}\,e^{-it/2}\right).\end{aligned}$$

It follows from (1.28) that

$$i\,\frac{\mathbf{z}_+(t)}{\mathbf{z}'(t)} = \frac{\sqrt{a+b}\,e^{it/2}+\sqrt{a+b}\,e^{-it/2}}{\sqrt{a+b}\,e^{it/2}-\sqrt{a+b}\,e^{-it/2}} = -i\,\frac{\mathbf{z}'(t)}{\mathbf{z}_-(t)},$$

and (ii) follows once one expresses the three complex numbers $\mathbf{z}_-(t), \mathbf{z}'(t), \mathbf{z}_-(t)$ in exponential form. ∎

Logarithmic spirals

A ***logarithmic spiral*** is parametrized by

$$\mathsf{logspiral}[a,b](t) = a\big(e^{bt}\cos t,\ e^{bt}\sin t\big). \tag{1.29}$$

Changing the sign of b makes the spiral unwind the other way, with $b = 0$ corresponding to the circle. If $b > 0$, it is evident that the curvature decreases as t increases. We check this by graphing the function

$$\boldsymbol{\kappa 2}[\mathsf{logspiral}[a,b]](t) = \frac{1}{a\,e^{bt}\sqrt{1+b^2}},$$

obtained from Notebook 1, in the case corresponding to Figure 1.10.

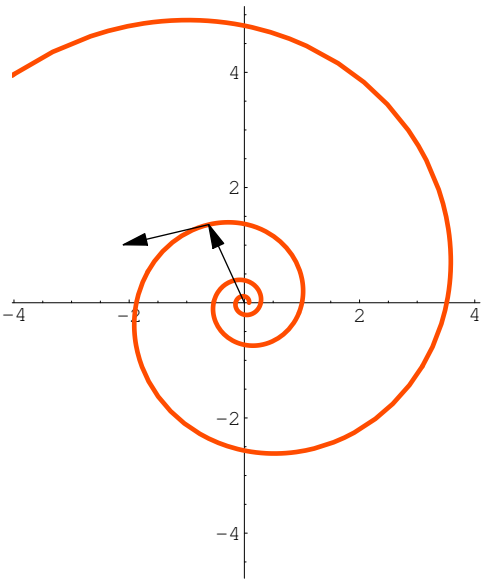


Figure 1.10: The curve $\mathsf{logspiral}[1, \frac{1}{5}]$

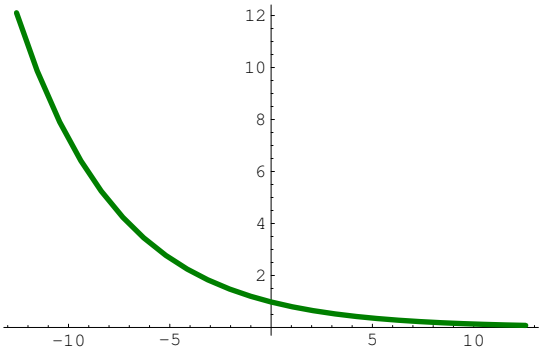


Figure 1.11: The function $\boldsymbol{\kappa 2}[\mathsf{logspiral}[1, \frac{1}{5}]](t) = 0.98e^{-0.2t}$

Logarithmic spirals were first discussed by Descartes in 1638 in connection with a problem from dynamics. He was interested in determining all plane curves $\boldsymbol{\alpha}$ with the property that the angle between $\boldsymbol{\alpha}(t)$ and the tangent vector $\boldsymbol{\alpha}'(t)$ is constant. Such curves turn out to be precisely the logarithmic spirals, as we proceed to prove.

Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a curve which does not pass through the origin. The ***tangent-radius angle*** of $\boldsymbol{\alpha}$ is defined to be the function $\boldsymbol{\phi} = \boldsymbol{\phi}[\boldsymbol{\alpha}]\colon (a,b) \to \mathbb{R}^2$ such that

$$\frac{\boldsymbol{\alpha}'(t) \cdot \boldsymbol{\alpha}(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\alpha}(t)\|} = \cos \boldsymbol{\phi}(t) \quad \text{and} \quad \frac{\boldsymbol{\alpha}'(t) \cdot J\boldsymbol{\alpha}(t)}{\|\boldsymbol{\alpha}'(t)\| \, \|\boldsymbol{\alpha}(t)\|} = \sin \boldsymbol{\phi}(t)$$

for $a < t < b$. The existence of $\boldsymbol{\phi}[\boldsymbol{\alpha}]$ follows from Lemma 1.23; see Exercise 6.

Lemma 1.29. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a curve which does not pass through the origin. The following conditions are equivalent:*

(i) *The tangent-radius angle $\boldsymbol{\phi}[\boldsymbol{\alpha}]$ is constant;*

(ii) *$\boldsymbol{\alpha}$ is a reparametrization of a logarithmic spiral.*

Proof. We write $\boldsymbol{\alpha}(t) = r\,e^{i\theta}$ with $r = r(t)$ and $\theta = \theta(t)$ both functions of t. Then $\boldsymbol{\alpha}'(t) = e^{i\theta}(r' + i\,r\,\theta')$, so that

$$\big|\boldsymbol{\alpha}(t)\big| = r, \quad \big|\boldsymbol{\alpha}'(t)\big| = \sqrt{r'^2 + r^2\theta'^2} \quad \text{and} \quad \overline{\boldsymbol{\alpha}(t)}\boldsymbol{\alpha}'(t) = r(r' + i\,r\,\theta').$$

Assume that (i) holds and let γ be the constant value of $\boldsymbol{\phi}[\boldsymbol{\alpha}](t)$. Equation (1.4) implies that

$$e^{i\gamma} = \frac{\boldsymbol{\alpha}'(t)\overline{\boldsymbol{\alpha}(t)}}{\big|\boldsymbol{\alpha}'(t)\big| \, \big|\boldsymbol{\alpha}(t)\big|} = \frac{r' + i\,r\,\theta'}{\sqrt{r'^2 + r^2\theta'^2}}. \tag{1.30}$$

From (1.30) we get

$$\cos\gamma = \frac{r'}{\sqrt{r'^2 + r^2\theta'^2}} \qquad \text{and} \qquad \sin\gamma = \frac{r\,\theta'}{\sqrt{r'^2 + r^2\theta'^2}},$$

so that

$$\frac{r'}{r} = \theta' \cot\gamma.$$

The solution of this differential equation is found to be

$$r = a\exp(\theta\cot\gamma),$$

where a is a constant, and we obtain

$$\boldsymbol{\alpha}(t) = a\exp\big((i+\cot\gamma)\theta(t)\big), \tag{1.31}$$

which is equivalent to (1.29).

Conversely, one can check that the tangent-radius angle of any logarithmic spiral is a constant γ. Under a reparametrization, the tangent-radius angle has the constant value $\pm\gamma$. ▌

1.7 The Semicubical Parabola and Regularity

It is informative to calculate lengths and curvatures by hand in a simple case.

We choose the ***semicubical parabola*** defined parametrically by

$$\mathsf{sc}(t) = \big(t^2,\, t^3\big). \tag{1.32}$$

The nonparametric version of sc is obviously $x^3 = y^2$. The semicubical parabola has a special place in the history of curves, because it was the first algebraic curve (other than a straight line) to be found whose arc length function is algebraic. See [Lock, pages 3–11] for an extensive discussion.

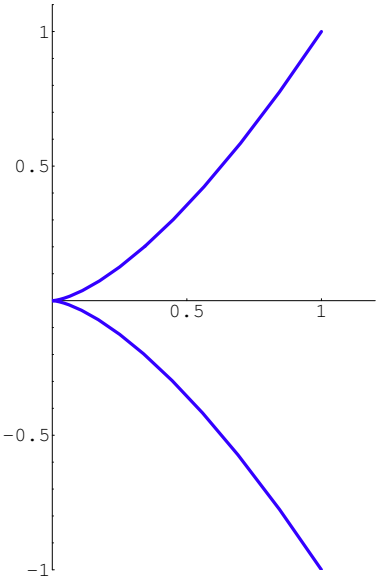


Figure 1.12: The semicubical parabola

We compute easily

$$\mathsf{sc}'(t) = (2t,\, 3t^2), \qquad J\,\mathsf{sc}'(t) = (-3t^2,\, 2t), \qquad \mathsf{sc}''(t) = (2,\, 6t),$$

so that

$$\boldsymbol{\kappa 2}[\mathsf{sc}](t) = \frac{(2,6t)\cdot(-3t^2,2t)}{\big\|(2t,3t^2)\big\|^3} = \frac{6}{|t|(4+9t^2)^{3/2}}.$$

Also, the length of sc over the interval $[0,t]$ is

$$s_{\mathrm{sc}}(t) = \int_0^t \big\|\mathsf{sc}'(u)\big\|\,du = \int_0^t u\sqrt{4+9\,u^2}\,du = \frac{1}{27}(4+9t^2)^{3/2} - \frac{8}{27}.$$

Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve, and let $a < t_0 < b$. There are two ways that $\boldsymbol{\alpha}$ can fail to be regular at t_0. On the one hand, regularity can fail because of the particular parametrization. For example, a horizontal straight line, when parametrized as $t \mapsto (t^3, 0)$, is not regular at 0. This kind of nonregularity can be avoided by changing the parametrization. A more fundamental kind of nonregularity occurs when it is impossible to find a regular reparametrization of a curve near a point.

Definition 1.30. *We say that a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ is **regular** at t_0 provided that it is possible to extend the function*

$$t \mapsto \frac{\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|}$$

*to be a differentiable function at t_0. Otherwise $\boldsymbol{\alpha}$ is said to be **singular** at t_0.*

It is easy to see that the definitions of regular and singular at t_0 are independent of the parametrization of $\boldsymbol{\alpha}$. The notion of ***regular curve*** $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$, as defined on page 6, implies regularity at each t_0 for $a < t_0 < b$, but is stronger, because it implies that $\boldsymbol{\alpha}'(t_0)$ must be nonzero.

The following result may easily be deduced from the previous definition:

Lemma 1.31. *A curve $\boldsymbol{\alpha}$ is regular at t_0 if and only if there exists a unit-speed parametrization of $\boldsymbol{\alpha}$ near t_0.*

For example, in seeking a unit-speed parametrization of sc, a computation shows that

$$\frac{\mathsf{sc}'(t)}{\big\|\mathsf{sc}'(t)\big\|} = \frac{(2t,3t^2)}{\sqrt{t^2(4+9t^2)}}.$$

However, this is discontinuous at the origin because

$$(-1,0) = \lim_{t\uparrow 0} \frac{(2t,3t^2)}{\sqrt{t^2(4+9t^2)}} \;\neq\; \lim_{t\downarrow 0} \frac{(2t,3t^2)}{\sqrt{t^2(4+9t^2)}} = (1,0),$$

reflecting the singularity of (1.32) at $t = 0$.

1.8 Exercises

1. Let $a > 0$. Find the arc length function and the curvature for each of the following curves, illustrated in Figure 1.13.

(a) $t \mapsto a\big(\cos t + t \sin t,\ \sin t - t \cos t\big)$;

(b) $t \mapsto \big(a \cosh \frac{t}{a},\ t\big)$;

(c) $t \mapsto a\big(\cos^3 t,\ \sin^3 t\big)$;

(d) $t \mapsto a\big(2t,\ t^2\big)$.

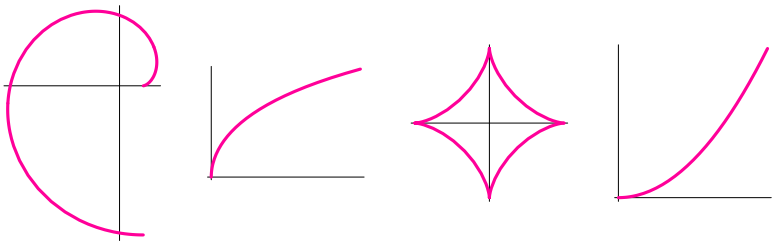


Figure 1.13: The curves of Exercise 1

2. Show that the velocity $\boldsymbol{\alpha}'$ of a differentiable curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ is given by

$$\boldsymbol{\alpha}'(t) = \lim_{\delta \to 0} \frac{\boldsymbol{\alpha}(t + \delta) - \boldsymbol{\alpha}(t)}{\delta}, \qquad a < t < b.$$

3. Let $\boldsymbol{\beta}\colon (a, b) \to \mathbb{R}^2$ be a circle of radius $r > 0$ centered at $\mathbf{q} \in \mathbb{R}^2$. Using the fact that $\big\|\boldsymbol{\beta}(t) - \mathbf{q}\big\| = r$, show that $\big|\boldsymbol{\kappa 2}[\boldsymbol{\beta}](t)\big| = 1/r$ for $a < t < b$.

4. Suppose that $\boldsymbol{\beta}\colon (a, b) \to \mathbb{R}^2$ is a unit-speed curve whose curvature is given by $\big|\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\big| = 1/r$ for all s, where $r > 0$ is a constant. Show that $\boldsymbol{\beta}$ is part of a circle of radius r centered at some point $\mathbf{q} \in \mathbb{R}^2$.

5. Verify the formula (1.14) for the curvature of a plane curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{C}$.

6. Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ be a regular curve which does not pass through the origin. Fix t_0 and choose ϕ_0 such that

$$\frac{\boldsymbol{\alpha}'(t_0) \cdot \boldsymbol{\alpha}(t_0)}{\big\|\boldsymbol{\alpha}'(t_0)\big\|\,\big\|\boldsymbol{\alpha}(t_0)\big\|} = \cos\phi_0 \quad \text{and} \quad \frac{\boldsymbol{\alpha}'(t_0) \cdot J\boldsymbol{\alpha}(t_0)}{\big\|\boldsymbol{\alpha}'(t_0)\big\|\,\big\|\boldsymbol{\alpha}(t_0)\big\|} = \sin\phi_0.$$

Establish the existence and uniqueness of a differentiable function $\boldsymbol{\phi} = \boldsymbol{\phi}[\boldsymbol{\alpha}]\colon (a, b) \to \mathbb{R}$ such that $\boldsymbol{\phi}(t_0) = \phi_0$,

$$\frac{\boldsymbol{\alpha}'(t) \cdot \boldsymbol{\alpha}(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|\,\big\|\boldsymbol{\alpha}(t)\big\|} = \cos\boldsymbol{\phi}(t) \quad \text{and} \quad \frac{\boldsymbol{\alpha}'(t) \cdot J\boldsymbol{\alpha}(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|\,\big\|\boldsymbol{\alpha}(t)\big\|} = \sin\boldsymbol{\phi}(t)$$

for $a < t < b$. Geometrically, $\boldsymbol{\phi}(t)$ represents the angle between the radius vector $\boldsymbol{\alpha}(t)$ and the tangent vector $\boldsymbol{\alpha}'(t)$.

7. Find a unit-speed parametrization of the semicubical parabola $t \mapsto (t^2,\ t^3)$, valid for $t > 0$.

8. Let $\boldsymbol{\gamma}_a \colon \mathbb{R} \to \mathbb{R}^2$ be the curve defined by

$$\boldsymbol{\gamma}_a(t) = \begin{cases} \left(t,\ |t|^a \sin\dfrac{1}{t}\right) & \text{if } t \neq 0, \\ (0,0) & \text{if } t = 0. \end{cases}$$

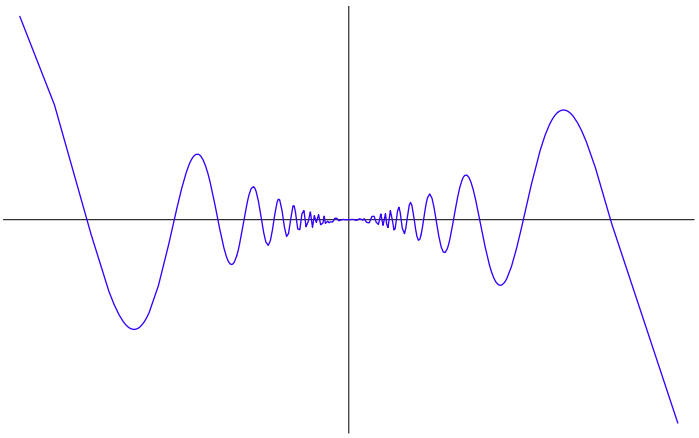


Figure 1.14: The curve $\boldsymbol{\gamma}_{3/2}$

(i) Show that $\boldsymbol{\gamma}_a$ is continuous if $a > 0$, but discontinuous if $a \leqslant 0$.

(ii) Show that $\boldsymbol{\gamma}_a$ is differentiable if $1 < a < 2$, but that the curve $(\boldsymbol{\gamma}_a)'$ is discontinuous.

Notebook 1

■ N1.0

This notebook explains how to use *Mathematica* to study curves in the plane. It permits easy access to a wide variety of curves for further study and experimentation. In the past, plane curves were drawn by ingenious but laborious methods that one can find, for example, in the book [Frost]. Computer methods allow us not only to reproduce any given curve quickly and accurately, but also to plot families of curves effortlessly.

As well as plotting curves, one can also use *Mathematica* to study them analytically. In this first notebook, we incorporate programs for computing the curvature and length of plane curves. Although one can compute the curvature of these curves symbolically, it is usually more informative to graph the curvature. The study of more complicated plane curves is relegated to Notebooks 2 and 3.

To illustrate a single curve effectively, it will often be necessary to add additional information to the overall plot, such as straight lines, arrows, and text. With this in mind, the first section explains how to define, combine and display individual graphics objects. The relevant commands form the basis of all the programs that follow, even if they are applied in a more sophisticated way in later notebooks.

■ N1.1

Ordered sets are represented using curly brackets in *Mathematica,* and can also be generated with the commands Table or Array. Vectors are expressed in this way, and can be added and multiplied by scalars:

```
v= {x,y,z}
Table[a[i],{i,3}]
w= Array[a,3]
2v+w
```

The printed versions of the notebooks display no output, but all the outputs can be inspected by evaluating cells in the electronic versions.

```
2v
Sin[v]
1+v
```

Other operators on vectors are also possible, and reflect the "listability" of many *Mathematica* functions. The next two products produce different outputs:

```
v w
v.w
```

We can use the dot product to define the norm of a vector. We shall accomplish this with a command denoted L (for "length") to avoid confusion with the built-in *Mathematica* function Norm. The definition of L also incorporates a simplification command:

```
L[v_]:= Sqrt[Simplify[v.v]]
```

This definition illustrates several important features of *Mathematica*. Square brackets are generally used to surround function arguments. The underscore is an important part of the syntax and v_ represents a generic vector, although its value on the right hand side crucially depends on whether the symbol = or := is used in the definition. The latter delays evaluation in order to use current values, which is often more appropriate. Had the colon been omitted, the norm of any future vector would always be output as $\sqrt{x^2 + y^2 + z^2}$, this being the norm of the particular vector v defined above.

The faint gray shading used in the previous cell will always accompany important definitions. Cells containing these form part of the initialization data that the user is invited to evaluate when the notebook is opened. Here is another example: the rotation operator J in the plane is defined by

```
J[{x_,y_}]:= {-y,x}
```

The form of this definition ensures that it can only be applied to vectors with two components. We can compute J squared as follows, the second and third lines demonstrating alternative ways of applying a function:

```
J[J[{x,y}]]
{x,y}//J//J
J@J@{x,y}
```

Mathematica plotting commands will be introduced in subsequent sections, but we make a start by introducing some of the so-called graphics primitives.

```
z1:= Graphics[Line[{{0,0},{4,2}}]]
z2:= Graphics[Point[{1,1}]]
z3:= Graphics[Circle[{1,1},.5]]
Show[z1,z2,z3]
```

Throughout the notebooks in this book, we shall often use z together with an index such as 1,2,3,... as a call symbol for a graphics object. This will enable us to list such objects separately, making it easier to modify and combine plots. For just a couple of objects, it will be easier to write z1 and z2, but when there are more it is more convenient to use automatic indexing of the form z[1],z[2],... The symbol z might occasionally be used on its own as an algebraic symbol, such as the third coordinate in space, but context will make the meaning clear.

The diagram can be improved by the use of graphics directives to emphasize each individual object, and graphics options with the command Show to improve overall features of the plot. In the present situation, one can equally well use = in place of := though semicolons are then needed to prevent echoing.

```
z[1]:= {Thickness[.01], Line[{{0,0},{4,2}}]}
z[2]:= {PointSize[.02], Point[{1,1}]}
z[3]:= {Hue[0], Circle[{1,1}, .5, {0,π}]}
z[4]:= Text["P", {1,1.1}]
Show[Graphics[Array[z,4]], Axes->True, AspectRatio->Automatic]
```

The Array command saves having to apply Graphics to each individual object. Setting the AspectRatio option corrects the scaling that is essential when displaying circles.

Graphics primitives will be used repeatedly to embellish the book's figures with points, lines, circles and text, and we shall carefully duplicate the programs used to draw each figure. However, these typically incorporate niceties that help improve the overall visual effect, and add nothing of substance; for example, a way of improving text can be found in the program for Figure 1.1 below. The code for more complicated figures is contained in cells, the first in N1.2.1, that are not visible in the printed book.

The next package is needed for Figure 1.1 and subsequent pictures. The command Arrow is then used much in the same way as Line above, but is only valid in 2 dimensions.

```
<<Graphics`Arrow`
```

In the program for Figure 1.1, graphics directives are added to thicken and color the vectors, and the TextStyle option in Show produces larger labels on the picture. The vector-valued function γ[t] represents a point of the unit circle, and is used to place arrows and text.

```
γ[t_]:= {Cos[t], Sin[t]}
z[1]:= {Hue[0], Thickness[.01], Arrow[{0,0}, γ[π/6]]}
z[2]:= {Hue[.7], Thickness[.01], Arrow[{0,0}, 1.5γ[π/3]]}
z[3]:= {Hue[0], Thickness[.01], Arrow[{0,0}, γ[2π/3]]}
z[4]:= Circle[{0,0}, 0.25, {π/6,π/3}]
z[5]:= {Text["θ", .3γ[π/4]], Text["q", 1.1γ[.53]],
  Text["p",1.6γ[1.05]], Text["Jq", 1.1γ[2]]}
ga:= Graphics[Array[z,5]]
Show[ga, Axes->True, Ticks->None, AspectRatio->Automatic,
  TextStyle->Section]
Clear[γ]
```

Figure 1.1

The last line saves us getting into difficulty when we try to redefine γ. The symbols z and ga are less problematic as they will not be used for functions.

■ N1.2

There are equivalent ways of computing derivatives of vector-valued functions, and these can be used to determine the velocity, speed and acceleration of a parametrized curve.

```
α[t_]:= {t, t^2, t^3}
D[α[t],t]
α'[u]
L[α'[t]]
```

```
D[α[t], {t,2}]
α''[t]
```

The chain rule and resulting Leibniz formula are illustrated next in various formats. The first line clears the previous definition of α.

```
Clear[α]
β[t_]:= α[f[t]]
Table[D[β[t],{t,k}], {k,4}]
ColumnForm[%]
Do[Print[D[β[t],{t,k}]], {k,5,6}]
```

Whilst ordinary functions of one variable are plotted with Plot, the main command in *Mathematica* for plotting plane curves is ParametricPlot.

```
α[t_]:={Cos[5t], Cos[7t]}
ParametricPlot[α[t]//Evaluate, {t,1,π}, AspectRatio->Automatic]
```

N1.2.1

N1.3

We begin by defining some simple curves:

```
circ[a_][t_]:= {a Cos[t], a Sin[t]}
parabola[a_][t_]:= {2a t, a t^2}
ellipse[a_,b_][t_]:= {a Cos[t], b Sin[t]}
```

Each is a function, depending on one or two parameters a,b that determine the curve's size or shape. The abbreviation circ is used to avoid a clash with *Mathematica*'s built-in command Circle.

It is more efficient to compute the derivative of the arc length function before computing the arc length itself. The command arclength[α] is an indefinite integral of arclengthprime[α].

```
arclengthprime[α_][t_]:= L[α'[t]]
arclength[α_][t_]:= Integrate[L[α'[tt]], tt]/.tt->t
```

We retain the notation of the book's first edition, whereby the symbol tt is reserved for a dummy variable. Here it is the variable of integration that, when complete, is

replaced by t; without its use, the computation of something like arclength[α][t^2] would fail. The command Module can be used to bind dummy variables, but for short definitions the above formalism is clearer. In practice, one also needs to add various simplification commands to obtain useful results.

```
Clear[a]
arclength[circ[a]][t]
%//Simplify//PowerExpand
```

```
arclength[parabola[a]][t];
%//Simplify//PowerExpand
```

The command PowerExpand converts $(ab)^s$ into a^s times b^s; sometimes this is correct mathematically, but sometimes it is not. For example, the square root of x^2 may not equal x, so PowerExpand must be used with care. To compute the length of a curve on some interval requires a definite integral, though the method may fail if L[α'[t]] is too complicated. A numerical value for the length can always be obtained using *Mathematica*'s numerical integrator.

```
length[a_,b_][α_]:= Integrate[L[α'[tt]], {tt,a,b},
  GenerateConditions->False]
lengthn[a_,b_][α_]:= NIntegrate[L[α'[tt]], {tt,a,b}]
```

The following example involves the semicubical parabola sc. The numerical computation tells us that a string stretched along the curve from sc[1] to sc[2] has length 7.63371.

```
sc[t_]:= {t^2, t^3}
```

```
length[a,b][sc]
%//PowerExpand
%//Simplify
```

```
lengthn[1,2][sc]
N[length[1,2][sc]]
```

The next figure uses the 3-dimensional plotting command ParametricPlot3D, but we shall not use this extensively until Notebook 7.

```
f[t_]:={Cos[t], Sin[t], t/3}
z1:= Graphics3D[{Hue[0], Thickness[.015],
  Line[Table[f[t], {t,0,4π,π/3}]]}]
ff[t_]:= Append[f[t], {Hue[.6], Thickness[.01]}]
z2:= ParametricPlot3D[ff[t], {t,0,4π}, DisplayFunction->Identity]
Show[z1, z2, DisplayFunction->$DisplayFunction]
```

Figure 1.3

The DisplayFunction options ensure that only the final combined plot is displayed. The Append command is used to define a new function ff representing a thickened 3-dimensional curve.

■ N1.4

The definition of the curvature of a plane curve translates easily into *Mathematica*. Once again, tt is used as a dummy variable. Examples follow.

```
κ2[α_][t_]:= α''[tt].J[α'[tt]]/L[α'[tt]]^3/.tt->t
```

```
κ2[circ[a]][t]
PowerExpand[Simplify[%]]
κ2[parabola[1]][t];
Simplify[%]
```

■ N1.4.1

■ N1.5

In this section, we display the relatively long program needed for Figure 1.6. The Module command is one of *Mathematica*'s ways of setting local variables; it ensures that the various definitions retain their validity solely for the course of the program. Definitions are usually given with = followed by a semicolon at the end of the line, though the colon before = is essential in line 5, so as to delay evaluation.

```
Module[{f,g,a,p,z},
  f[t_]= {1+Sin[1.3t], Sin[.7t]};
  g[t_]= 2{Cos[1.3t], Cos[.7t]} ;
  a= 2.8;
  p[f_,h_]:= ParametricPlot[f[t]//Evaluate, {t,.5,8},
    PlotStyle->{Hue[h], Thickness[.01]},
    DisplayFunction->Identity];
  z[1]= {Thickness[.007], Arrow[f[a], f[a]+f'[a]]};
  z[2]= {Thickness[.007], Arrow[g[a], g[a]+g'[a]]};
  z[3]= {Thickness[.007], Arrow[g[a], g[a]+f'[a]]};
  z[4]= {Text["α'(t)", f[a]+1.3f'[a]], Text["β'(t)", g[a]+1.15g'[a]]};
  z[5]= {Hue[.7], Thickness[.007], Circle[g[a], .2, {-.75,3.3}]};
  ga= Graphics[Array[z,5]];
  Show[p[f,.6], p[g,0], ga, AspectRatio->Automatic,
    Axes->False, PlotRange->All, TextStyle->Section,
    DisplayFunction->$DisplayFunction, ImageSize->450]
]
```

Figure 1.6

The whole thing can be replotted after changing the value of the parameter `a`, though `z[5]` requires a more sophisticated definition if the arc of the circle indicating the angle is to transform correctly.

■ N1.5.1

■ N1.6

The following program illustrates the power of the `Table` command.

```
cir[x_]:= {Hue[x/5], Circle[{x,0}, Abs[x]]}
family:= Table[Graphics[cir[x]], {x,-50,50}]
Show[family, AspectRatio->Automatic]
```

The logarithmic spiral is defined by

```
logspiral[a_,b_][t_]:= a Exp[b t]{Cos[t], Sin[t]}
```

The one plotted has `b=0.2`, and its curvature can be computed as follows:

```
α:= logspiral[1,.2]
PowerExpand[Simplify[κ2[α][t]]]
```

Open cells below show how to plot both the spiral and its curvature. The next few lines compute the curvature $\kappa 2$ of an ellipse, and then plot it:

```
ϵ:= ellipse[3,4]
κ2[ϵ][t]//Simplify
Plot[κ2[ϵ][t]//Evaluate,{t,0,2π}, PlotRange->{{0,2π},{-.2,.75}},
  AspectRatio->1/3, PlotStyle->{Thickness[.01], RGBColor[0,.5,0]},
  ImageSize->450]
```

Figure 1.9

The next program plots a spiral together with tangent and radial vector as a function of the variable a.

```
spiral[a_]:=
Module[{f,ep},
  f= logspiral[1,.2];
  ep= {Arrow[{0,0}, f[a]], Arrow[f[a], f[a]+f'[a]]};
  ParametricPlot[f[t]//Evaluate, {t,-4π,4π},
    PlotStyle->{Hue[.05],Thickness[.01]},
    AspectRatio->Automatic, Epilog->ep, ImageSize->300]
]
spiral[2]
```

Figure 1.10

A program of this sort can be used to make multiple plots. Clicking on one of the resulting graphics windows will produce an animation. Effects such as timing can be customized using the menu Edit->Preferences->Graphics Options.

```
Do[spiral[a/3], {a,20}]
```

The plot for Figure 1.11 is more straightforward.

```
Plot[κ2[logspiral[1,.2]][t], {t,-4π,4π}, ImageSize->400,
   PlotStyle->{Thickness[.01], RGBColor[0,.5,0]}]
```

Figure 1.11

■ N1.7

The semicubical parabola sc was defined in N1.3, and is now plotted.

```
ParametricPlot[sc[t]//Evaluate, {t,-1,1},
  PlotStyle->{Thickness[.01],Hue[.7]},
  PlotRange->{{0,1.2},{-1,1.1}}, AspectRatio->Automatic,
  Ticks->{{.5,1},Automatic}, ImageSize->250]
```

Figure 1.12

■ N1.8

The following curves from Exercise 1 are plotted in N1.8.1.

```
circleinvolute[a_][t_]:= circ[a][t] - t circ[a]'[t]
catenary[a_][t_]:= {a Cosh[t/a], t}
astroid[a_][t_]:= a{Cos[t]^3, Sin[t]^3}
```

Here the Table command is combined with ParametricPlot:

```
p[a_,b_,h_]:= ParametricPlot[logspiral[a,b][t]//Evaluate,
  {t,-4π,4π}, PlotStyle->{Hue[h], Thickness[.01]},
  AspectRatio->Automatic, DisplayFunction->Identity]
```

```
Show[Table[p[n/10,.2,.1], {n,20,35}],
  Table[p[-n/10,.2,.7], {n,20,35}],
  PlotRange->{{-20,20}, {-20,20}},
  DisplayFunction->$DisplayFunction]
```

■ N1.8.1

Chapter 2

Famous Plane Curves

Plane curves have been a subject of much interest beginning with the Greeks. Both physical and geometric problems frequently lead to curves other than ellipses, parabolas and hyperbolas. The literature on plane curves is extensive. Diocles studied the *cissoid* in connection with the classic problem of doubling the cube. Newton[1] gave a classification of cubic curves (see [Newton] or [BrKn]). Mathematicians from Fermat to Cayley often had curves named after them. In this chapter, we illustrate a number of historically-interesting plane curves.

Cycloids are discussed in Section 2.1, *lemniscates of Bernoulli* in Section 2.2 and *cardioids* in Section 2.3. Then in Section 2.4 we derive the differential equation for a *catenary*, a curve that at first sight resembles the parabola. We shall present a geometrical account of the *cissoid of Diocles* in Section 2.5, and an analysis of the *tractrix* in Section 2.6. Section 2.7 is devoted to an illustration of *clothoids*, though the significance of these curves will become



[1] Sir Isaac Newton (1642–1727). English mathematician, physicist, and astronomer. Newton's contributions to mathematics encompass not only his fundamental work on calculus and his discovery of the binomial theorem for negative and fractional exponents, but also substantial work in algebra, number theory, classical and analytic geometry, methods of computation and approximation, and probability. His classification of cubic curves was published as an appendix to his book on optics; his work in analytic geometry included the introduction of polar coordinates.

As Lucasian professor at Cambridge, Newton was required to lecture once a week, but his lectures were so abstruse that he frequently had no auditors. Twice elected as a member of Parliament for the university, Newton was appointed warden of the mint; his knighthood was awarded primarily for his services in that role. In *Philosophiæ Naturalis Principia Mathematica*, Newton set forth fundamental mathematical principles and then applied them to the development of a world system. This is the basis of the Newtonian physics that determined how the universe was perceived until the twentieth-century work of Einstein.

clearer in Chapter 5. Finally, *pursuit curves* are discussed in Section 2.8.

There are many books on plane curves. Four excellent classical books are those of Cesàro [Ces], Gomes Teixeira[2] [Gomes], Loria[3] [Loria1], and Wieleitner [Wiel2]. In addition, Struik's book [Stru2] contains much useful information, both theoretical and historical. Modern books on curves include [Arg], [BrGi], [Law], [Lock], [Sav], [Shikin], [vonSeg], [Yates] and [Zwi].

2.1 Cycloids

The general cycloid is defined by

$$\mathsf{cycloid}[a,b](t) = \big(a\,t - b\sin t,\ a - b\cos t\big).$$

Taking $a = b$ gives $\mathsf{cycloid}[a,a]$, which describes the locus of points traced by a point on a circle of radius a which rolls without slipping along a straight line.

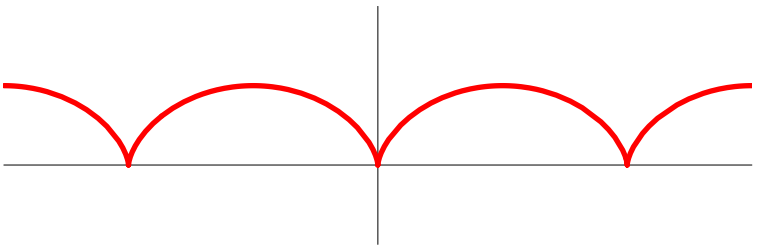


Figure 2.1: The cycloid $t \mapsto (t - \sin t,\ 1 - \cos t)$

In Figure 2.2, we plot the graph of the curvature $\boldsymbol{\kappa 2}(t)$ of $\mathsf{cycloid}[1,1](t)$ over the range $0 \leqslant t \leqslant 2\pi$. Note that the horizontal axis represents the variable t, and not the x-coordinate $t - \sin t$ of Figure 2.1. A more faithful representation of the curvature function is obtained by expressing the parameter t as a function of x. In Figure 2.3, the curvature of a given point $(x(t), y(t))$ of $\mathsf{cycloid}[1, \frac{1}{2}]$ is represented by the point $(x(t), \boldsymbol{\kappa 2}(t))$ on the graph vertically above or below it.

---

2 

Francisco Gomes Teixeira (1851–1933). A leading Portuguese mathematician of the last half of the 19th century. There is a statue of him in Porto.

3 

Gino Loria (1862–1954). Italian mathematician, professor at the University of Genoa. In addition to his books on curve theory, [Loria1] and [Loria2], Loria wrote several books on the history of mathematics.

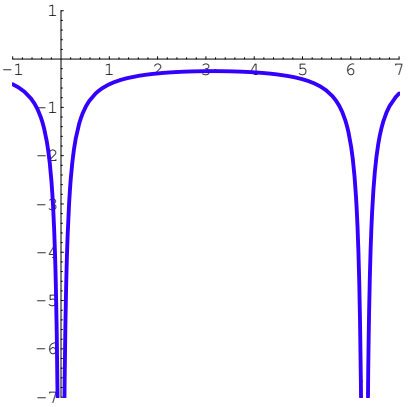


Figure 2.2: Curvature of cycloid$[1, 1]$

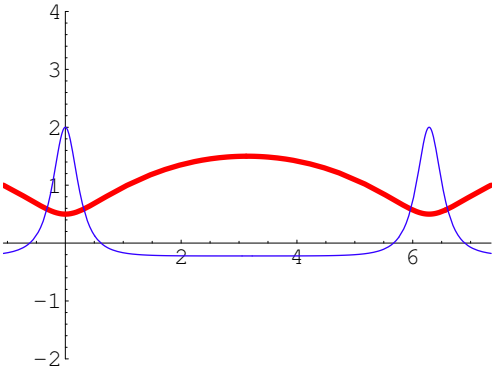


Figure 2.3: cycloid$[1, \frac{1}{2}]$ together with its curvature

Consider now the case in which a and b are not necessarily equal. The curve cycloid$[a, b]$ is that traced by a point on a circle of radius b when a concentric circle of radius a rolls without slipping along a straight line. The cycloid is ***prolate*** if $a < b$ and ***curtate*** if $a > b$.

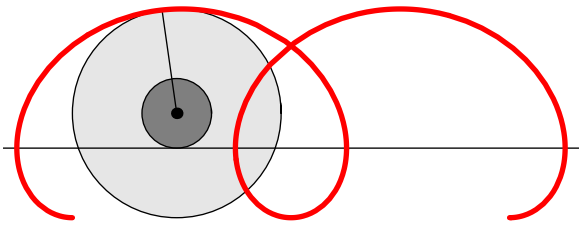


Figure 2.4: The prolate cycloid $t \mapsto (t - 3\sin t,\ 1 - 3\cos t)$

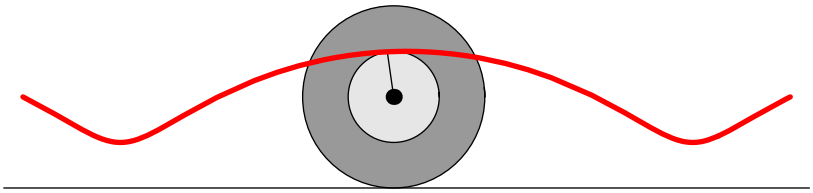


Figure 2.5: The curtate cycloid $t \mapsto (2t - \sin t,\ 2 - \cos t)$

The final figure in this section demonstrates the fact that the tangent and normal to the cycloid always intersect the vertical diameter of the generating circle on the circle itself. A discussion of this and other properties of cycloids can be found in [Lem, Chapter 4] and [Wagon, Chapter 2].

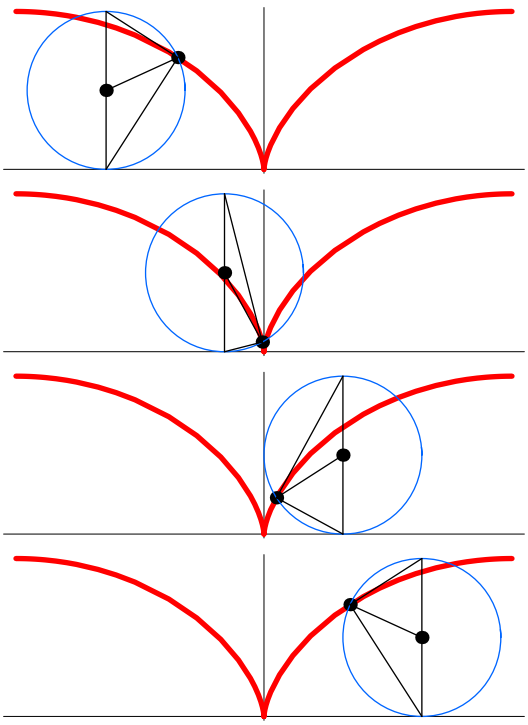


Figure 2.6: Properties of tangent and normal

Instructions for animating Figures 2.4–2.6 are given in Notebook 2.

2.2 Lemniscates of Bernoulli

Each curve in the family

$$\mathsf{lemniscate}[a](t) = \left(\frac{a\cos t}{1+\sin^2 t}, \frac{a\sin t\cos t}{1+\sin^2 t} \right) \tag{2.1}$$

is called a ***lemniscate of Bernoulli***.[4] Like an ellipse, a lemniscate has foci F_1 and F_2, but the lemniscate is the locus $\mathscr{L}$ of points P for which the *product* of distances from F_1 and F_2 is a certain constant f^2. More precisely,

$$\mathscr{L} = \{\ (x,y)\ \mid \mathsf{distance}\big((x,y),F_1\big)\,\mathsf{distance}\big((x,y),F_2\big) = f^2\ \},$$

where $\mathsf{distance}(F_1,F_2) = 2f$. This choice ensures that the midpoint of the segment connecting F_1 with F_2 lies on the curve $\mathscr{L}$.

Let us derive (2.1) from the focal property. Let the foci be $(\pm f, 0)$ and let $\mathscr{L}$ be a set of points containing $(0,0)$ such that the product of the distances from $F_1 = (-f,0)$ and $F_2 = (f,0)$ is the same for all $P \in \mathscr{L}$. Write $P = (x,y)$. Then the condition that P lie on $\mathscr{L}$ is

$$\big((x-f)^2+y^2\big)\big((x+f)^2+y^2\big) = f^4, \tag{2.2}$$

or equivalently

$$(x^2+y^2)^2 = 2f^2(x^2-y^2). \tag{2.3}$$

The substitutions $y = x\sin t$, $f = a/\sqrt{2}$ transform (2.3) into (2.1).

Figure 2.7 displays four lemniscates. Starting from the largest, each successively smaller one passes through the foci of the previous one. Figure 2.8 plots the curvature of one of them as a function of the x-coordinate.

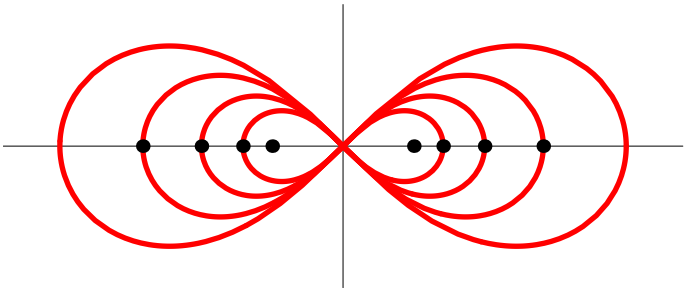


Figure 2.7: A family of lemniscates



[4] Jakob Bernoulli (1654–1705). Jakob and his brother Johann were the first of a Swiss mathematical dynasty. The work of the Bernoullis was instrumental in establishing the dominance of Leibniz's methods of exposition. Jakob Bernoulli laid basic foundations for the development of the calculus of variations, as well as working in probability, astronomy and mathematical physics. In 1694 Bernoulli studied the lemniscate named after him.

Although a lemniscate of Bernoulli resembles the figure eight curve parametrized in the simper way by

$$\mathsf{eight}[a](t) = \big(\sin t,\ \sin t \cos t\big), \tag{2.4}$$

a comparison of the graphs of their respective curvatures shows the difference between the two curves: the curvature of a lemniscate has only one maximum and one minimum in the range $0 \leqslant t < 2\pi$, whereas that of (2.4) has three maxima, three minima, and two inflection points.

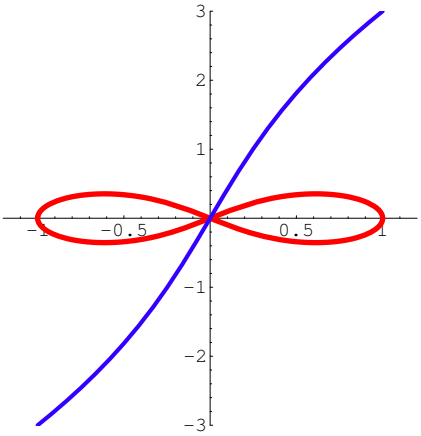


Figure 2.8: A lemniscate and its curvature

In Section 6.1, we shall define the ***total signed curvature*** by integrating $\boldsymbol{\kappa 2}$ through a full turn. It is zero for both (2.1) and (2.4); for the former, this follows because the curvature graphed in Figure 2.8 is an odd function.

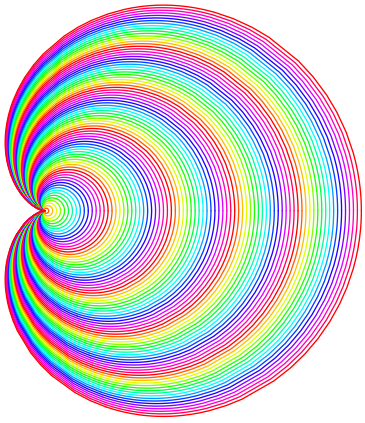


Figure 2.9: A family of cardioids

2.3 Cardioids

A ***cardioid*** is the locus of points traced out by a point on a circle of radius a which rolls without slipping on another circle of the same radius a. Its parametric equation is

$$\mathsf{cardioid}[a](t) = \big(2a(1+\cos t)\cos t,\ 2a(1+\cos t)\sin t\big).$$

Figure 2.9 shows the effect of varying a, which acts as an overall scaling factor.

The curvature of the cardioid can be simplified by hand to get

$$\boldsymbol{\kappa 2}[\mathsf{cardioid}[a]](t) = \frac{3}{8|a\cos(t/2)|}.$$

The result is illustrated below using the same method as in Figures 2.3 and 2.8. The curvature consists of two branches that meet at one of two points where the value of $\boldsymbol{\kappa 2}$ coincides with the cardioid's y-coordinate.

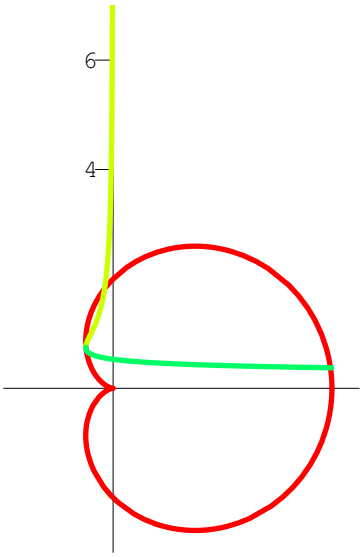


Figure 2.10: A cardioid and its curvature

2.4 The Catenary

In 1691, Jakob Bernoulli gave a solution to the problem of finding the curve assumed by a flexible inextensible cord hung freely from two fixed points; Leibniz has called such a curve a ***catenary*** (which stems from the Latin word *catena*, meaning chain). The solution is based on the differential equation

$$\frac{d^2y}{dx^2} = \frac{1}{a}\sqrt{1+\left(\frac{dy}{dx}\right)^2}. \tag{2.5}$$

To derive (2.5), we consider a portion $\overline{\mathbf{pq}}$ of the cable between the lowest point $\mathbf{p}$ and an arbitrary point $\mathbf{q}$. Three forces act on the cable: the weight of the portion $\overline{\mathbf{pq}}$, as well as the tensions $\mathbf{T}$ and $\mathbf{U}$ at $\mathbf{p}$ and $\mathbf{q}$. If w is the linear density and s is the length of $\overline{\mathbf{pq}}$, then the weight of the portion $\overline{\mathbf{pq}}$ is ws.

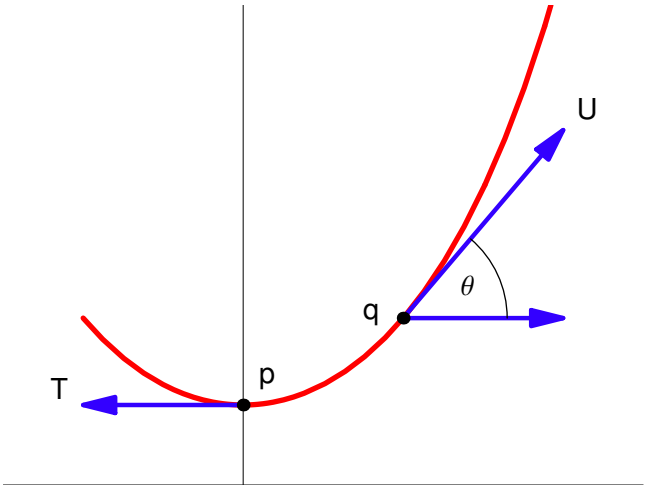


Figure 2.11: Definition of a catenary

Let $|\mathbf{T}|$ and $|\mathbf{U}|$ denote the magnitudes of the forces $\mathbf{T}$ and $\mathbf{U}$, and write

$$\mathbf{U} = \big(|\mathbf{U}|\cos\theta,\ |\mathbf{U}|\sin\theta\big),$$

with θ the angle shown in Figure 2.11. Because of equilibrium we have

$$|\mathbf{T}| = |\mathbf{U}|\cos\theta \qquad \text{and} \qquad w\,s = |\mathbf{U}|\sin\theta. \tag{2.6}$$

Let $\mathbf{q} = (x, y)$, where x and y are functions of s. From (2.6) we obtain

$$\frac{dy}{dx} = \tan\theta = \frac{w\,s}{|\mathbf{T}|}. \tag{2.7}$$

Since the length of $\overline{\mathbf{pq}}$ is

$$s = \int_0^x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx,$$

the fundamental theorem of calculus tells us that

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}. \tag{2.8}$$

When we differentiate (2.7) and use (2.8), we get (2.5) with $a = w/|T|$.

Although at first glance the catenary looks like a parabola, it is in fact the graph of the hyperbolic cosine. A solution of (2.5) is given by

$$y = a\cosh\frac{x}{a}. \tag{2.9}$$

The next figure compares a catenary and a parabola having the same curvature at $x = 0$. The reader may need to refer to Notebook 2 to see which is which.

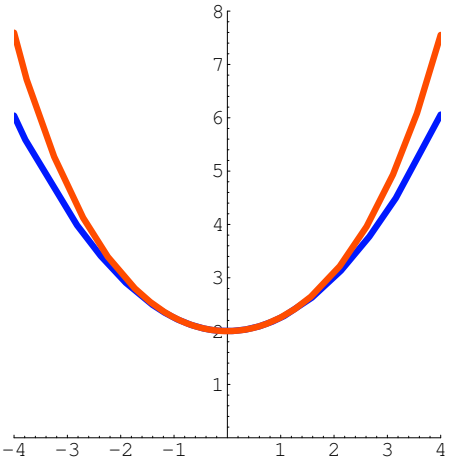


Figure 2.12: Catenary and parabola

We rotate the graph of (2.9) to define

$$\mathsf{catenary}[a](t) = \left(a\cosh\frac{t}{a},\ t\right), \tag{2.10}$$

where without loss of generality, we assume that $a > 0$. A catenary is one of the few curves for which it is easy to compute the arc length function. Indeed,

$$|\mathsf{catenary}[a]'(t)| = \cosh\frac{t}{a},$$

and it follows that a unit-speed parametrization of a catenary is given by

$$s \mapsto a\left(\sqrt{1+\frac{s^2}{a^2}},\ \operatorname{arcsinh}\frac{s}{a}\right).$$

In Chapter 15, we shall use the catenary to construct an important minimal surface called the ***catenoid***.

2.5 The Cissoid of Diocles

The ***cissoid of Diocles*** is the curve defined nonparametrically by

$$x^3 + xy^2 - 2ay^2 = 0, \tag{2.11}$$

where $a \neq 0$. To find a parametrization of the cissoid, we substitute $y = xt$ in (2.11) and obtain

$$x = \frac{2at^2}{1+t^2}, \qquad y = \frac{2at^3}{1+t^2}.$$

Thus we define

$$\mathsf{cissoid}[a](t) = \left(\frac{2at^2}{1+t^2}, \frac{2at^3}{1+t^2}\right). \tag{2.12}$$

The Greeks used the cissoid of Diocles to try to find solutions to the problems of doubling a cube and trisecting an angle. For more historical information and the definitions used by the Greeks and Newton see [BrKn, pages 9–12], [Gomes, volume 1, pages 1–25] and [Lock, pages 130–133]. Cissoid means "ivy-shaped." Observe that $\mathsf{cissoid}[a]'(0) = 0$ so that $\mathsf{cissoid}$ is not regular at 0. In fact, a cissoid has a cusp at 0, as can be seen in Figure 2.14.

The definition of the cissoid used by the Greeks and by Newton can best be explained by considering a generalization of the cissoid. Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be two curves and A a fixed point. Draw a line ℓ through A cutting $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ at points Q and R respectively, and find a point P on ℓ such that the distance from A to P equals the distance from Q to R. The locus of such points P is called the ***cissoid of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with respect to*** A.

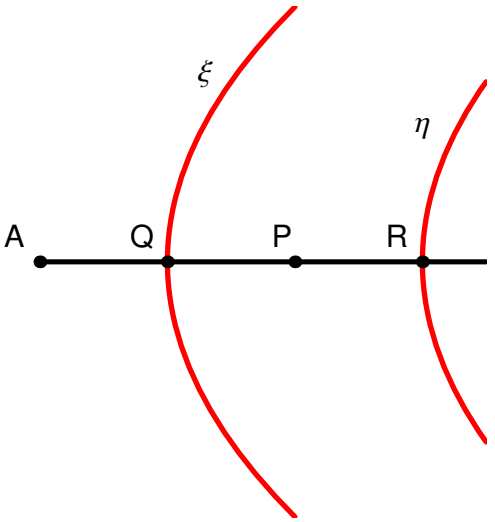


Figure 2.13: The cissoid of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with respect to the point A

Then $\mathsf{cissoid}[a]$ is precisely the cissoid of a circle of radius a and one of its tangent lines with respect to the point diametrically opposite to the tangent line, as in Figure 2.14. Let us derive (2.11). Consider a circle of radius a centered at $(a, 0)$. Let (x, y) be the coordinates of a point P on the cissoid. Then $2a = \mathsf{distance}(A, S)$, so by the Pythagorean theorem we have

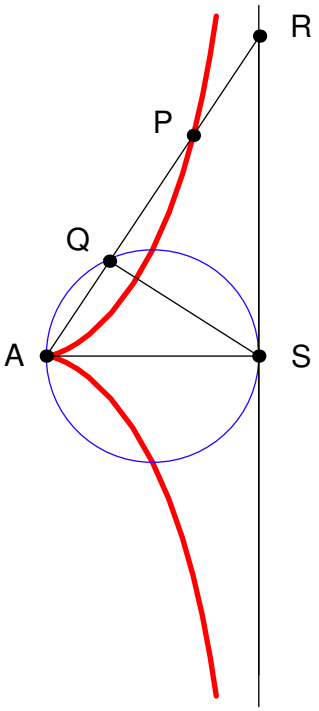


Figure 2.14: AR moves so that $\mathsf{distance}(A,P) = \mathsf{distance}(Q,R)$

$$\begin{aligned} \mathsf{distance}(Q,S)^2 &= \mathsf{distance}(A,S)^2 - \mathsf{distance}(A,Q)^2 \\ &= 4a^2 - \big(\mathsf{distance}(A,R) - \mathsf{distance}(Q,R)\big)^2. \end{aligned} \tag{2.13}$$

The definition of the cissoid says that

$$\mathsf{distance}(Q,R) = \mathsf{distance}(A,P) = \sqrt{x^2+y^2}.$$

By similar triangles, $\mathsf{distance}(A,R)/(2a) = \mathsf{distance}(A,P)/x$. Therefore, (2.13) becomes

$$\mathsf{distance}(Q,S)^2 = 4a^2 - \left(\frac{2a}{x} - 1\right)^2 (x^2+y^2). \tag{2.14}$$

On the other hand,

$$\begin{aligned} \mathsf{distance}(Q,S)^2 &= \mathsf{distance}(R,S)^2 - \mathsf{distance}(Q,R)^2 \\ &= \mathsf{distance}(R,S)^2 - \mathsf{distance}(A,P)^2 \\ &= \left(\frac{2ay}{x}\right)^2 - (x^2+y^2), \end{aligned} \tag{2.15}$$

since by similar triangles, $\mathsf{distance}(R,S)/(2a) = y/x$. Then (2.11) follows easily by equating the right-hand sides of (2.14) and (2.15).

From Notebook 2, the curvature of the cissoid is given by

$$\boldsymbol{\kappa 2}[\mathsf{cissoid}[a]](t) = \frac{3}{a|t|(4+t^2)^{3/2}},$$

and is everywhere strictly positive.

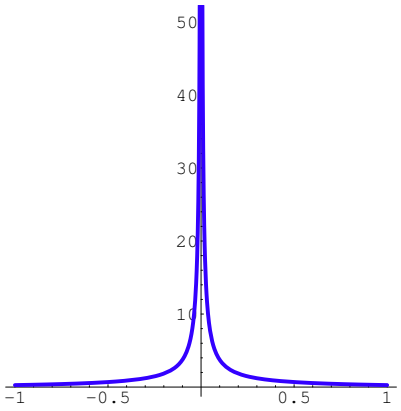


Figure 2.15: Curvature of the cissoid

2.6 The Tractrix

A ***tractrix*** is a curve $\boldsymbol{\alpha}$ passing through the point $A = (a, 0)$ on the horizontal axis with the property that the length of the segment of the tangent line from any point on the curve to the vertical axis is constant, as shown in Figure 2.16.

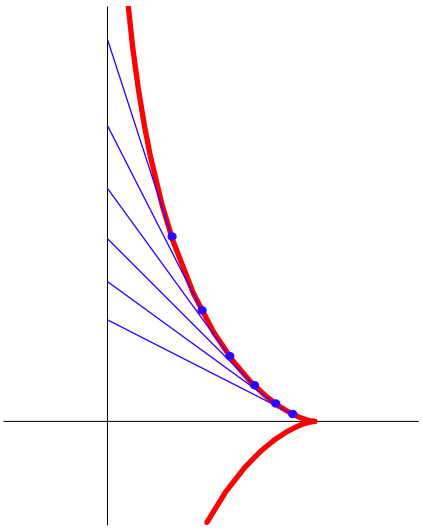


Figure 2.16: Tangent segments have equal lengths

The German word for tractrix is the more descriptive *Hundekurve*. It is the path that an obstinate dog takes when his master walks along a north-south path. One way to parametrize the curve is by means of

$$\text{tractrix}[a](t) = a\Big(\sin t,\ \cos t + \log\big(\tan\tfrac{t}{2}\big)\Big)\,. \tag{2.16}$$

It approaches the vertical axis asymptotically as $t \to 0$ or $t \to \pi$, and has a cusp at $t = \pi/2$.

To find the differential equation of a tractrix, write $\text{tractrix}[a](t) = (x(t), y(t))$. Then dy/dx is the slope of the curve, and the differential equation is therefore

$$\frac{dy}{dx} = -\frac{\sqrt{a^2 - x^2}}{x}. \tag{2.17}$$

It can be checked (with the help of (2.19) overleaf) that the differential equation is satisfied by the components $x(t) = a\sin t$ and $y(t) = a\big(\cos t + \log(\tan(t/2))\big)$ of (2.16), and that both sides of (2.17) equal $\cot t$.

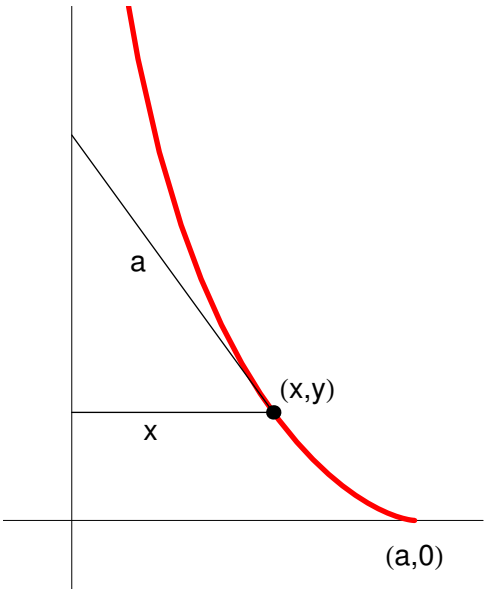


Figure 2.17: Finding the differential equation

The curvature of the tractrix is

$$\boldsymbol{\kappa 2}[\text{tractrix}[1]](t) = -|\tan t\,|.$$

In particular, it is everywhere negative and approaches $-\infty$ at the cusp. Figure 2.18 graphs $\boldsymbol{\kappa 2}$ as a function of the x-coordinate $\sin t$, so that the curvature at a given point on the tractrix is found by referring to the point on the curvature graph vertically below.

For future use let us record

Lemma 2.1. *A unit-speed parametrization of the tractrix is given by*

$$\boldsymbol{\alpha}(s) = \begin{cases} \left(ae^{-s/a},\ \displaystyle\int_0^s \sqrt{1-e^{-2t/a}}\,dt\right) & \text{for } 0 \leqslant s < \infty, \\ \left(ae^{s/a},\ \displaystyle\int_0^s \sqrt{1-e^{2t/a}}\,dt\right) & \text{for } -\infty < s < 0. \end{cases} \tag{2.18}$$

Note that

$$\int_0^s \sqrt{1-e^{-2t/a}}\,dt = a\,\operatorname{arctanh}\left(\sqrt{1-e^{-2s/a}}\right) - a\sqrt{1-e^{-2s/a}}.$$

Proof. First, we compute

$$\text{tractrix}[a]'(\phi) = a\left(\cos\phi,\ -\sin\phi + \frac{1}{\sin\phi}\right). \tag{2.19}$$

Define $\phi(s) = \pi - \arcsin(e^{-s/a})$ for $s \geqslant 0$. Then $\sin\phi(s) = e^{-s/a}$; furthermore, $\pi/2 \leqslant \phi(s) < \pi$ for $s \geqslant 0$, so that $\cos\phi(s) = -\sqrt{1-e^{-2s/a}}$. Hence

$$\phi'(s) = \frac{e^{-s/a}}{a\sqrt{1-e^{-2s/a}}} = -\frac{\sin\phi(s)}{a\cos\phi(s)}. \tag{2.20}$$

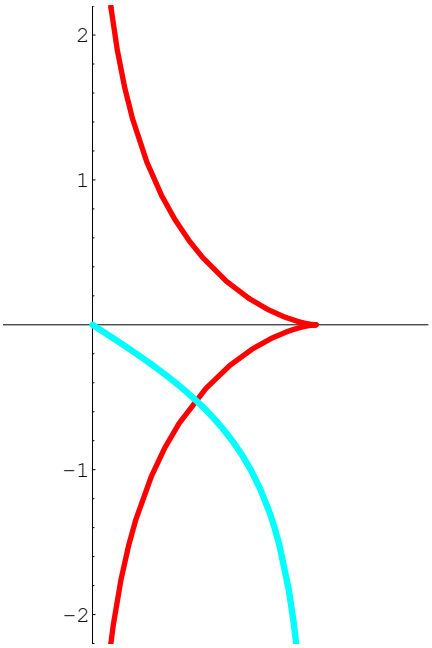


Figure 2.18: A tractrix and its curvature

Therefore, if we define a curve $\boldsymbol{\beta}$ by $\boldsymbol{\beta}(s) = \mathsf{tractrix}[a]\big(\phi(s)\big)$, it follows from (2.19) and (2.20) that

$$\begin{aligned}\boldsymbol{\beta}'(s) &= \mathsf{tractrix}[a]'\big(\phi(s)\big)\phi'(s)\\ &= a\Big(\cos\phi(s), -\sin\phi(s) + \frac{1}{\sin\phi(s)}\Big)\Big(-\frac{\sin\phi(s)}{a\cos\phi(s)}\Big)\\ &= \big(-\sin\phi(s), -\cos\phi(s)\big) = \Big(-e^{-s/a}, \sqrt{1-e^{-2s/a}}\Big) = \boldsymbol{\alpha}'(s).\end{aligned}$$

Also, $\boldsymbol{\beta}(0) = (a,0) = \boldsymbol{\alpha}(0)$. Thus $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ coincide for $0 \leqslant s < \infty$, so that $\boldsymbol{\alpha}$ is a reparametrization of a tractrix in that range. The proof that $\boldsymbol{\alpha}$ is a reparametrization of $\mathsf{tractrix}[a]$ for $-\infty < s \leqslant 0$ is similar. Finally, an easy calculation shows that $\boldsymbol{\alpha}$ has unit speed. ∎

2.7 Clothoids

One of the most elegant of all plane curves is the ***clothoid*** or ***spiral of Cornu***.[5] We give a generalization of the clothoid by defining

$$\mathsf{clothoid}[n,a](t) = a\Big(\int_0^t \sin\Big(\frac{u^{n+1}}{n+1}\Big)du,\ \int_0^t \cos\Big(\frac{u^{n+1}}{n+1}\Big)du\Big).$$

Clothoids are important curves used in freeway and railroad construction (see [Higg] and [Roth]). For example, a clothoid is needed to make the gradual transition from a highway, which has zero curvature, to the midpoint of a freeway exit, which has nonzero curvature. A clothoid is clearly preferable to a path consisting of straight lines and circles, for which the curvature is discontinuous.

The standard clothoid is $\mathsf{clothoid}[1,a]$, represented by the larger one of two plotted in Figure 2.19. The quantities

$$\int_0^t \sin\Big(\frac{\pi u^2}{2}\Big)du \qquad \text{and} \qquad \int_0^t \cos\Big(\frac{\pi u^2}{2}\Big)du$$

are called ***Fresnel***[6] ***integrals***; $\mathsf{clothoid}[1,a]$ is expressible in terms of them. Since

$$\int_0^{\pm\infty} \sin\Big(\frac{u^2}{2}\Big)du = \int_0^{\pm\infty} \cos\Big(\frac{u^2}{2}\Big)du = \pm\frac{\sqrt{\pi}}{2}$$

[5] Marie Alfred Cornu (1841–1902). French scientist, who studied the clothoid in connection with diffraction. The clothoid was also known to Euler and Jakob Bernoulli. See [Gomes, volume 2, page 102–107] and [Law, page 190].

[6] 

Augustin Jean Fresnel (1788–1827). French physicist, one of the founders of the wave theory of light.

(as is easily checked by computer), the ends of the clothoid$[1, a]$ curl around the points $\pm\frac{1}{2}a\sqrt{\pi}(1, 1)$. The first clothoid is symmetric with respect to the origin, but the second one (smaller in Figure 2.19) is symmetric with respect to the horizontal axis. The odd clothoids have shapes similar to clothoid$[1, a]$, while the even clothoids have shapes similar to clothoid$[2, a]$.

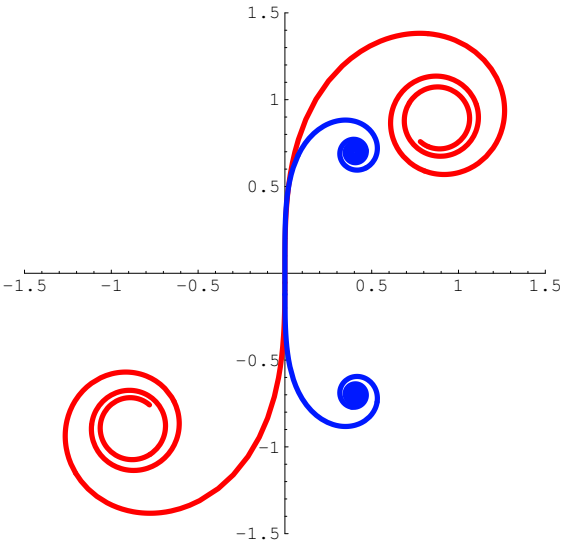


Figure 2.19: clothoid$[1, 1]$ and clothoid$[2, \frac{1}{2}]$

Although the definition of clothoid$[n, a]$ is quite complicated, its curvature is simple:

$$\boldsymbol{\kappa}\mathbf{2}[\text{clothoid}[n, a]](t) = -\frac{t^n}{a}. \tag{2.21}$$

In Chapter 5, we shall show how to define the clothoid as a numerical solution to a differential equation arising from (2.21).

2.8 Pursuit Curves

The problem of pursuit probably originated with Leonardo da Vinci. It is to find the curve by which a vessel moves while pursuing another vessel, supposing that the speeds of the two vessels are always in the same ratio. Let us formulate this problem mathematically.

Definition 2.2. *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be plane curves parametrized on an interval $[a, b]$. We say that $\boldsymbol{\alpha}$* ***is a pursuit curve of*** *$\boldsymbol{\beta}$ provided that*

(i) *the velocity vector $\boldsymbol{\alpha}'(t)$ points towards the point $\boldsymbol{\beta}(t)$ for $a < t < b$; that is, $\boldsymbol{\alpha}'(t)$ is a multiple of $\boldsymbol{\alpha}(t) - \boldsymbol{\beta}(t)$;*

(ii) *the speeds of* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *are related by* $\|\boldsymbol{\alpha}'\| = k\|\boldsymbol{\beta}'\|$, *where* k *is a positive constant. We call* k *the* ***speed ratio***.

A ***capture point*** *is a point* $\mathbf{p}$ *for which* $\mathbf{p} = \boldsymbol{\alpha}(t_1) = \boldsymbol{\beta}(t_1)$ *for some* t_1.

In Figure 2.20, $\boldsymbol{\alpha}$ is the curve of the pursuer and $\boldsymbol{\beta}$ the curve of the pursued.

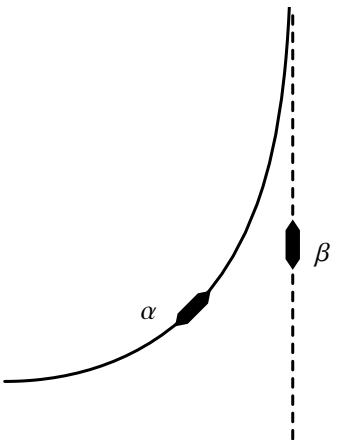


Figure 2.20: A pursuit curve

When the speed ratio k is larger than 1, the pursuer travels faster than the pursued. Although this would usually be the case in a physical situation, it is not a necessary assumption for the mathematical analysis of the problem.

We derive differential equations for pursuit curves in terms of coordinates.

Lemma 2.3. *Write* $\boldsymbol{\alpha} = (x, y)$ *and* $\boldsymbol{\beta} = (f, g)$, *and assume that* $\boldsymbol{\alpha}$ *is a pursuit curve of* $\boldsymbol{\beta}$. *Then*

$$x'^2 + y'^2 = k^2(f'^2 + g'^2) \tag{2.22}$$

and

$$x'(y - g) - y'(x - f) = 0. \tag{2.23}$$

Proof. Equation (2.22) is the same as $\|\boldsymbol{\alpha}'\| = k\|\boldsymbol{\beta}'\|$. To prove (2.23), we observe that $\boldsymbol{\alpha}(t) - \boldsymbol{\beta}(t) = \big(x(t) - f(t), y(t) - g(t)\big)$ and $\boldsymbol{\alpha}'(t) = \big(x'(t), y'(t)\big)$. Note that the vector $J\big(\boldsymbol{\alpha}(t)-\boldsymbol{\beta}(t)\big) = \big(-y(t)+g(t), x(t)-f(t)\big)$ is perpendicular to $\boldsymbol{\alpha}(t)-\boldsymbol{\beta}(t)$. The condition that $\boldsymbol{\alpha}'(t)$ is a multiple of $\boldsymbol{\alpha}(t)-\boldsymbol{\beta}(t)$ is conveniently expressed by saying that $\boldsymbol{\alpha}'(t)$ is perpendicular to $J\big(\boldsymbol{\alpha}(t) - \boldsymbol{\beta}(t)\big)$, which is equivalent to (2.23). ∎

Next, we specialize to the case when the curve of the pursued is a straight line. Assume that the curve $\boldsymbol{\beta}$ of the pursued is a vertical straight line passing through the point $(a, 0)$, and that the speed ratio k is larger than 1. We want to find the curve $\boldsymbol{\alpha}$ of the pursuer, assuming the initial conditions $\boldsymbol{\alpha}(0) = (0, 0)$ and $\boldsymbol{\alpha}'(0) = (1, 0)$.

We can parametrize $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta}(t) = \big(a,\, g(t)\big).$$

Furthermore, the curve $\boldsymbol{\alpha}$ of the pursuer can be parametrized as

$$\boldsymbol{\alpha}(t) = \big(t,\, y(t)\big).$$

The condition (2.22) becomes

$$1 + y'^2 = k^2 g'^2, \tag{2.24}$$

and (2.23) reduces to

$$(y - g) - y'(t - a) = 0.$$

Differentiation with respect to t yields

$$-y''(t - a) = g'. \tag{2.25}$$

From (2.24) and (2.25) we get

$$1 + y'^2 = k^2 (a - t)^2 y''^2. \tag{2.26}$$

Let $p = y'$; then (2.26) can be rewritten as

$$\frac{k\,dp}{\sqrt{1 + p^2}} = \frac{dt}{a - t}.$$

This separable first-order equation has the solution

$$\operatorname{arcsinh} p = -\frac{1}{k} \log\left(\frac{a - t}{a}\right), \tag{2.27}$$

when we make use of the initial condition $y'(0) = 0$. Then (2.27) can be rewritten as

$$\begin{aligned} y' \;=\; p \;=\; \sinh\big(\operatorname{arcsinh} p\big) \;&=\; \tfrac{1}{2}\left(e^{\operatorname{arcsinh} p} - e^{-\operatorname{arcsinh} p}\right) \\ &=\; \frac{1}{2}\left(\left(\frac{a - t}{a}\right)^{-1/k} - \left(\frac{a - t}{a}\right)^{1/k}\right). \end{aligned}$$

Integrating, with the initial condition $y(0) = 0$, yields

$$y = \frac{a\,k}{k^2 - 1} + \frac{1}{2}\left(\frac{a\,k}{k + 1}\left(\frac{a - t}{a}\right)^{1 + 1/k} - \frac{a\,k}{k - 1}\left(\frac{a - t}{a}\right)^{1 - 1/k}\right). \tag{2.28}$$

The curve of the pursuer is then $\boldsymbol{\alpha}(t) = (t,\, y(t))$, where y is given by (2.28). Since $\boldsymbol{\alpha}(t_1) = \boldsymbol{\beta}(t_1)$ if and only if $t_1 = a$, the capture point is

$$\mathbf{p} = \left(a,\, \frac{a\,k}{k^2 - 1}\right). \tag{2.29}$$

The graph below depicts the case when $a = 1$ and k has the values $2, 3, 4, 5$. As the speed ratio k becomes smaller and smaller, the capture point goes higher and higher.

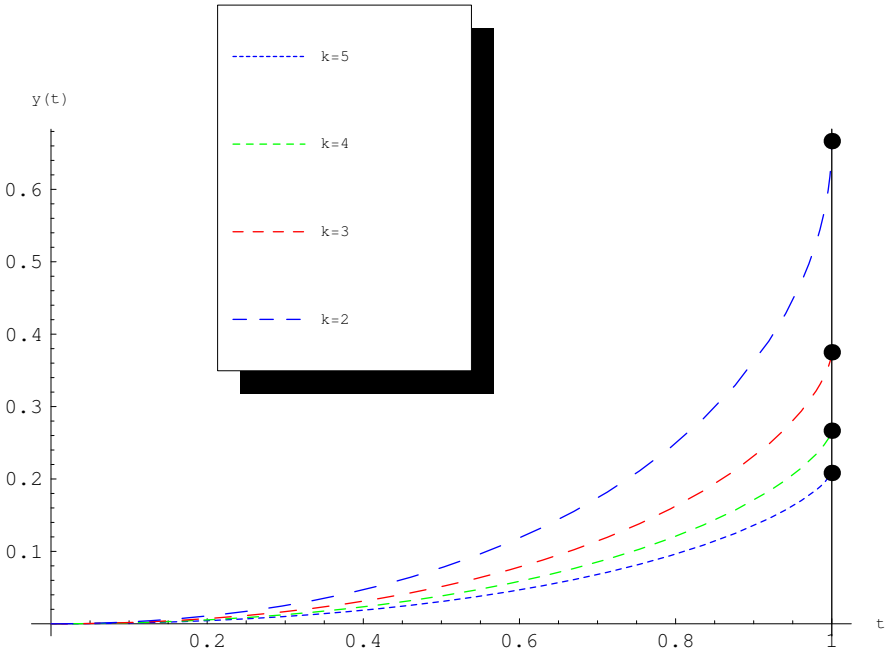


Figure 2.21: The case in which the pursued moves in a straight line

2.9 Exercises

M 1. Graph the curvatures of the cycloids illustrated in Figures 2.4 and 2.5. Find the formula for the curvature $\boldsymbol{\kappa 2}$ of the general cycloid $\mathsf{cycloid}[a, b]$. Then define and draw ordinary, prolate, and curtate cycloids together with the defining circle such as those on page 42.

M 2. A ***deltoid*** is defined by

$$\mathsf{deltoid}[a](t) = \big(2a\cos t(1+\cos t) - a,\ 2a\sin t(1-\cos t)\big).$$

The curve is so named because it resembles a Greek capital delta. It is a particular case of a curve called ***hypocycloid*** (see Exercise 13 of Chapter 6). Plot as one graph the deltoids $\mathsf{deltoid}[a]$ for $a = 1, 2, 3, 4$. Graph the curvature of the first deltoid.

M **3.** The ***Lissajous***[7] or ***Bowditch curve***[8] is defined by

$$\mathsf{lissajous}[n, d, a, b](t) = \big(a\sin(nt + d),\ b\sin t\big).$$

Draw several of these curves and plot their curvatures. (One is shown in Figure 11.19 on page 349.)

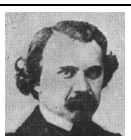
M **4.** The ***limaçon***, sometimes called ***Pascal's snail***, named after Étienne Pascal, father of Blaise Pascal,[9] is a generalization of the cardioid. It is defined by

$$\mathsf{limacon}[a, b](t) = (2a\cos t + b)\big(\cos t,\ \sin t\big).$$

Find the formula for the curvature of the limaçon, and plot several of them.

5. Consider a circle with center $C = (0, a)$ and radius a. Let ℓ be the line tangent to the circle at $(0, 2a)$. A line from the origin $O = (0, 0)$ intersecting ℓ at a point A intersects the circle at a point Q. Let x be the first coordinate of A and y the second coordinate of Q, and put $P = (x, y)$. As A varies along ℓ the point P traces out a curve called ***versiera***, in Italian, and misnamed in English as the ***witch of Agnesi***.[10] Verify that a parametrization of the Agnesi versiera is

$$\mathsf{agnesi}[a](t) = \big(2a\tan t,\ 2a\cos^2 t\big)\,.$$



7 Jules Antoine Lissajous (1822–1880). French physicist, who studied similar curves in 1857 in connection with his optical method for studying vibrations.



8 Nathaniel Bowditch (1773–1838). American mathematician and astronomer. His *New American Practical Navigator*, written in 1802, was highly successful. Bowditch also translated and annotated *Laplace's Mécanique Céleste*. His study of pendulums in 1815 included the figures named after him. Preferring his post as president of the Essex Fire and Marine Insurance Company from 1804 to 1823, Bowditch refused chairs of mathematics at several universities.



9 Blaise Pascal (1623–1662). French mathematician, philosopher and inventor. Pascal was an early investigator in projective geometry and invented the first mechanical device for performing addition and subtraction.



10 Maria Gaetana Agnesi (1718–1799). Professor at the University of Bologna. She was the first woman to occupy a chair of mathematics. Her widely used calculus book *Instituzioni Analitiche* was translated into French and English.

M **6.** Define the curve

$$\mathsf{tschirnhausen}[n,a](t) = a\left(\frac{\cos t}{(\cos(t/3))^n},\ \frac{\sin t}{(\cos(t/3))^n}\right).$$

When $n = 1$, this curve is attributed to Tschirnhausen.[11] Find the formula for the curvature of $\mathsf{tschirnhausen}[n,a][t]$ and make a simultaneous plot of the curves for $1 \leqslant n \leqslant 8$.

7. In the special case that the speed ratio is 1, show that the pursuit curve is $\boldsymbol{\alpha}(t) = (t,\ y(t))$, where

$$y(t) = \frac{a}{4}\left(\left(\frac{a-t}{a}\right)^2 - 1 - 2\log\left(\frac{a-t}{a}\right)\right),$$

and that the pursuer never catches the pursued.

M **8.** Equation (2.28) defines the function

$$y(t) = \frac{ak}{k^2-1} + \frac{k(a-t)^{1+1/k}}{2a^{1/k}(1+k)} - \frac{a^{1/k}k(a-t)^{1-1/k}}{2(k-1)}.$$

Plot a pursuit curve with $a = 1$ and $k = 1.2$.

[11] Ehrenfried Walter Tschirnhausen (1651–1708). German mathematician, who tried to solve equations of any degree by removing all terms except the first and last. He contributed to the rediscovery of the process for making hard-paste porcelain. Sometimes the name is written *von Tschirnhaus*.

Notebook 2

■ N2.0

This notebook is primarily concerned with the plotting of special curves and their curvature. Often, we shall plot two or more functions simultaneously in order to compare members of the same family of curves, or to visualize a given curve sharing a common coordinate with its curvature.

Based on the experience from Chapter 1, we shall proceed to set default options that will apply to the command ParametricPlot throughout Notebook 2. In particular, setting AspectRatio->Automatic will generally ensure that x,y have the same scale, whereas setting Compiled->False will assist evaluation of the function to be plotted.

```
SetOptions[ParametricPlot,
  Compiled->False, AspectRatio->Automatic,
  PlotStyle->{{RGBColor[1,0,0], AbsoluteThickness[2]}}
]
```

The PlotStyle option allows us to use thick red as the default, and double brackets bind together the color and thickness commands so that they apply equally to multiple plots. The output summarizes all the new defaults, many of which are unaffected, but can be suppressed by adding a semicolon after the final closing bracket.

We next take the opportunity to reload some definitions made in Notebook 1, in order to make the present notebook self-contained. There is a similar subsection Nn.0.1 in Notebook n for n>2, in which we have placed relevant programs from previous notebooks that have no need of further comment. The contents of these subsections are not displayed in print, but can be inspected and used in the electronic version, and many of the encapsulated programs are automatically run when the initialization cells are evaluated.

▪ N2.0.1

The command L computes the norm or "length" of a vector. The pairing of Unprotect and Protect ensures that the symbol L cannot be used in a different sense, though we shall only adopt this device for single upper case letters that might accidentally be used in another context.

```
Unprotect[L]
  L[v_]:= Sqrt[Simplify[v.v]]
Protect[L]
```

We also redefine the complex structure J with the same device, even though there is less danger that J will be used in a different sense.

```
Unprotect[J]
  J[{x_,y_}]:= {-y,x}
Protect[J]
```

The curvature $\kappa 2$ of a plane curve was defined in N1.4 by the single line

```
κ2[α_][t_]:= α''[tt].J[α'[tt]]/L[α'[tt]]^3/.tt->t
```

■ N2.1

The parametric equation of a cycloid depending on two parameters is

```
cycloid[a_,b_][t_]:= {a t - b Sin[t], a - b Cos[t]}
```

The acceleration vector (when applied to a fixed point) moves in a circle of radius b.

```
cycloid[a,b]'[t]
cycloid[a,b]''[t]
```

The simplest cycloid has a=b=1:

```
α:= cycloid[1,1]
ParametricPlot[α[t], {t,-3π,3π},
  PlotRange->{{-3π,3π},{-2,4}}, Ticks->None]
```

Figure 2.1

With α continuing to represent this particular cycloid, we can easily compute and plot its curvature, which we abbreviate with the symbol k:

```
k[t_]:= FullSimplify[κ2[α][t]]
k[t]
Plot[k[t], {t,-π,3π}, AspectRatio->1, PlotRange->{{-1,7},{-7,1}},
  PlotStyle->{Hue[.7],Thickness[.01]}]
```

Figure 2.2

Observe that the curvature is approximately constant except very close to the cusps. The AspectRatio and PlotRange options can be set by trial and error. Given that we have chosen the horizontal and vertical intervals to be equal, AspectRatio->1 is equivalent to AspectRatio->Automatic.

The next program numerically computes the value g[x] of the parameter t, as a function of the cycloid's first coordinate x. Thus g is the inverse of the function f given by f[t] = t - Sin[t]/2, and its graph is the reflection of that of f in the line y=x.

```
α:=cycloid[1, 1/2]
f[t_]:= α[t][[1]]
f[t]
g[x_]:= t /. FindRoot[f[t]==x, {t,1}]
Plot[{f[s], s, g[s]}, {s,0,1}, PlotStyle->{Hue[0],Hue[.5],Hue[.7]},
  AspectRatio->1, PlotRange->{{0,1},{0,1}}]
```

Using definitions from the previous cell, we can plot the curvature in such a way that the curvature of a point on the cycloid is represented by the vertical coordinate of the point on the blue curve with the same horizontal coordinate.

```
Off[FindRoot::cvnwt]
{a,b}= {-π/2, 5π/2};
y[1]:= ParametricPlot[α[t],{t,a,b}, DisplayFunction->Identity]
y[2]:= Plot[κ2[α][g[x]], {x,f[a],f[b]}, PlotStyle->Hue[.7],
  PlotPoints->100, DisplayFunction->Identity]
Show[y[1], y[2], PlotRange->{{f[a],f[b]},{-2,4}},
  AspectRatio->Automatic, DisplayFunction->$DisplayFunction]
```

Figure 2.3

```
Clear[α,f,g,k,a,b,y]
```

It is best to clear the variables that were introduced in the course of the previous two cells in a separate third cell. In the future, we shall often avoid this problem by arranging each new program within the command Module.

The prolate cycloid is animated by the inclusion of the Do command in the following miniprogram. In the definition of z[1], it is important to color or shade the disks in *decreasing* size so as not to mask previous assignments.

```
Module[{α,c,C,z},
  α= cycloid[1,3];
  C= {c,1};
  z[1]= {{GrayLevel[.9], Disk[C,3]}, {GrayLevel[.5], Disk[C,1]}};
  z[2]= {PointSize[.02], Point[C]};
  z[3]= {Circle[C,1], Circle[C,3]};
  z[4]= {Line[{C,α[c]}], Line[{{-5,0},{20,0}}]};
  Do[ParametricPlot[α[t], {t,0,4π}, Axes->None,
    Prolog->Array[z,4]], {c,3,9,3}]
]
```

Figure 2.4

N2.1.1

N2.2

In the following definition, the parameter a may seem redundant since it can in theory be added later, but this way it is easier to handle lemniscates of different sizes. The same remark applies to many other curves, and such a parameter was already used in N1.3 to define circle and parabola.

```
lemniscate[a_][t_]:= a{Cos[t], Sin[t]Cos[t]}/(1+Sin[t]^2)
```

```
lemniscate[a][t] - a lemniscate[1][t]
```

Inclusion of a parameter in the definition is convenient, since the following syntax fails:

```
lem:= a lemniscate[1]
lem[t]
```

The next calculation verifies the relationship between implicit and parametric forms of the lemniscate.

```
t1= (x+f)^2+y^2
t2= t1/.f->-f
t3= t1 t2 - f^4
```

```
t4= Expand[t3]/.{y->x Sin[t], f->a Sqrt[2]}
Simplify[Solve[t4==0, x]]//PowerExpand
```

Next, we plot multiple lemniscates simultaneously by inserting a Table command inside ParametricPlot. In this case, our default option Compiled->False will not prevent compilation problems that result from the multiple plotting, and it is essential to add Evaluate. The function f has been defined to make it easier to ensure that each lemniscate passes through the previous one's foci.

```
Module[{f,y,z},
  f[a_]= (1/Sqrt[2])^a;
  y[a_]= lemniscate[f[a]];
  z[a_]= {PointSize[.02],Point[{-f[a],0}],Point[{f[a],0}]};
  ParametricPlot[Table[y[a][t],{a,0,3}]//Evaluate, {t,0,2π},
    Epilog->Array[z,4], Ticks->None,
    PlotRange->{{-1.2,1.2},{-.5,.5}},
    ImageSize->400]
]
```

Figure 2.7

The curvature of a lemniscate is described by

```
Module[{β,k},
  β:= lemniscate[1];
  k[t_]= FullSimplify[κ2[β][t]];
  Print[k[t]];
  Plot[k[t], {t,0,2π}, AspectRatio->Automatic,
    PlotStyle->Hue[.7]]
]
```

N2.2.1

N2.3

The definition of the cardioid employs the same parameter-handling device explained at the start of N2.2.

```
cardioid[a_][t_]:= 2a(1+Cos[t]){Cos[t], Sin[t]}
```

In the following multiple plot, colors have been added by inserting a Table with a different Hue for each curve. The coloring is determined by n,r,p. The number of plots is n, while r is the number of times the color scheme is repeated, and p the fraction of the spectrum used. The values chosen give a smooth 3-dimensional maze effect. We prefer to append the Evaluate command on the right as it is helpful to think of it as an "afterthought" rather than part of the mathematics.

```
Module[{n,r,p,cc},
  n= 80;
  r= 5;
  p= 1;
  cc[t_]:= Table[cardioid[a][t], {a,n}];
  ParametricPlot[cc[t]//Evaluate, {t,0,2π},
    PlotStyle->Table[Hue[a p r/n],{a, n/r}], Axes->None]
]
```

Figure 2.9

The curvature of the cardioid is given by

```
γ:=cardioid[1]
Simplify[κ2[γ][t]]
```

N2.3.1

N2.4

We list some of the standard conic definitions that form part of Notebook 2's initialization data:

```
parabola[a_][t_]:= a{2t, t^2}
ellipse[a_,b_][t_]:= {a Cos[t], b Sin[t]}
hyperbola[a_,b_][t_]:= {a Cosh[t], b Sinh[t]}
```

The catenary is essentially the graph of the hyperbolic cosine, though with the first coordinate expressed as a function of the second. It can be rotated in one of two ways:

```
catenary[a_][t_]:= {a Cosh[t/a], t}
revcatenary[a_][t_]:= J[catenary[a][-t]]
```

```
revcatenary[a][t]
Reverse[catenary[a][t]]
```

We now load a package to draw the arrows in Figure 2.11, relating to the catenary.

```
<<Graphics`Arrow`
```

The following command is used for placing text in Figure 2.11. Care has to be taken to choose the font size carefully, as it does not scale if the image is resized, though the size of each graphics window can be adjusted manually or using the command ImageSize. For our purposes, 18pt is a good compromise though we shall modify the command tx in later notebooks.

```
tx[A_,v_]:= Text[FontForm[A, {"Helvetica",18}], v]
```

```
Module[{r,ar,y},
  r= revcatenary[2];
  ar[t_,v_]:= Arrow[r[t], r[t]+v];
  y[1]= {Thickness[.007], Hue[.7],
    ar[0,-2r'[0]], ar[2,2r'[2]], ar[2, 2{r'[2][[1]],0}]};
  y[2]= {PointSize[.02], Point[r[0]], Point[r[2]]};
  y[3]= Circle[r[2], 1.3, {0,Apply[ArcTan,r'[2]]}];
  y[4]= {tx["θ",{2.8,3.5}], tx["T",{-2.3,2.2}],
    tx["U",{4.3,5.7}], tx["p",{.3,2.4}], tx["q",{1.6,3.2}]};
  ParametricPlot[r[t], {t,-2,4}, PlotRange->{{-3,5},{1,7}},
    Ticks->None, Epilog->Array[y,4], ImageSize->400]
]
```

Figure 2.11

Here is a comparison of the curvatures of the parabola and catenary:

```
κ2[parabola[a]][0]
κ2[catenary[a]][0]//PowerExpand
```

This enables us to plot the catenary and parabola with equal curvature at their vertices. The parabola has `Hue[.65]` and is therefore blue!

```
ps[h_]:= {Thickness[.015], Hue[h]}
ParametricPlot[Evaluate[{parabola[1][t]+{0,2},
  revcatenary[2][t]}], {t,-50,50}, PlotRange->{{-4,4},{0,8}},
  PlotStyle->{ps[.65],ps[.05]}]
```

Figure 2.12

```
f:= catenary[a]
ff:= PowerExpand[L[f'[#]]] &
Integrate[ff[tt],{tt,0,t}]
f[t/.Solve[s==%, t]]
```

■ N2.5

In the next definition, the variable `t` can assume any real value:

```
cissoid[a_][t_]:=2a{t^2, t^3}/(1+t^2)
```

```
Clear[a]
γ:= cissoid[a]
γ[t]
```

```
γ'[t]
%//Simplify
```

```
ParametricPlot[cissoid[1][t], {t,-2,2}];
```

```
y[1]:= {Thickness[.01], Line[{{-1,0},{3,0}}]}
y[2]:= {PointSize[.025],
  Point[{-1,0}], Point[{0,0}], Point[{1,0}], Point[{2,0}]}
y[3]:= {tx["A",{-1.2,.2}], tx["Q",{-.2,.2}],
  tx["P",{.9,.2}], tx["R",{1.8,.2}],
  tx["ξ",{.3,1.5}], tx["η",{2,1.1}]}
ParametricPlot[{Reverse[parabola[1][t]],
  Reverse[parabola[1][t]]+{2,0}}, {t,-1,1},
  PlotRange->{{-1.5,2.5},{-2,2}},
  Epilog->Array[y,3], Axes->None, ImageSize->300]
```

Figure 2.13

The following computation verifies the implicit form of the cissoid.

```
rr:= x^2+y^2
Solve[4a^2 - (2a/x-1)^2 rr == (2a y/x)^2 - rr, y]
Simplify[y^2/.%[[1]]]
```

■ N2.5.1

■ N2.6

By contrast to the cissoid in N2.5, the variable `t` is now restricted to the open interval $(0,\pi)$:

```
tractrix[a_][t_]:= a{Sin[t], Cos[t]+Log[Tan[t/2]]}
```

```
τ:= tractrix[1]
ParametricPlot[τ[t], {t,0,π}, PlotRange->{{-1,2},{-2,2}}];
```

The curvature of the tractrix should be everywhere negative, and the graph is correct.

```
k:=κ2[τ]
FullSimplify[k[t]]
PowerExpand[%]
Plot[-Abs[k[t]], {t,0,π}, PlotStyle->Hue[.7]]
```

The next command draws a tangent line of length s to a curve at an assigned point, and is used in Figure 2.13.

```
tangentline[α_][s_,t_]:= Line[{α[t], α[t] + s α'[t]/L[α'[t]]}]
```

```
Module[{τ,a,h,p,q,y},
  τ= tractrix[1];
  a=.6π;
  h=.05π;
  p[i_]= Point[τ[a+i h]];
  q[i_]= tangentline[τ][1,a+i h];
  y[1]= {Hue[.7], Array[q,6]};
  y[2]= {Hue[.7], PointSize[.02], Array[p,6]};
  ParametricPlot[τ[t], {t,.5,π}, Epilog->Array[y,2],
    PlotRange->{{-.5,1.5},{-.5,2}}, Ticks->None]
]
```

Figure 2.16

The next computation helps one to find a valid parametrization of the tractrix.

```
DSolve[y'[x]==-Sqrt[a^2-x^2]/x, y[x], x]
```

■ N2.6.1

■ N2.7

Here are the definitions needed to plot clothoids:

```
clothoidprime[n_,a_][t_]:=
  a{Sin[t^(n+1)/(n+1)], Cos[t^(n+1)/(n+1)]}
clothoid[n_,a_][t_]:= Integrate[clothoidprime[n,a][tt], {tt,0,t}]
```

Mathematica has internal commands FresnelS and FresnelC that define the Fresnel integrals. In fact, clothoid[1,a] will be converted automatically into an equivalent expression using these two functions.

```
s:= Sqrt[π]
f[t_]:= a s{FresnelS[t/s], FresnelC[t/s]}
f'[t]
Clear[s,f]
```

```
ParametricPlot[clothoid[1,1][t], {t,-10,10}, PlotPoints->10]
```

```
Integrate[clothoidprime[1,1][tt], {tt,0,Infinity}]
```

The ouput gives the coordinates around which the top right end of the first clothoid curls. Here is another example:

```
ParametricPlot[Evaluate[clothoid[2,3][t]], {t,-4,4}, PlotPoints->2]
```

Here is a verification of a clothoid's curvature:

```
κ2[clothoid[1,a]][t]
PowerExpand[Simplify[%]]
```

The command clothoid can be extremely slow because *Mathematica* uses numerical integration. In the cell below, we use a different definition of the clothoid curve which calls NDSolve, so that the curve is plotted much more quickly.

```
clothoidnds[min_,max_][n_,a_][s_]:=
Module[{so,ss},
  so=NDSolve[{
    x'[ss]==clothoidprime[n,a][ss][[1]],
    y'[ss]==clothoidprime[n,a][ss][[2]],
    x[0]==0,y[0]==0}, {x,y}, {ss, min, max}];
  First[{x[s],y[s]} /.so]
]
```

Since numerical solution of a differential equation requires the specification of an interval, clothoidnds will only be defined on the interval (min,max). We use this method to draw Figure 2.19, and then show what goes wrong when the solution is defined on too small an interval.

```
m:= 6
r:= 1.5
th:= Thickness[.01]
cc[t_]:= {clothoidnds[-m,m][1,1][t],
  clothoidnds[-m,m][2,5/8][t]}
zz:= ParametricPlot[cc[t]//Evaluate, {t,-6,6},
  PlotRange->{{-r,r},{-r,r}},
  PlotStyle->{{Hue[0],th},{Hue[.65],th}},
  AspectRatio->Automatic, ImageSize->400]
zz
```

Figure 2.19

```
m:= 4
zz
```

■ N2.8

Special packages are used to produce the sketch for Figure 2.20. The first one requires no comment, though we shall have occasion to use Rotations in a subsequent notebook.

```
<<Graphics`Legend`
<<Geometry`Rotations`
```

The following definition is copied from the text, but we must not forget to manufacture the "boat."

```
linearpursuit[a_,k_][t_]:=
a k/(k^2-1) + k(a - t)^(1+1/k)/(2a^(1/k)(1+k)) -
  k a^(1/k)(a-t)^(1-1/k)/(2(k-1))
```

```
boat[c_,d_,e_,f_]:=
  {{-c,d},{0,d-f},{1,d},{1,d+e},{0,d+e+f},{-1,d+e}}
```

```
Module[{b,z},
  b= boat[1,1,4,1.5]/40;
  z[1]= {Dashing[{.02}], Thickness[.007], Line[{{1,-1},{1,2}}]};
  z[2]= {Polygon[Map[{1,.4}+#&, b]],
    Polygon[Map[{.6,.2}+Rotate2D[#,.25π]&,b]]};
  z[3]= {tx["β",{1.1,.45}], tx["α",{.5,.25}]};
  Plot[linearpursuit[1,1.2][t], {t,0,1},
    Axes->None, PlotStyle->Thickness[.007],
    PlotRange->{{-.2,1.3},{-.2,1.3}},
    Epilog->Array[z,3], AspectRatio->1]
]
```

Figure 2.20

N2.8.1

N2.9

A triple of cycloids is achieved by setting up a function whose values are plots whose shape and color depends on the given parameter. Another way of plotting multiple curves is given in Section 2.2.

```
Module[{y},
  y[a_]:=ParametricPlot[cycloid[1,a][t], {t,-π,3π},
    PlotStyle->{Hue[1.5/a],Thickness[.008]},
    DisplayFunction->Identity];
  Show[y[1], y[2], y[.5], DisplayFunction->$DisplayFunction]
]
```

We finish with some further definitions of curves cited in the Exercises.

```
deltoid[a_][t_]:= {2a Cos[t](1+Cos[t])-a, 2a Sin[t](1-Cos[t])}
lissajous[n_,d_,a_,b_][t_]:= {a Sin[n t+d], b Sin[t]}
limacon[a_,b_][t_]:= (2a Cos[t]+b){Cos[t], Sin[t]}
agnesi[a][t_]:= 2a{Tan[t], Cos[t]^2}
```

Chapter 3

Alternative Ways of Plotting Curves

So far, we have mainly been plotting curves in *parametric* form. In Section 3.1, we discuss implicitly defined curves, and contrast them with those parametrized in the previous chapters. A curve is defined implicitly as the "zero set" or set of zeros of a differentiable function of two variables. In particular cases, this is the more natural way of defining the curve. Even so, if one is given the nonparametric form of a curve, it is sometimes easier to find a suitable parametrization before attempting to plot it. In any case, it can be important to be able to switch between the two representations of the curve.

As an initial example, after lines and circles, we recall that the cissoid was first defined as the set of zeros of a cubic equation. Similar techniques can be applied to another cubic curve, the folium of Descartes. The latter is studied in Section 3.2, which provides an example of what can happen to a curve as one deforms its initial equation. Cassinian ovals, which we consider in Section 3.3, form a further class of curves that are easily plotted implicitly. This generalizes the lemniscate by requiring that the product of the distances from a point on the curve to two fixed foci be constant without regard to the distance between the foci; the resulting equation is quartic.

As the second topic of this chapter, we shall use polar coordinates to describe curves in Section 3.4. We use this approach to give a simple generalization of a cardioid and limaçon, and to study related families of closed curves. We also compute lengths and curvature in polar coordinates, using the formalism of complex numbers to prove our results.

The use of polar coordinates leads to an investigation of different types of spirals in Section 3.5. A number of these have especially simple polar equations. Further examples of implicit and polar plots can be found in this chapter's exercises and notebook.

3.1 Implicitly Defined Plane Curves

We now consider a way of representing a plane curve that does not involve an explicit parametrization.

Definition 3.1. *Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be any function. The* ***set of zeros*** *of F is*

$$F^{-1}(0) = \big\{\, \mathbf{p} \in \mathbb{R}^2 \mid F(\mathbf{p}) = 0 \,\big\}.$$

If no restrictions are placed on F, then not much can be said about its set of zeros, so we make an additional assumption.

Definition 3.2. *An* ***implicitly defined curve*** *in $\mathbb{R}^2$ is the set of zeros of a differentiable function $F\colon \mathbb{R}^2 \to \mathbb{R}$. Frequently, we refer to the set of zeros as "the curve $F(x, y) = 0$."*

We point out that the analogous theory for nonparametrically defined surfaces will be presented in Section 10.6.

Even when F is assumed to be differentiable, the set of zeros of F may have cusps, and hence appear to be nondifferentiable. Moreover, a theorem of Whitney[1] states that any closed subset of $\mathbb{R}^2$ is the set of zeros of some differentiable function (see [BrLa, page 56]). However, there is an important case when it is possible to find a parametrized curve whose trace is the set of zeros of F.

Theorem 3.3. *Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be a differentiable function, and let $\mathbf{q} = (q_1, q_2)$ be a point such that $F(\mathbf{q}) = 0$. Assume that at least one of the partial derivatives F_x, F_y is nonzero at $\mathbf{q}$. Then there is a neighborhood $\mathcal{U}$ of $\mathbf{q}$ in $\mathbb{R}^2$ and a parametrized curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ such that the trace of $\boldsymbol{\alpha}$ is precisely*

$$\big\{\, \mathbf{p} \in \mathcal{U} \mid F(\mathbf{p}) = 0 \,\big\}.$$

Proof. Suppose, for example, that $F_v(\mathbf{q}) \neq 0$. The Implicit Function Theorem states that there is a differentiable real-valued function g defined on a small neighborhood of q_1 in $\mathbb{R}$ such that $g(q_1) = q_2$ and $t \mapsto F(t, g(t))$ vanishes identically. Then we define $\boldsymbol{\alpha}(t) = (t,\ g(t))$. ∎

Definition 3.4. *Let $\mathscr{C}$ be a subset of $\mathbb{R}^2$. If $F\colon \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function with $F^{-1}(0) = \mathscr{C}$, we say that the equation $F(x, y) = 0$ is a* ***nonparametric form*** *or* ***implicit form*** *of $\mathscr{C}$. If $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ is a curve whose trace is $\mathscr{C}$, we say that $t \mapsto \boldsymbol{\alpha}(t)$ is a* ***parametrization*** *or* ***parametric form*** *of $\mathscr{C}$.*

---

[1]  Hassler Whitney (1907–1989). Influential American differential-topologist.

To start with some examples using the operations of vector calculus, let $\mathbf{p}, \mathbf{v} \in \mathbb{R}^2$ with $\mathbf{v} \neq 0$, and define functions

$$F[\mathbf{p}, \mathbf{v}],\ G[\mathbf{p}, \mathbf{v}] \colon \mathbb{R}^2 \longrightarrow \mathbb{R}$$

by

$$\begin{aligned} F[\mathbf{p}, \mathbf{v}](\mathbf{q}) &= (\mathbf{q} - \mathbf{p}) \cdot \mathbf{v} \\ G[\mathbf{p}, \mathbf{v}](\mathbf{q}) &= \|\mathbf{q} - \mathbf{v}\|^2 - \|\mathbf{p} - \mathbf{v}\|^2. \end{aligned} \tag{3.1}$$

Then $F[\mathbf{p}, \mathbf{v}]^{-1}(0)$ is the straight line through $\mathbf{p}$ perpendicular to $\mathbf{v}$ and the curve $G[\mathbf{p}, \mathbf{v}]^{-1}(0)$ is the circle with center $\mathbf{v}$ that contains $\mathbf{p}$.

Some curves are more naturally defined implicitly. This was the case of the cissoid, for which it is easy to pass between the two representations. For given the parametric form (2.12), we have no difficulty in spotting that $t = y/x$ and reversing the derivation on page 47. Substituting this expression for t into the equation $x = 2at/(1 + t^2)$ yields the nonparametric form

$$x = \frac{2a\,y^2}{x^2 + y^2}$$

of the cissoid, which is equivalent to the cubic equation (2.11).

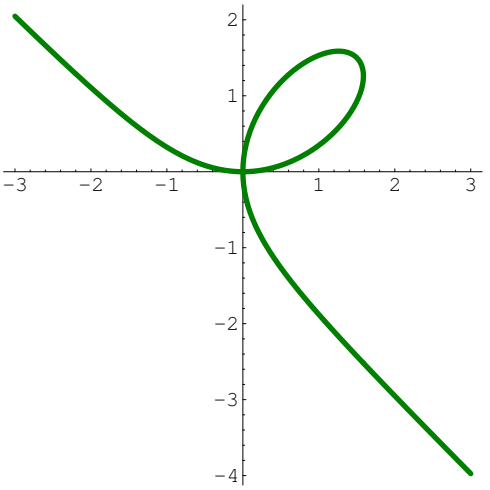


Figure 3.1: The folium of Descartes, $x^3 + y^3 = 3xy$

The trace of a curve $\boldsymbol{\alpha} \colon (a, b) \to \mathbb{R}^2$ is always connected. This is a special case of the theorem from topology that states that the image under a continuous map of a connected set remains connected [Kelley]. It can easily happen however that the set of zeros of a differentiable function $F \colon \mathbb{R}^2 \to \mathbb{R}$ is *disconnected*.

This means that the notions of parametric form and nonparametric form are essentially different.

We give an example of a disconnected zero set immediately below, by perturbing a slightly different cubic equation, that used to plot Figure 3.1.

3.2 The Folium of Descartes

The ***folium of Descartes***[2] is defined nonparametrically by the equation

$$x^3 + y^3 = 3xy. \tag{3.2}$$

In order to formalize this definition, we consider the family F_ε of functions defined by

$$F_\varepsilon(x, y) = x^3 + y^3 - 3\,x\,y - \varepsilon,$$

and their associated curves

$$\mathscr{F}_\varepsilon = \{(x, y) \in \mathbb{R}^2 \mid F_\varepsilon(x, y) = 0\}.$$

Thus, $\mathscr{F}_\varepsilon = F_\varepsilon^{-1}(0)$ is the set of zeros of the function F_ε, and $\mathscr{F}_0$ is the folium of Descartes. Whilst the latter is connected, $\mathscr{F}_\varepsilon$ is disconnected for negative values of ε; Figure 3.2 illustrates the case $\varepsilon = -\frac{1}{10}$.

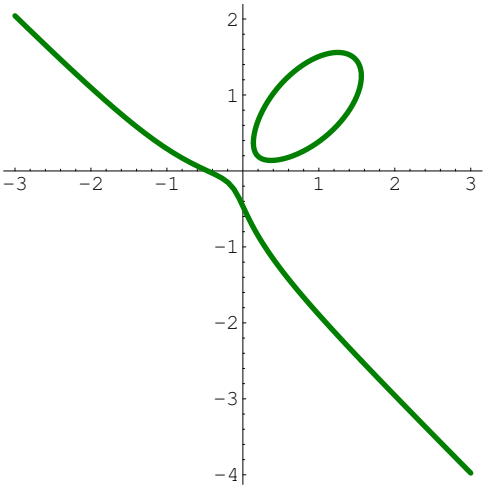


Figure 3.2: Perturbed folium, $\mathscr{F}_{-0.1}$

[2]It was Huygens who first drew the curve correctly. See [Still1, pages 67–68].

We can find a parametrization for $\mathscr{F}_0$, by observing that F_0 is close to a homogeneous polynomial. Dividing the equation $F_0 = 0$ by x^3, we obtain

$$1 + t^3 = \frac{3}{x}t,$$

where $t = y/x$ (as for the cissoid, in the previous section). Solving for x, and then $y = tx$, yields the parametrization

$$\mathsf{folium}(t) = \left(\frac{3t}{1+t^3}, \frac{3t^2}{1+t^3}\right). \tag{3.3}$$

Such a ***rational*** parametrization cannot be found for the curve $\mathscr{F}_\varepsilon$, unless ε equals 0 or 1. This follows from the fact that $\mathscr{F}_\varepsilon$ is a nonsingular cubic curve if $\varepsilon \neq 0, 1$, and from standard results on the projective geometry of cubic curves. We explain briefly how this theory applies to our example, but refer the reader to [BrKn, Kir] for more details.

The equation $F_\varepsilon = 0$ can be fully "homogenized" by inserting powers of a third variable z so that each term has total degree equal to 3; the result is

$$x^3 + y^3 - 3xyz - \varepsilon z^3 = 0. \tag{3.4}$$

This equation actually defines a curve not in $\mathbb{R}^2$, but in the *real projective plane* $\mathbb{RP}^2$, a surface that we define in Chapter 11. Linear transformations in the variables x, y, z determine *projective transformations* of $\mathbb{RP}^3$, and can be used to find an equivalent form of the equation of the cubic. For example, applying the linear change of coordinates

$$\begin{cases} x = \frac{1}{2}X + Y, \\ y = \frac{1}{2}X - Y, \\ z = Z - X, \end{cases} \tag{3.5}$$

converts (3.4) into

$$(1+\varepsilon)^2X^3 - (3\varepsilon + \frac{3}{4})X^2Z + 3\varepsilon XZ^2 - \varepsilon Z^3 + 3Y^2Z = 0 \tag{3.6}$$

(Exercise 8). Finally, we set $Z = 1$ in order to return to an equation in two variables, namely

$$Y^2 = -\frac{1}{3}(1+\varepsilon)^2X^3 + (\varepsilon + \frac{1}{4})X^2 - \varepsilon X + \frac{1}{3}\varepsilon. \tag{3.7}$$

This is a type of "canonical equation" of the cubic curve $\mathscr{F}_\varepsilon$.

If we write (3.7) as $Y^2 = p(X)$, then p is a cubic polynomial provided $\varepsilon \neq -1$; if $\varepsilon = -1$, (3.6) is divisible by Z, and p arises from the remaining quadratic factor. The cubic curve $\mathscr{F}_\varepsilon$ is said to be *nonsingular* if p has three distinct roots (two of which may be complex), and it can be checked in Notebook 3

that this is the case provided $\varepsilon \neq 0, 1$. If $\varepsilon = 0$, then the origin $(0, 0)$ counts as a singular point in the implicit theory because there is no unique tangent line there (see Figure 3.3). The cissoid is a cubic curve with a different type of singularity. To parametrize a nonsingular cubic curve, one needs the Weierstrass $\wp$ function, a special case of which is defined (albeit with complex numbers) in Section 22.7. The analysis there should make it clear that there is no elementary parametrization of the curve $\mathscr{F}_\varepsilon$ in general.

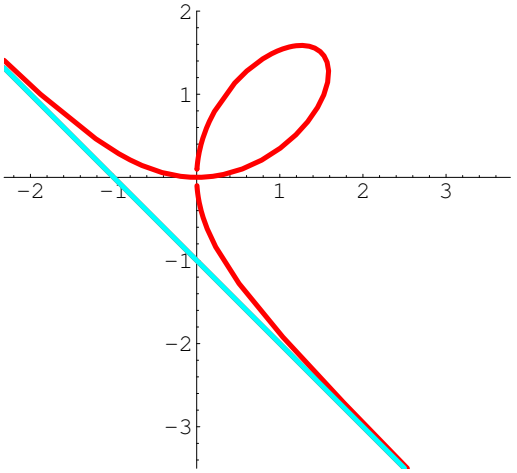


Figure 3.3: Folium of Descartes with asymptote

We conclude this section with some remarks of a more practical nature. The vector-valued function (3.3) was used to produce Figure 3.3, in which the asymptote is generated automatically by the plotting program. The latter is unable to distinguish a pair of points far apart in opposite quadrants from the many pairs of adjacent points selected on the curve. This phenomenon is one possible disadvantage of parametric plotting, though we explain in Notebook 3 how the problem can be overcome to advantage.

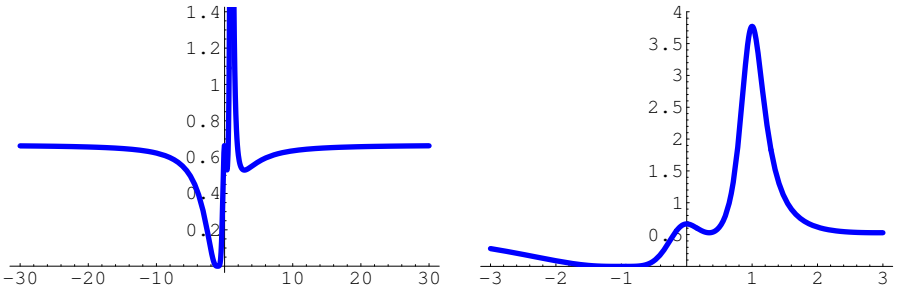


Figure 3.4: Folium curvature

The curvature of $\mathscr{F}_0$ can be obtained from the parametrization (3.3). To understand the resulting function $t \mapsto \boldsymbol{\kappa 2}[\mathsf{folium}](t)$, we plot it first over the range $-30 < t < 30$ and then over the range $-3 < t < 3$. These two plots constitute Figure 3.4. The determination of the maxima and minima of the curvature can be carried out by computer, and is explained in Notebook 3.

3.3 Cassinian Ovals

A ***Cassinian oval***[3] is a generalization of the lemniscate of Bernoulli, and therefore of the ellipse. It is the locus

$$\mathscr{C}_{a,b} = \big\{(x,y) \mid \mathsf{distance}\big((x,y),\, F_1\big)\mathsf{distance}\big((x,y),\, F_2\big) = b^2\big\},$$

where F_1, F_2 are two fixed points with $\mathsf{distance}(F_1, F_2) = 2a$. The constants a, b are unrelated, though setting $a = b = f$, the curve $\mathscr{C}_{f,f}$ coincides with the lemniscate $\mathscr{L}$ defined on page 43.

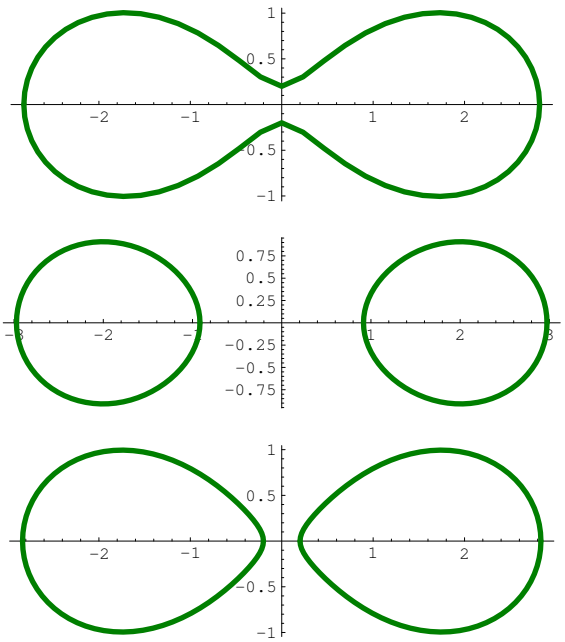


Figure 3.5: Ovals with $b = 2$ and $200a = 199, 200, 201$ respectively



3 Gian Domenico Cassini (1625–1712). Italian astronomer, who did his most important work in France. He proposed the fourth degree curves now called ovals of Cassini to describe planetary motion.

To study Cassinian ovals, we first define

$$\mathsf{cassiniimplicit}[a,b](x,y) = (x^2+y^2+a^2)^2 - b^4 - 4a^2x^2.$$

Generalizing the computations of Section 2.2, one can prove that $\mathscr{C}_{a,b}$ is the zero set of this function. This is the next result, whose proof we omit.

Lemma 3.5. *An oval of Cassini with* $(\pm a, 0)$ *as foci is the implicitly defined curve*

$$(x^2+y^2)^2 + 2a^2(y^2-x^2) = b^4 - a^4.$$

The zero set of

$$\mathsf{cassiniimplicit}[f,f](x,y) = (x^2+y^2)^2 + 2f^2(y^2-x^2)$$

is the lemniscate $\mathscr{L} = \mathscr{C}_{f,f}$, while $\mathsf{cassiniimplicit}[0,b]$ is of course a circle.

The horizontal intercepts of $\mathsf{cassiniimplicit}[a,b]$ are $(\pm\sqrt{a^2 \pm b^2},\ 0)$. When $a^2 > b^2$, there are four horizontal intercepts, but when $a^2 < b^2$ there are only two horizontal intercepts, since we must exclude imaginary solutions. Similarly, $\mathsf{cassiniimplicit}[a,b]$ has no vertical intercepts when $a^2 > b^2$, whereas for $a^2 < b^2$ the vertical intercepts are $(0,\ \pm\sqrt{b^2-a^2})$. This is visible in Figures 3.5 and 3.6.

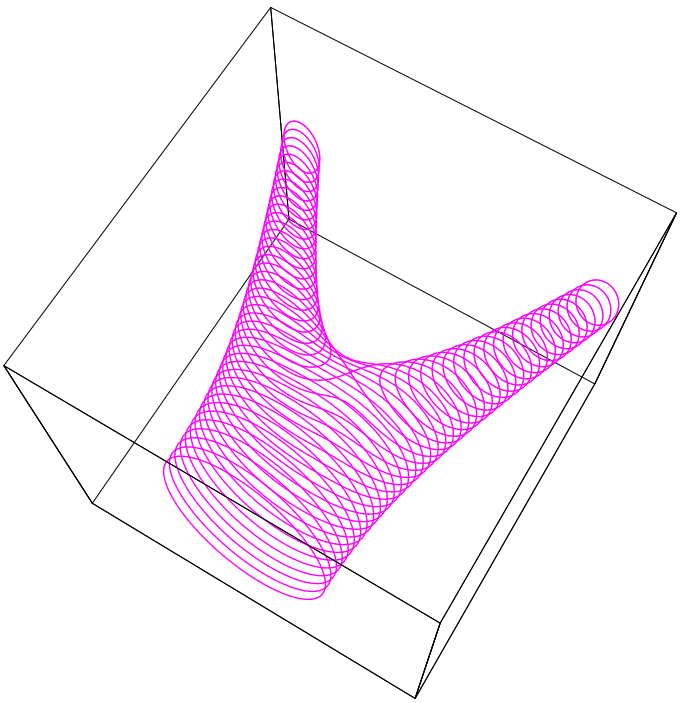


Figure 3.6: A Cassinian family

3.4 Plane Curves in Polar Coordinates

In this section we show how to study and compute the length and curvature of a plane curve using polar coordinates.

Definition 3.6. *A* ***polar parametrization*** *is a curve* $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^2$ *of the form*

$$\boldsymbol{\gamma}(\theta) = \mathbf{r}[\boldsymbol{\gamma}](\theta)\,(\cos\theta,\ \sin\theta), \tag{3.8}$$

where $\mathbf{r}[\boldsymbol{\gamma}](\theta) \geqslant 0$ *for* $a < \theta < b$. *We call* $\mathbf{r}[\boldsymbol{\gamma}]$ *the* ***radius function*** *of the curve* $\boldsymbol{\gamma}$, *and abbreviate it to* $\mathbf{r}$ *when there is no danger of confusion.*

The radius function completely determines the polar parametrization, so usually a curve is described in polar coordinates simply by giving the definition of $\mathbf{r}$.

A polar parametrization of a plane curve is often very simple. Its description is simplified further by writing just the definition of the radius function of the curve. We shall see shortly that there are formulas for the arc length and curvature of a polar parametrization in terms of the radius function alone.

Let us give a generalization of the cardioid and limaçon (see Exercise 4 of Chapter 2) by defining the radius function

$$\mathsf{limaconpolar}[n,a,b](\theta) = 2a\cos n\theta + b.$$

Observe that $\mathsf{limaconpolar}[1,a,2a](\theta) = 2a(\cos\theta + 1)$ is the radius function of a standard cardioid. More complicated examples are exhibited in Figure 3.7. An analogous polar parametrization is

$$\mathsf{pacman}[n](\theta) = 1 + \cos^n\theta, \tag{3.9}$$

and its name is justified by the example shown in Figure 3.13 and Exercise 10.

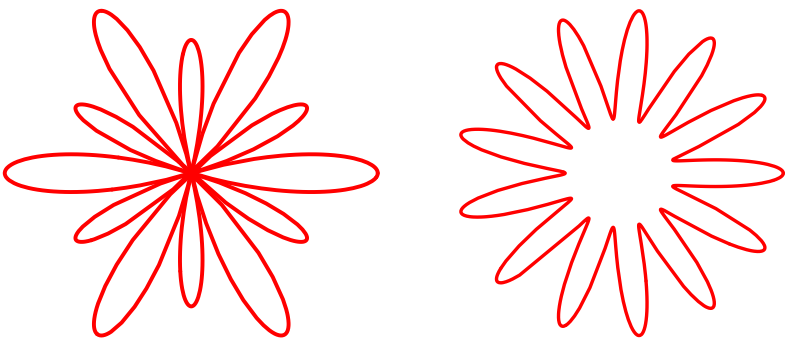


Figure 3.7: The polar limaçons $\mathsf{limaconpolar}[6,3,1]$ and $\mathsf{limaconpolar}[13,\frac{1}{2},2]$

Next, we derive the polar coordinate formulas for arc length and curvature.

Lemma 3.7. *The length and curvature of the polar parametrization* (3.8) *are given in terms of the radius function* $\mathbf{r} = \mathbf{r}[\boldsymbol{\gamma}]$ *by the formulas*

$$\mathsf{length}[\boldsymbol{\gamma}] \;=\; \int_a^b \sqrt{\mathbf{r}'(\theta)^2 + \mathbf{r}(\theta)^2}\, d\theta, \tag{3.10}$$

$$\boldsymbol{\kappa 2}[\boldsymbol{\gamma}] \;=\; \frac{-\mathbf{r}''\mathbf{r} + 2\mathbf{r}'^2 + \mathbf{r}^2}{(\mathbf{r}'^2 + \mathbf{r}^2)^{3/2}}. \tag{3.11}$$

Proof. The calculations can be carried out most easily using complex numbers. Equation (3.8) can be written more succinctly as

$$\boldsymbol{\gamma}(\theta) = \mathbf{r}(\theta) e^{i\theta}.$$

We get

$$\begin{aligned} \boldsymbol{\gamma}'(\theta) &= \big(\mathbf{r}'(\theta) + i\mathbf{r}(\theta)\big)e^{i\theta} \\ \boldsymbol{\gamma}''(\theta) &= \big(\mathbf{r}''(\theta) + 2i\mathbf{r}'(\theta) - \mathbf{r}(\theta)\big)e^{i\theta}. \end{aligned}$$

Therefore

$$\|\boldsymbol{\gamma}'(\theta)\|^2 = \mathbf{r}'(\theta)^2 + \mathbf{r}(\theta)^2, \tag{3.12}$$

and (3.10) follows immediately from the definition of length. Furthermore, using Lemma 1.2 we find that

$$\begin{aligned} \boldsymbol{\gamma}''(\theta) \cdot J\boldsymbol{\gamma}'(\theta) \;&=\; \mathfrak{Re}\Big\{ \big(\mathbf{r}''(\theta) + 2i\mathbf{r}'(\theta) - \mathbf{r}(\theta)\big)e^{i\theta} \big(\overline{i(\mathbf{r}'(\theta) + i\mathbf{r}(\theta))e^{i\theta}} \big) \Big\} \\ &=\; \mathfrak{Re}\Big\{ \big(\mathbf{r}''(\theta) + 2i\mathbf{r}'(\theta) - \mathbf{r}(\theta)\big) \big(- i\mathbf{r}'(\theta) - \mathbf{r}(\theta)\big) \Big\} \\ &=\; -\mathbf{r}''(\theta)\mathbf{r}(\theta) + 2\mathbf{r}'(\theta)^2 + \mathbf{r}(\theta)^2. \end{aligned}$$

Then (3.11) now follows from (3.12) and (1.12). ∎

3.5 A Selection of Spirals

A good example of a polar parametrization is the logarithmic spiral, that was studied in some detail towards the end of Chapter 1. There, we saw that its radius function is given by

$$\mathsf{logspiralpolar}[a, b](\theta) = a\, e^{b\theta}.$$

The curve is characterized by Lemma 1.29 on page 24, and for this reason is sometimes called the ***equiangular spiral*** [Coxeter, 8.7].

A somewhat more naïve equation for a spiral in polar coordinates is

$$\mathbf{r}^n = a^n \theta, \tag{3.13}$$

where n is a nonzero integer. Several of the resulting curves have been given particular names, and we now plot a few of them.

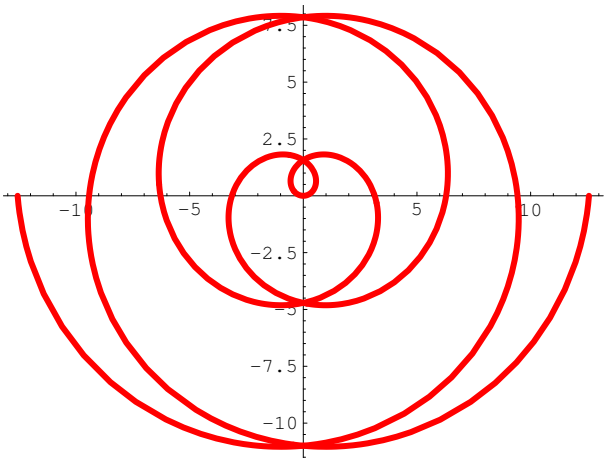


Figure 3.8: The spiral of Archimedes,[4] $r = \theta$

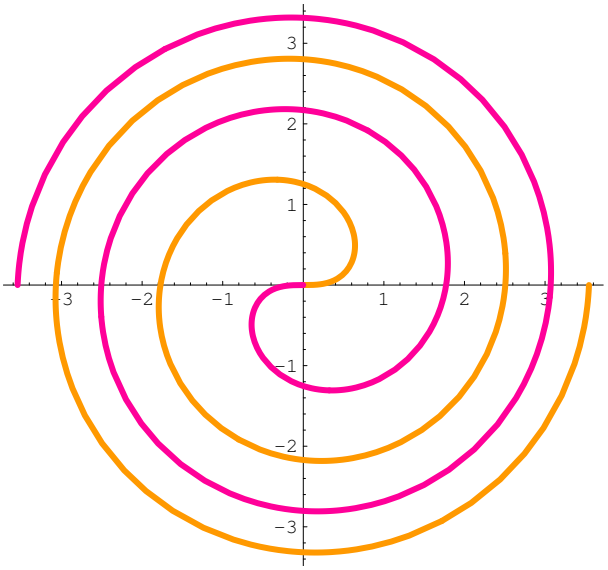


Figure 3.9: Fermat's spiral, $r^2 = \theta$



4 Archimedes of Syracuse (287–212 BC). Archimedes is credited with the creation of many mechanical devices such as compound pulleys, water clocks, catapults and burning mirrors. Legend has it that he was killed by a Roman soldier as he traced mathematical figures in the sand during the siege of Syracuse, at that time a Greek colony on what is now Sicily. His mathematical work included finding the area of a circle and the area under a parabola using the method of exhaustion.

The spiral (3.13) has two branches, the second obtained by allowing the radius function to be negative. The two branches join up provided the exponent n is positive. For Archimedes' spiral, the two branches are reflections of each other in the vertical y-axis. By contrast, for Fermat's spiral[5] in Figure 3.9, the two branches are rotations of each other by 180°.

Similar behavior is seen in Figure 3.10. However, when n is negative, θ cannot be allowed to pass through zero, and this explains why the two branches fail to join up.

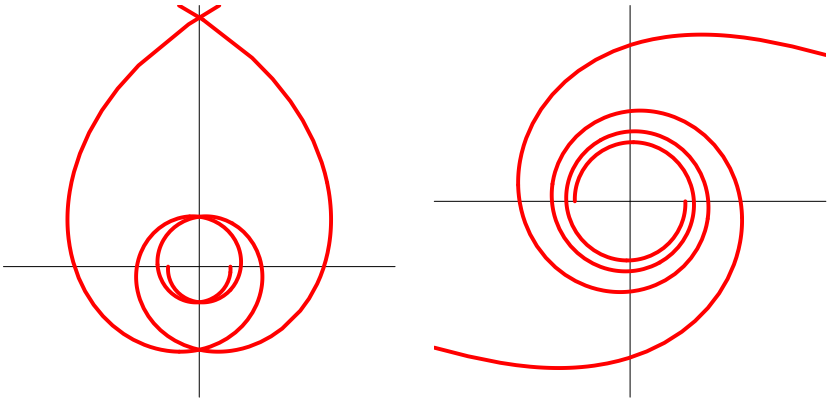


Figure 3.10: The hyperbolic spiral $r = 3/\theta$ and lituus $r^2 = 1/\theta$

Finally, we use the formula (3.11) to compute the curvature of a spiral $\boldsymbol{\gamma}$ given by (3.13), as a function of the angle θ. The result,

$$\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](\theta) = \frac{n\theta^{1-a/n}(1+n+n^2\theta^2)}{a(1+n^2\theta^2)^{3/2}},$$

is finite for all θ. In the special case $n = -1$ and $a = 3$, we get

$$\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](\theta) = -\frac{\theta^6}{3(1+\theta^2)^{3/2}},$$

corresponding to Figure 3.10 (left).



5 Pierre de Fermat (1601–1665). Fermat, like his contemporary Descartes, was trained as a lawyer. Fermat generalized the work of Archimedes on spirals. Like Archimedes he had the germs of the ideas of both differential and integral calculus — using the new techniques of analytic geometry — but failed to see the connections. Fermat is most remembered for stating the theorem that $x^n + y^n = z^n$ has no solution in integers for $n > 2$.

3.6 Exercises

1. Find the nonparametric form of the following curves:

(a) $\boldsymbol{\alpha}(t) = (t^{15},\, t^6)$.

(b) The hyperbola-like curve $\boldsymbol{\gamma}(t) = (t^3,\, t^{-4})$.

(c) The ***strophoid*** defined by

$$\mathsf{strophoid}[a](t) = a\left(\frac{t^2-1}{t^2+1},\, \frac{t(t^2-1)}{t^2+1}\right).$$

2. The ***devil's curve*** is defined nonparametrically as the zeros of the function

$$\mathsf{devilimplicit}[a,b](x,y) = y^2(y^2-b^2) - x^2(x^2-a^2).$$

Find its intercepts, and compare the result with Figure 3.11.

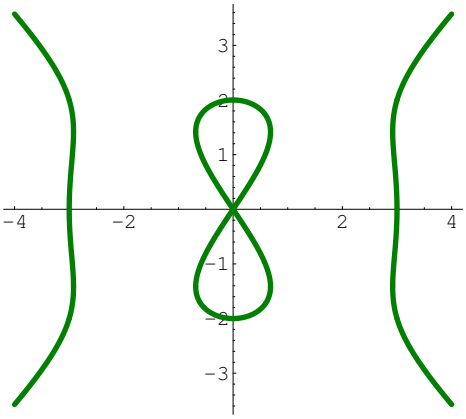


Figure 3.11: The devil's curve

M **3.** ***Kepler's folium***[6] is the curve defined nonparametrically as the set of zeros of the function

$$\mathsf{keplerimplicit}[a,b](x,y) = \big((x-b)^2+y^2\big)\big(x(x-b)+y^2\big) - 4a(x-b)y^2.$$

Plot Kepler's folium and some of its perturbations.

---

6 

Johannes Kepler (1571–1630). German astronomer, who lived in Prague. By empirical observations Kepler showed that a planet moves around the sun in an elliptical orbit having the sun at one of its two foci, and that a line joining the planet to the sun sweeps out equal areas in equal times as the planet moves along its orbit.

M **4.** Plot the following curves with interesting singularities ([Walk, page 56]):

(a) $x^3 - x^2 + y^2 = 0$,

(b) $x^4 + x^2y^2 - 2x^2y - xy^2 + y^2 = 0$,

(c) $2x^4 - 3x^2y + y^2 - 2y^3 - y^4 = 0$,

(d) $x^3 - y^2 = 0$,

(e) $(x^2 + y^2)^2 + 3x^2y - y^3 = 0$,

(f) $(x^2 + y^2)^3 - 4x^2y^2 = 0$.

M **5.** The parametric equation for the ***trisectrix of Maclaurin***[7] is

$$\mathsf{trisectrix}[a](t) = \big(a(4\cos^2 t - 3),\ a(4\cos^2 t - 1)\tan t\big)\,.$$

Plot it, using its simpler polar equation

$$\mathsf{polartrisectrix}[a](\theta) = a\sec\frac{\theta}{3}.$$

6. A ***nephroid*** is the kidney-shaped curve

$$\mathsf{epicycloid}[2b, b] = b\big(3\cos t - \cos 3t,\ 2\sin t - \sin 2t\big),$$

that results from setting $a = 2b$ in Exercise 10 on page 144. Show that its equation in polar coordinates is

$$r = 2b\left(\left(\sin\frac{\theta}{2}\right)^{\frac{2}{3}} + \left(\cos\frac{\theta}{2}\right)^{\frac{2}{3}}\right)^{\frac{3}{2}}.$$

M **7.** The general semicubical parabola is defined implicitly by

$$y^2 = ax^3 + bx^2 + cx + d.$$

Find appropriate values of a, b, c, d so as to duplicate Figure 3.12.



7 Colin Maclaurin (1698–1746). Scottish mathematician. His *Geometria organica, sive discriptio linearum curvarum universalis* dealt with general properties of conics and higher plane curves. In addition to Maclaurin's own results, this book contained the proofs of many important theorems that Newton had given without proofs. Maclaurin is also known for his work on power series and the defense of his book *Theory of Fluxions* against the religious attacks of Bishop George Berkeley.

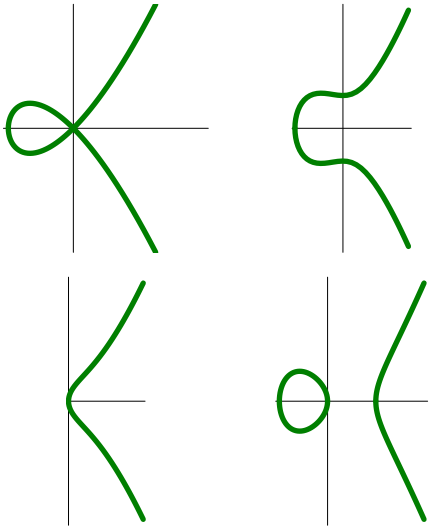


Figure 3.12: Semicubical parabolas

M **8.** Verify that the transformation (3.5) does indeed convert the equation (3.4) on page 77 into (3.6) and (3.7).

9. Show that, up to sign, the signed curvature of the curve $g(x, y) = 0$ is

$$\boldsymbol{\kappa 2}(x, y) = \frac{g_{xx}\, g_y^2 - 2 g_{xy}\, g_x\, g_y + g_{yy}\, g_x^2}{\left(g_x^2 + g_y^2\right)^{3/2}},$$

where subscripts indicate partial derivatives.

M **10.** Draw a selection of the curves (3.9) for n between 1 and 1000, paying attention to the difference in behavior when n is even or odd.

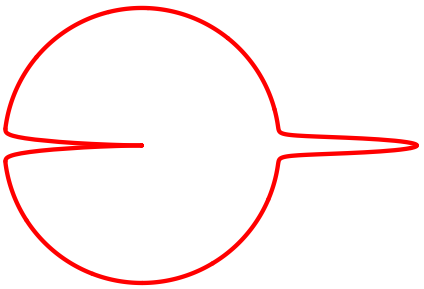


Figure 3.13: pacman[999]

Show that pacman$[n]$ has finite curvature at $\theta = 0$ but infinite curvature at $\theta = \pi$.

Notebook 3

■ N3.0

This notebook will focus on the command ImplicitPlot, used to plot construct curves that are defined in an implicit or nonparametric way. While this is an effective tool, there are a number of technical problems that we will need to mention in the process. Notebook 3 also provides many examples of curves plotted using PolarPlot, and we give particular attention to spirals in N3.5. Animations are included of Fermat's spiral and the "pacman" curve.

The following curves, all defined previously, will be needed in N3.1 and N3.3.

```
cissoid[a_][t_]:= 2a{t^2, t^3}/(1+t^2)
lemniscate[a_][t_]:= a{Cos[t], Sin[t]Cos[t]}/(1+Sin[t]^2)
agnesi[a][t_]:= 2a{Tan[t], Cos[t]^2}
```

We next load various standard *Mathematica* packages. ImplicitPlot enables us to use the command of the same name to visualize the zero set of many real-valued functions defined in the plane.

```
<<Graphics`ImplicitPlot`
```

The next will enable us to experiment with plots using polar coordinates in N3.3.

```
<<Graphics`Graphics`
```

The final 3-dimensional package will be used extensively when we start describing curves and surfaces in space (this will be done following Chapters 6 and 10 respectively). In the present notebook, it is only used to draw Figure 3.6 in N3.3.

```
<<Graphics`Graphics3D`
```

In N2.0, we set options for ParametricPlot. Based on what we have in mind below, we do the same for Plot and ImplicitPlot.

```
SetOptions[Plot,
  Compiled->False,
  PlotStyle->{{RGBColor[0,0,1], AbsoluteThickness[2]}}
];
```

While blue is our choice for standard plots, we adopt dark green to indicate that a plot has been constructed implicitly.

```
SetOptions[ImplicitPlot,
  AspectRatio->Automatic,
  PlotStyle->{{RGBColor[0,.5,0], AbsoluteThickness[2]}}
];
```

The hidden subsection N3.0.1 contains the default options for ParametricPlot adopted in N2.0, and the programs J, L, κ2.

■ N3.0.1

■ N3.1

The following *Mathematica* command will attempt to convert the parametric form of a curve α to the implicit form. The command will usually fail if the components of α are not rational functions.

```
param2imp[α_][x_,y_]:= Eliminate[{x,y}==α[t], t]
```

We have used *Mathematica* to eliminate variables between a set of simultaneous equations. For example, we can find nonparametric equations of the cissoid and versiera, though fuller information is obtained by replacing Eliminate by Reduce.

```
param2imp[cissoid[a]][x,y]
param2imp[agnesi[a]][x,y]
```

Here is a parametrization of the folium of Descartes:

```
folium[t_]:= {3t/(1+t^3), 3t^2/(1+t^3)}
```

```
foliumimplicit= param2imp[folium][x,y]
```

There are two ways that the command ImplicitPlot can be used. The first attempts to plot the curve F(x,y)=a by finding solutions at each point in the x range. Because some equations may be too complicated symbolically for *Mathematica*, this version does not always work. But there are many interesting curves that can be handled nicely using this method.

In the following cell, ImageSize has been set so that the size of the numbers labelling the axes is in proportion to the overall plot. This is necessary since (as mentioned in N2.4) text does not change size when images are rescaled, although the size of images can of course be adjusted manually.

```
α:= foliumimplicit
ImplicitPlot[α, {x,-3,3}, ImageSize->300]
```

Figure 3.1

■ N3.2

The syntax with ImplictPlot is typical: the equation, a horizontal range and possibly some plotting options need to be specified. Any symbol (other than x) can be used in place of y in the equation, provided it has not already been assigned a value. This version of ImplicitPlot produces a type of graphics called simply Graphics.

```
ImplicitPlot[α[[1]]-α[[2]]==-0.1, {x,-3,3}, ImageSize->300]
Clear[α]
```

Figure 3.2

There is a second version of ImplicitPlot that produces a different type of graphics, called ContourGraphics. *Mathematica* produces several different kinds of graphics; we shall also come across SurfaceGraphics and Graphics3D. The syntax for the second version of ImplicitPlot is the same as the first version, except that a vertical range is also specified. Compare the following two examples:

```
eq:= (y - x^2)(y - 2x^2)==0
ImplicitPlot[eq, {x,-1,1}]
ImplicitPlot[eq, {x,-1,1}, {y,-1,1}]
Clear[eq]
```

In contrast to the first version, this second version always works, though the resulting plot may be more jagged.

The *Mathematica* command to draw the parametric version of the folium of Descartes incorrectly displays the asymptote. This problem occurs because as the parameter passes through the singular value -1, the curve jumps across quadrants:

```
y1= ParametricPlot[folium[t], {t,-30,30},
  PlotRange->{Automatic,{-3.5,2}}]
```

An effective way to get rid of the asymptote is to use the following trick, due to Michael Mezzino (http://math.cl.uh.edu).

```
rem:= {x__, {a_,b_}, {c_,d_},y__}/; Negative[b d]->{x,Null,y}
y2:= y1/.rem
Show[y2]
```

The same idea can be modified so as to display the asymptote in a different color. In the following plot, the asymptote is actually plotted twice.

```
y[h_]:= ParametricPlot[folium[t], {t,-30,30},
  PlotRange->{Automatic,{-3.5,2}},
  PlotStyle->{Hue[h], Thickness[.01]},
  DisplayFunction->Identity];
rm:= {x__, {a_,b_}, {c_,d_},y__}/; Negative[b d]->{{a,b},{c,d}}
Show[y[0], y[.5]//.rm, DisplayFunction->$DisplayFunction]
```

Figure 3.3

The command rem worked by repeatedly searching InputForm[y1] for any pair {a,b},{c,d} of coordinates for which there is the strict inequality bd<0, and then eliminating it. The subject of the search can be inspected by removing the semicolon in the next cell:

```
InputForm[y1];
```

```
y1:= Plot[κ2[folium][t], {t,-30,30}, DisplayFunction->Identity]
y2:= Plot[κ2[folium][t], {t,-3,3}, PlotRange->{Automatic,{0,4}},
  DisplayFunction->Identity]
Show[GraphicsArray[{y1,y2}], GraphicsSpacing->0,
  DisplayFunction->$DisplayFunction, ImageSize->650]
```

Figure 3.4

The formula for the curvatures plotted in Figure 3.4 is produced by

```
κ2[folium][t]
%//PowerExpand
%//Simplify
```

The graphs of the curvature of the folium show several local maxima and minima. The values of t that give rise to these local maxima and minima can be found using NSolve. The second line picks out those solutions with zero imaginary component:

```
sol= NSolve[κ2[folium]'[t]==0, t]
Select[t /. sol, Im[#]==0 &]
```

■ N3.3

We cheat by giving a ready-made equation describing Cassinian ovals. But we can easily verify that this function is correct by use of the substitution denoted by su in the second cell.

```
cassiniimplicit[a_,b_][x_,y_]:=
  (x^2 + y^2 + a^2)^2 - b^4 - 4 a^2 x^2 == 0
```

```
lem:= lemniscate[a]
su:= {x->Sqrt[2]lem[t][[1]], y->Sqrt[2]lem[t][[2]]}
cassiniimplicit[a,a][x,y]/.su
%//Simplify
```

ImplicitPlot can now be used to show Cassinian ovals, as in the following cell. We display the output using GraphicsArray, in order that the three plots constitute a single file when exported from the notebook. In general, this command is capable of

displaying a rectangular matrix of separate objects, though in this case the array ar consists of a single column:

```
Module[{z},
  z[a_,b_]:= ImplicitPlot[cassiniimplicit[a,b][x,y], {x,-3,3},
    DisplayFunction->Identity];
  ar:= {{z[199/100,2]}, {z[2.2,2]}, {z[201/100,2]}};
  Show[GraphicsArray[ar], ImageSize->450,
    DisplayFunction->$DisplayFunction]
]
```

Figure 3.5

```
ta:= Table[cassiniimplicit[a,2][x,y], {a,0,4,.2}]
ImplicitPlot[ta//Evaluate, {x,-5,5}]
```

The command Solve can be used to find the intercepts of a curve (that is, where the curve meets the horizontal and vertical axes):

```
Solve[cassiniimplicit[a,b][x,0], x]
Solve[cassiniimplicit[a,b][0,y], y]
```

This output yields the horizontal and vertical intercepts of cassiniimplicit[a,b]. When $a^2 > b^2$, there are four horizontal intercepts, but when $a^2 < b^2$ there are only two horizontal intercepts, since we must exclude imaginary solutions. Similarly, there are no vertical intercepts when $a^2 > b^2$.

Mathematica's command StackGraphics can be used to good effect to display Cassinian ovals in 3 dimensions. This command is available from the package Graphics3D, loaded in N3.0, and PlotStyle overrides the default option from N3.0. In the first line we create, but do not display, the Cassinian ovals.

```
cassinis:= Table[ImplicitPlot[cassiniimplicit[a,2][x,y], {x,-5,5},
  DisplayFunction->Identity, PlotStyle->RGBColor[1,0,1]],
  {a,0,4,1/10}]
Show[StackGraphics[cassinis], ViewPoint->{.6,-1,3}, Axes->False];
```

Figure 3.6

The formula for the curvature of a plane curve involves a quotient of the form A/B. The following miniprogram uses abbreviations for the various first and second derivatives,

and then computes the numerator and denominator before evaluating the final quotient.

```
κ2implicit[g_][x_,y_]:=
Module[{u,v,A,B},
  gu= D[g[u,v],u];
  gv= D[g[u,v],v];
  guu= D[g[u,v],u,u];
  guv= D[g[u,v],u,v];
  gvv= D[g[u,v],v,v];
  A= guu gv^2 - 2guv gu gv + gvv gu^2;
  B= (gu^2 + gv^2)^(3/2);
  A/B /. {u->x, v->y}
];
```

We compute the curvature of a Cassinian oval:

```
h[x_,y_]= Part[cassiniimplicit[2,2][x,y], 1]
k:= κ2implicit[h]
k[x,y]//Factor
Plot3D[k[x,y], {x,-1,1}, {y,.1,1}]
```

■ N3.4

The command PolarPlot contained in the Graphics package loaded in N3.0 plots a curve using the radius function. In fact, PolarPlot is a *Mathematica* command to convert the radius function of a polar parametrization into a form that is required for ParametricPlot. Here is a slightly simplified version:

```
PolarPlot[r_, {t_,min_,max_},opts___]:=
  ParametricPlot[{r Cos[t], r Sin[t]}, {t,min,max}, opts,
  AspectRatio->Automatic]
```

The triple underscore symbol "___" allows the insertion of zero or more options of the type that one would use with ParametricPlot. It is easy to use PolarPlot to graph the radius function of the limaçon:

```
limaconpolar[n_,a_,b_][θ_]:= 2a Cos[n θ] + b
```

The parameters n,a,b below were chosen by trial and error in order to illustrate two particular limaçon polar curves.

```
Module[{z},
  z[n_,a_,b_]:= PolarPlot[limaconpolar[n,a,b][θ], {θ,0,2π},
    Axes->None, DisplayFunction->Identity];
  Show[GraphicsArray[{z[6,3,1], z[13,1/2,2]}],
    DisplayFunction->$DisplayFunction]
]
```

Figure 3.7

```
PolarPlot[2+Cos[50 θ], {θ,0,2π}]
```

A similar formula defines the following curve in polar coordinates. Its essential form depends on the parity of n.

```
pacman[n_][θ_]:= Cos[θ]^n + 1
```

The next animation illustrates pacman's evolution in the case in which n is an increasing odd integer. A large odd value is illustrated by Figure 3.13 in N3.5.

```
Do[PolarPlot[pacman[4n-1][θ], {θ,0,2π}, Axes->False,
  PlotRange->{{-2,2},{-2,2}}, PlotPoints->500], {n,30}]
```

The *Mathematica* formula for the curvature of a polar parametrization is

```
κ2polar[ρ_][t_]:= (-D[ρ[tt],{tt,2}]ρ[tt] + 2D[ρ[tt],tt]^2 + ρ[tt]^2)/
  (D[ρ[tt],tt]^2 + ρ[tt]^2)^(3/2) /. tt->t
```

■ N3.5

Here is the polar parametrization of the logarithmic spiral already encountered in N1.6:

```
logspiralpolar[a_,b_][θ_]:= a Exp[b θ]
```

```
PolarPlot[logspiralpolar[1,0.08][θ], {θ,0,12π}]
```

```
archimedesspiralpolar[n_,a_][θ_]:= a θ^(1/n)
```

```
arc:= archimedesspiralpolar[8,1]
PolarPlot[arc[θ], {θ,0,12π}]
```

The following simplification commands seem exaggerated but, applied one after the other, they achieve the desired result.

```
κ2polar[archimedesspiralpolar[n,a]][θ]
%//Simplify
%//PowerExpand
%//Simplify
```

Here is the spiral representing the simplest equation in polar coordinates, namely r=θ .

```
PolarPlot[θ, {θ,0,4π}]
```

Adding negative values of θ has the effect of adding the reflection of this curve in the vertical axis:

```
PolarPlot[θ, {θ,-4π,4π}, ImageSize->400]
```

Figure 3.8

To plot Fermat's spiral $r^2=\theta$, PolarPlot reverts to its built-in definition since the first double argument does not fit the pattern in our definition above, and so our default plotting options no longer apply. We plot separately the two branches corresponding to r>0 and r<0, choosing a different hue for each.

```
ps[h_]:= {Thickness[.01], Hue[h]};
PolarPlot[{Sqrt[θ],-Sqrt[θ]}, {θ,0,4π}, PlotStyle->{ps[.1],ps[.9]},
  ImageSize->400]
```

Figure 3.9

The same curve can be "grown" using the animation that follows.

```
Module[{f,z},
  f[θ_]= {Sqrt[θ], -Sqrt[θ]};
  z[a_]:= PolarPlot[f[θ]//Evaluate, {θ,0,a π}, Axes->False,
    PlotRange->{{-4,4},{-4,4}}, PlotStyle->{ps[.8]}];
  Do[z[a], {a, 1/20, 4, 1/20}]
]
```

Similar techniques, which we skip over, are used to plot the spirals in Figure 3.10 in N3.5.1. Each of these has two separate branches that need to be plotted separately.

■ N3.5.1

■ N3.6

We begin a study of implicit plots with the following example:

```
strophoid[a_][t_]:= a(t^2 - 1)/(t^2 + 1){1,t}
```

```
param2imp[strophoid[a]][x,y]
```

In the next two cells, we use the alternative notation for functions in which the symbol # stands for the variable. The final & is essential, but can be remembered by thinking of it as short for <--#, i.e., "mapsto" in reverse.

```
f:= {#^15, #^6} &
param2imp[f][x,y]
```

```
g:= {#^3, 1/#^4} &
param2imp[g][x,y]
```

A number of curves are now defined implicitly and then plotted in the plane, beginning with the "diabolical" case:

```
devilimplicit[a_,b_][x_,y_]:= y^2(y^2-b^2) - x^2(x^2-a^2) == 0
```

```
ImplicitPlot[devilimplicit[3,2][x,y], {x,-4,4}]
```

Figure 3.11

```
keplerimplicit[a_,b_][x_,y_]:=
  ((x - b)^2 + y^2)(x(x - b) + y^2) - 4a(x - b)y^2 == 0
```

```
ImplicitPlot[Table[keplerimplicit[a,4][x,y], {a,1,4}]//Evaluate,
  {x,-4,8}]
```

```
trisectrixpolar[a_][θ_]:= a/Cos[θ/3]
```

Here is an array of cubic curves. The error message can be ignored.

```
semicubicimplicit[a_,b_,c_,d_][x_,y_]:=
  a x^3 + b x^2 + c x + d - y^2 == 0
```

```
Module[{z},
  z[a_,b_,c_,d_]:= ImplicitPlot[semicubicimplicit[a,b,c,d][x,y],
    {x,-2,2}, Ticks->None, DisplayFunction->Identity];
  Show[GraphicsArray[{{z[1,1,0,0], z[1,1,0,1]},
    {z[1,0,1,0], z[1,0,-1,0]}}], GraphicsSpacing->{.2,.1},
    DisplayFunction->$DisplayFunction]
]
```

Figure 3.12

The option GraphicsSpacing is used to fix both the horizontal and vertical spacing in the array.

N3.6.1

Chapter 4

New Curves from Old

Any plane curve gives rise to other plane curves through a variety of general constructions. Each such construction can be thought of as a *function* which assigns one curve to another, and we shall discover some new curves in this manner. Four classic examples of constructing one plane curve from another are studied in the present chapter: *evolutes* in Sections 4.1 and 4.2, *involutes* in Section 4.3, *parallel curves* in Section 4.5 and *pedal curves* in Section 4.6.

Along the way, we show how to construct normal and tangent lines to a curve, and *osculating circles* to curves. We explain in Section 4.4 that the circle through three points on a plane curve tends to the osculating circle as the three points become closer and coincide. For the same reason, the evolute of a plane curve can be visualized by plotting a sufficient number of normal lines to the curve, as illustrated by the well-known design in Figure 4.12 on page 111.

4.1 Evolutes

A point $\mathbf{p} \in \mathbb{R}^2$ is called a ***center of curvature*** at $\mathbf{q}$ of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$, provided that there is a circle $\boldsymbol{\gamma}$ with center $\mathbf{p}$ which is tangent to $\boldsymbol{\alpha}$ at $\mathbf{q}$ such that the curvatures of the curves $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, suitably oriented, are the same at $\mathbf{q}$. We shall see that this implies that there is a line ℓ from $\mathbf{p}$ to $\boldsymbol{\alpha}$ which meets $\boldsymbol{\alpha}$ perpendicularly at $\mathbf{q}$, and the distance from $\mathbf{p}$ to $\mathbf{q}$ is the radius of curvature of $\boldsymbol{\alpha}$ at $\mathbf{q}$, as defined on page 14. An example is shown in Figure 4.2.

The centers of curvature form a new plane curve, called the evolute of $\boldsymbol{\alpha}$, whose precise definition is as follows.

Definition 4.1. *The* ***evolute*** *of a regular plane curve* $\boldsymbol{\alpha}$ *is the curve given by*

$$\mathsf{evolute}[\boldsymbol{\alpha}](t) = \boldsymbol{\alpha}(t) + \frac{1}{\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)} \frac{J\boldsymbol{\alpha}'(t)}{\left\|\boldsymbol{\alpha}'(t)\right\|}. \tag{4.1}$$

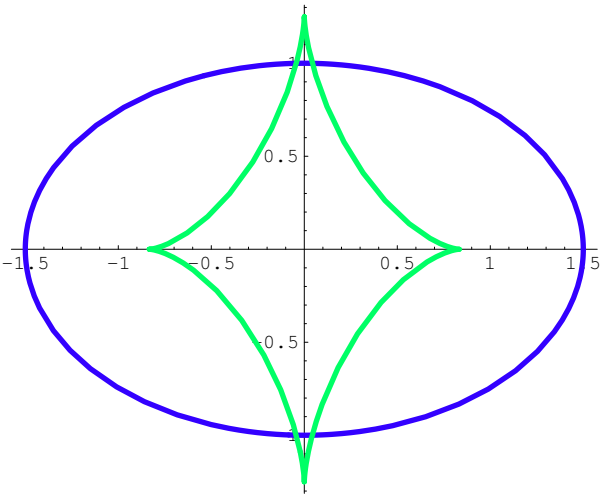


Figure 4.1: The ellipse and its evolute

It turns out that a circle with center $\mathsf{evolute}[\boldsymbol{\alpha}](t)$ and radius $1/|\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)|$ will be tangent to the plane curve $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$. This is the circle, called the ***osculating circle***, that best approximates $\boldsymbol{\alpha}$ near $\boldsymbol{\alpha}(t)$; it is studied in Section 4.4.

Using Formula (1.12), page 14, we see that the formula for the evolute can be written more succinctly as

$$\mathsf{evolute}[\boldsymbol{\alpha}] = \boldsymbol{\alpha} + \frac{\|\boldsymbol{\alpha}'\|^2}{\boldsymbol{\alpha}'' \cdot J(\boldsymbol{\alpha}')} J\boldsymbol{\alpha}'. \tag{4.2}$$

An easy consequence of (4.1) and (1.15) is the following important fact.

Lemma 4.2. The definition of evolute of a curve $\boldsymbol{\alpha}$ is independent of parametrization, so that

$$\mathsf{evolute}[\boldsymbol{\alpha} \circ h] = \mathsf{evolute}[\boldsymbol{\alpha}] \circ h,$$

for any differentiable function $h\colon (c,d) \to (a,b)$.

The evolute of a circle consists of a single point. The evolute of any plane curve $\boldsymbol{\gamma}$ can be described physically. Imagine light rays starting at all points of the trace of $\boldsymbol{\gamma}$ and propagating down the normals of $\boldsymbol{\gamma}$. In the case of a circle, these rays focus perfectly at the center, so for $\boldsymbol{\gamma}$ the focusing occurs along the centers of best fitting circles, that is, along the evolute of $\boldsymbol{\gamma}$.

A more interesting example is the evolute of the ellipse $x^2/a^2 + y^2/b^2 = 1$, parametrized on page 21. This evolute is the curve $\boldsymbol{\gamma}$ defined by

$$\boldsymbol{\gamma}(t) = \left(\frac{(a^2 - b^2)\cos^3 t}{a}, \ \frac{(a^2 - b^2)\sin^3 t}{b} \right), \tag{4.3}$$

and is an astroid[1] (see Exercise 1 of Chapter 1). Of course, setting $a = b$ confirms that the evolute of a circle is simply its center-point. Figure 4.1 represents an ellipse and its evolute simultaneously, though the focusing property is best appreciated by viewing Figure 4.12.

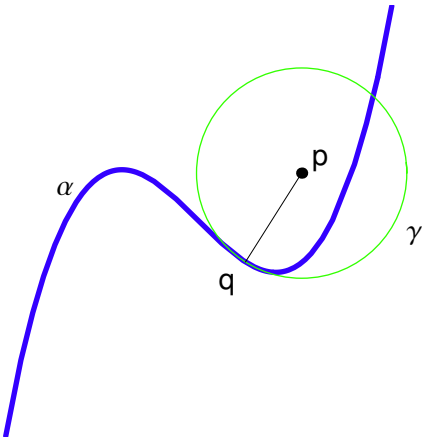


Figure 4.2: A center of curvature on a cubic curve

The notions of tangent line and normal line to a curve are clear intuitively; here is the mathematical definition.

Definition 4.3. *The* ***tangent line*** *and* ***normal line*** *to a curve* $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ *at* $\boldsymbol{\alpha}(t)$ *are the straight lines passing through* $\boldsymbol{\alpha}(t)$ *with velocity vectors equal to* $\boldsymbol{\alpha}'(t)$ *and* $J\boldsymbol{\alpha}'(t)$*, respectively.*

Next, we obtain a characterization of the evolute of a curve in terms of tangent lines and normal lines, and also determine the singular points of the evolute.

Theorem 4.4. *Let* $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ *be a unit-speed curve. Then*

(i) *the evolute of* $\boldsymbol{\beta}$ *is the unique curve of the form* $\boldsymbol{\gamma} = \boldsymbol{\beta} + fJ\boldsymbol{\beta}'$ *for some function* f *for which the tangent line to* $\boldsymbol{\gamma}$ *at each point* $\boldsymbol{\gamma}(s)$ *coincides with the normal line to* $\boldsymbol{\beta}$ *at* $\boldsymbol{\beta}(s)$*.*

(ii) *Suppose that* $\boldsymbol{\kappa 2}[\boldsymbol{\beta}]$ *is nowhere zero. The singular points of the evolute of* $\boldsymbol{\beta}$ *occur at those values of* s *for which* $\boldsymbol{\kappa 2}[\boldsymbol{\beta}]'(s) = 0$*.*

Proof. When we differentiate (4.1) and use Lemma 1.21, page 16, we obtain

$$\mathsf{evolute}[\boldsymbol{\beta}]' = -\frac{\boldsymbol{\kappa 2}[\boldsymbol{\beta}]'}{\big(\boldsymbol{\kappa 2}[\boldsymbol{\beta}]\big)^2} J\boldsymbol{\beta}'. \tag{4.4}$$

[1]Note that an *astroid* is a four-cusped curve, but that an *asteroid* is a small planet. This difference persists in English, French, Spanish and Portugese, but curiously in Italian the word for both notions is *asteroide*.

Hence the tangent line to evolute$[\boldsymbol{\beta}]$ at evolute$[\boldsymbol{\beta}](s)$ coincides with the normal line to $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s)$.

Conversely, suppose $\boldsymbol{\gamma} = \boldsymbol{\beta} + fJ\boldsymbol{\beta}'$. Again, using Lemma 1.21, we compute

$$\boldsymbol{\gamma}' = \big(1 - f\boldsymbol{\kappa 2}[\boldsymbol{\beta}]\big)\boldsymbol{\beta}' + f'J\boldsymbol{\beta}'.$$

If the tangent line to $\boldsymbol{\gamma}$ at each point $\boldsymbol{\gamma}(s)$ coincides with the normal line to $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s)$, then $f = 1/\boldsymbol{\kappa 2}[\boldsymbol{\beta}]$, and so $\boldsymbol{\gamma}$ is the evolute of $\boldsymbol{\beta}$.

This proves (i); then (ii) is a consequence of (4.4). ∎

For the parametrization ellipse$[a, b]$, it is verified in Notebook 4 that

$$\boldsymbol{\kappa 2}'(t) = \frac{3ab(b^2 - a^2)\sin 2t}{2(a^2\sin^2 t + b^2\cos^2 t)^{5/2}}.$$

It therefore follows that the evolute of an ellipse is singular when t assumes one of the four values 0, $\pi/2$, π, $3\pi/2$. This is confirmed by both differentiating (4.3) and inspecting the plot in Figure 4.1.

4.2 Iterated Evolutes

An obvious problem is to see what happens when one applies the evolute construction repeatedly. Computations are carried out in Notebook 4 to find the evolute of an evolute of a curve, and to iterate the procedure.

As an example of this phenomenon, Figure 4.3 plots the first three evolutes of the cissoid, parametrized on page 48. In spite of the fact that the cissoid has a cusp, its first evolute (passing through $(-2, 0)$) and the third evolute (passing through $(0, 0)$) do not.

Another example of a curve with a cusp whose evolute has no cusp is the tractrix, parametrized on page 51. It turns out that the evolute of a tractrix is a catenary. By direct computation, the evolute of tractrix$[a]$ is the curve

$$t \mapsto a\left(\frac{1}{\sin t},\ \log\Big(\tan\frac{t}{2}\Big)\right).$$

To see that this curve is actually a catenary, let $\tau = \tan(t/2)$ and $u = \log\tau$. Then

$$e^u + e^{-u} = \frac{\tau^2 + 1}{\tau} = \frac{2}{\sin t}.$$

Thus the evolute of the tractrix can be reparametrized as $u \mapsto a(\cosh u,\ u)$, which is a multiple of catenary$[1]$ as defined on page 47.

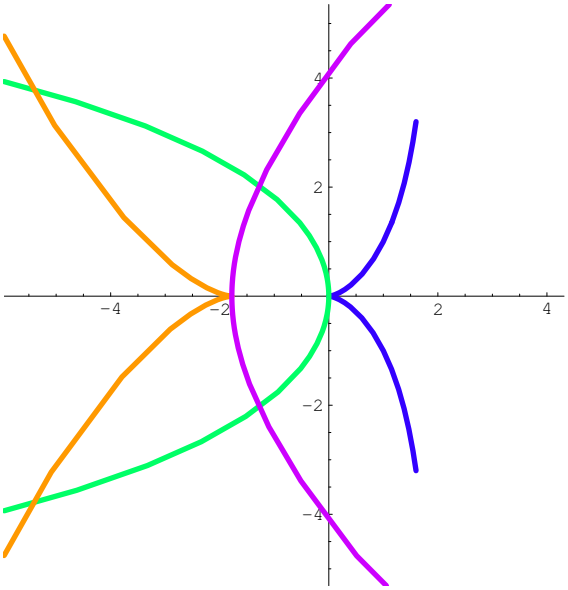


Figure 4.3: Iterated evolutes of cissoid[1]

For Figure 4.4, the evolute of the tractrix was computed and plotted automatically; it is the smooth curve that passes through the cusp. A catenary was then added to the picture off-center, but close enough to emphasize that it is the same curve as the evolute.

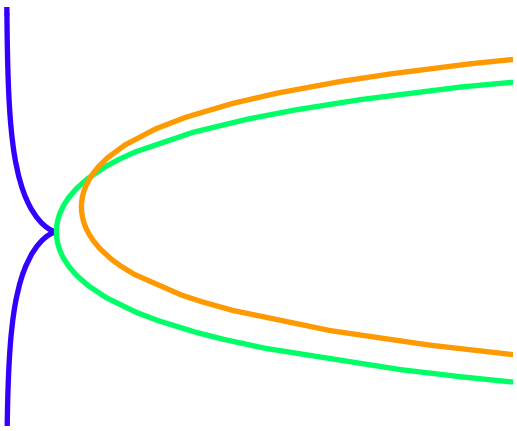


Figure 4.4: tractrix[1] and its catenary evolute

4.3 Involutes

The involute is a geometrically important operation that is inverse to the map $\boldsymbol{\alpha} \mapsto \mathsf{evolute}[\boldsymbol{\alpha}]$ that associates to a curve its evolute. In fact, the evolute is related to the involute in the same way that differentiation is related to indefinite integration. Just like the latter, the operation of taking the involute depends on an arbitrary constant. Furthermore, we shall prove (Theorem 4.9) that the evolute of the involute of a curve $\boldsymbol{\gamma}$ is again $\boldsymbol{\gamma}$; this corresponds to the fact that the derivative of the indefinite integral of a function f is again f.

We first give the definition of the involute of a unit-speed curve.

Definition 4.5. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed curve, and let $a < c < b$. The* ***involute*** *of $\boldsymbol{\beta}$ starting at $\boldsymbol{\beta}(c)$ is the curve given by*

$$\mathsf{involute}[\boldsymbol{\beta}, c](s) = \boldsymbol{\beta}(s) + (c-s)\boldsymbol{\beta}'(s). \tag{4.5}$$

Whereas the evolute of a plane curve $\boldsymbol{\beta}$ is a linear combination of $\boldsymbol{\beta}$ and $J\boldsymbol{\beta}'$, an involute of $\boldsymbol{\beta}$ is a linear combination of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$. Note that although we use s as the arc length parameter of $\boldsymbol{\beta}$, it is *not* necessarily an arc length parameter for the involute of $\boldsymbol{\beta}$.

The formula for the involute of an arbitrary-speed curve needs the arc length function defined on page 12.

Lemma 4.6. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular arbitrary-speed curve. Then the involute of $\boldsymbol{\alpha}$ starting at c (where $a < c < b$) is given by*

$$\mathsf{involute}[\boldsymbol{\alpha}, c](t) = \boldsymbol{\alpha}(t) + \big(s_\alpha(c) - s_\alpha(t)\big)\frac{\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|}, \tag{4.6}$$

where $t \mapsto s_\alpha(t)$ denotes the arc length of $\boldsymbol{\alpha}$ measured from an arbitary point.

The involute of a curve can be described geometrically.

Theorem 4.7. *An involute of a regular plane curve $\boldsymbol{\beta}$ is formed by unwinding a taut string which has been wrapped around $\boldsymbol{\beta}$.*

This result is illustrated in Figure 4.5, in which the string has been "cut" at the point $\boldsymbol{\beta}(c)$ on the curve, and gradually unwound from that point.

Proof. Without loss of generality, we may suppose that $\boldsymbol{\beta}$ has unit speed. Then

$$\mathsf{involute}[\boldsymbol{\beta}, c](s) - \boldsymbol{\beta}(s) = (c-s)\boldsymbol{\beta}'(s),$$

so that

$$\big\|\mathsf{involute}[\boldsymbol{\beta}, c](s) - \boldsymbol{\beta}(s)\big\| = |s-c|. \tag{4.7}$$

Here $|s-c|$ is the distance from $\boldsymbol{\beta}(s)$ to $\boldsymbol{\beta}(c)$ measured along the curve $\boldsymbol{\beta}$, while the left-hand side of (4.7) is the distance from $\mathsf{involute}[\boldsymbol{\beta}, c](s)$ to $\boldsymbol{\beta}(s)$ measured along the tangent line to $\boldsymbol{\beta}$ emanating from $\boldsymbol{\beta}(s)$. ∎

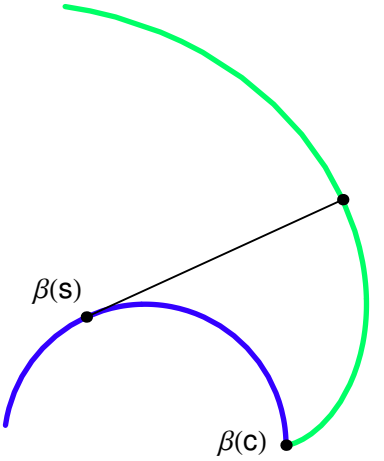


Figure 4.5: Definition of an involute

The most famous involute is that of a circle. With reference to (1.25),

$$\mathsf{involute}[\mathsf{circle}[a], b](t) = a\big(\cos t + (-b+t)\sin t,\ (b-t)\cos t + \sin t\big).$$

An example is visible in Figure 4.7 on page 107, in which the operation of taking the involute has been iterated. (This is carried out analytically in Exercise 6.)

The involute of the figure eight curve (2.4), page 44, requires a complicated integral, which is computed numerically in Notebook 4 to obtain Figure 4.6 and a selection of normal lines.

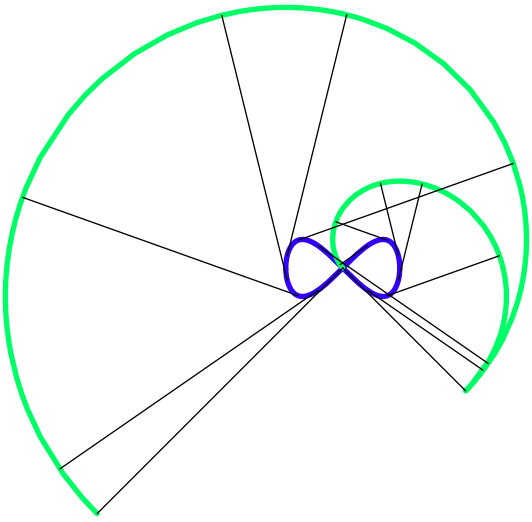


Figure 4.6: Involute of a figure eight

Next, we find a useful relation between the curvature of a curve and that of its involute.

Lemma 4.8. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed curve, and let $\boldsymbol{\gamma}$ be the involute of $\boldsymbol{\beta}$ starting at c, where $a < c < b$. Then the curvature of $\boldsymbol{\gamma}$ is given by*

$$\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](s) = \frac{\operatorname{sign}\big(\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\big)}{|s-c|}. \tag{4.8}$$

Proof. First, we use (4.5) and Lemma 1.21 to compute

$$\boldsymbol{\gamma}'(s) = (c-s)\boldsymbol{\beta}''(s) = (c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\, J\boldsymbol{\beta}'(s), \tag{4.9}$$

and

$$\begin{aligned} \boldsymbol{\gamma}''(s) \;&=\; -\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\, J\boldsymbol{\beta}'(s) + (c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}]'(s)\, J\boldsymbol{\beta}'(s) \\ &\qquad\qquad +(c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\, J\boldsymbol{\beta}''(s) \\ &=\; \Big(-\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s) + (c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}]'(s)\Big) J\boldsymbol{\beta}'(s) \\ &\qquad\qquad -(c-s)\big(\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\big)^2\boldsymbol{\beta}'(s). \end{aligned} \tag{4.10}$$

From (4.9) and (4.10), we get

$$\boldsymbol{\gamma}''(s) \cdot J\boldsymbol{\gamma}'(s) = (c-s)^2\big(\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\big)^3. \tag{4.11}$$

Now (4.8) follows from (4.9), (4.11) and the definition of $\boldsymbol{\kappa 2}[\boldsymbol{\gamma}]$. ∎

Lemma 4.8 implies that the absolute value of the curvature of the involute of a curve is always decreasing as a function of s in the range $s \geqslant c$. This can be seen clearly in Figures 4.6 and 4.7.

Theorem 4.9. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed curve and let $\boldsymbol{\gamma}$ be the involute of $\boldsymbol{\beta}$ starting at c, where $a < c < b$. Then the evolute of $\boldsymbol{\gamma}$ is $\boldsymbol{\beta}$.*

Proof. By definition the evolute of $\boldsymbol{\gamma}$ is the curve $\boldsymbol{\zeta}$ given by

$$\boldsymbol{\zeta}(s) = \boldsymbol{\gamma}(s) + \frac{1}{\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](s)} \frac{J\boldsymbol{\gamma}'(s)}{\big\|\boldsymbol{\gamma}'(s)\big\|}. \tag{4.12}$$

When we substitute (4.5), (4.8) and (4.9) into (4.12), we get

$$\begin{aligned} \boldsymbol{\zeta}(s) \;&=\; \boldsymbol{\beta}(s) + (c-s)\boldsymbol{\beta}'(s) + \frac{|c-s|}{\operatorname{sign}\big(\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s)\big)} \frac{(c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s) J^2\boldsymbol{\beta}'(s)}{\big\|(c-s)\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s) J\boldsymbol{\beta}'(s)\big\|} \\ &=\; \boldsymbol{\beta}(s). \end{aligned}$$

Thus $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ coincide. ∎

The previous result is consistent with thinking of "evolution" as differentiation, and "involution" as integration, as explained at the start of this section. That being the case, one would expect a sequel to Theorem 4.9 to assert that the involute of the evolute of a curve $\boldsymbol{\beta}$ is the same as $\boldsymbol{\beta}$ "up to a constant." The appropriate notion is contained in Definition 4.13 below: it can be shown that the involute of the evolute of $\boldsymbol{\beta}$ is actually a *parallel curve* to $\boldsymbol{\beta}$. Examples are given in Notebook 4.

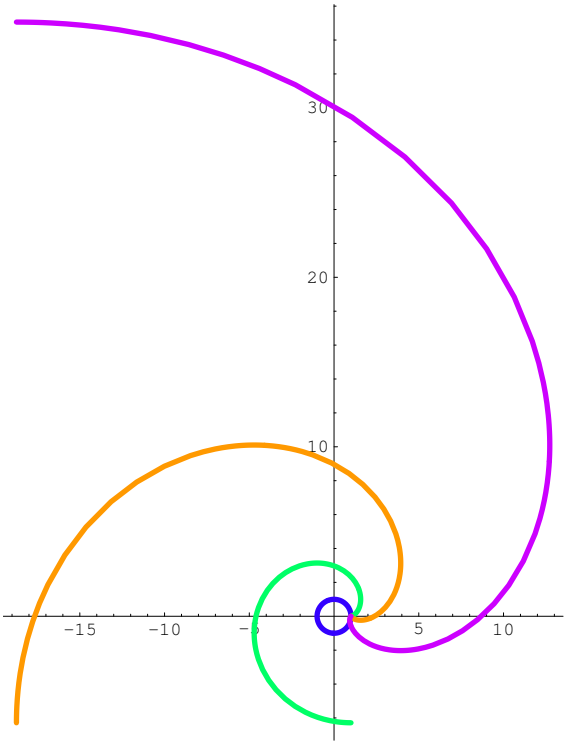


Figure 4.7: A circle and three successive involutes

4.4 Osculating Circles to Plane Curves

Just as the tangent is the best *line* that approximates a curve at one of its points $\mathbf{p}$, the osculating circle is the best *circle* that approximates the curve at $\mathbf{p}$.

Definition 4.10. *Let $\boldsymbol{\alpha}$ be a regular plane curve defined on an interval (a, b), and let $a < t < b$ be such that $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) \neq 0$. Then the* ***osculating circle*** *to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$ is the circle of radius $1/|\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)|$ and center*

$$\boldsymbol{\alpha}(t) + \frac{1}{\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)} \frac{J\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|}.$$

The dictionary definition of "osculating" is "kissing." In fact, the osculating circle at a point $\mathbf{p}$ on a curve approximates the curve much more closely than the tangent line. Not only do $\boldsymbol{\alpha}$ and its osculating circle at $\boldsymbol{\alpha}(t)$ have the same tangent line and normal line, but also the same curvature.

It is easy to see from equation (4.1) that

Lemma 4.11. *The centers of the osculating circles to a curve form the evolute to the curve.*

The osculating circles to a logarithmic spiral are a good example of close approximation to the curve. Figure 4.8 represents these circles without the spiral itself.

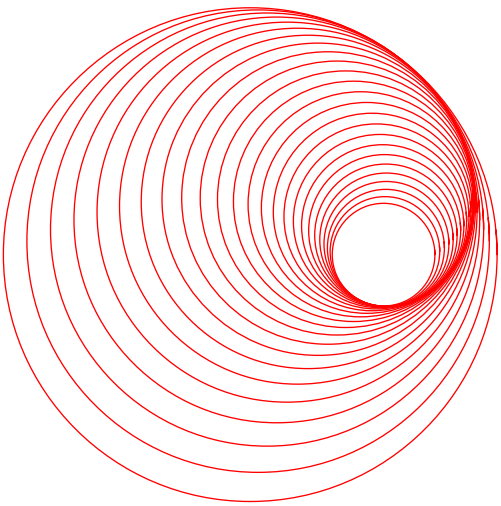


Figure 4.8: Osculating circles to logspiral$[1, -1.5]$

Next, we show that an osculating circle to a plane curve is the limit of circles passing through three points of the curve as the points tend to the point of contact of the osculating circle.

Theorem 4.12. *Let $\boldsymbol{\alpha}$ be a plane curve defined on an interval (a, b), and let $a < t_1 < t_2 < t_3 < b$. Denote by $\mathscr{C}(t_1, t_2, t_3)$ the circle passing through the points $\boldsymbol{\alpha}(t_1), \boldsymbol{\alpha}(t_2), \boldsymbol{\alpha}(t_3)$ provided these points are distinct and do not lie on the same straight line. Assume that $t_0 \in (a, b)$ and $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t_0) \neq 0$. Then the osculating circle to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_0)$ is the circle*

$$\mathscr{C} = \lim_{\substack{t_1 \to t_0 \\ t_2 \to t_0 \\ t_3 \to t_0}} \mathscr{C}(t_1, t_2, t_3).$$

Proof. Denote by $\mathbf{p}(t_1, t_2, t_3)$ the center of $\mathscr{C}(t_1, t_2, t_3)$, and define $f\colon (a, b) \to \mathbb{R}$ by

$$f(t) = \big\|\boldsymbol{\alpha}(t) - \mathbf{p}(t_1, t_2, t_3)\big\|^2.$$

Then

$$\left\{ \begin{array}{rcl} f'(t) &=& 2\boldsymbol{\alpha}'(t) \cdot \big(\boldsymbol{\alpha}(t) - \mathbf{p}(t_1, t_2, t_3)\big), \\ f''(t) &=& 2\boldsymbol{\alpha}''(t) \cdot \big(\boldsymbol{\alpha}(t) - \mathbf{p}(t_1, t_2, t_3)\big) + 2\big\|\boldsymbol{\alpha}'(t)\big\|^2. \end{array} \right. \tag{4.13}$$

Since f is differentiable and $f(t_1) = f(t_2) = f(t_3)$, there exist u_1 and u_2 with $t_1 < u_1 < t_2 < u_2 < t_3$ such that

$$f'(u_1) = f'(u_2) = 0. \tag{4.14}$$

Similarly, there exists v with $u_1 < v < u_2$ such that

$$f''(v) = 0. \tag{4.15}$$

(Equations (4.14) and (4.15) follow from Rolle's theorem.[2] See for example, [Buck, page 90].) Clearly, as t_1, t_2, t_3 tend to t_0, so do u_1, u_2, v. Equations (4.13)–(4.15) imply that

$$\left\{ \begin{array}{rcl} \boldsymbol{\alpha}'(t_0) \cdot (\boldsymbol{\alpha}(t_0) - \mathbf{p}) &=& 0, \\ \boldsymbol{\alpha}''(t_0) \cdot (\boldsymbol{\alpha}(t_0) - \mathbf{p}) &=& -\big\|\boldsymbol{\alpha}'(t_0)\big\|^2, \end{array} \right. \tag{4.16}$$

where

$$\mathbf{p} = \lim_{\substack{t_1 \to t_0 \\ t_2 \to t_0 \\ t_3 \to t_0}} \mathbf{p}(t_1, t_2, t_3).$$

It follows from (4.16) and the definition of $\boldsymbol{\kappa 2}$ that

$$\boldsymbol{\alpha}(t_0) - \mathbf{p} = \frac{-1}{\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t_0)} \frac{J\boldsymbol{\alpha}'(t_0)}{\big\|\boldsymbol{\alpha}'(t_0)\big\|}.$$

Thus, by definition, $\mathscr{C}$ is the osculating circle to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$. ∎

Figure 4.9 plots various circles, each passing through three points on the parabola $4y = x^2$. As the three points converge to the vertex of the parabola, the circle through the three points converges to the osculating circle at the vertex. Since the curvature of the parabola at the vertex is $1/2$, the osculating circle at the vertex has radius exactly 2.

---

[2] Michel Rolle (1652–1719). French mathematician, who resisted the infinitesimal techniques of calculus.

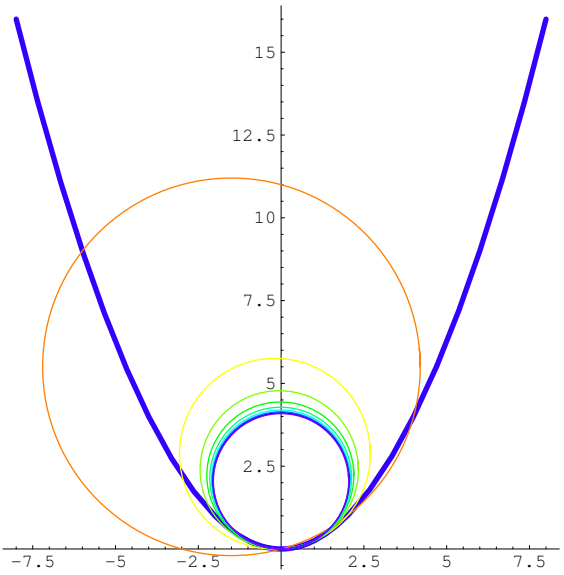


Figure 4.9: Circles converging to an osculating circle of a parabola

4.5 Parallel Curves

It is appropriate to begin this section by illustrating the concept of tangent and normal lines. The tangent line to a plane curve at a point $\mathbf{p}$ on the curve is the best linear approximation to the curve at $\mathbf{p}$, and the normal line is the tangent line rotated by $\pi/2$.

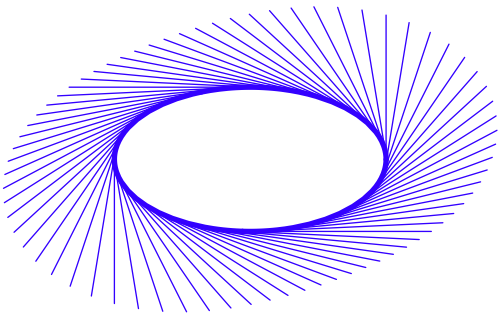


Figure 4.10: Tangent lines to ellipse$[\frac{3}{2}, 1]$

Having drawn some short tangent lines to an ellipse (parametrized on page 21), as if to give it fur, we draw normal lines to the same ellipse in Figure 4.11. Longer normal lines may intersect one another, as we see from Figure 4.12, in which we can clearly distinguish the evolute of the ellipse.

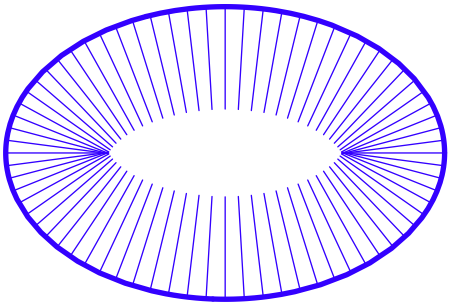


Figure 4.11: Normal lines to ellipse$[\frac{3}{2}, 1]$

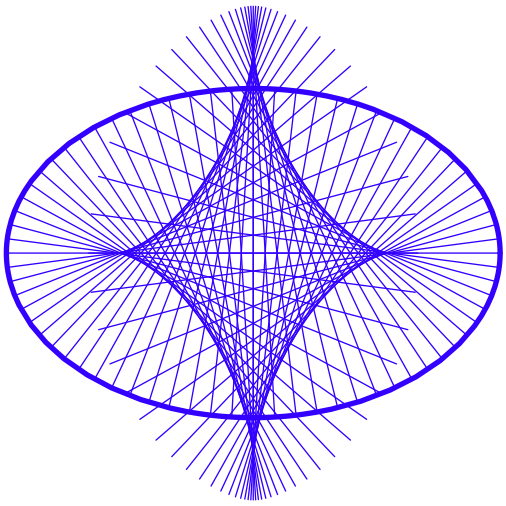


Figure 4.12: Intersecting normals to the same ellipse

We shall now construct a curve $\boldsymbol{\gamma}$ at a fixed distance $r > 0$ from a given curve $\boldsymbol{\alpha}$, where r is not too large. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ be defined on an interval (a, b); then we require

$$\|\boldsymbol{\gamma}(t) - \boldsymbol{\alpha}(t)\| = r \qquad \text{and} \qquad \big(\boldsymbol{\gamma}(t) - \boldsymbol{\alpha}(t)\big) \cdot \boldsymbol{\alpha}'(t) = 0,$$

for $a < t < b$. This leads us to the next definition.

Definition 4.13. *A* ***parallel curve*** *to a regular plane curve* $\boldsymbol{\alpha}$ *at a distance* r *is the plane curve given by*

$$\text{parcurve}[\boldsymbol{\alpha}, r](t) = \boldsymbol{\alpha}(t) + \frac{r\, J\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|}. \tag{4.17}$$

Actually, we can now allow r in (4.17) to be either positive or negative, in order to obtain parallel curves on either side of $\boldsymbol{\alpha}$, without changing t. The definition of parallel curve does not depend on the choice of positive reparametrization. In fact, it is not hard to prove

Lemma 4.14. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a plane curve, and let $h\colon (c,d) \to (a,b)$ be differentiable. Then*

$$\mathsf{parcurve}[\boldsymbol{\alpha} \circ h,\, r](u) = \mathsf{parcurve}[\boldsymbol{\alpha},\, r \operatorname{sign} h']\big(h(u)\big).$$

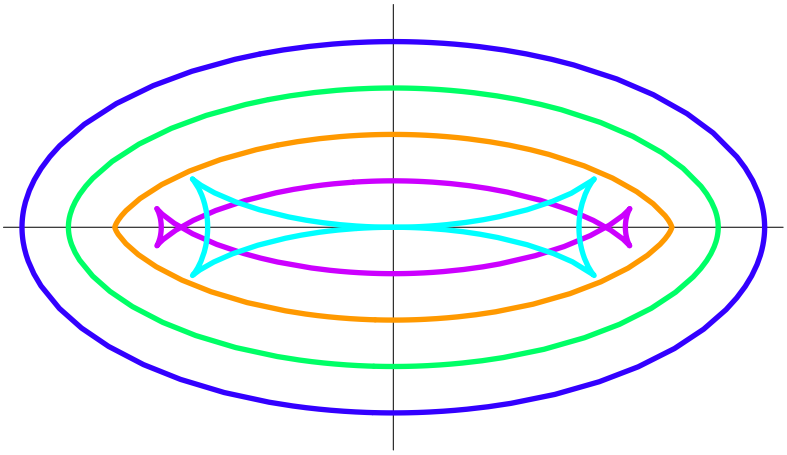


Figure 4.13: Four parallel curves to $\mathsf{ellipse}[2,1]$

Figure 4.13 illustrates some parallel curves to an ellipse. It shows that if $|r|$ is too large, a parallel curve may intersect itself and $|r|$ will not necessarily represent distance to the original curve. We can estimate when this first happens, and at the same time compute the curvature of the parallel curve.

Lemma 4.15. *Let $\boldsymbol{\alpha}$ be a regular plane curve. Then the curve $\mathsf{parcurve}[\boldsymbol{\alpha}, r]$ is regular at those t for which $1 - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t) \neq 0$. Furthermore, its curvature is given by*

$$\boldsymbol{\kappa 2}[\mathsf{parcurve}[\boldsymbol{\alpha}, r]](t) = \frac{\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)}{\big|1 - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)\big|}.$$

Proof. By Theorem 1.20, page 16, and Lemma 4.14 we can assume that $\boldsymbol{\alpha}$ has unit speed. Write $\boldsymbol{\beta}(t) = \mathsf{parcurve}[\boldsymbol{\alpha}, r](t)$. Then $\boldsymbol{\beta} = \boldsymbol{\alpha} + r\,J\boldsymbol{\alpha}'$ so that by Lemma 1.21 we have

$$\boldsymbol{\beta}' = \boldsymbol{\alpha}' + r\,J\boldsymbol{\alpha}'' = \boldsymbol{\alpha}' + r\,J^2\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\boldsymbol{\alpha}' = \big(1 - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big)\boldsymbol{\alpha}'.$$

The regularity statement follows. Also, we compute

$$\boldsymbol{\beta}'' = \big(1 - r\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big)\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]J\boldsymbol{\alpha}' - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]'\boldsymbol{\alpha}'.$$

Hence

$$\boldsymbol{\kappa 2}[\boldsymbol{\beta}] = \frac{\boldsymbol{\beta}'' \cdot J\boldsymbol{\beta}'}{\|\boldsymbol{\beta}'\|^3} = \frac{\big(1 - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big)^2 \boldsymbol{\kappa 2}[\boldsymbol{\alpha}]}{\big|1 - r\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big|^3} = \frac{\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]}{\big|1 - r\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big|},$$

as stated. ∎

4.6 Pedal Curves

Let $\boldsymbol{\alpha}$ be a curve in the plane, and let $\mathbf{p} \in \mathbb{R}^2$. The locus of base points $\boldsymbol{\beta}(t)$ of a perpendicular line let down from $\mathbf{p}$ to the tangent line to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t)$ is called the pedal curve to $\boldsymbol{\alpha}$ with respect to $\mathbf{p}$. It follows that $\boldsymbol{\beta}(t) - \mathbf{p}$ is the projection of $\boldsymbol{\alpha}(t) - \mathbf{p}$ in the $J\boldsymbol{\alpha}'(t)$ direction, as shown in Figure 4.15. This enables us to give a more formal definition:

Definition 4.16. *The* ***pedal curve*** *of a regular curve* $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ *with respect to a point* $\mathbf{p} \in \mathbb{R}^2$ *is defined by*

$$\mathsf{pedal}[\boldsymbol{\alpha}, \mathbf{p}](t) = \mathbf{p} + \frac{\big(\boldsymbol{\alpha}(t) - \mathbf{p}\big) \cdot J\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|^2} J\boldsymbol{\alpha}'(t).$$

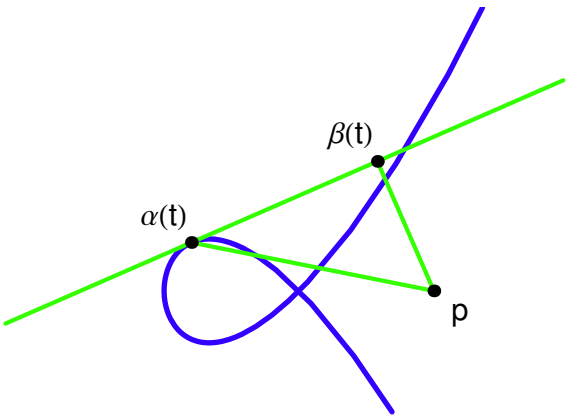


Figure 4.14: The definition of a pedal curve

The proof of the following lemma follows the model of Theorem 1.20; see Exercise 13.

Lemma 4.17. *The definition of the pedal curve of* $\boldsymbol{\alpha}$ *is independent of the parametrization of* $\boldsymbol{\alpha}$*, so that*

$$\mathsf{pedal}[\boldsymbol{\alpha} \circ h,\ \mathbf{p}] = \mathsf{pedal}[\boldsymbol{\alpha}, \mathbf{p}] \circ h.$$

Other examples are:

(i) The pedal curve of a parabola with respect to its vertex is a cissoid; see Exercise 12.

(ii) The pedal curve of a circle with respect to any point $\mathbf{p}$ other than the center of the circle is a limaçon. If $\mathbf{p}$ lies on the circumference of the circle, the limaçon reduces to a cardioid. The pedal curve of a circle with respect to its center is the circle itself; see Exercise 17.

(iii) The pedal curve of $\mathsf{cardioid}[a]$ (see page 45) with respect to its cusp point is called ***Cayley's sextic***.[3] Its equation is

$$\mathsf{cayleysextic}[a](t) = 4a\left((2\cos t - 1)\cos^4\frac{t}{2},\ \sin\frac{3t}{2}\cos^3\frac{t}{2}\right).$$

Figure 4.16 plots the cardioid and Cayley's sextic simultaneously.

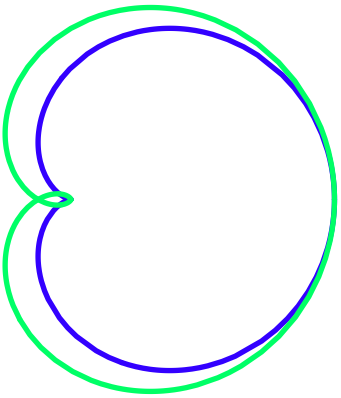


Figure 4.15: Cardioid and (lighter) its pedal curve

(iv) The pedal curve of

$$\boldsymbol{\alpha}(t) = \left(t,\ \frac{t^3}{3} + \frac{1}{2}\right),$$

illustrated in Figure 4.14, has a singularity, discussed in [BGM, page 5].

---

[3] 

Arthur Cayley (1821–1895). One of the leading English mathematicians of the 19th century; his complete works fill many volumes. Particularly known for his work on matrices, elliptic functions and nonassociative algebras. His first 14 professional years were spent as a lawyer; during that time he published over 250 papers. In 1863 Cayley was appointed Sadleirian professor of mathematics at Cambridge with a greatly reduced salary.

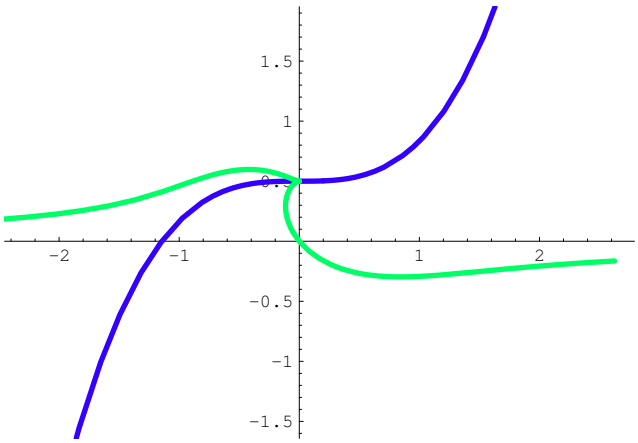


Figure 4.16: The cubic $y = \frac{1}{3}x^3 + \frac{1}{2}$ and (lighter) its pedal curve

For more information about pedal curves see [Law, pages 46–49], [Lock, pages 153–160], and [Zwi, pages 150–158]. We shall return to pedal curves, and parametrize them with an angle, as part of the study of ovals in Section 6.8.

4.7 Exercises

M **1.** Plot evolutes of a cycloid, a cardioid, and a logarithmic spiral. Show that the evolute of a cardioid is another cardioid, and that the evolute of a logarithmic spiral is another logarithmic spiral.

2. Determine conditions under which the evolute of a cycloid is another cycloid.

3. Prove Lemma 4.2.

M **4.** Plot normal lines to a cardioid, making the lines sufficiently long so that they intersect.

M **5.** Plot as one graph four parallel curves to a lemniscate. Do the same for a cardioid and a deltoid (see page 57).

6. Show that the curve defined by

$$\boldsymbol{\gamma}(t) = a\, e^{it} \sum_{k=0}^{n} \frac{(-it)^k}{k!}$$

is the nth involute starting at $(a, 0)$ of a circle of radius a.

7. A ***strophoid*** of a curve $\boldsymbol{\alpha}$ with respect to a point $\mathbf{p} \in \mathbb{R}^2$ is a curve $\boldsymbol{\gamma}$ such that

$$\big\|\boldsymbol{\alpha}(t) - \boldsymbol{\gamma}(t)\big\| = \big\|\boldsymbol{\alpha}(t) - \mathbf{p}\big\| \qquad \text{and} \qquad \boldsymbol{\gamma}(t) = s\,\boldsymbol{\alpha}(t),$$

for some s. Show that $\boldsymbol{\alpha}$ has two strophoids and find the equations for them. (A special case was defined in Exercise 1 on page 85.)

8. Show that the involute of a catenary is a tractrix.

M **9.** A ***point of inflection*** of a plane curve $\boldsymbol{\alpha}$ is defined to be a point $\boldsymbol{\alpha}(t_0)$ for which $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t_0) = 0$; a ***strong inflection point*** is a point $\boldsymbol{\alpha}(t_0)$ for which there exists $\varepsilon > 0$ such that $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)$ is negative for $t_0 - \varepsilon < t < t_0$ and positive for $t_0 < t < t_0 + \varepsilon$, or vice versa.

(a) Show that if $\boldsymbol{\alpha}(t_0)$ is a strong inflection point, and if $t \mapsto \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)$ is continuous at t_0, then $\boldsymbol{\alpha}(t_0)$ is an inflection point.

(b) For the curve $t \mapsto (t^3, t^5)$ show that $(0,0)$ is a strong inflection point which is not an inflection point. Plot the curve.

10. Let $\boldsymbol{\alpha}(t_0)$ be a strong inflection point of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$. Show that any involute of $\boldsymbol{\alpha}$ must have discontinuous curvature at t_0. This accounts for the cusps on the involutes of a figure eight and of a cubic parabola.

M **11.** Find and draw the pedal curve of (i) an ellipse with respect to its center, and (ii) a catenary with respect to the origin.

12. Find the parametric form of the pedal curve of the parabola $t \mapsto (2at,\ t^2)$ with respect to the point $(0,b)$, where $a^2 \neq b$. Show that the nonparametric form of the pedal curve is

$$(x^2 + y^2)y + (a^2 - b)x^2 - 2by^2 + b^2y = 0.$$

13. Prove Lemma 4.17.

14. An ordinary pendulum swings back and forth in a circular arc. The oscillations are not isochronous. To phrase it differently, the time it takes for the (circular) pendulum to go from its starting point to its lowest point will be almost, but not quite, independent of the height from which the pendulum is released. In 1680 Huygens[4] observed two important facts:



[4] Christiaan Huygens (1629–1695). A leading Dutch scientist of the 17th century. As a mathematician he was a major precursor of Leibniz and Newton. His astronomical contributions include the discovery of the rings of Saturn. In 1656, Huygens patented the first pendulum clock.

(a) a pendulum that swings back and forth in an inverted cycloidial arc *is* isochronous, and (b) the involute of a cycloid is a cycloid. Combined, these facts show that the involute of an inverted cycloidial arc can be used to constrain a pendulum so that it is isochronous. Prove (b). For (a), see the end of Notebook 4.

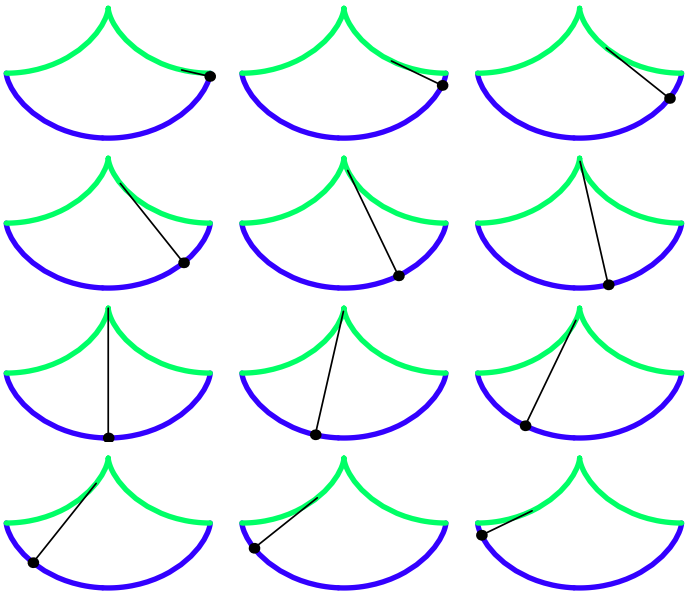


Figure 4.17: A cycloidal pendulum in action

15. Show that the formula for the pedal curve of $\boldsymbol{\alpha}$ with respect to the origin can be written as

$$\mathsf{pedal}[\boldsymbol{\alpha}, \mathbf{0}] = \frac{\boldsymbol{\alpha}\overline{\boldsymbol{\alpha}'} - \overline{\boldsymbol{\alpha}}\boldsymbol{\alpha}'}{2\overline{\boldsymbol{\alpha}'}}.$$

See [Zwi, page 150].

16. The ***contrapedal*** of a plane curve $\boldsymbol{\alpha}$ is defined analogously to the pedal of $\boldsymbol{\alpha}$. It is the locus of bases of perpendicular lines let down from a point $\mathbf{p}$ to a variable *normal* line to $\boldsymbol{\alpha}$. Prove that the exact formula is

$$\mathsf{contrapedal}[\boldsymbol{\alpha}, \mathbf{p}](t) = \mathbf{p} + \frac{\big(\boldsymbol{\alpha}(t) - \mathbf{p}\big) \cdot \boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|^2}\boldsymbol{\alpha}'(t).$$

M **17.** Show that the pedal curve of a circle with respect to a point on the circle is a cardioid. Plot pedal curves of a circle with respect to points inside the circle, and then with respect to points outside the circle.

Notebook 4

■ N4.0

The computer can be used to carry out many of the constructions of Chapter 4, and it is particularly adapted for studying iterated operations and illustrating complicated examples of curves. This notebook uses *Mathematica* programs to define and construct evolutes, involutes, parallel curves, and pedal curves. Programs are used both to draw the new curve and to study the resulting curve analytically, for example, by computing curvature. An animation of a cycloidal pendulum is given in N4.7.

Below, we shall often plot different curves together using common axes. For such presentations, which are somewhat special to Chapter 4, we make the following declaration in order to provide a default plotting option of up to ten different colors for multiple curves.

```
SetOptions[ParametricPlot,
  AspectRatio->Automatic, Compiled->False,
  PlotStyle->Table[{Hue[.7n], AbsoluteThickness[2]}, {n,10}]
]
```

We next list explicitly the curves that will provide the main examples of the evolute and related constructions.

```
cissoid[a_][t_]:= 2a{t^2, t^3}/(1+t^2)
tractrix[a_][t_]:= a{Sin[t], Cos[t]+Log[Tan[t/2]]}
catenary[a_][t_]:= {a Cosh[t/a], t}
cardioid[a_][t_]:= 2a(1+Cos[t]){Cos[t], Sin[t]}
logspiral[a_,b_][t_]:= a{Exp[b t]Cos[t], Exp[b t]Sin[t]}
cycloid[a_,b_][t_]:= {a t - b Sin[t], a - b Cos[t]}
```

The next command, similar to the one in N2.4 but with an offset vector (null by default), allows the quick labelling of diagrams (for example, on Figure 4.1).

```
tx[A_,v_, o_:{0,0}]:= Text[FontForm[A,{"Helvetica",18}], v, o]
```

Standard conic curves, including the circle, are given in N4.0.1. This subsection also includes the definitions of the norm L, rotation operator J, signed curvature $\kappa 2$, and various arc length functions. All these commands will be loaded in this notebook's initialization process.

■ N4.0.1

■ N4.1

The definition of the evolute involves the radius of curvature $1/\kappa 2$, and a unit normal vector. Rather than importing the definition of the curvature $\kappa 2$, it is simpler to give a direct formula for the evolute of a curve:

```
evolute[α_][t_]:= α[t] + α'[t].α'[t]/(α''[t].J[α'[t]]) J[α'[t]]
```

```
evolute[ellipse[a,b]][t]
%//Simplify
```

The first expression for the evolute of the ellipse can be simplified further. In these circumstances, we freely use *Mathematica*'s powerful commands for manipulating algebraic expressions. These include Expand, Factor, Simplify, Together, and if necessary FullSimplify, Cancel. Trigonometric simplifications are often automatic. Frequently, however, it is not clear how to simplify complicated expressions, and a graph can best explain the situation.

At times, it is appropriate to use PowerExpand, but care is needed as this command ignores negative roots, and can lead to sign errors:

```
f[x_]:= PowerExpand[Sqrt[x^2]]
Plot[f[x], {x,-1,1}]
```

No simplification is required to verify that the evolute of a circle consists of a single point!

```
evolute[circ[a]][t]
```

Plotting a curve together with its evolute is now an easy matter. The options set in N4.0 ensure that the two curves are automatically colored blue and green respectively.

```
Module[{α,ϵ},
  α= ellipse[1.5,1];
  ϵ= evolute[α];
  ParametricPlot[{α[t],ϵ[t]}, {t,0,2π}, ImageSize->400];
]
```

Figure 4.1

Figure 4.2 illustrates the center of curvature, and is hidden below.

N4.1.1

N4.2

Iterated evolutes may be constructed using the Nest command, though it may be easier in practice to take a more direct approach.

```
evolute[α_,n_]:= Nest[evolute, α, n]
```

The cases n=0,1,2,3 correspond to the definitions

```
α:= cissoid[1]
β:= evolute[α]
γ:= evolute[β]
δ:= evolute[γ]
```

We can append the command Timing, so as to compare the computation time that each requires.

```
δ[t]//Simplify//Timing
evolute[α,3][t]//Simplify//Timing
```

```
ParametricPlot[Evaluate[{α[t], β[t],γ[t], δ[t]}], {t,-2,2},
  ImageSize->400]
```

Figure 4.3

The next figure confirms that the evolute of a tractrix is a catenary.

```
Module[{α,β,ϵ},
  α= tractrix[1];
  β[t_]= catenary[1][t]+{1,1}/2;
  ϵ= evolute[α];
  ParametricPlot[{α[t], ϵ[t], β[α[t][[2]]]}, {t,0,π}, Axes->None];
]
```

Figure 4.4

■ N4.3

Our involute command is more complicated as it involves the arc length function and a definite integral.

```
involute[α_,c_:0][t_]:=
Module[{d,u},
  d= D[α[u],u];
  α[u] - length[c,u][α] d/L[d]/.u->t
]
```

The construction c_:0 represents a default assignment. If no second argument is inserted into involute (as will often be the case below), then the symbol c is assigned the value 0 in the program.

```
α:= circ[1]
involute[α,b][t]//Simplify
involute[α][t]//Simplify
```

```
τ:= tractrix[1]
length[1,2][τ]
```

```
f[t_]= involute[τ, .1][t]
g[t_]= involute[τ, π/2][t]
```

A singularity in the curve can cause the involute to "jump." In such case, variants of the involute can be produced by modifying the definition.

```
ParametricPlot[{τ[t],f[t]}, {t,.1,π}]
ParametricPlot[{τ[t],g[t]}, {t,.1,π}]
```

```
Clear[α, z, ι]
α:= circ[1]
ι= involute[α];
s:= 2
z[1]:= {Thickness[.005], Line[{α[s],ι[s]}]}
z[2]:= {PointSize[.03], Point[α[s]], Point[ι[s]], Point[α[0]]}
z[3]:= {tx["β(s)",α[s],{1,-1.6}], tx["β(c)",α[0],{1.5,-.3}]}
ParametricPlot[{α[t], ι[t]}, {t,0,3}, Epilog->Array[z,3],
  Axes->False]
```

Figure 4.5

```
p= 15;
η= involute[α, π/2];
ParametricPlot[{α[t], ι[t], η[t+π/2]}, {t,0, p},
  PlotRange->{{-p,p},{-p,p}}]
```

The definition of the eight curve is

```
eight[t_]:= {Sin[t], Sin[t]Cos[t]}
```

```
involute[eight][t]//Simplify
```

```
lengthn[0,1][eight]//N
```

N4.3.1

N4.4

Here is the definition of the circle that best approximates a curve at a given parameter value. It uses the *Mathematica* expression Circle[v,r] that represents a circle of radius r and center v, drawn by adding the command Graphics.

```
osculatingcircle[α_][t_]:= Circle[evolute[α][t], 1/κ2[α][t]]
```

```
Clear[α]
osculatingcircle[α][t]
```

It is an easy matter to draw a family of osculating circles to a given curve, which is then highlighted by their presence.

```
Module[{oc,ga},
  α= logspiral[1,1/2];
  oc[n_]= osculatingcircle[α][n π/24];
  ga= Graphics[{Hue[0], Table[oc[n],{n,-12,12}]}];
  Show[ga, AspectRatio->Automatic, PlotRange->All]
]
```

Figure 4.8

N4.4.1

N4.5

The following definitions call upon both the rotation operator J and the norm function L defined in N4.0.1.

```
tangentline[α_][s_,t_]:= Line[{α[t], α[t]+s α'[t]/L[α'[t]]}]
normalline[α_][s_,t_]:= Line[{α[t], α[t]+s J[α'[t]]/L[α'[t]]}]
```

```
α:= ellipse[1.5,1]
tl[n_]:= tangentline[α][2, n π/36]
ga:= Graphics[{Hue[.7], Array[tl,72]}]
pl:= ParametricPlot[α[t], {t,0,2π}, DisplayFunction->Identity]
Show[ga,pl, DisplayFunction->$DisplayFunction]
```

Figure 4.10

The next program constructs normal lines to the ellipse. We see that if these normals are long enough, they highlight the ellipse's evolute.

```
normal2ellipse[s_]:=
Module[{α,nl,ga,pl},
  α= ellipse[1.5,1];
  nl[n_]= normalline[α][s, n π/36];
  ga= Graphics[{Hue[.7], Array[nl,72]}];
  pl= ParametricPlot[α[t], {t,0,2π}, DisplayFunction->Identity];
  Show[ga, pl, DisplayFunction->$DisplayFunction,
    AspectRatio->Automatic];
]
```

```
normal2ellipse[.7]
```

Figure 4.11

```
normal2ellipse[2.5]
```

Figure 4.12

The next tiny program produces, for a given curve, a "parallel" one formed by travelling a distance r along the normal direction from each point.

```
parcurve[α_,r_][t_]:=
Module[{d},
  d= D[α[u],u];
  α[u] + r J[d]/L[d]/.u->t
]
```

```
α:= ellipse[2,1]
ta:= Table[parcurve[α,s][t], {s,0,1,.25}]
ParametricPlot[Evaluate[ta], {t,0,2π},
  PlotRange->{{-2.1,2.1},{-1.2,1.2}}, Ticks->False,
  ImageSize->600]
```

Figure 4.13

The next two cells verify that the involute of the evolute of a given curve is the same as a parallel curve.

```
invev[α_,r_][t_]:=
Module[{ϵ},
  ϵ= evolute[α];
  involute[ϵ,0][t] - parcurve[α,r][t]//PowerExpand//Simplify
]
```

```
α[t_]:= {Cos[t], 2Sin[t]}
β[t_]:= {t, t^2}
invev[α,4][t]
invev[β,r][t]
```

■ N4.6

Here is the figure illustrating the definition of a pedal curve:

```
Module[{α,c,p,v,d,w,z},
  α[t_]={t^2-2, -t^3+t};
  c=.45; r= 2.1;
  p= {0,0}; v= α[c]; d=α'[c];
  w= v - (v.d)d/(d.d);
  z[1]= {Thickness[.005], Hue[.3], Line[{p,v}], Line[{p,w}],
    Line[{2v-w,2w-v}]};
  z[2]= {PointSize[.015], Point[p], Point[v], Point[w]};
  z[3]= {tx["p",p,{-2,1}], tx["α(t)",v,{1,-1.5}],
    tx["β(t)",w,{1,-1.5}]};
  ParametricPlot[α[t], {t,-2,1.3}, Epilog->Array[z,3],
    Axes->False, PlotRange->{{-2r,r},{-r,r}},
    ImageSize->500];
]
```

Figure 4.14

The construction of the pedal curve also involves the addition of a suitable multiple of the unit normal vector, by default starting from the origin:

```
pedal[α_,v_:{0,0}][t_]:=
Module[{d,u},
  d= D[α[u],u];
  v+(α[u]-v).J[d]/L[d]^2 J[d]/.u->t
]
```

```
pedal[parabola[a]][t]
```

```
α:= cardioid[1]
γ:= pedal[α]
γ[t]//Simplify
ParametricPlot[{α[t],γ[t]}, {t,0,2π}, Axes->False]
```

Figure 4.15

▪ N4.6.1

■ N4.7

The next cell contains an animation of the cycloidal pendulum. The two adjacent Do commands ensure that it goes back and forth (and hits someone on the head!). The use of Epilog is convenient here, as it saves having to display a line and point separately. However, it has the disadvantage that the plot range is determined by the curves, and should ideally be extended to take account of the extras.

```
Module[{δ,ϵ,y,z},
  δ= cycloid[-1,-1];
  ϵ= evolute[δ];
  y[a_]= {Thickness[.008], Line[{δ[a],ϵ[a]}], PointSize[.05],
    Point[δ[a]]};
  z[a_]:= ParametricPlot[{δ[t],ϵ[t]}//Evaluate, {t,-.001,2π+.001},
    Epilog->y[a],  Axes->None];
  Do[z[a], {a, .1, 2π+.1, π/4}];
  Do[z[2π-a], {a, .1, 9π/4+.1, π/4}]
]
```

The same program is adapted in the final hidden subsection to draw Figure 4.17.

▪ N4.7.1

Chapter 5

Determining a Plane Curve from Its Curvature

In this chapter we confront the following question: *to what extent does curvature determine a plane curve?* This question has two parts: *when do two curves have the same curvature*, and *when can a curve be determined from its curvature?*

To answer the first part, we begin by asking: *which transformations of the plane preserve curvature*? We have seen (in Theorem 1.20 on page 16) that the curvature of a plane curve is independent of the parametrization, at least up to sign. Therefore, without loss of generality, we can assume in this chapter that all curves have unit speed.

It is intuitively clear that the image of a plane curve $\boldsymbol{\alpha}$ under a rotation or translation has the same curvature as the original curve $\boldsymbol{\alpha}$. Rotations and translations are examples of Euclidean motions of $\mathbb{R}^2$, meaning those maps of $\mathbb{R}^2$ onto itself which do not distort distances. We discuss Euclidean motions in general in Section 5.1, and introduce the important concept of a *group*. We specialize Euclidean motions to the plane in Section 5.2, while the case of orthogonal transformations and rotations of $\mathbb{R}^3$ will be taken up in Chapter 23.

The invariance of curvature under Euclidean motions is established in Section 5.3, in which we also prove the *Fundamental Theorem of Plane Curves*. This important theorem states that two unit-speed curves in $\mathbb{R}^2$ that have the same curvature differ only by a Euclidean motion.

The second part of our question now becomes: *can a unit-speed curve in* $\mathbb{R}^2$ *be determined up to a Euclidean motion from its curvature*? In Section 5.3, we give an explicit system of differential equations for determining a plane curve from its curvature. In simple cases this system can be solved explicitly, but in the general case only a numerical solution is possible. Such solutions are illustrated in Section 5.4. So the answer to the second part of the question is effectively "*yes*."

5.1 Euclidean Motions

In order to discuss the invariance of the curvature of plane curves under rotations and translations, we need to make precise what we mean by "invariance." This requires discussing various kinds of maps of $\mathbb{R}^2$ to itself. We begin by defining transformations of $\mathbb{R}^n$, before restricting to $n = 2$.

First, we define several important types of maps that are linear, a concept introduced in Chapter 1. The linearity property

$$A(a\,\mathbf{p} + b\,\mathbf{q}) = a\,A\mathbf{p} + b\,A\mathbf{q},$$

for $a, b \in \mathbb{R}$, implies that each of these maps takes the origin into the origin.

Definition 5.1. *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be a nonsingular linear map.*

(i) *We say that A is* ***orientation-preserving*** *if* $\det A$ *is positive, or* ***orientation-reversing*** *if* $\det A$ *is negative.*

(ii) *A is called an* ***orthogonal transformation*** *if*

$$A\mathbf{p} \cdot A\mathbf{q} = \mathbf{p} \cdot \mathbf{q}$$

for all $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$.

(iii) *A* ***rotation*** *of $\mathbb{R}^n$ is an orientation-preserving orthogonal transformation.*

(iv) *A* ***reflection*** *of $\mathbb{R}^n$ is an orthogonal transformation of the form $\mathsf{refl}_{\mathbf{q}}$, where*

$$\mathsf{refl}_{\mathbf{q}}(\mathbf{p}) = \mathbf{p} - \frac{2(\mathbf{p} \cdot \mathbf{q})}{\|\mathbf{q}\|^2}\mathbf{q} \tag{5.1}$$

for all $\mathbf{p} \in \mathbb{R}^n$ and some fixed $\mathbf{q} \neq \mathbf{0}$.

Lemma 5.2. *Let $A\colon \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal transformation. Then*

$$\det A = \pm 1.$$

Proof. This well-known result follows from the fact that, if we represent A by a matrix relative to an orthonormal basis, and $\mathbf{p}, \mathbf{q}$ by column vectors, then the orthogonality property is equivalent to asserting that

$$(A\mathbf{p})^T(A\mathbf{q}) = \mathbf{p}^T\mathbf{q}, \tag{5.2}$$

where A^T denotes the transpose of A. It follows that A is orthogonal if and only if A^TA is the identity matrix, and

$$1 = \det(A^TA) = \det(A^T)\det A = (\det A)^2. \quad \blacksquare$$

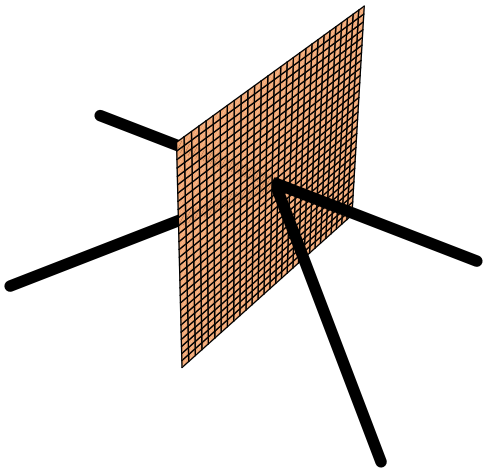


Figure 5.1: Reflection in 3 dimensions

We can characterize reflections geometrically.

Lemma 5.3. *Let* $\mathsf{refl}_{\mathbf{q}}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be a reflection,* $\mathbf{q} \neq \mathbf{0}$. *Then* $\det(\mathsf{refl}_{\mathbf{q}}) = -1$. *Furthermore,* $\mathsf{refl}_{\mathbf{q}}$ *restricts to the identity map on the orthogonal complement* Π *of* $\mathbf{q}$ (*the hyperplane of* $\mathbb{R}^n$ *of dimension* $n-1$ *of the vectors orthogonal to* $\mathbf{q}$) *and reverses directions parallel to* $\mathbf{q}$. *Conversely, any hyperplane gives rise to a reflection* $\mathsf{refl}_{\mathbf{q}}$ *which restricts to the identity map on* Π *and reverses directions on the orthogonal complement of* Π.

Proof. Let Π be the orthogonal complement of $\mathbf{q}$, that is

$$\Pi = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} \cdot \mathbf{q} = \mathbf{0}\}.$$

Clearly, $\dim(\Pi) = n-1$, and it is easy to check that $\mathsf{refl}_{\mathbf{q}}(\mathbf{p}) = \mathbf{p}$ for all $\mathbf{p} \in \Pi$. Since $\mathsf{refl}_{\mathbf{q}}(\mathbf{q}) = -\mathbf{q}$, it follows that $\mathsf{refl}_{\mathbf{q}}$ reverses directions on the orthogonal complement of Π.

To show that $\det(\mathsf{refl}_{\mathbf{q}}) = -1$, choose a basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_{n-1}\}$ of Π. It follows that $\{\mathbf{q}, \mathbf{f}_1, \ldots, \mathbf{f}_{n-1}\}$ is a basis of $\mathbb{R}^n$. Since the matrix of $\mathsf{refl}_{\mathbf{q}}$ with respect to this basis is

$$\begin{pmatrix} -1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix},$$

we have $\det(\mathsf{refl}_{\mathbf{q}}) = -1$.

Conversely, given a hyperplane Π, let $\mathbf{q}$ be a nonzero vector perpendicular to Π. Then the reflection $\mathsf{refl}_{\mathbf{q}}$ defined by (5.1) has the required properties. ∎

Definition 5.4. *Let* $\mathbf{q} \in \mathbb{R}^n$.

(i) *An* ***affine transformation*** *of* $\mathbb{R}^n$ *is a map* $F\colon \mathbb{R}^n \to \mathbb{R}^n$ *of the form*

$$F(\mathbf{p}) = A\mathbf{p} + \mathbf{q}$$

for all $\mathbf{p} \in \mathbb{R}^n$, *where* A *is a linear transformation of* $\mathbb{R}^n$. *We shall call* A *the* ***linear part*** *of* F. *An affine transformation* F *is said to be* ***nonsingular*** *if* A *is, and* F *is then* ***orientation-preserving*** *if* $\det A$ *is positive, or* ***orientation-reversing*** *if* $\det A$ *is negative.*

(ii) *A* ***translation*** *is an affine map as above with* A *the identity, that is a mapping* $\mathsf{tran}_{\mathbf{q}}\colon \mathbb{R}^n \to \mathbb{R}^n$ *of the form*

$$\mathsf{tran}_{\mathbf{q}}(\mathbf{p}) = \mathbf{p} + \mathbf{q}$$

for all $\mathbf{p} \in \mathbb{R}^n$.

(iii) *A* ***Euclidean motion*** *of* $\mathbb{R}^n$ *is an affine transformation whose linear part is an* orthogonal *transformation.*

(iv) *An* ***isometry*** *of* $\mathbb{R}^n$ *is a map* $F\colon \mathbb{R}^n \to \mathbb{R}^n$ *that preserves distance, so that*

$$\|F(\mathbf{p}_1) - F(\mathbf{p}_2)\| = \|\mathbf{p}_1 - \mathbf{p}_2\|$$

for all $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^n$. *It is easy to see that an isometry must be an injective map; we shall see below that it is necessarily bijective.*

Given an affine transformation, we can study its effect on any geometrical object in $\mathbb{R}^n$. Figure 5.2 illustrates a logarithmic spiral in $\mathbb{R}^2$ and its image under an affine transformation F, a process formalized in Definition 5.11 below. Notice that F is orientation-reversing, as the inward spiralling is changed from clockwise to counterclockwise.

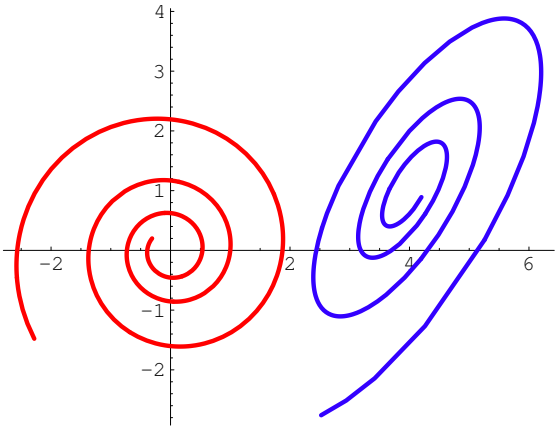


Figure 5.2: The effect of an affine transformation

To put these notions in their proper context, let us recall the following abstract algebra.

Definition 5.5. *A* ***group*** *consists of a nonempty set* $\mathscr{G}$ *and a multiplication*

$$\circ\colon \mathscr{G} \times \mathscr{G} \longrightarrow \mathscr{G}$$

with the following properties.

(i) *There exists an* ***identity element*** $\mathbf{1} \in \mathscr{G}$*, that is, an element* $\mathbf{1}$ *such that*

$$a \circ \mathbf{1} = \mathbf{1} \circ a = a$$

for all $a \in \mathscr{G}$*.*

(ii) *Multiplication is* ***associative****; that is,*

$$(a \circ b) \circ c = a \circ (b \circ c)$$

for all $a, b, c \in \mathscr{G}$*.*

(iii) *Every element* $a \in \mathscr{G}$ *has an* ***inverse****; that is, for each* $a \in \mathscr{G}$ *there exists an element* $a^{-1} \in \mathscr{G}$ *(that can be proved to be unique) such that*

$$a \circ a^{-1} = a^{-1} \circ a = \mathbf{1}.$$

Moreover, a ***subgroup*** *of a group* $\mathscr{G}$ *is a nonempty subset* $\mathscr{S}$ *of* $\mathscr{G}$ *closed under multiplication and inverses, that is*

$$a \circ b \in \mathscr{S} \qquad \text{and} \qquad a^{-1} \in \mathscr{S},$$

for all $a, b \in \mathscr{S}$*.*

It is easy to check that a subgroup of a group is itself a group. See any book on abstract algebra for more details on groups, for example [Scott] or [NiSh].

Many sets of maps form groups. Recall that if

$$F\colon \mathbb{R}^m \longrightarrow \mathbb{R}^n \qquad \text{and} \qquad G\colon \mathbb{R}^n \longrightarrow \mathbb{R}^p$$

are maps, their ***composition*** $G \circ F\colon \mathbb{R}^m \to \mathbb{R}^p$ is defined by

$$(G \circ F)(\mathbf{p}) = G\big(F(\mathbf{p})\big)$$

for $\mathbf{p} \in \mathbb{R}^m$. Denote by $\mathsf{GL}(n)$ the set of all nonsingular linear maps of $\mathbb{R}^n$ into itself. It is easy to verify that composition makes $\mathsf{GL}(n)$ into a group. Similarly, we may speak of the following groups:

- the group $\mathsf{Affine}(\mathbb{R}^n)$ of nonsingular affine transformations of $\mathbb{R}^n$,

- the group $\mathsf{Orthogonal}(\mathbb{R}^n)$ of orthogonal transformations of $\mathbb{R}^n$,
- the group $\mathsf{Translation}(\mathbb{R}^n)$ of translations of $\mathbb{R}^n$,
- the group $\mathsf{Isometry}(\mathbb{R}^n)$ of isometries of $\mathbb{R}^n$,
- the group $\mathsf{Euclidean}(\mathbb{R}^n)$ of Euclidean motions of $\mathbb{R}^n$.

The verification that each is a group is easy. Moreover:

Theorem 5.6. *A map $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isometry of $\mathbb{R}^n$ if and only if it is a composition of a translation and an orthogonal transformation of $\mathbb{R}^n$. As a consequence, the group $\mathsf{Euclidean}(\mathbb{R}^n)$ coincides with the group $\mathsf{Isometry}(\mathbb{R}^n)$.*

Proof. It is easy to check that any orthogonal transformation or translation of $\mathbb{R}^n$ preserves the distance function. Hence the composition of an orthogonal transformation and a translation also preserves the distance function. This shows that $\mathsf{Euclidean}(\mathbb{R}^n) \subseteq \mathsf{Isometry}(\mathbb{R}^n)$.

Conversely, let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be an isometry, and define $G\colon \mathbb{R}^n \to \mathbb{R}^n$ by $G(\mathbf{p}) = F(\mathbf{p}) - F(\mathbf{0})$. Then G is an isometry and $G(\mathbf{0}) = \mathbf{0}$. To show that G preserves the scalar product of $\mathbb{R}^n$, we calculate

$$\begin{aligned} -2G(\mathbf{p}) \cdot G(\mathbf{q}) &= \|G(\mathbf{p}) - G(\mathbf{q})\|^2 - \|G(\mathbf{p}) - G(\mathbf{0})\|^2 - \|G(\mathbf{q}) - G(\mathbf{0})\|^2 \\ &= \|\mathbf{p} - \mathbf{q}\|^2 - \|\mathbf{p} - \mathbf{0}\|^2 - \|\mathbf{q} - \mathbf{0}\|^2 \\ &= -2\mathbf{p} \cdot \mathbf{q}. \end{aligned}$$

To show that G is linear, let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be an orthonormal basis of $\mathbb{R}^n$. By definition, each $\mathbf{e}_i$ has unit length and is perpendicular to every other $\mathbf{e}_j$. Then $\{G(\mathbf{e}_1), \ldots, G(\mathbf{e}_n)\}$ is also an orthonormal basis of $\mathbb{R}^n$. Thus, for any $\mathbf{p} \in \mathbb{R}^n$ we have

$$G(\mathbf{p}) = \sum_{j=1}^{n} \big(G(\mathbf{p}) \cdot G(\mathbf{e}_j)\big) G(\mathbf{e}_j) = \sum_{j=1}^{n} (\mathbf{p} \cdot \mathbf{e}_j) G(\mathbf{e}_j)$$

(see Exercise 1). Because each mapping $\mathbf{p} \mapsto \mathbf{p} \cdot \mathbf{e}_j$ is linear, so is G. Hence G is an orthogonal transformation of $\mathbb{R}^n$, so that F is the composition of the orthogonal transformation G and the translation $\mathsf{tran}_{F(\mathbf{0})}$. We have proved that $\mathsf{Isometry}(\mathbb{R}^n) \subseteq \mathsf{Euclidean}(\mathbb{R}^n)$. Since we have already proved the reverse inclusion, it follows that $\mathsf{Isometry}(\mathbb{R}^n) = \mathsf{Euclidean}(\mathbb{R}^n)$. ∎

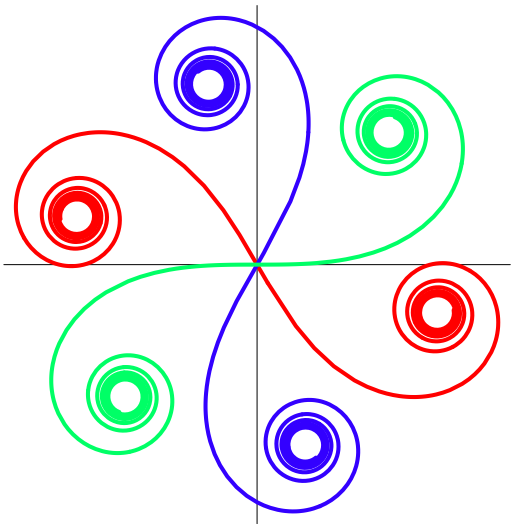


Figure 5.3: Bouquet of clothoids

5.2 Isometries of the Plane

We now specialize to $\mathbb{R}^2$. First, we give a characterization of transformations of $\mathbb{R}^2$ in terms of the linear map J, defined on page 3, represented by the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 5.7. *Let A be an orthogonal transformation of $\mathbb{R}^2$. Then*

$$AJ = \varepsilon JA, \tag{5.3}$$

$$\det A = \varepsilon, \tag{5.4}$$

where

$$\varepsilon = \begin{cases} +1 & \text{if } A \text{ is orientation-preserving,} \\ -1 & \text{if } A \text{ is orientation-reversing.} \end{cases}$$

Furthermore, $\det A = -1$ *if and only if A is a reflection.*

Proof. Let $\mathbf{p} \in \mathbb{R}^2$ be a nonzero vector. Then $JA\mathbf{p} \cdot A\mathbf{p} = \mathbf{0}$; since A is orthogonal,

$$AJ\mathbf{p} \cdot A\mathbf{p} = J\mathbf{p} \cdot \mathbf{p} = \mathbf{0}.$$

Thus both $AJ\mathbf{p}$ and $JA\mathbf{p}$ are perpendicular to $A\mathbf{p}$, so that for some $\lambda \in \mathbb{R}$ we have $AJ\mathbf{p} = \lambda JA\mathbf{p}$. Since both A and J are orthogonal transformations, it is

easy to see that $|\lambda| = 1$. One can now obtain both (5.3) and (5.4) by computing the matrix of A with respect to the basis $\{\mathbf{p}, J\mathbf{p}\}$ of $\mathbb{R}^2$.

Lemma 5.3 implies that the determinant of any reflection of $\mathbb{R}^2$ is -1. Conversely, an orthogonal transformation $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ with $\det A = -1$ must have 1 and -1 as its eigenvalues. Let $\mathbf{q} \in \mathbb{R}^2$ be such that $A(\mathbf{q}) = -\mathbf{q}$ and $\|\mathbf{q}\| = 1$. Then $A = \mathsf{refl}_{\mathbf{q}}$. ∎

The following theorem is due to Chasles.[1]

Theorem 5.8. *Every isometry of $\mathbb{R}^2$ is the composition of translations, reflections and rotations.*

Proof. Theorem 5.6 implies that an isometry F of $\mathbb{R}^2$ is a Euclidean motion. There is an orthogonal transformation $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{q} \in \mathbb{R}^2$ such that $F(\mathbf{p}) = A(\mathbf{p}) + \mathbf{q}$. Otherwise said,

$$F = \mathsf{tran}_{\mathbf{q}} \circ A.$$

From Lemma 5.7 we know that A is a rotation if $\det A = 1$ or a reflection if $\det A = -1$. ∎

A standard procedure for simplifying the equation

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0$$

of a conic into "normal form" corresponds to first rotating the vector (x, y) so as to eliminate the term in xy (in the new coordinates) and then translating it to eliminate the terms in x and y. But it is sometimes convenient to eliminate the linear terms first, so as determine the center, and then rotate. These two contrasting methods are illustrated in the Figure 5.4, starting with the ellipse in (say) the first quadrant.

By expressing an arbitrary vector $\mathbf{p} \in \mathbb{R}^2$ as a linear combination of two nonzero vectors $\mathbf{q}$, $J\mathbf{q}$, and using (5.3), we may deduce

Lemma 5.9. *Let A be a rotation of $\mathbb{R}^2$. Then the numbers*

$$\frac{A\mathbf{p} \cdot \mathbf{p}}{\|\mathbf{p}\|^2} \qquad \textit{and} \qquad \frac{AJ\mathbf{p} \cdot \mathbf{p}}{\|\mathbf{p}\|^2}$$

are each independent of the nonzero vector $\mathbf{p}$.



1 Michel Chasles (1793–1880). French geometer and mathematical historian. He worked on algebraic and projective geometry.

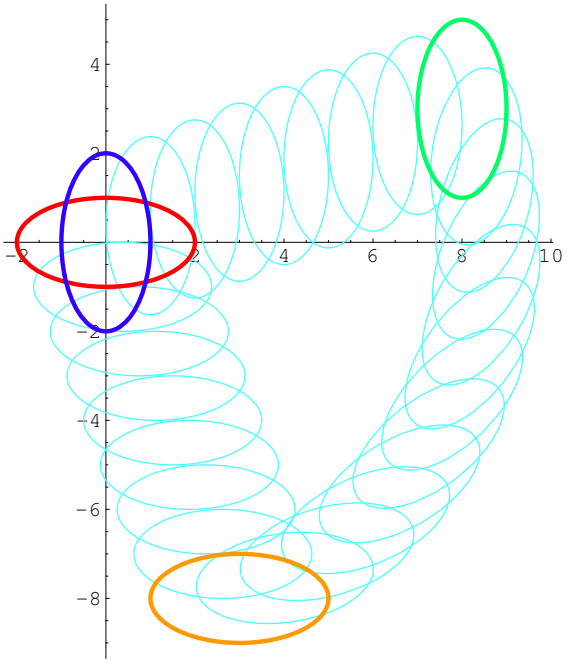


Figure 5.4: Translations and rotations of a conic

Definition 5.10. *Let A be a rotation of $\mathbb{R}^2$. The* ***angle of rotation*** *of A is any number θ such that*

$$\cos\theta = \frac{A\mathbf{p}\cdot\mathbf{p}}{\|\mathbf{p}\|^2} \qquad \textit{and} \qquad \sin\theta = \frac{AJ\mathbf{p}\cdot\mathbf{p}}{\|\mathbf{p}\|^2}$$

for all nonzero $\mathbf{p} \in \mathbb{R}^2$. Hence using the identification (1.1) *of $\mathbb{C}$ with $\mathbb{R}^2$, we can write*

$$A\mathbf{p} = (\cos\theta)\mathbf{p} + (\sin\theta)J\mathbf{p} = e^{i\theta}\mathbf{p}$$

for all $\mathbf{p} \in \mathbb{R}^2$.

It is clear intuitively that any mapping of $\mathbb{R}^2$ into itself maps a curve into another curve. Here is the exact definition:

Definition 5.11. *Let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a map, and let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a curve. The* ***image of $\boldsymbol{\alpha}$ under*** *F is the curve $F \circ \boldsymbol{\alpha}$.*

If for example F is an affine transformation with linear part represented by a matrix A, then the image of $\boldsymbol{\alpha}(t)$ is found by *pre*multiplying the column vector $\boldsymbol{\alpha}(t)$ by A and then adding the translation vector. This fact is the basis of the practical implementation of rotations in Notebooks 5 and 23. It was used to produce Figure 5.3, which shows a clothoid and its images under two rotations.

5.3 Intrinsic Equations for Plane Curves

Since a Euclidean motion F of the plane preserves distance, it cannot stretch or otherwise distort a plane curve. We now show that F preserves the curvature up to sign.

Theorem 5.12. *The absolute value of the curvature and the derivative of arc length of a curve are invariant under Euclidean motions of $\mathbb{R}^2$. The signed curvature $\boldsymbol{\kappa 2}$ is preserved by an orientation-preserving Euclidean motion of $\mathbb{R}^2$ and changes sign under an orientation-reversing Euclidean motion.*

Proof. Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a curve, and let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a Euclidean motion. Let A denote the linear part of F, so that for all $\mathbf{p} \in \mathbb{R}^2$ we have $F(\mathbf{p}) = A\mathbf{p} + F(0)$. Define a curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^2$ by $\boldsymbol{\gamma} = F \circ \boldsymbol{\alpha}$; then for $a < t < b$ we have

$$\boldsymbol{\gamma}(t) = A\boldsymbol{\alpha}(t) + F(0).$$

Hence $\boldsymbol{\gamma}'(t) = A\boldsymbol{\alpha}'(t)$ and $\boldsymbol{\gamma}''(t) = A\boldsymbol{\alpha}''(t)$. Let s_α and s_γ denote the arc length functions with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Since A is an orthogonal transformation, we have

$$s'_\gamma(t) = \big\|\boldsymbol{\gamma}'(t)\big\| = \big\|A\boldsymbol{\alpha}'(t)\big\| = \big\|\boldsymbol{\alpha}'(t)\big\| = s'_\alpha(t).$$

We compute the curvature of $\boldsymbol{\gamma}$ using (5.3):

$$\begin{aligned}\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](t) \;&=\; \frac{\boldsymbol{\gamma}''(t)\cdot J\boldsymbol{\gamma}'(t)}{\big\|\boldsymbol{\gamma}'(t)\big\|^3} \;=\; \frac{A\boldsymbol{\alpha}''(t)\cdot \varepsilon A\, J\boldsymbol{\alpha}'(t)}{\big\|A\boldsymbol{\alpha}'(t)\big\|^3} \\ &= \varepsilon\frac{\boldsymbol{\alpha}''(t)\cdot J\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|^3} = \varepsilon\,\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t). \;\blacksquare\end{aligned}$$

Next, we prove the converse of Theorem 5.12.

Theorem 5.13. (Fundamental Theorem of Plane Curves, Uniqueness) *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ be unit-speed regular curves in $\mathbb{R}^2$ defined on the same interval (a,b), and having the same signed curvature. Then there is an orientation-preserving Euclidean motion F of $\mathbb{R}^2$ mapping $\boldsymbol{\alpha}$ into $\boldsymbol{\gamma}$.*

Proof. Fix $s_0 \in (a,b)$. Clearly, there exists a translation of $\mathbb{R}^2$ taking $\boldsymbol{\alpha}(s_0)$ into $\boldsymbol{\gamma}(s_0)$. Moreover, we can find a rotation of $\mathbb{R}^2$ that maps $\boldsymbol{\alpha}'(s_0)$ into $\boldsymbol{\gamma}'(s_0)$. Thus there exists an orientation-preserving Euclidean motion F of $\mathbb{R}^2$ such that

$$F\big(\boldsymbol{\alpha}(s_0)\big) = \boldsymbol{\gamma}(s_0) \qquad \text{and} \qquad F\big(\boldsymbol{\alpha}'(s_0)\big) = \boldsymbol{\gamma}'(s_0).$$

To show that $F \circ \boldsymbol{\alpha}$ coincides with $\boldsymbol{\gamma}$, we define a real-valued function f by

$$f(s) = \big\|(F\circ\boldsymbol{\alpha})'(s) - \boldsymbol{\gamma}'(s)\big\|^2$$

for $a < s < b$. The derivative of f is easily computed to be

$$\begin{aligned} f'(s) &= 2\big((F\circ\boldsymbol{\alpha})''(s) - \boldsymbol{\gamma}''(s)\big)\cdot\big((F\circ\boldsymbol{\alpha})'(s) - \boldsymbol{\gamma}'(s)\big) \\ &= 2\big((F\circ\boldsymbol{\alpha})''(s)\cdot(F\circ\boldsymbol{\alpha})'(s) + \boldsymbol{\gamma}''(s)\cdot\boldsymbol{\gamma}'(s) \\ &\qquad -(F\circ\boldsymbol{\alpha})''(s)\cdot\boldsymbol{\gamma}'(s) - (F\circ\boldsymbol{\alpha})'(s)\cdot\boldsymbol{\gamma}''(s)\big). \end{aligned} \tag{5.5}$$

Since both $F\circ\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have unit speed, it follows that

$$(F\circ\boldsymbol{\alpha})''(s)\cdot(F\circ\boldsymbol{\alpha})'(s) = \boldsymbol{\gamma}''(s)\cdot\boldsymbol{\gamma}'(s) = 0.$$

Hence (5.5) reduces to

$$f'(s) = -2\big((F\circ\boldsymbol{\alpha})''(s)\cdot\boldsymbol{\gamma}'(s) + (F\circ\boldsymbol{\alpha})'(s)\cdot\boldsymbol{\gamma}''(s)\big). \tag{5.6}$$

Let A denote the linear part of the motion F. Since $(F\circ\boldsymbol{\alpha})'(s) = A\boldsymbol{\alpha}'(s)$ and $(F\circ\boldsymbol{\alpha})''(s) = A\boldsymbol{\alpha}''(s)$, we can rewrite (5.6) as

$$f'(s) = -2\big(A\boldsymbol{\alpha}''(s)\cdot\boldsymbol{\gamma}'(s) + A\boldsymbol{\alpha}'(s)\cdot\boldsymbol{\gamma}''(s)\big). \tag{5.7}$$

Now we use the assumption that $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}] = \boldsymbol{\kappa 2}[\boldsymbol{\gamma}]$ and (5.7) to get

$$\begin{aligned} f'(s) &= -2\big(\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](s)A\,J\boldsymbol{\alpha}'(s)\cdot\boldsymbol{\gamma}'(s) + A\boldsymbol{\alpha}'(s)\cdot\boldsymbol{\kappa 2}[\boldsymbol{\gamma}](s)J\boldsymbol{\gamma}'(s)\big) \\ &= -2\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](s)\big(JA\boldsymbol{\alpha}'(s)\cdot\boldsymbol{\gamma}'(s) + A\boldsymbol{\alpha}'(s)\cdot J\boldsymbol{\gamma}'(s)\big) = 0. \end{aligned}$$

Since $f(s_0) = 0$, we conclude that $f(s) = 0$ for all s. Hence $(F\circ\boldsymbol{\alpha})'(s) = \boldsymbol{\gamma}'(s)$ for all s, and so there exists $\mathbf{q}\in\mathbb{R}^2$ such that $(F\circ\boldsymbol{\alpha})(s) = \boldsymbol{\gamma}(s) - \mathbf{q}$ for all s. In fact, $\mathbf{q} = 0$ because $F\big(\boldsymbol{\alpha}(s_0)\big) = \boldsymbol{\gamma}(s_0)$. Thus the Euclidean motion F maps $\boldsymbol{\alpha}$ into $\boldsymbol{\gamma}$. ∎

Next, we turn to the problem of explicitly determining a plane curve from its curvature. Theory from Chapter 1 enables us to prove

Theorem 5.14. (Fundamental Theorem of Plane Curves, Existence) *A unit-speed curve $\boldsymbol{\beta}\colon (a,b)\to\mathbb{R}^2$ whose curvature is a given piecewise-continuous function $k\colon (a,b)\to\mathbb{R}$ is parametrized by*

$$\begin{cases} \boldsymbol{\beta}(s) &= \Big(\displaystyle\int\cos\theta(s)\,ds + c,\ \int\sin\theta(s)\,ds + d\Big), \\ \theta(s) &= \displaystyle\int k(s)\,ds + \theta_0, \end{cases} \tag{5.8}$$

where c, d, θ_0 are constants of integration.

Proof. We define $\boldsymbol{\beta}$ and θ by (5.8); it follows that

$$\begin{cases} \boldsymbol{\beta}'(s) &= \big(\cos\theta(s),\ \sin\theta(s)\big), \\ \theta'(s) &= k(s). \end{cases} \tag{5.9}$$

Thus $\boldsymbol{\beta}$ has unit speed, so Corollary 1.27 on page 20 tells us that the curvature of $\boldsymbol{\beta}$ is k. ∎

One classical way to describe a plane curve $\boldsymbol{\alpha}$ is by means of a ***natural equation***, which expresses the curvature $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ (at least implicitly) in terms of the arc length s of $\boldsymbol{\alpha}$. Although such an equation may not be the most useful equation for calculational purposes, it shows clearly how curvature changes with arc length and is obviously invariant under translations and rotations. To find the natural equation of a curve $t \mapsto \boldsymbol{\alpha}(t)$, one first computes the curvature and arc length as functions of t and then one tries to eliminate t. Conversely, the ordinary equation can be found (at least in principle) by solving the natural equation and using Theorem 5.14.

The following table lists some examples, while Figure 5.5 plots variants of its last two entries.

Curve	Natural equation
straight line	$\boldsymbol{\kappa 2} = 0$
circle of radius a	$\boldsymbol{\kappa 2} = \dfrac{1}{a}$
catenary$[a]$	$\boldsymbol{\kappa 2} = -\dfrac{a}{a^2 + s^2}$
clothoid$[n, a]$	$\boldsymbol{\kappa 2} = -\dfrac{s^n}{a^{n+1}}$
logspiral$[a, b]$	$\boldsymbol{\kappa 2} = \dfrac{1}{bs}$
involute[circle$[a]$]	$\boldsymbol{\kappa 2} = \dfrac{1}{\sqrt{2as}}$

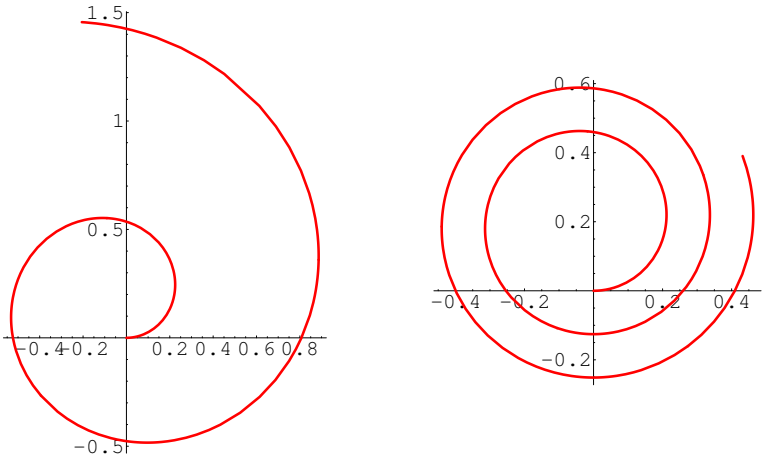


Figure 5.5: Curves with $\boldsymbol{\kappa 2}(s) = 5/(s+1)$ and $5/\sqrt{s+1}$

Natural equations arose as a reaction against the use of Cartesian and polar coordinates, which, despite their utility, were considered arbitrary. In 1736, Euler[2] proposed the use of the arc length s and the radius of curvature $1/|\boldsymbol{\kappa 2}|$ (see [Euler1]). Natural equations were also studied by Lacroix[3] and Hill[4], and excellent discussions of them are given in [Melz, volume 2, pages 33–41] and [Ces]. The book [Ces] of Cesàro[5] uses natural equations as its starting point; it contains many plots (quite remarkable for a book published at the end of the 19th century) of plane curves given by means of natural equations.

For example, let us find the natural equation of the catenary defined by

$$\mathsf{catenary}[a](t) = \left(a\cosh\frac{t}{a},\ t\right). \tag{5.10}$$

A computation tells us that the curvature and arc length are given by

$$\boldsymbol{\kappa 2}(t) = -\frac{\operatorname{sech}(t/a)^2}{a} \qquad \text{and} \qquad s(t) = a\sinh\frac{t}{a}. \tag{5.11}$$

Then (5.11) implies that

$$(s^2 + a^2)\boldsymbol{\kappa 2} = -a. \tag{5.12}$$

Although (5.10) and (5.12) both describe a catenary, (5.12) is considered more natural because it does not depend on the choice of parametrization of the catenary.

---

2 

Leonhard Euler (1707–1783). Swiss mathematician. Euler was a geometer in the broad sense in which the term was used during his time. Not only did he contribute greatly to the evolution and systematization of analysis — in particular to the founding of the calculus of variation and the theories of differential equations, functions of complex variables, and special functions — but he also laid the foundations of number theory as a rigorous discipline. Moreover, he concerned himself with applications of mathematics to fields as diverse as lotteries, hydraulic systems, shipbuilding and navigation, actuarial science, demography, fluid mechanics, astronomy, and ballistics.

[3]Sylvestre Francois Lacroix (1765–1843). French writer of mathematical texts who was a student of Monge.

[4]Thomas Hill (1818–1891). American mathematician who became president of Harvard University.

5 

Ernesto Cesàro (1859–1906). Italian mathematician, born in Naples, professor in Palermo and Rome. Although his most important contribution was his monograph *Lezioni di Geometria Intrinseca*, he is also remembered for his work on divergent series.

5.4 Examples of Curves with Assigned Curvature

A curve $\boldsymbol{\beta}$ parametrized by arc length s whose curvature is a given function $\boldsymbol{\kappa 2} = k(s)$ can be found by solving the system (5.9). Notebook 5 finds numerical solutions of the resulting ordinary differential equations. In this way, we may readily draw plane curves with assigned curvature.

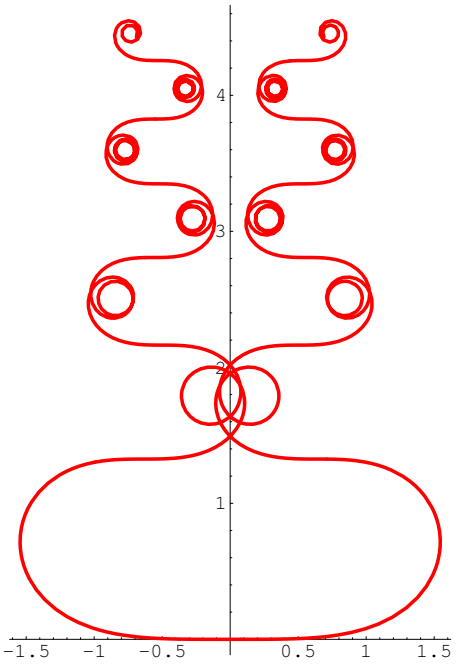


Figure 5.6: A curve with $\boldsymbol{\kappa 2}(s) = s \sin s$

Quite simple curvature functions can be used to produce very interesting curves. We have already seen in Section 2.7 that the curve clothoid$[1, 1]$ has curvature given by $\boldsymbol{\kappa 2}(s) = -s$, where s denotes arc length. The clothoids in Figure 5.3 were in fact generated by using this program.

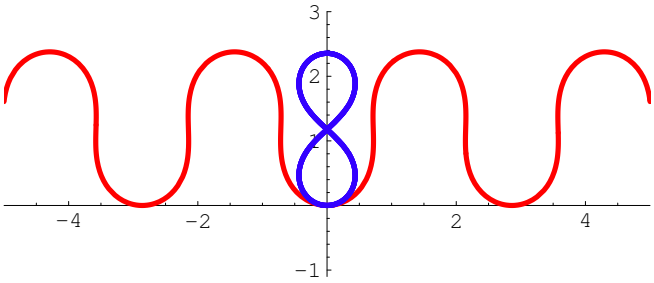


Figure 5.7: Curves with $\boldsymbol{\kappa 2}(s) = 1.6 \sin s$ and $\boldsymbol{\kappa 2}(s) = 2.4 \sin s$

Figure 5.6 displays the case $\boldsymbol{\kappa 2}(s) = s \sin s$. But setting $\boldsymbol{\kappa 2}$ to be a mere multiple of $\sin s$ presents an intriguing situation. Figure 5.7 plots simultaneously two curves whose curvature has the form $\boldsymbol{\kappa 2}(s) = c \sin s$ with c constant; the one in the middle with $c = 2.4$ appears to be a closed curve with this property. Notebook 5 includes an animation of all the intermediate values of c.

Finally, we consider several functions that oscillate more wildly, namely those of the form $s \mapsto s f(s)$, where f is one of the Bessel[6] functions J_n. As well as solving a natural differential equation, these functions occur as coefficients of the Laurent series of the function $\exp\left(\frac{1}{2}t(z - z^{-1})\right)$. This and other useful expressions for them can be found in books on complex analysis, such as [Ahlf]. Below we display both the function f and the resulting curve, omitting the axis scales for visual clarity.

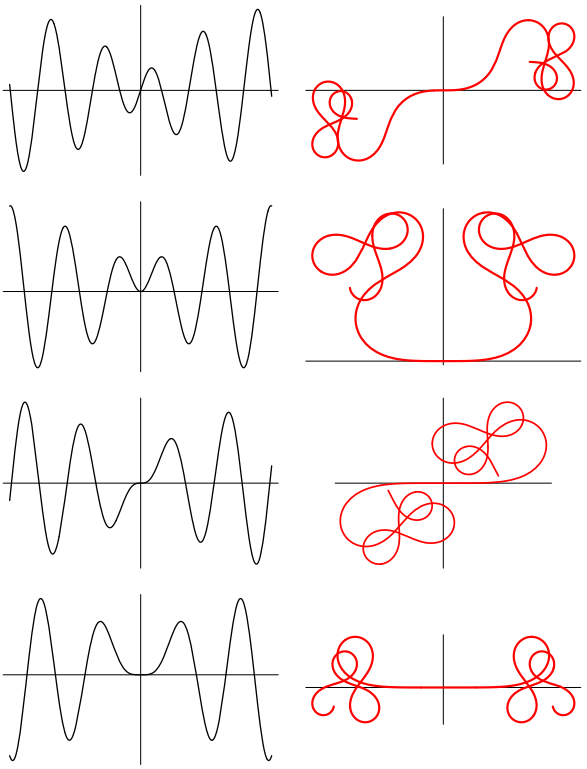


Figure 5.8: Curves with $\boldsymbol{\kappa 2}(s) = s J_n(s)$ with $n = 1, 2, 3, 4$



6 Friedrich Wilhelm Bessel (1784–1846). German astronomer and friend of Gauss. In 1800, Bessel was appointed director of the observatory at Königsberg, where he remained for the rest of his life. Bessel functions were introduced in 1817 in connection with a problem of Kepler of determining the motion of three bodies moving under mutual gravitation.

5.5 Exercises

1. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Show that any vector $\mathbf{v} \in \mathbb{R}^n$ can be written as

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1 + \cdots + (\mathbf{v} \cdot \mathbf{e}_n)\mathbf{e}_n.$$

2. A ***homothety*** of $\mathbb{R}^n$ is a map $F\colon \mathbb{R}^n \to \mathbb{R}^n$ for which there exists a constant $\lambda > 0$ such that

$$\|F(\mathbf{p}_1) - F(\mathbf{p}_2)\| = \lambda \|\mathbf{p}_1 - \mathbf{p}_2\|$$

for all $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^n$. Show that a homothety is a bijective affine transformation.

3. Show that a Euclidean motion $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ preserves evolutes, involutes, parallel curves and pedal curves. More precisely, if $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ is a curve, show that the following formulas hold for $a < t < b$:

$$\begin{aligned}
\mathsf{evolute}[F \circ \boldsymbol{\alpha}](t) &= \big(F \circ \mathsf{evolute}[\boldsymbol{\alpha}]\big)(t), \\
\mathsf{involute}[F \circ \boldsymbol{\alpha}, c](t) &= \big(F \circ \mathsf{involute}[\boldsymbol{\alpha}, c]\big)(t), \\
\mathsf{pedal}[F(\mathbf{p}), F \circ \boldsymbol{\alpha}](t) &= \big(F \circ \mathsf{pedal}[\mathbf{p}, \boldsymbol{\alpha}]\big)(t),
\end{aligned}$$

and

$$\mathsf{parcurve}[F \circ \boldsymbol{\alpha}][s](t) = \begin{cases} \big(F \circ \mathsf{parcurve}[\boldsymbol{\alpha}][s]\big)(t) \\ \qquad\qquad \text{if } F \text{ is orientation-preserving,} \\ \big(F \circ \mathsf{parcurve}[\boldsymbol{\alpha}][-s]\big)(t) \\ \qquad\qquad \text{if } F \text{ is orientation-reversing.} \end{cases}$$

4. An ***algebraic curve*** is an implicitly-defined curve of the form $P(x, y) = 0$, where P is a polynomial in x and y. The ***order*** of an algebraic curve is the order of the polynomial P. Show that an affine transformation preserves the order of an algebraic curve.

5. Find the natural equations of the first three involutes of a circle.

6. ***Nielsen's spiral***[7] is defined as

$$\mathsf{nielsenspiral}[a](t) = a\big(\mathrm{Ci}(t),\ \mathrm{Si}(t)\big),$$

[7] Niels Nielsen (1865–1931). Danish mathematician.

where the sine and cosine integrals are defined by

$$\mathrm{Si}(t) = \int_0^t \frac{\sin u}{u} du, \qquad \mathrm{Ci}(t) = -\int_t^\infty \frac{\cos u}{u} du$$

for $t > 0$. It is illustrated in Figure 5.9.

(a) Explain why the spiral is asymptotic to the x-axis as $t \to -\infty$, and find its limit as $t \to \infty$.

(b) Show that the function $\mathrm{Ci}(t)$ can be defined in the equivalent way

$$\mathrm{Ci}(t) = \int_0^t \frac{(\cos u - 1)}{u} du + \log t + \gamma,$$

again for $t > 0$, where γ is Euler's constant $0.5772\ldots$.

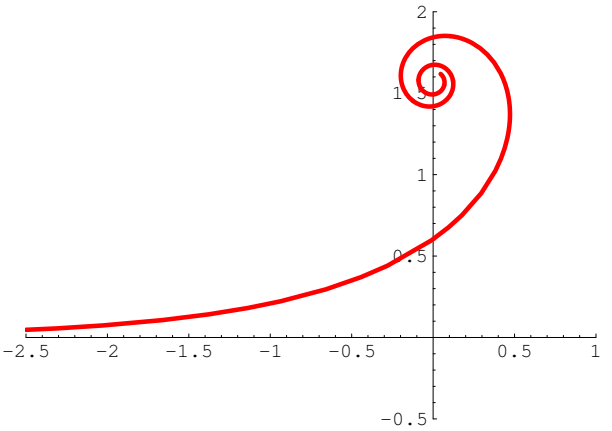


Figure 5.9: Nielsen's spiral plotted with $-2.5 < t < 15$

7. Show that the natural equation of the spiral in the previous exercise is

$$\boldsymbol{\kappa 2}(s) = \frac{e^{s/a}}{a}.$$

M **8.** Plot curves whose curvature is $n\, s \sin s$ for $n = \pm 2, \pm 2.5, \pm 3$.

M **9.** Plot a curve whose curvature is $s \mapsto \Gamma(s + \frac{1}{2})$, where $-5 < s < 5$, and the ***Gamma function*** is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt.$$

M **10.** An ***epicycloid*** is the curve that is traced out by a point $\mathbf{p}$ on the circumference of a circle (of radius b) rolling *outside* another circle (of radius a). Similarly, a ***hypocycloid*** is the curve that is traced out by a point $\mathbf{p}$ on the circumference of a circle (of radius b) rolling *inside* another circle (of radius a).

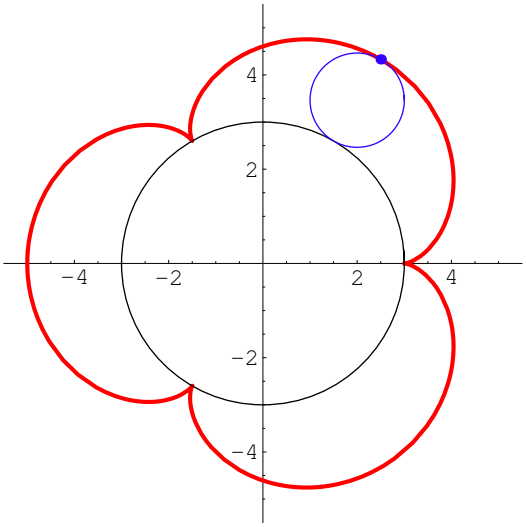


Figure 5.10: An epicycloid with 3 cusps

These curves are parametrized by

$$\begin{aligned}\mathsf{epicycloid}[a,b](t) &= \left((a+b)\cos t - b\cos\left(\frac{(a+b)t}{b}\right),\right.\\ &\qquad \left.(a+b)\sin t - b\sin\left(\frac{(a+b)t}{b}\right)\right),\\ \mathsf{hypocycloid}[a,b](t) &= \left((a-b)\cos t + b\cos\left(\frac{(a-b)t}{b}\right),\right.\\ &\qquad \left.(a-b)\sin t - b\sin\left(\frac{(a-b)t}{b}\right)\right),\end{aligned}$$

and illustrated in Figures 5.10 and 5.11. Show that the natural equation of an epicycloid is

$$a^2s^2 + \left(\frac{a+2b}{\boldsymbol{\kappa 2}}\right)^2 = 16b^2(a+b)^2, \tag{5.13}$$

and find that of the hypocycloid.

M **11.** Let $\boldsymbol{\alpha}_n\colon \mathbb{R} \to \mathbb{R}^2$ be a curve whose curvature is $n\sin^2 s$, where n is an integer and s denotes arc length. Show that $\boldsymbol{\alpha}_n$ is a closed curve if and only if n is *not* divisible by 4. Plot $\boldsymbol{\alpha}_n$ for $1 \leqslant n \leqslant 12$.

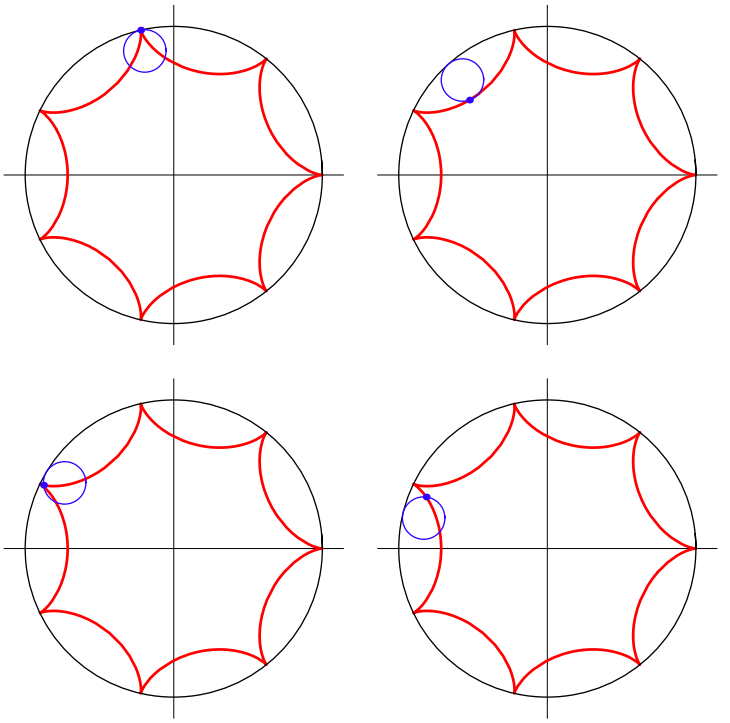


Figure 5.11: A hypocycloid with 7 cusps

12. Show that the natural equation of a tractrix, defined in Section 2.6, is

$$\boldsymbol{\kappa 2} = -\frac{1}{a\sqrt{e^{2s/a}-1}}.$$

M **13.** The ***Airy***[8] ***functions***, denoted AiryAi and AiryBi, are linearly independent solutions of the differential equation $y''(t) - t\,y(t) = 0$. Plot curves whose signed curvature is respectively $s \mapsto \mathsf{AiryAi}(s)$ and $s \mapsto \mathsf{AiryBi}(s)$.

[8]  Sir George Biddel Airy (1801–1892). Royal astronomer of England.

Notebook 5

■ N5.0

Mathematica can be used effectively to graph a unit-speed plane curve with an assigned curvature function. The notebook associated to this chapter provides such a miniprogram, presented in N5.3. Many strange yet interesting curves can be constructed numerically in this way, and examples are given in N5.4.

The notebook begins with instructions for translating, rotating and reflecting curves, and ends with animations of epicycloids and hypocycloids.

We first list a few curves that will provide examples of the constructions of this chapter. Now that we are familar with them, we choose to use the alternative notation for functions mentioned in N3.4, in which # stands for the variable. This has the advantage that the left-hand sides of the definitions are less cluttered with parameters. The final & is essential, but can be remembered by thinking of it as short for <-- #, or "maps to" written from right to left.

```
parabola[a_]:= a{2#, #^2} &
ellipse[a_,b_]:= {a Cos[#], b Sin[#]} &
```

We may of course define a circle as a special case of an ellipse:

```
circ[a_]:= ellipse[a,a]
```

The logarithmic spiral below was one of the very first examples of curves to be investigated in Notebook 1. Another spiral will be defined in N5.5.

```
logspiral[a_,b_]:= a Exp[b#]{Cos[#], Sin[#]} &
```

The hidden part of this initial section once again contains a plotting option and the basic commands J, L, κ2, from which others are manufactured.

■ N5.0.1

■ N5.1

Translation of a curve is achieved by merely adding the vector representing the translation to the vector-valued function describing the curve. The command tran is only introduced to enable a translated curve to be defined without referring to a parameter.

```
tran[α_, v_][t_]:= α[t] + v
```

```
v:= {2,1}
α:= ellipse[2,1]
β:= tran[α, v]
β[t]
```

```
ParametricPlot[{α[t], β[t]}, {t,0,2π}]
```

Rotations are best defined with the appropriate matrix. This could easily be written down by hand, but we may as well extract it from the built-in package

```
<<Geometry`Rotations`
```

```
M[θ_]:= RotationMatrix2D[-θ]
MatrixForm[M[θ]]
```

Note the minus sign that suits our convention, whereby the following command rotates a vector α counterclockwise by an angle θ. The subsequent computation confirms that this operation coincides with premultiplication by the matrix M[θ].

```
rot[α_,θ_][t_]:= Rotate2D[α[t], -θ]
```

```
Clear[α]
α[t_]:= {x[t], y[t]}
rot[α,θ][t] - M[θ].α[t]
```

It is now an easy matter to rotate an ellipse through various angles and plot the result. Below, ellipse number n (ranging from 0 to 9) is colored with Hue[.7n] according to

the option set in N5.0 (the prime 7 was chosen on purpose!). The idea behind Figure 5.3 is similar, although this code makes use of the program intrinsic defined in an initialization cell in N5.3 below.

```
β:= ellipse[2,1]
ga:= Table[rot[β, n π/10][t], {n,0,9}]
ParametricPlot[Evaluate[ga], {t,0,2π}];
```

Reflection of a curve α in a plane (or indeed, hypersurface) orthogonal to the vector v is computed with

```
refl[α_,v_][t_]:= α[t] - 2v(α[t].v)/L[v]^2
```

```
α[t_]:= {x[t], y[t]}
refl[α,{0,1}][t] - {1,-1}α[t]
```

```
Clear[x,y,z,β]
β[t_]:= {x[t], y[t], z[t]}
refl[β, {1,1,1}][t]
Map[Expand, %]
```

The command Map on the line above enables the function Simplify to be applied to each component in turn.

```
Module[{α,β},
  α= logspiral[1, .2];
  β= tran[refl[α, {1,1}], {1,1}];
  ParametricPlot[{α[t], β[t]}, {t,-8,8}];
]
```

■ N5.1.1

■ N5.2

The "bouquet of clothoids" also previews the program intrinsic about to be sold in N5.3.

```
Module[{r,α,ga},
  r= 9;
  α= intrinsic[#&, 0, {0,0,0}, {-9,9}];
  ga= Table[rot[α, n 2π/3][t], {n,3}];
  ParametricPlot[Evaluate[ga], {t,-9,9}, Ticks->None]
];
```

Figure 5.3

N5.2.1

N5.3

The following program numerically solves the system of differential equations obtained by assigning the curvature function f. The quantity eqic represents the "equations and initial conditions."

```
intrinsic[f_, a_:0, {c_:0, d_:0, e_:0}, {min_:-10, max_:10},
  opts___][t_]:=
Module[{x,y,θ,s},
  eqic= {x'[s]==Cos[θ[s]], y'[s]==Sin[θ[s]], θ'[s]==f[s],
         x[a]==c, y[a]==d, θ[a]==e};
  sol= NDSolve[eqic, {x,y, θ}, {s,min,max}, opts];
  {x[t],y[t]}/.sol[[1]]
]
```

Here is its basic operation. The symbols 2& stand for the constant function 2 (the "&" symbol is essential here), so that α below represents a circle of radius 1/2. This is confirmed by its plot.

```
α:= intrinsic[2&, 0, {0,0,0}, {0,20}]
ParametricPlot[Evaluate[α[t]], {t,0,20}]
```

Here is a more complicated example; as the values of the function f approach 1, so the curve resembles more a circle of radius 1.

```
f[s_]:= s/(s+1)
α:= intrinsic[f, 0, {0,0,0}, {0,20}]
ParametricPlot[Evaluate[α[t]], {t,0,20}]
```

The following plot with the secant (reciprocal of cosine) function is made for |s|<π/2 to avoid singularities:

```
g[s_]:= Sec[s]
r:= 1.56
β:= intrinsic[g, 0, {0,0,0}, {-r,r}]
ParametricPlot[Evaluate[β[t]], {t,-r,r}]
```

■ N5.3.1

■ N5.4

We start with an exotic example corresponding to κ2(s)=s sin(s). Subsequent plots analyse functions based on sin(s) in more detail.

```
α:= intrinsic[# Sin[#] &, 0, {0,0,0}, {-20,20}]
ParametricPlot[Evaluate[α[t]], {t,-20,20}, ImageSize->300]
```

Figure 5.6

```
Clear[pl]
β[k_]:= intrinsic[k Sin[#] &, π/2, {0,0,0}, {-5π,5π}]
pl[k_,opts___]:= ParametricPlot[Evaluate[β[k][t]], {t,-5π,5π},
  PlotRange->{{-5,5},{-1.1,3}}, ImageSize->400, opts];
Show[{pl[1.6], pl[2.405, PlotStyle->{Hue[.7],Thickness[.008]}]}]
```

Figure 5.7

The animation can be viewed by clicking on any one image in the output produced by the iteration

```
Do[pl[k], {k, 1, 8, .2}]
```

The next program will be used in general to display a given curvature function f together with a curve realizing f as its curvature.

```
compare[f_,a_,b_]:=
Module[{z1,z2,α},
  z1= Plot[f[s], {s,a,b}, DisplayFunction->Identity, Ticks->None];
  α= intrinsic[f, 0, {0,0,0}, {a,b}];
  z2= ParametricPlot[Evaluate[α[t]], {t,a,b}, Ticks->None,
    DisplayFunction->Identity];
  Show[GraphicsArray[{z1,z2}], DisplayFunction->$DisplayFunction]
]
```

A list of examples now provides the array.

```
h[1][s_]:= Sin[s]
h[2][s_]:= s Cos[s]
h[3][s_]:= s^2 Sin[s]
h[4][s_]:= s Sin[s]^2
Table[compare[h[i],-2π,2π], {i,4}]
```

■ N5.4.1

■ N5.5

We next define and plot Nielsen's spiral, using the built-in sine and cosine integral functions.

```
nielsenspiral[a_][t_]:= a{CosIntegral[t], SinIntegral[t]}
```

```
ParametricPlot[nielsenspiral[1][t], {t,0.001,15},
  PlotRange->{{-2.5,1},{-.5,2}}, ImageSize->400]
```

Figure 5.9

The two equivalent definitions of the cosine integral can be compared:

```
x=100;
Integrate[(Cos[t]-1)/t, {t,0,x}]+Log[x]+EulerGamma//N
CosIntegral[x]//N
```

The following two definitions are given for the sake of completeness, even though the second curve can be derived from the first, or vice versa.

```
epicycloid[a_,b_][t_]:= {(a+b)Cos[t] - b Cos[((a+b)/b)t],
  (a+b)Sin[t] - b Sin[((a+b)/b)t]}
hypocycloid[a_,b_][t_]:= {(a-b)Cos[t] + b Cos[((a-b)/b)t],
  (a-b)Sin[t] - b Sin[((a-b)/b)t]}
```

```
hypocycloid[a,b][t] - epicycloid[a,-b][t]
```

The next miniprogram constructs a circle touching a fixed point of an epicycloid.

```
roll[s_]:=
Module[{ϵ,γ,p,cp},
  ϵ= epicycloid[3,1];
  γ= circ[4];
  p= 5.5;
  cp:= {Circle[{0,0},3], Hue[.7], Circle[γ[s],1],
    PointSize[.02], Point[ϵ[s]]};
  ParametricPlot[ϵ[t], {t,0,2π}, PlotRange->{{-p,p},{-p,p}},
    Epilog->cp]
]
```

With its definition, the figure and associated animation becomes easy.

```
roll[π/3]
```

Figure 5.10

```
Do[roll[s], {s, 0, 2π, π/20}]
```

The same technique is used for Figure 5.11.

■ N5.5.1

Chapter 6

Global Properties of Plane Curves

The geometry of plane curves that we have been studying in the previous chapters has been local in nature. For example, the curvature of a plane curve describes the bending of that curve, point by point. In this chapter, we consider *global* properties that are concerned with the curve as a whole.

In Section 6.1, we define the *total signed curvature* of a plane curve $\boldsymbol{\alpha}$, by integrating its signed curvature $\boldsymbol{\kappa 2}$. The total signed curvature is an overall measure of curvature, directly related to the turning angle that was defined in Chapter 1. For a regular closed curve, the total signed curvature gives rise to a *turning number*, which is an integer. These concepts are well illustrated with reference to epitrochoids and hypotrochoids, curves formed by rolling wheels, described in Section 6.2. An example is given of relevant curvature functions.

Section 6.3 is devoted to the *rotation index* of a closed curve, which is defined topologically as the degree of an associated continuous mapping. We then show that it coincides with the turning number. The concept of *homotopy* for maps from a circle is introduced, and used to show that the turning number of a *simple* closed curve has absolute value 1.

Convex plane curves are considered in Section 6.4, where it is shown that a closed plane curve is convex if and only if its signed curvature does not change sign. We prove the *Four Vertex Theorem* for such curves in Section 6.5, and illustrate it with the sine oval, parametrized using an iterated sine function.

Section 6.6 is concerned with more general *ovals*, closed curves for which $\boldsymbol{\kappa 2}$ not only does not change sign, but is either strictly positive or negative. We also establish basic facts about curves of *constant width*, and then give two classes of examples in Section 6.7. The formula for an oval in terms of its *support function* is derived in Section 6.8, using the envelope of a family of straight lines. A number of examples are investigated in the text and subsequent exercises.

6.1 Total Signed Curvature

The signed curvature $\boldsymbol{\kappa 2}$ of a plane curve $\boldsymbol{\alpha}$ was defined on page 14, and measures the bending of the curve at each of its points. A measure of the total bending of $\boldsymbol{\alpha}$ is given by an integral involving $\boldsymbol{\kappa 2}$.

Definition 6.1. *The* ***total signed curvature*** *of a curve* $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^2$ *is*

$$\mathsf{TSC}[\boldsymbol{\alpha}] = \int_a^b \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)\big\|\,\boldsymbol{\alpha}'(t)\big\|\,dt,$$

where $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ *denotes the signed curvature of* $\boldsymbol{\alpha}$.

Although the definition depends upon the endpoints a, b, we assume that the mapping $\boldsymbol{\alpha}$ is in fact defined on an open interval I containing $[a,b]$.

First, let us check that the total signed curvature is a geometric concept.

Lemma 6.2. *The total signed curvature of a plane curve remains unchanged under a positive reparametrization, but changes sign under a negative one.*

Proof. Let $\boldsymbol{\gamma} = \boldsymbol{\alpha} \circ h$ where $\boldsymbol{\gamma}\colon (c,d) \to \mathbb{R}^2$ and $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ are curves. We do the case when $h'(u) > 0$ for all u. Using Theorem 1.20 on page 16, and the formula from calculus for the change of variables in an integral, we compute

$$\begin{aligned}
\mathsf{TSC}[\boldsymbol{\alpha}] \;=\; \int_a^b \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)\big\|\boldsymbol{\alpha}'(t)\big\|dt \;&=\; \int_c^d \boldsymbol{\kappa 2}[\boldsymbol{\alpha}]\big(h(u)\big)\big\|\boldsymbol{\alpha}'\big(h(u)\big)\big\|h'(u)\,du \\
&= \int_c^d \boldsymbol{\kappa 2}[\boldsymbol{\gamma}](u)\big\|\boldsymbol{\gamma}'(u)\big\|du \\
&= \mathsf{TSC}[\boldsymbol{\gamma}].
\end{aligned}$$

The proof that $\mathsf{TSC}[\boldsymbol{\gamma}] = -\mathsf{TSC}[\boldsymbol{\alpha}]$ when $\boldsymbol{\gamma}$ is a negative reparametrization of $\boldsymbol{\alpha}$ is similar. ∎

There is a simple relation linking the total signed curvature and the turning angle of a curve defined in Section 1.5.

Lemma 6.3. *The total signed curvature can be expressed in terms of the turning angle* $\boldsymbol{\theta}[\boldsymbol{\alpha}]$ *of* $\boldsymbol{\alpha}$ *by*

$$\mathsf{TSC}[\boldsymbol{\alpha}] = \boldsymbol{\theta}[\boldsymbol{\alpha}](b) - \boldsymbol{\theta}[\boldsymbol{\alpha}](a). \tag{6.1}$$

Proof. Equation (6.1) results when (1.22), page 20, is integrated. ∎

Since $\boldsymbol{\theta}[\boldsymbol{\alpha}]$ represents the direction of a unit tangent vector to the curve, $\mathsf{TSC}[\boldsymbol{\alpha}]$ measures the rotation of this vector. This is illustrated by the curve in Figure 6.2 for which $\boldsymbol{\theta}[\boldsymbol{\alpha}]$ varies between $-\pi/4$ and $5\pi/4$.

The total signed curvature of a *closed* curve is especially important. First, we define carefully the notion of closed curve.

Definition 6.4. *A regular curve* $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ *is* ***closed*** *provided there is a constant* $c > 0$ *such that*

$$\boldsymbol{\alpha}(t+c) = \boldsymbol{\alpha}(t) \tag{6.2}$$

for all t*. The least such number* c *is called the* ***period*** *of* $\boldsymbol{\alpha}$*.*

Equation 6.2 expresses closure in the topological sense, as it ensures that $\boldsymbol{\alpha}$ determines a continuous mapping from a circle into $\mathbb{R}^n$, though we shall generally only consider closed curves that are regular. Exercise 1 provides a particular example of a nonregular curve satisfying (6.2).

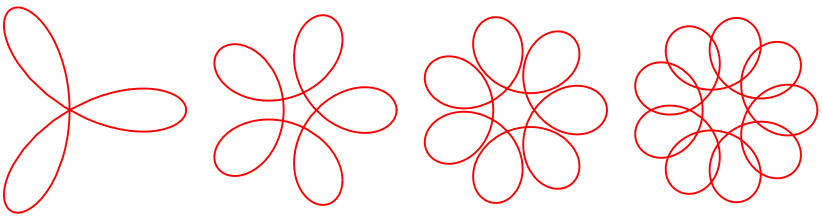


Figure 6.1: Hypotrochoids with turning numbers 2,4,6,8

Clearly, we can use (6.2) to define $\boldsymbol{\alpha}(t)$ for all t; that is, the domain of definition of a closed curve can be extended from (a,b) to $\mathbb{R}$. Just as in the case of a circle, the trace $\mathscr{C}$ of a regular closed curve $\boldsymbol{\alpha}$ is covered over and over again by $\boldsymbol{\alpha}$. Intuitively, when we speak of the length of a closed curve, we mean the length of the trace $\mathscr{C}$. Therefore, in the case of a closed curve we can use either $\mathbb{R}$, or a closed interval $[a,b]$, for the domain of definition of the curve, where $b-a$ is the period. Furthermore, we modify the definition of length given on page 9 as follows:

Definition 6.5. *Let* $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^n$ *be a regular closed curve with period* c*. By the* ***length*** *of* $\boldsymbol{\alpha}$ *we mean the length of the restriction of* $\boldsymbol{\alpha}$ *to* $[0,c]$*, namely,*

$$\int_0^c \left\|\boldsymbol{\alpha}'(t)\right\| dt.$$

Next, we find the relation between the period and length of a closed curve.

Lemma 6.6. *Let* $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^n$ *be a closed curve with period* c*, and let* $\boldsymbol{\beta}\colon \mathbb{R} \to \mathbb{R}^n$ *be a unit-speed reparametrization of* $\boldsymbol{\alpha}$*. Then* $\boldsymbol{\beta}$ *is also closed; the period of* $\boldsymbol{\beta}$ *is* L*, where* L *is the length of* $\boldsymbol{\alpha}$*.*

Proof. Let s denote the arc length function of $\boldsymbol{\alpha}$ starting at 0; we can assume (by Lemma 1.17, page 13) that $\boldsymbol{\alpha}(t) = \boldsymbol{\beta}\big(s(t)\big)$. We compute

$$\begin{aligned} s(t+c) \;&=\; \int_0^{c+t} \big\|\boldsymbol{\alpha}'(u)\big\|\,du \;=\; \int_0^{c} \big\|\boldsymbol{\alpha}'(u)\big\|\,du + \int_c^{c+t} \big\|\boldsymbol{\alpha}'(u)\big\|\,du \\ &=\; L + \int_c^{c+t} \big\|\boldsymbol{\alpha}'(u)\big\|\,du \\ &=\; L + \int_0^{t} \big\|\boldsymbol{\alpha}'(u)\big\|\,du \;=\; L + s(t). \end{aligned}$$

Thus

$$\boldsymbol{\beta}\big(s(t)+L\big) = \boldsymbol{\beta}\big(s(t+c)\big) = \boldsymbol{\alpha}(t+c) = \boldsymbol{\alpha}(t) = \boldsymbol{\beta}\big(s(t)\big),$$

for all $s(t)$; it follows that $\boldsymbol{\beta}$ is closed. Furthermore, since c is the least positive number such that $\boldsymbol{\alpha}(t+c) = \boldsymbol{\alpha}(c)$ for all t, it must be the case that L is the least positive number such that $\boldsymbol{\beta}(s+L) = \boldsymbol{\beta}(s)$ for all $\boldsymbol{\beta}$. ∎

We return to plane curves, that is, to the case $n = 2$.

Definition 6.7. *The* ***turning number*** *of a closed curve* $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^2$ *is*

$$\mathsf{Turn}[\boldsymbol{\alpha}] = \frac{1}{2\pi}\int_0^c \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)\big\|\boldsymbol{\alpha}'(t)\big\|\,dt,$$

where c *denotes the period of* $\boldsymbol{\alpha}$.

Thus the turning number of a closed curve $\boldsymbol{\alpha}$ is just the total signed curvature of $\boldsymbol{\alpha}$ divided by 2π, so that

$$\mathsf{TSC}[\boldsymbol{\alpha}] = 2\pi\mathsf{Turn}[\boldsymbol{\alpha}].$$

For example, let

$$\boldsymbol{\alpha}\colon [0, 2n\pi] \longrightarrow \mathbb{R}^2$$

be the function $t \mapsto a(\cos t, \sin t)$, so that $\boldsymbol{\alpha}$ covers a circle n times. It is easy to compute

$$\boldsymbol{\kappa 2}(\boldsymbol{\alpha}) = \frac{1}{a} \qquad \text{and} \qquad \big\|\boldsymbol{\alpha}'(t)\big\| = a.$$

Hence

$$\mathsf{TSC}[\boldsymbol{\alpha}] = 2n\pi \qquad \text{and} \qquad \mathsf{Turn}[\boldsymbol{\alpha}] = n$$

are independent of a.

More generally, if $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^2$ is a regular closed curve with trace $\mathscr{C}$, then the mapping

$$t \mapsto \frac{\boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}'(t)\big\|} \tag{6.3}$$

gives rise to a mapping $\mathbf{\Phi}$ from $\mathscr{C}$ to the unit circle $S^1(1)$ of $\mathbb{R}^2$. Note that $\mathbf{\Phi}(\mathbf{p})$ is just the end point of the unit tangent vector to $\boldsymbol{\alpha}$ at $\mathbf{p}$. It is intuitively clear that when a point $\mathbf{p}$ goes around $\mathscr{C}$ once, its image $\mathbf{\Phi}(\mathbf{p})$ goes around $S^1(1)$ an integral number of times. The turning number of a *regular closed* curve should therefore be an integer; this will be proved rigorously in Section 6.3.

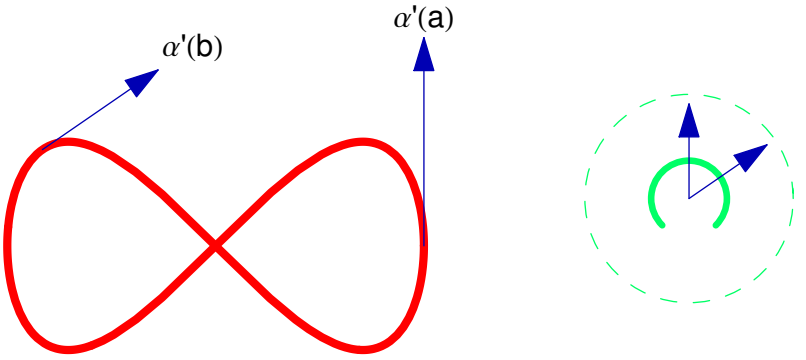


Figure 6.2: Turning of the tangent direction

If $\boldsymbol{\alpha}$ denotes the figure eight in Figure 6.2, then $\mathsf{Turn}[\boldsymbol{\alpha}] = 0$ because the "clock hand" never reaches 6 pm. By contrast, each curve in Figure 6.1 has turning number equal to one less than the number of loops; a more explicit representation of a different example is displayed in Figure 6.4, after we have discussed the trochoids in more detail.

6.2 Trochoid Curves

The ***epitrochoid*** and ***hypotrochoid*** are good curves to illustrate turning number. They are defined by

$$\mathsf{epitrochoid}[a,b,h](t) = \left((a+b)\cos t - h\cos\left(\frac{(a+b)t}{b}\right),\ (a+b)\sin t - h\sin\left(\frac{(a+b)t}{b}\right)\right)$$

$$\mathsf{hypotrochoid}[a,b,h](t) = \left((a-b)\cos t + h\cos\left(\frac{(a-b)t}{b}\right),\ (a-b)\sin t - h\sin\left(\frac{(a-b)t}{b}\right)\right).$$

One can describe $\mathsf{epitrochoid}[a,b,h]$ as the parametrized curve that is traced out by a point $\mathbf{p}$ fixed relative to a circle of radius b rolling outside a fixed circle of

radius a (see Figure 6.18 at the end of the chapter). Here h denotes the distance from $\mathbf{p}$ to the center of the rolling circle. In the case that $h = b$, the loops degenerate into points, and the resulting curve is a ***epicycloid*** (parametrized in Exercise 10 of Chapter 5).

Similarly, hypotrochoid$[a, b, h]$ is the curve traced out by a point $\mathbf{p}$ fixed relative to one circle (of radius b) rolling *inside* another circle (of radius a). The curves in Figure 6.1 are hypotrochoid$[2k-1, 1, k]$ for $k = 2, 3, 4, 5$, respectively. A ***hypocycloid*** is a hypotrochoid for which the loops degenerate to points, which again occurs for $h = b$. Abstractly speaking, the parametrization of the hypotrochoid is obtained from that of the epitrochoid by simultaneously changing the signs of b and h (Exercise 6).

Both epitrochoid$[a, b, h]$ and hypotrochoid$[a, b, h]$ are precisely contained in a circle of radius $a + b + h$. Each has a/b loops as t ranges over the interval $[0, 2\pi]$ provided a/b is an integer. More generally, if a, b are integers, it can be verified experimentally that the number of loops is $a/\gcd(a, b)$, where $\gcd(a, b)$ denotes the greatest common denominator (or highest common factor) of a, b. Whilst the curve will eventually close up provided a/b is rational, it never closes if a/b is irrational; this can be exploited to draw pictures like Figure 6.3.

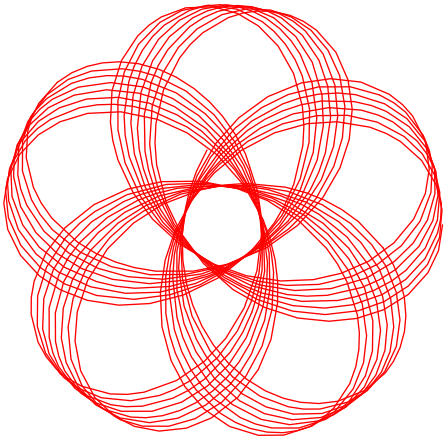


Figure 6.3: $t \mapsto$ epitrochoid$[7, \sqrt{2}, 8](t)$ with $0 \leqslant t \leqslant 16\pi$

As an example to introduce the next section, consider an epitrochoid with 5 inner loops, given by

$$\text{epitrochoid}[5, 1, 3](t) = \big(6\cos t - 3\cos 6t,\ 6\sin t - 3\sin 6t\big).$$

It is clear from Figure 6.4 that the curvature has constant sign. But it varies considerably, attaining its greatest (absolute) value during the inner loops. On

the other hand, its total signed curvature (plotted rising steadily in Figure 6.5) is surprisingly linear, since the fluctuations of

$$t \mapsto \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t)\|\boldsymbol{\alpha}'(t)\|$$

(the base curve in Figure 6.5) are relatively small.

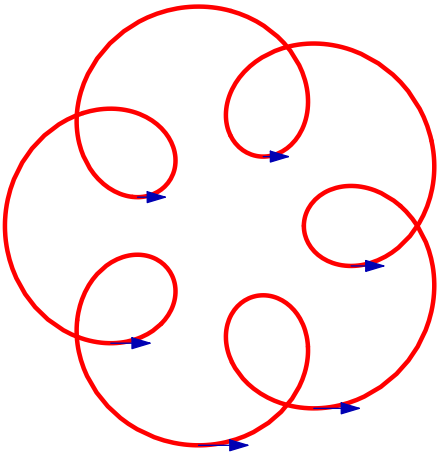


Figure 6.4: $\boldsymbol{\alpha} =$ epitrochoid$[5, 1, 3]$

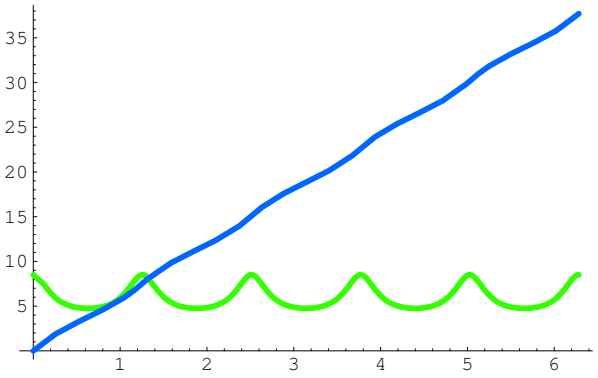


Figure 6.5: TSC$[\boldsymbol{\alpha}]$ and (lighter) its integrand

Using the turning angle interpretation from Lemma 6.3, it is already clear from Figure 6.4 that $\boldsymbol{\alpha}$ has turning number 6. Indeed, the mini-arrows were positioned by solving an equation asserting that the tangent vector points in the direction of the positive x-axis, and they show that the tangent vector undergoes 6 full turns in traversing the curve once. We shall now prove that the turning number of a regular closed curve is always an integer.

6.3 The Rotation Index of a Closed Curve

In Section 6.1, we gave the definition of turning number in terms of the total signed curvature. In this section, we determine the turning number more directly in terms of the mapping (6.3). First, an auxiliary definition.

Definition 6.8. *Let $\phi\colon S^1(1) \to S^1(1)$ be a continuous function, where $S^1(1)$ denotes a circle of radius 1 and center the origin in $\mathbb{R}^2$. Let $\widetilde{\phi}\colon \mathbb{R} \to \mathbb{R}$ be a continuous function such that*

$$\phi(\cos t,\, \sin t) = \big(\cos\widetilde{\phi}(t),\, \sin\widetilde{\phi}(t)\big) \tag{6.4}$$

*for all t. The **degree** of ϕ is the integer n such that $\widetilde{\phi}(2\pi) - \widetilde{\phi}(0) = 2\pi n$.*

The choice of $\widetilde{\phi}$ is not unique, since we can certainly add integer multiples of 2π to its value without affecting (6.4). But we have

Lemma 6.9. *The definition of degree is independent of the choice of $\widetilde{\phi}$.*

Proof. Let $\widehat{\phi}\colon \mathbb{R} \to \mathbb{R}$ be another continuous function satisfying the same conditions as $\widetilde{\phi}$. Then we have

$$\widehat{\phi}(t) - \widetilde{\phi}(t) = 2\pi\, n(t)$$

where $n(t)$ is an integer. Since $n(t)$ is continuous, it must be constant. Thus $\widetilde{\phi}(2\pi) - \widetilde{\phi}(0) = \widehat{\phi}(2\pi) - \widehat{\phi}(0)$. ∎

We next use the notion of degree of a map to define the rotation index of a curve. Let $\boldsymbol{\alpha}\colon \mathbb{R} \to \mathbb{R}^2$ be a regular closed curve. Rescale the parameter so that its period is 2π and $\boldsymbol{\alpha}$ determines a mapping with domain $S^1(1)$.

Definition 6.10. *The **rotation index** of $\boldsymbol{\alpha}$ is the degree of the corresponding mapping $\boldsymbol{\Phi}[\boldsymbol{\alpha}]\colon S^1(1) \to S^1(1)$ defined by*

$$\boldsymbol{\Phi}[\boldsymbol{\alpha}](t) = \frac{\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|}. \tag{6.5}$$

Notice that the rotation index of a curve is defined *topologically*. By contrast, the turning number is defined *analytically*, as an integral. It is now easy for us to prove that these integers are the same.

Theorem 6.11. *The rotation index of a regular closed curve $\boldsymbol{\alpha}$ coincides with the turning number of $\boldsymbol{\alpha}$.*

Proof. Without loss of generality, we can assume that $\boldsymbol{\alpha}$ has unit speed. Then by (5.9) on page 137,

$$\boldsymbol{\Phi}[\boldsymbol{\alpha}](s) = \boldsymbol{\alpha}'(s) = \big(\cos\boldsymbol{\theta}(s),\, \sin\boldsymbol{\theta}(s)\big),$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\alpha}]$ denotes the turning angle of $\boldsymbol{\alpha}$. Therefore, we can choose $\widetilde{\boldsymbol{\Phi}}[\boldsymbol{\alpha}] = \boldsymbol{\theta}$, and appeal to Lemma 6.3 and Corollary 1.27:

$$\begin{aligned} \mathsf{degree}\,\boldsymbol{\Phi}[\boldsymbol{\alpha}] &= \frac{1}{2\pi}\big(\boldsymbol{\theta}(2\pi) - \boldsymbol{\theta}(0)\big) &&= \frac{1}{2\pi}\int_0^{2\pi} \frac{d\boldsymbol{\theta}(s)}{ds}\,ds \\ &&&= \frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{\kappa 2}[\boldsymbol{\alpha}](s)\,ds \\ &&&= \mathsf{Turn}[\boldsymbol{\alpha}]. \ \blacksquare \end{aligned}$$

Those closed curves which do not cross themselves form an important subclass.

Definition 6.12. *A regular closed curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ with period c is* ***simple*** *provided $\boldsymbol{\alpha}(t_1) = \boldsymbol{\alpha}(t_2)$ only if $t_1 - t_2$ is an integer multiple of c.*

For example, an ellipse is a simple closed curve, but the curves in Figure 6.1 are not.

It is clear intuitively that the rotation index of a simple closed curve is ± 1. To prove this rigorously, we need the important topological notion of homotopy. Let

$$\boldsymbol{\alpha}, \boldsymbol{\gamma}\colon [0, L] \longrightarrow X$$

be two continuous mappings such that

$$\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}(L) \quad \text{and} \quad \boldsymbol{\gamma}(0) = \boldsymbol{\gamma}(L). \tag{6.6}$$

Here, X can be any topological space, although in our setting it suffices to take X to equal $\mathbb{R}^2$, so that $\boldsymbol{\alpha}, \boldsymbol{\gamma}$ may be thought of as plane curves. (Actually, in the proof below, the traces of the two curves will both be unit circles.) What is more, the condition (6.6) ensures that the domain of these curves can also be regarded as a circle (recall (6.2)).

Definition 6.13. *The mappings $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are* ***homotopic*** *(as maps from a circle) if there exists a continuous mapping*

$$F\colon [0,1] \times [0, L] \longrightarrow X,$$

such that

(i) $F(0,t) = \boldsymbol{\alpha}(t)$ *for each* t,

(ii) $F(1,t) = \boldsymbol{\gamma}(t)$ *for each* t,

(iii) $F(u,0) = F(u,L)$ *for each* u.

The map F is called a ***homotopy*** *between $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.*

Applying Lemma 6.8, we may define the degree of the mapping $t \mapsto F(u,t)$ for each fixed u. But this map varies continuously with u, so its degree must in fact be constant, as in Lemma 6.9. In conclusion, *homotopic curves have the same degree.* This is used to prove the following theorem, due to H. Hopf[1] [Hopf1].

Theorem 6.14. *The turning number of a simple closed plane curve* $\boldsymbol{\alpha}$ *is* ± 1.

Proof. Fix a point $\mathbf{p}$ on the trace $\mathscr{C}$ of $\boldsymbol{\alpha}$ with the property that $\mathscr{C}$ lies entirely to one side of the tangent line at $\mathbf{p}$. This is always possible: choose a line that does not meet $\mathscr{C}$ and then translate it until it becomes tangent to $\mathscr{C}$. The idea of the proof is then to construct a homotopy between the unit tangent map $\boldsymbol{\Phi}[\boldsymbol{\alpha}]$ of (6.5), whose degree we know to be $\mathsf{Turn}[\boldsymbol{\alpha}]$, and a mapping of degree ± 1 defined by secants passing through $\mathbf{p}$.

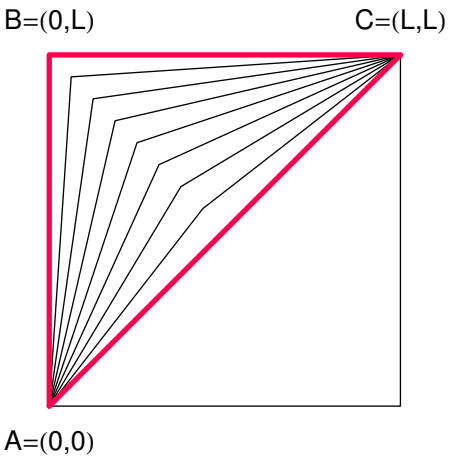


Figure 6.6: A homotopy square

Denote by L the length of $\boldsymbol{\alpha}$ and consider the triangular region

$$\mathcal{T} = \{\, (t_1, t_2) \;\mid\; 0 \leqslant t_1 \leqslant t_2 \leqslant L \,\}$$

shown in Figure 6.6. Let $\boldsymbol{\beta}$ be a reparametrization of $\boldsymbol{\alpha}$ with $\boldsymbol{\beta}(0) = \mathbf{p}$. The ***secant map*** $\Sigma\colon \mathcal{T} \to S^1(1)$ is defined by

---

[1] 

Heinz Hopf (1894–1971). Professor at the Eidgenössische Technische Hochschule in Zürich. The greater part of his work was in algebraic topology, motivated by an exceptional geometric intuition. In 1931, Hopf studied homotopy classes of maps from the sphere S^3 to the sphere S^2 and defined what is now known as the Hopf invariant.

$$\Sigma(t_1,t_2)=\begin{cases}\dfrac{\boldsymbol{\beta}'(t)}{\|\boldsymbol{\beta}'(t)\|} & \text{for } t_1=t_2=t,\\[2ex] -\dfrac{\boldsymbol{\beta}'(0)}{\|\boldsymbol{\beta}'(0)\|} & \text{for } t_1=0 \text{ and } t_2=L\\[2ex] \dfrac{\boldsymbol{\beta}(t_2)-\boldsymbol{\beta}(t_1)}{\|\boldsymbol{\beta}(t_2)-\boldsymbol{\beta}(t_1)\|} & \text{otherwise.}\end{cases}$$

Since $\boldsymbol{\beta}$ is regular and simple, Σ is continuous. Let $A=(0,0)$, $B=(0,L)$ and $C=(L,L)$ be the vertices of $\mathcal{T}$, as in Figure 6.6. Because the restriction of Σ to the side AC is $\boldsymbol{\beta}'/\|\boldsymbol{\beta}'\|$, the degree of this restriction is the turning number of $\boldsymbol{\beta}$. Thus, by construction, $\boldsymbol{\beta}'/\|\boldsymbol{\beta}'\|$ is homotopic to the restriction of Σ to the path consisting of the sides AB and BC joined together. We must show that the degree of the latter map is ± 1.

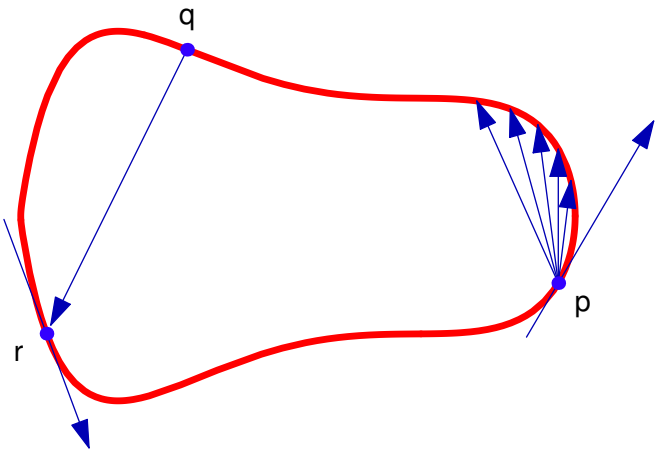


Figure 6.7: Secants and tangents

Assume that $\boldsymbol{\beta}$ is oriented with respect to $\mathbb{R}^2$, so that the angle from $\boldsymbol{\beta}'(0)$ to $-\boldsymbol{\beta}'(0)$ is π. The restriction of Σ to AB is represented by the family of unit vectors parallel to those (partially) shown emanating from $\mathbf{p}$ in Figure 6.7, and covers one half of $S^1(1)$, by the judicious choice of $\mathbf{p}$. Similarly, the restriction of Σ to BC covers the other half of $S^1(1)$. Hence the degree of Σ restricted to AB and BC is $+1$. Reversing the orientation, we obtain -1 for the degree. This completes the proof. ∎

Corollary 6.15. If $\boldsymbol{\beta}$ is a simple closed unit-speed curve of period L, then the map $s \mapsto \boldsymbol{\beta}'(s)$ maps the interval $[0,L]$ onto all of the unit circle $S^1(1)$.

There is actually a far-reaching generalization of Theorem 6.14:

Theorem 6.16. (Whitney-Graustein) *Two curves that have the same turning number are homotopic.*

For a proof of this theorem see [BeGo, page 325].

6.4 Convex Plane Curves

Any straight line ℓ divides $\mathbb{R}^2$ into two half-planes H_1 and H_2 such that

$$H_1 \cup H_2 = \mathbb{R}^2 \qquad \text{and} \qquad H_1 \cap H_2 = \ell.$$

We say that a curve $\mathscr{C}$ *lies on one side of* ℓ provided either $\mathscr{C}$ is completely contained in H_1 or $\mathscr{C}$ is completely contained in H_2.

Definition 6.17. *A plane curve is* ***convex*** *if it lies on one side of each of its tangent lines.*

Since the half-planes are closed, a straight line is certainly convex. We shall however be more concerned with closed curves in this section. Obviously, any ellipse is a convex curve, though we shall encounter many other examples. For a characterization of convex curves in terms of curvature, one needs the notion of a monotone function.

Definition 6.18. *Let $f\colon (a,b) \to \mathbb{R}$ be a function, not necessarily continuous. We say that f is* ***monotone increasing*** *provided that $s \leqslant t$ implies $f(s) \leqslant f(t)$, and* ***monotone decreasing*** *provided that $s \leqslant t$ implies $f(s) \geqslant f(t)$. If f is either monotone decreasing or monotone increasing, we say that f is* ***monotone****.*

It is easy to find examples of noncontinuous monotone functions, and also of continuous monotone functions that are not differentiable. In the differentiable case we have the following well-known result:

Lemma 6.19. *A function $f\colon (a,b) \to \mathbb{R}$ is monotone if and only if the derivative f' does not change sign on (a,b). More precisely, $f' \geqslant 0$ implies monotone increasing and $f' \leqslant 0$ implies monotone decreasing.*

A glance at any simple closed convex curve $\mathscr{C}$ convinces us that the signed curvature of $\mathscr{C}$ does not change sign. We now prove this rigorously.

Theorem 6.20. *A simple closed regular plane curve $\mathscr{C}$ is convex if and only if its curvature $\boldsymbol{\kappa 2}$ has constant sign; that is, $\boldsymbol{\kappa 2}$ is either always nonpositive or always nonnegative.*

Proof. Parametrize $\mathscr{C}$ by a unit-speed curve $\boldsymbol{\beta}$ whose turning angle is $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\beta}]$ (see Section 1.5). Since $\boldsymbol{\theta}' = \boldsymbol{\kappa}\mathbf{2}[\boldsymbol{\beta}]$, we must show that $\boldsymbol{\theta}$ is monotone if and only if $\boldsymbol{\beta}$ is convex, and then use Lemma 6.19.

Suppose that $\boldsymbol{\theta}$ is monotone, but that $\boldsymbol{\beta}$ is not convex. Then there exists a point $\mathbf{p}$ on $\boldsymbol{\beta}$ for which $\boldsymbol{\beta}$ lies on both sides of the tangent line ℓ to $\boldsymbol{\beta}$ at $\mathbf{p}$. Since $\boldsymbol{\beta}$ is closed, there are points $\mathbf{q}_1$ and $\mathbf{q}_2$ on opposite sides of ℓ that are farthest from ℓ.

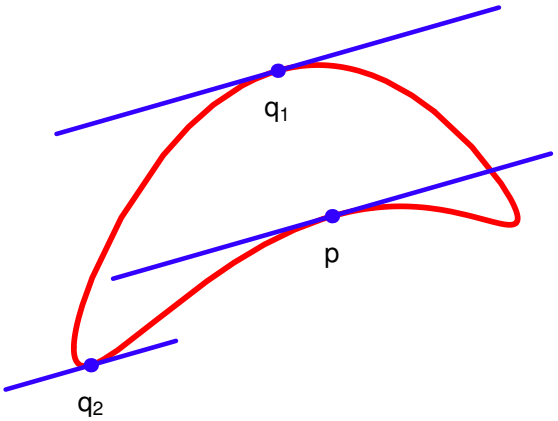


Figure 6.8: Parallel tangent lines on a nonconvex curve

The tangent lines ℓ_1 at $\mathbf{q}_1$ and ℓ_2 at $\mathbf{q}_2$ must be parallel to ℓ; see Figure 6.8. If this were not the case, we could construct a line $\tilde{\ell}$ through $\mathbf{q}_1$ (or $\mathbf{q}_2$) parallel to ℓ. Since $\tilde{\ell}$ would pass through $\mathbf{q}_1$ (or $\mathbf{q}_2$) but would not be tangent to $\boldsymbol{\beta}$, there would be points on $\boldsymbol{\beta}$ on both sides of $\tilde{\ell}$. There would then be points on $\boldsymbol{\beta}$ more distant from ℓ than $\mathbf{q}_1$ (or $\mathbf{q}_2$).

Two of the three points $\mathbf{p}, \mathbf{q}_1, \mathbf{q}_2$ must have tangents pointing in the same direction. In other words, if $\mathbf{p} = \boldsymbol{\beta}(s_0)$, $\mathbf{q}_1 = \boldsymbol{\beta}(s_1)$ and $\mathbf{q}_2 = \boldsymbol{\beta}(s_2)$, then there exist s_i and s_j with $s_i < s_j$ such that

$$\boldsymbol{\beta}'(s_i) = \boldsymbol{\beta}'(s_j) \qquad \text{and} \qquad \boldsymbol{\theta}(s_j) = \boldsymbol{\theta}(s_i) + 2n\pi,$$

for some integer n. Since $\boldsymbol{\theta}$ is monotone, Theorem 6.14 implies that $n = 0$, 1 or -1. If $n = 0$, then $\boldsymbol{\theta}(s_i) = \boldsymbol{\theta}(s_j)$, and the monotonicity of $\boldsymbol{\theta}$ implies that $\boldsymbol{\theta}$ is constant on the interval $[s_i, s_j]$. If $n = \pm 1$, then $\boldsymbol{\theta}$ is constant on the intervals $[0, s_i]$ and $[s_j, L]$. In either case, one of the segments of $\boldsymbol{\beta}$ between $\boldsymbol{\beta}(s_i)$ and $\boldsymbol{\beta}(s_j)$ is a straight line. Hence the tangent lines at $\boldsymbol{\beta}(s_i)$ and $\boldsymbol{\beta}(s_j)$ coincide. But ℓ, ℓ_1 and ℓ_2 are distinct. Thus we reach a contradiction, and so $\boldsymbol{\beta}$ must be convex.

To prove the converse, assume that $\boldsymbol{\beta}$ is convex, but that the turning angle $\boldsymbol{\theta}$ is not monotone. Then we can find s_0, s_1 and s_2 such that $s_1 < s_0 < s_2$ with

$\boldsymbol{\theta}(s_1) = \boldsymbol{\theta}(s_2) \neq \boldsymbol{\theta}(s_0)$. Corollary 6.15 says that $s \mapsto \boldsymbol{\beta}'(s)$ maps the interval $[0, L]$ onto all of the unit circle $S^1(1)$; hence there is s_3 such that $\boldsymbol{\beta}'(s_3) = -\boldsymbol{\beta}'(s_1)$. If the tangent lines at $\boldsymbol{\beta}(s_1)$, $\boldsymbol{\beta}(s_2)$ and $\boldsymbol{\beta}(s_3)$ are distinct, they are parallel, and one lies between the other two. This cannot be the case, since $\boldsymbol{\beta}$ is convex. Thus two of the tangent lines coincide, and there are points $\mathbf{p}$ and $\mathbf{q}$ of $\boldsymbol{\beta}$ lying on the same tangent line.

We show that the curve $\boldsymbol{\beta}$ is a straight line connecting $\mathbf{p}$ and $\mathbf{q}$. Let $\ell(\mathbf{p}, \mathbf{q})$ denote the straight line segment from $\mathbf{p}$ to $\mathbf{q}$. Suppose that some point $\mathbf{r}$ of $\ell(\mathbf{p}, \mathbf{q})$ is not on $\boldsymbol{\beta}$. Let $\widehat{\ell}$ be the straight line perpendicular to $\ell(\mathbf{p}, \mathbf{q})$ at $\mathbf{r}$. Since $\boldsymbol{\beta}$ is convex, $\widehat{\ell}$ is nowhere tangent to $\boldsymbol{\beta}$. Thus $\widehat{\ell}$ intersects $\boldsymbol{\beta}$ in at least two points $\mathbf{r}_1$ and $\mathbf{r}_2$. If $\mathbf{r}_1$ denotes the point closer to $\mathbf{r}$, then the tangent line to $\boldsymbol{\beta}$ at $\mathbf{r}_1$ has $\mathbf{r}_2$ on one side and one of $\mathbf{p}, \mathbf{q}$ on the other, contradicting the assumption that $\boldsymbol{\beta}$ is convex.

Hence $\mathbf{r}$ cannot exist, and so the straight line segment $\ell(\mathbf{p}, \mathbf{q})$ is contained in the trace of $\boldsymbol{\beta}$. Thus $\mathbf{p}$ and $\mathbf{q}$ are $\boldsymbol{\beta}(s_1)$ and $\boldsymbol{\beta}(s_2)$, so that the restriction of $\boldsymbol{\beta}$ to the interval $[s_1, s_2]$ is a straight line. Therefore, $\boldsymbol{\theta}$ is constant on $[s_1, s_2]$, and the assumption that $\boldsymbol{\theta}$ is not monotone leads to a contradiction. It follows that $\boldsymbol{\kappa 2}[\boldsymbol{\beta}]$ has constant sign. ∎

The proof of Theorem 6.20 contains that of the following result.

Corollary 6.21. Let $\boldsymbol{\alpha}$ be a regular simple closed curve with turning angle $\boldsymbol{\theta}[\boldsymbol{\alpha}]$. If $\boldsymbol{\theta}[\boldsymbol{\alpha}](t_1) = \boldsymbol{\theta}[\boldsymbol{\alpha}](t_2)$ with $t_1 < t_2$, then the restriction of $\boldsymbol{\alpha}$ to the interval $[t_1, t_2]$ is a straight line.

6.5 The Four Vertex Theorem

In this section, we prove a celebrated global theorem about plane curves. To understand the result, let us first consider an example.

The ***sine oval curve*** is defined by

$$\mathsf{sinoval}[n, a](t) = \big(a\cos t,\ a\sin^{(n)}(t)\big),$$

where $\sin^{(n)}(t)$ denotes the application

$$\underbrace{\sin(\sin\cdots(\sin\ t))}_{n}$$

of the sine function n times. (This iterated function is easily computable, as we shall see in Notebook 6.) Clearly, $\mathsf{sinoval}[1, a]$ is a circle of radius a. Next, $\mathsf{sinoval}[2, a]$ is the curve

$$t \mapsto \big(a\cos t,\ a\sin(\sin t)\big),$$

and so forth. As n increases, the top and bottom of $\mathsf{sinoval}[n, a]$ are pushed together more and more. A typical plot is shown in Figure 6.9.

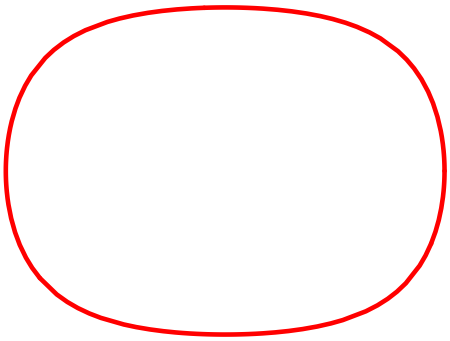


Figure 6.9: sinoval$[3, 1]$

The curvature of the sine oval has four maxima (all absolute) and four minima (two absolute and two local), as shown in the curvature graph in Figure 6.10.

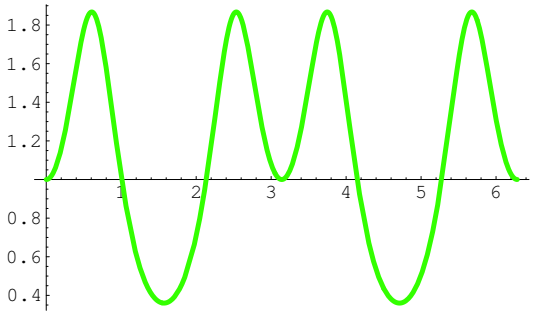


Figure 6.10: Curvature of sinoval$[3, 1]$

A simpler example is an ellipse. The curvature of an ellipse has exactly 2 maxima and 2 minima (see Exercise 9). Further examples lead naturally to the conjecture that the signed curvature of any simple closed convex curve has at least two maxima and two minima.

To prove this conjecture, we first make the following definition.

Definition 6.22. A ***vertex*** *of a regular plane curve is a point where the signed curvature has a relative maximum or minimum.*

On any closed curve, the continuous function $\boldsymbol{\kappa 2}$ must attain a maximum and a minimum, so there are at least two vertices, and they come in pairs. To find the vertices of a simple closed convex curve, we must determine those points where the derivative of the curvature vanishes. It follows from differentiating equation (1.15), page 16, that the definition of vertex is independent of the choice of regular parametrization.

We need an elementary lemma:

Lemma 6.23. *Let ℓ be a line in the plane. Then there exist constant vectors $\mathbf{a}, \mathbf{c} \in \mathbb{R}^2$ with $\mathbf{c} \neq 0$ such that $\mathbf{z} \in \ell$ if and only if $(\mathbf{z} - \mathbf{a}) \cdot \mathbf{c} = 0$.*

Proof. If we parametrize ℓ by $\boldsymbol{\alpha}(t) = \mathbf{p} + t\mathbf{q}$, we can take $\mathbf{a} = \mathbf{p}$ and $\mathbf{c} = J\mathbf{q}$. ∎

Finally, we are ready to prove the conjecture.

Theorem 6.24. (Four Vertex Theorem) *A simple closed convex curve $\boldsymbol{\alpha}$ has at least four vertices.*

Proof. The derivative $\boldsymbol{\kappa 2}'$ vanishes at each vertex of $\boldsymbol{\alpha}$. If $\boldsymbol{\kappa 2}$ is constant on any segment of $\boldsymbol{\alpha}$, then every point on the segment is a vertex, and we are done. We can therefore assume that $\boldsymbol{\alpha}$ contains neither circular arcs nor straight line segments, and that $\boldsymbol{\alpha}$ has at least two distinct vertices $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2$; without loss of generality, $\boldsymbol{\alpha}(0) = \mathbf{p}$. We now show that the assumption that $\mathbf{p}$ and $\mathbf{q}$ are the only vertices leads to a contradiction. Because vertices come in pairs, this will complete the proof of the theorem.

Let ℓ be the straight line joining $\mathbf{p}$ and $\mathbf{q}$; then ℓ divides $\boldsymbol{\alpha}$ into two segments. Since we have assumed that there are exactly two vertices, it must be the case that $\boldsymbol{\kappa 2}'$ is positive on one segment of $\boldsymbol{\alpha}$ and negative on the other. Lemma 6.23 says that there are constant vectors $\mathbf{a}$ and $\mathbf{c} \neq 0$ such that $\mathbf{z} \in \ell$ if and only if $(\mathbf{z} - \mathbf{a}) \cdot \mathbf{c} = 0$. Because $\boldsymbol{\alpha}$ is convex, $(\mathbf{z} - \mathbf{a}) \cdot \mathbf{c}$ is positive on one segment of $\boldsymbol{\alpha}$ and negative on the other.

It can be checked case by case that $\boldsymbol{\kappa 2}'(s)\big(\boldsymbol{\alpha}(s) - \mathbf{a}\big) \cdot \mathbf{c}$ does not change sign on $\boldsymbol{\alpha}$. Hence there must be an s_0 for which $\boldsymbol{\kappa 2}'(s_0)\big(\boldsymbol{\alpha}(s_0) - \mathbf{a}\big) \cdot \mathbf{c} \neq 0$, and so the integral of $\boldsymbol{\kappa 2}'(s)\big(\boldsymbol{\alpha}(s_0) - \mathbf{a}\big) \cdot \mathbf{c}$ from 0 to L is nonzero, where L is the length of $\mathscr{C}$. Integrating by parts, we obtain

$$\begin{aligned} 0 &\neq \int_0^L \boldsymbol{\kappa 2}'(s)\big(\boldsymbol{\alpha}(s) - \mathbf{a}\big) \cdot \mathbf{c}\, ds \\ &= \boldsymbol{\kappa 2}(s)\big(\boldsymbol{\alpha}(s) - \mathbf{a}\big) \cdot \mathbf{c} \Big|_0^L - \int_0^L \boldsymbol{\kappa 2}(s)\big(\boldsymbol{\alpha}'(s) \cdot \mathbf{c}\big) ds \\ &= \int_0^L \big(-\boldsymbol{\kappa 2}(s)\boldsymbol{\alpha}'(s) \cdot \mathbf{c}\big) ds. \end{aligned} \tag{6.7}$$

Lemma 1.21, page 16, implies that $J\boldsymbol{\alpha}''(s) = -\boldsymbol{\kappa 2}(s)\boldsymbol{\alpha}'(s)$, and so the last integral of (6.7) can be written as

$$\int_0^L \big(\boldsymbol{\alpha}''(s) \cdot \mathbf{c}\big) ds = \boldsymbol{\alpha}'(s) \cdot \mathbf{c} \Big|_0^L.$$

But $\boldsymbol{\alpha}'(L) = \boldsymbol{\alpha}'(0)$, so we reach a contradiction. It follows that the assumption that $\mathscr{C}$ had only two vertices is false. ∎

It turns out that any simple closed curve, convex or not, has at least four vertices; this is the result of Mukhopadhyaya [Muk]. On the other hand, it is easy to find nonsimple closed curves with only two vertices; see Exercise 5.

6.6 Curves of Constant Width

Why is a manhole cover (at least in the United States) round? Probably a square manhole cover would be easier to manufacture. But a square manhole cover, when rotated, could slip through the manhole; however, a circular manhole cover can never slip underground. The reason is that a circular manhole cover has constant width, but the width of a square manhole cover varies between a and $a\sqrt{2}$, where a is the length of a side.

Are there other curves of constant width? This question was answered affirmatively by Euler [Euler4] over two hundred years ago. A city governed by a mathematician might want to use manhole covers in the shape of a Reuleaux triangle or the involute of a deltoid, both to be discussed in Sections 6.7.

In this section we concentrate on the basic theory of curves of constant width. The literature on this subject is large: see, for example, [Bar], [Bieb, pages 27–29], [Dark], [Euler4], [Fischer, chapter 4], [HC-V, page 216], [MiPa, pages 66–71], [RaTo1, pages 137–150], [Strub, volume 1, pages 120–124] and [Stru2, pages 47–51].

Definition 6.25. *An **oval** is a simple closed plane curve for which the signed curvature $\boldsymbol{\kappa 2}$ is always strictly positive or always strictly negative.*

By Theorem 6.20, an oval is convex. The converse is false; for example, the curve $x^4 + y^4 = 1$ has points with vanishing curvature (see Notebook 6). The signed curvature of an oval can be assumed to be positive, for if the signed curvature of $\boldsymbol{\alpha}$ is always negative, one may consider instead $t \mapsto \boldsymbol{\alpha}(-t)$.

Let $\boldsymbol{\beta}\colon \mathbb{R} \to \mathbb{R}^2$ be a unit-speed oval and let $\boldsymbol{\beta}(s)$ be a point on $\boldsymbol{\beta}$. Corollary 6.15 implies that there is a point $\boldsymbol{\beta}(s^{\circ})$ on $\boldsymbol{\beta}$ for which $\boldsymbol{\beta}'(s^{\circ}) = -\boldsymbol{\beta}'(s)$, and the reasoning in the proof of Theorem 6.20 implies that $\boldsymbol{\beta}(s^{\circ})$ is unique. We call $\boldsymbol{\beta}(s^{\circ})$ the ***point opposite to*** $\boldsymbol{\beta}(s)$. Let $\widehat{\boldsymbol{\beta}}\colon \mathbb{R} \to \mathbb{R}^2$ be the curve such that $\widehat{\boldsymbol{\beta}}(s) = \boldsymbol{\beta}(s^{\circ})$ for all s. Let us write $\mathbf{T}(s) = \boldsymbol{\beta}'(s)$. Since $\mathbf{T}(s)$ and $J\mathbf{T}(s)$ are linearly independent, we can write

$$\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s) = \lambda(s)\mathbf{T}(s) + \mu(s)J\mathbf{T}(s), \tag{6.8}$$

where $\lambda, \mu\colon \mathbb{R} \to \mathbb{R}$ are piecewise-differentiable functions. Note that even though $\boldsymbol{\beta}$ has unit speed, $\widehat{\boldsymbol{\beta}}$ may not have unit speed. We put $\widehat{\mathbf{T}}(s) = \widehat{\boldsymbol{\beta}}'(s)/\|\widehat{\boldsymbol{\beta}}'(s)\|$.

Definition 6.26. *The **spread** and **width** of a unit-speed oval $\boldsymbol{\beta}\colon \mathbb{R} \to \mathbb{R}^2$ are the functions λ and μ. We say that an oval has **constant width** if μ is constant.*

Figure 6.11 shows that $|\lambda(s)|$ (respectively $|\mu(s)|$) is the distance between the normal (respectively, tangent) lines at $\boldsymbol{\beta}(s)$ and $\widehat{\boldsymbol{\beta}}(s)$. Moreover, $\sqrt{\lambda(s)^2 + \mu(s)^2}$ is the distance between the two opposite points.

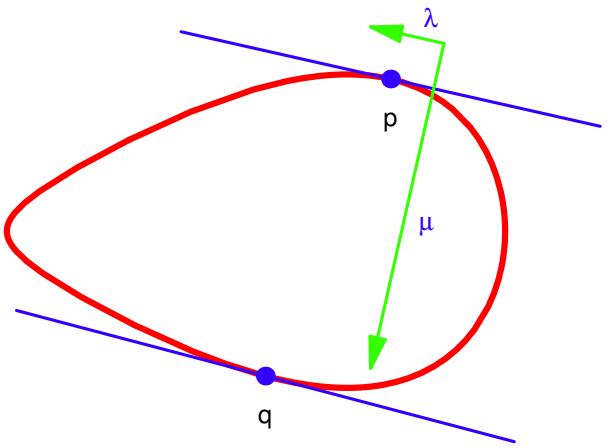


Figure 6.11: Spread and width

Next, we prove a fundamental theorem about constant width ovals.

Theorem 6.27. (Barbier[2]) *An oval of constant width w has length πw.*

Proof. We parametrize the oval by a unit-speed curve $\boldsymbol{\beta}\colon \mathbb{R} \to \mathbb{R}^2$. From (6.8) and Lemma 1.21, it follows that

$$\widehat{\boldsymbol{\beta}}' = \left(1 + \frac{d\lambda}{ds} - \mu\boldsymbol{\kappa 2}\right)\mathbf{T} + \left(\lambda\boldsymbol{\kappa 2} + \frac{d\mu}{ds}\right)J\mathbf{T}. \tag{6.9}$$

On the other hand, if $\widehat{s}$ denotes the arc length function of $\widehat{\boldsymbol{\beta}}$ we have

$$\frac{\widehat{\boldsymbol{\beta}}'}{\|\widehat{\boldsymbol{\beta}}'\|} = -\mathbf{T} \qquad \text{and} \qquad \frac{d\widehat{s}}{ds} = \|\widehat{\boldsymbol{\beta}}'\|,$$

so that

$$\widehat{\boldsymbol{\beta}}' = -\frac{d\widehat{s}}{ds}\mathbf{T}. \tag{6.10}$$

From (6.9) and (6.10) we obtain

$$\left(1 + \frac{d\widehat{s}}{ds} + \frac{d\lambda}{ds} - \mu\boldsymbol{\kappa 2}\right)\mathbf{T} + \left(\lambda\boldsymbol{\kappa 2} + \frac{d\mu}{ds}\right)J\mathbf{T} = 0. \tag{6.11}$$

Let $\vartheta = \boldsymbol{\theta}[\boldsymbol{\beta}]$ denote the turning angle, so that

$$\boldsymbol{\kappa 2} = \frac{d\vartheta}{ds}.$$

---

[2]Joseph Émile Barbier (1839–1889). French mathematician who wrote many excellent papers on differential geometry, number theory and probability.

We can use this to rewrite (6.11) as

$$\left(\frac{d(s+\widehat{s})}{ds}+\frac{d\lambda}{ds}-\mu\frac{d\vartheta}{ds}\right)\mathbf{T}+\left(\lambda\frac{d\vartheta}{ds}+\frac{d\mu}{ds}\right)J\mathbf{T}=0,$$

from which we conclude that

$$\frac{d(s+\widehat{s})}{ds}+\frac{d\lambda}{ds}-\mu\frac{d\vartheta}{ds}=0 \qquad \text{and} \qquad \lambda\frac{d\vartheta}{ds}+\frac{d\mu}{ds}=0. \tag{6.12}$$

Now suppose that μ has the constant value w. Since the curvature of $\boldsymbol{\beta}$ is always positive, the second equation of (6.12) tells us that $\lambda = 0$, and so the first equation of (6.12) reduces to

$$\frac{d(s+\widehat{s})}{ds}-w\frac{d\vartheta}{ds}=0. \tag{6.13}$$

Fix a point $\mathbf{p}$ on the oval, and let s_0 and s_1 be such that $\boldsymbol{\beta}(s_0) = \mathbf{p}$ and $\boldsymbol{\beta}(s_1) = \widehat{\mathbf{p}}$, where $\widehat{\mathbf{p}}$ is the point on the oval opposite to $\mathbf{p}$. If L denotes the length of the oval, we have from (6.13) that

$$L=\int_{s_0}^{s_1}\frac{d(s+\widehat{s})}{ds}\,ds=\int_{s_0}^{s_1}w\frac{d\vartheta}{ds}\,ds=w\int_0^{\pi}d\vartheta=w\,\pi. \qquad \blacksquare$$

An elegant generalization of Barbier's theorem, giving formulas for the width and spread of a general oval, has been proved by Mellish.[3]

Theorem 6.28. (Mellish) *The width μ of an oval parametrized by a unit-speed curve $\boldsymbol{\beta}$, as a function of ϑ, is a solution of the differential equation*

$$\frac{d^2\mu}{d\vartheta^2}+\mu=f(\vartheta), \tag{6.14}$$

where

$$f(\vartheta)=\frac{1}{\boldsymbol{\kappa}\mathbf{2}(\vartheta)}+\frac{1}{\boldsymbol{\kappa}\mathbf{2}(\vartheta+\pi)}.$$

Moreover, if we set

$$U(c)=\int_0^c f(t)\cos t\,dt, \qquad V(c)=\int_0^c f(t)\sin t\,dt,$$

then

$$\mu(\vartheta)=\big(U(\vartheta)-\tfrac{1}{2}U(\pi)\big)\sin\vartheta-\big(V(\vartheta)-\tfrac{1}{2}V(\pi)\big)\cos\vartheta.$$

Proof. We can rewrite (6.12) in terms of differentials:

$$ds+d\widehat{s}+d\lambda-\mu\,d\vartheta=0 \qquad \text{and} \qquad \lambda\,d\vartheta+d\mu=0.$$

[3] Arthur Preston Mellish (1905–1930). Canadian mathematician.

Clearly, this implies that

$$\frac{ds}{d\vartheta} + \frac{d\widehat{s}}{d\vartheta} + \frac{d\lambda}{d\vartheta} - \mu = 0 \qquad \text{and} \qquad \lambda + \frac{d\mu}{d\vartheta} = 0. \tag{6.15}$$

But

$$\frac{ds}{d\vartheta} = \frac{1}{\boldsymbol{\kappa}\mathbf{2}(\vartheta)} \qquad \text{and} \qquad \frac{d\widehat{s}}{d\vartheta} = \frac{1}{\boldsymbol{\kappa}\mathbf{2}(\vartheta+\pi)},$$

so that (6.15) becomes

$$f(\vartheta) + \frac{d\lambda}{d\vartheta} - \mu = 0 \qquad \text{and} \qquad \lambda + \frac{d\mu}{d\vartheta} = 0. \tag{6.16}$$

Elimination of λ in (6.16) yields (6.14).

The general solution of (6.14) is

$$\mu(\vartheta) = \sin\vartheta \left(\int_0^\vartheta f(t)\cos t\, dt + C_1 \right) - \cos\vartheta \left(\int_0^\vartheta f(t)\sin t\, dt + C_2 \right),$$

and the arbitrary constants C_1 and C_2 can be determined by observing that μ, λ, f are all periodic functions of ϑ with period π. ∎

There is a similar formula for the spread $\lambda(\vartheta)$ of the oval (see Exercise 13).

6.7 Reuleaux Polygons and Involutes

If we relax the condition that our closed curves be regular, simple examples with constant width can be constructed from regular polygons with odd numbers of sides. Let $\mathsf{P}[n, a]$ be a regular polygon with $2n + 1$ sides, where a denotes the length of any side. Corresponding to each vertex $\mathbf{p}$, there is a side of $\mathsf{P}[n, a]$ that is most distant from $\mathbf{p}$. Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be its vertices, and let $\widehat{\mathbf{p}_1\mathbf{p}_2}$ be the arc of the circle with center $\mathbf{p}$ connecting $\mathbf{p}_1$ and $\mathbf{p}_2$.

Definition 6.29. *The* ***Reuleaux***[4] ***polygon*** *is the curve* $\mathsf{R}[n, a]$ *made up of the circular arcs* $\widehat{\mathbf{p}_1\mathbf{p}_2}$ *formed when* $\mathbf{p}$ *ranges over the vertices of* $\mathsf{P}[n, a]$.

The Reuleaux polygon $\mathsf{R}[n, a]$ consists of $2n + 1$ arcs of a circle of radius a, each subtending an angle of $\pi/(2n+1)$; thus the length of $\mathsf{R}[n, a]$ equals πa. If we parametrize $\mathsf{R}[n, a]$ by arc length s, the associated mapping $\boldsymbol{\Phi}\colon S^1(1) \to S^1(1)$ (recall (6.3)) is undefined at the points $s = k\pi a/(2n + 1)$ with $k = 0, 1, \ldots, 2n$. Ignoring this finite number of points, the images of $\boldsymbol{\Phi}$ and $-\boldsymbol{\Phi}$ exactly cover the circle, and *no* point has an opposite in the sense of the previous section! Despite this defect, we may still assert that the width of the convex curve $\mathsf{R}[n, a]$ is constant and equal to a. For however we orient the manhole, it will always fit snugly into a pipe of diameter a.

[4] Franz Reuleaux (1829–1905). German professor of machine design.

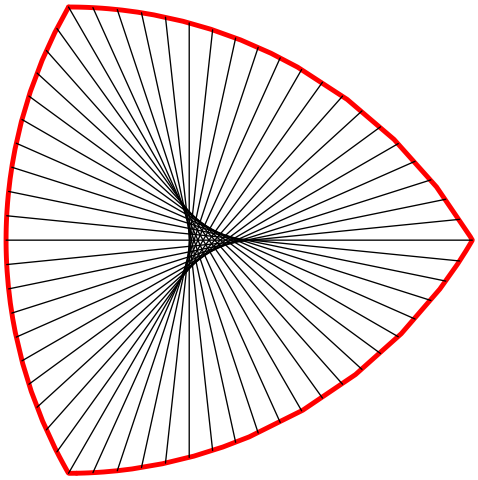


Figure 6.12: The Reuleaux triangle $\mathsf{R}[1,1]$ and a family of lines

The unit Reuleaux triangle is shown in Figure 6.12, together with straight lines joining points of the curve an arc length $\pi/2$ apart. The curve $\mathsf{R}[1,1]$ is the model for a cross section of the rotor in the Wankel[5] engine.

Curves with constant width can be effectively constructed as the involutes of a suitable curve. We shall illustrate this in terms of the deltoid, a special hypocycloid first defined in Exercise 2 of Chapter 2 on page 57.

Let $\mathscr{D}$ be a deltoid with vertices $\mathbf{a}, \mathbf{b}, \mathbf{c}$, together with a flexible cord attached to the curved side $\widehat{\mathbf{bc}}$. Keep the cord attached to a point $\mathbf{p}$ on $\widehat{\mathbf{bc}}$, and let both ends of the cord unwind. The end of the cord that was originally attached to $\mathbf{b}$ traces out a curve, part of $\mathscr{D}$'s involute, and similarly for the end of the cord attached to $\mathbf{c}$. Let $\mathbf{b}$ move to $\mathbf{q}$ and $\mathbf{c}$ to $\mathbf{r}$, and denote by $\overline{\mathbf{pq}}$ and $\overline{\mathbf{pr}}$ the line segments from $\mathbf{p}$ to $\mathbf{q}$ and from $\mathbf{p}$ to $\mathbf{r}$. By definition of the involute,

$$\mathsf{length}\,\overline{\mathbf{pq}} = \mathsf{length}\,\widehat{\mathbf{pb}} \quad \text{and} \quad \mathsf{length}\,\overline{\mathbf{pr}} = \mathsf{length}\,\widehat{\mathbf{pc}},$$

where $\widehat{\mathbf{pb}}$ and $\widehat{\mathbf{pc}}$ denote arcs of $\mathscr{D}$.

---

[5]Felix Heinrich Wankel (1902–1988). German engineer. The Wankel engine differs greatly from conventional engines. It retains the familiar intake, compression, power, and exhaust cycle but uses a rotor in the shape of a Reuleaux triangle, instead of a piston, cylinder, and mechanical valves. The Wankel engine has 40% fewer parts and roughly one third the bulk and weight of a comparable reciprocating engine. Within the Wankel, three chambers are formed by the sides of the rotor and the wall of the housing. The shape, size, and position of these chambers are constantly altered by the rotor's clockwise rotation. The engine is unique in that the power impulse is spread over approximately 270° of crank shaft rotation, as compared to 180° for the conventional reciprocating two-stroke engine.

Since the line segments $\overline{\mathbf{pq}}$ and $\overline{\mathbf{pr}}$ are both tangent to the deltoid at $\mathbf{p}$, they are part of the straight line segment $\overline{\mathbf{qr}}$ connecting $\mathbf{q}$ to $\mathbf{r}$. Consequently,

$$\text{length } \overline{\mathbf{qr}} = \text{length along the deltoid of } \widehat{\mathbf{bc}}.$$

Therefore, $\overline{\mathbf{qr}}$ has the same length, no matter the position of $\mathbf{p}$. It follows that the involute of $\mathscr{D}$ has constant width. This construction was originally carried out by Euler [Euler4] in more generality.

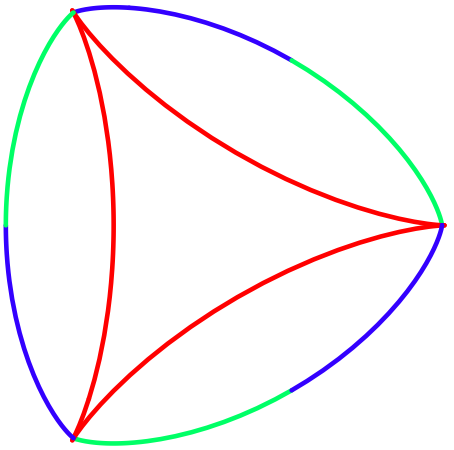


Figure 6.13: A deltoid and (on the outside) its involute

The involute of the deltoid parametrized on page 57 is given by

$$t \mapsto \frac{a}{3}\left(8\cos\frac{t}{2} + 2\cos t - \cos 2t,\ -8\sin\frac{t}{2} + 2\sin t + \sin 2t\right)$$

(refer to Notebook 6). Unlike any Reuleaux polygon, this involute is a *regular* curve of constant width. It is plotted in Figure 6.13.

6.8 The Support Function of an Oval

Given a simple closed curve $\mathscr{C}$, choose a point $\mathbf{o}$ inside $\mathscr{C}$. Let m denote the tangent line to $\mathscr{C}$ at a point $\mathbf{r}$ on $\mathscr{C}$. Let ℓ denote the line through $\mathbf{o}$ meeting m perpendicularly at a point $\mathbf{p}$. As $\mathbf{r}$ traces out $\mathscr{C}$, so $\mathbf{p}$ traces out the associated pedal curve defined in Section 4.6. Take $\mathbf{o}$ to be the origin of coordinates, and let ψ be the angle between ℓ and the x-axis. Set

$$p = \text{length } \overline{\mathbf{op}};$$

the diagram is Figure 6.17 on page 178.

The point $\mathbf{r}$ is given in polar coordinates as $re^{i\theta}$, where $r = \text{length } \overline{\mathbf{or}}$, and θ is the angle between $\overline{\mathbf{or}}$ and the x-axis. Then $p = r\cos(\psi - \theta)$, so that the line m is given by

$$\begin{aligned} p = r\cos(\psi-\theta) &= r\cos\psi\cos\theta + r\sin\psi\sin\theta \\ &= x\cos\psi + y\sin\psi. \end{aligned} \tag{6.17}$$

Since x, y can themselves be expressed in terms of ψ, we ultimately obtain p as a function of ψ. Conversely, let $p(\psi)$ be a given function of ψ; for each value of ψ, (6.17) defines a straight line, and so we obtain a family of lines. This family is defined by

$$F(x, y, \psi) = 0, \tag{6.18}$$

where $F(x, y, \psi) = p(\psi) - x\cos\psi - y\sin\psi$.

The discourse that follows applies in greater generality to a family of straight lines (or indeed, curves). When a given ψ is replaced by $\psi + \delta$, we obtain a new line implicitly defined by

$$F(x, y, \psi + \delta) = 0. \tag{6.19}$$

The set of points that belong to both curves satisfies

$$\frac{F(x, y, \psi + \delta) - F(x, y, \psi)}{\delta} = 0. \tag{6.20}$$

When we take the limit as δ tends to zero in (6.20), we obtain

$$\frac{\partial F(x, y, \psi)}{\partial \psi} = 0. \tag{6.21}$$

One calls the curve implicitly defined by eliminating ψ from (6.18) and (6.21) the ***envelope*** of the family of lines. In imprecise but descriptive language, we say that the envelope consists of those points which belong to each pair of infinitely near curves in the family (6.18). A similar argument shows that the evolute of a plane curve is the envelope of its normals, exhibited in Figure 4.12 on page 111.

In the case at hand, (6.18) and (6.21) become

$$\begin{cases} x\cos\psi + y\sin\psi &= p(\psi), \\ -x\sin\psi + y\cos\psi &= p'(\psi). \end{cases} \tag{6.22}$$

Their solution (6.22) is

$$\begin{cases} x &= p(\psi)\cos\psi - p'(\psi)\sin\psi, \\ y &= p(\psi)\sin\psi + p'(\psi)\cos\psi, \end{cases} \tag{6.23}$$

and the parameter ψ can be used to parametrize $\mathscr{C}$.

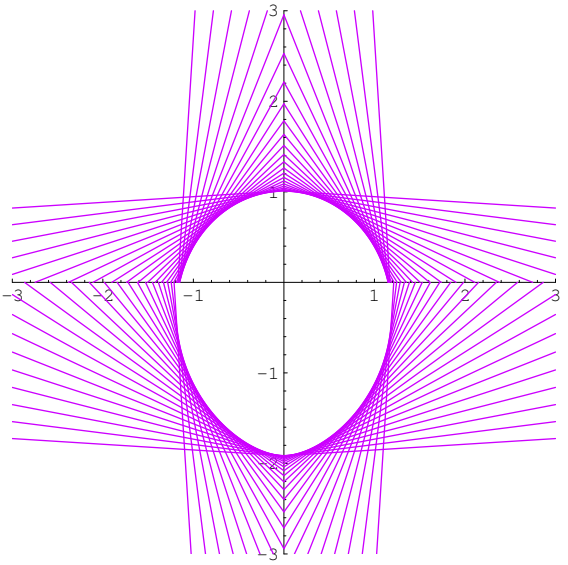


Figure 6.14: The family of straight lines for $p(\psi) = \csc(1 + c\sin\psi)$

Definition 6.30. *Let $\mathscr{C}$ be a plane curve and $\mathbf{o}$ a point. The* ***support function*** *of $\mathscr{C}$ with respect to $\mathbf{o}$ is the function p defined by* (6.17). *The* ***pedal parametrization*** *of $\mathscr{C}$ with respect to $\mathbf{o}$ is*

$$\text{oval}[p](\psi) = \big(p(\psi)\cos\psi - p'(\psi)\sin\psi,\ p(\psi)\sin\psi + p'(\psi)\cos\psi\big). \tag{6.24}$$

This formula defines a curve for any function p. However, it will only be closed if p is periodic; this fails in the example illustrated by Figure 6.16. The curve defined by (6.24) will be an oval if and only if it is a simple closed curve and $\boldsymbol{\kappa 2}$ is never zero. This is certainly the case in Figure 6.15, the envelope determined by the lines in Figure 6.14.

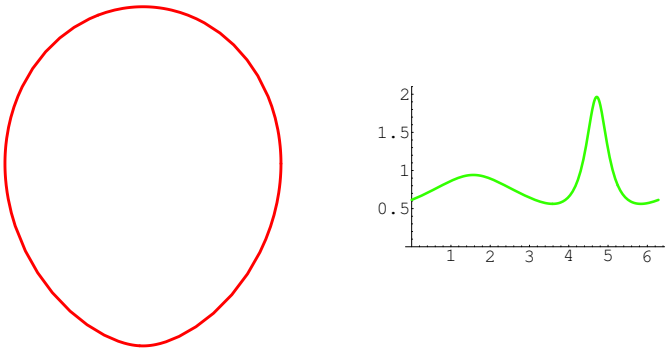


Figure 6.15: An egg and its curvature

Lemma 6.31. *Let $p\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function. The curvature of* oval$[p]$ *is given by*

$$\boldsymbol{\kappa 2}(\psi) = \frac{1}{p(\psi) + p''(\psi)}. \tag{6.25}$$

As a consequence, oval$[p]$ *is an oval if and only if it is a simple closed curve and $p(\psi) + p''(\psi)$ is never zero.*

Proof. Equation (6.25) is an easy calculation from (6.24) and the definition of $\boldsymbol{\kappa 2}$. It can also be checked by computer; see Exercise 10. ∎

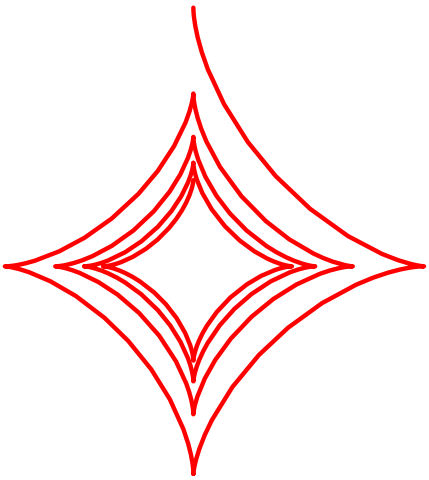


Figure 6.16: The nonclosed curve with $p(\psi) = \sin(2\psi)/\psi$

There is a formula relating the width and support functions of an oval.

Lemma 6.32. *Let p be the support function of an oval $\mathscr{C}$ with respect to some point $\mathbf{o}$ inside $\mathscr{C}$, and let μ denote the width function of $\mathscr{C}$. Then*

$$\mu(\psi) = p(\psi) + p(\psi + \pi) \tag{6.26}$$

for $0 \leqslant \psi \leqslant 2\pi$.

Proof. Let $\mathbf{r}$ be a point on $\mathscr{C}$, and let $\widehat{\mathbf{r}}$ denote the point on $\mathscr{C}$ opposite $\mathbf{r}$. Suppose that the ray from $\mathbf{o}$ to $\mathbf{p}$ meets the tangent line m to $\mathscr{C}$ at $\mathbf{r}$ perpendicularly, as in Figure 6.17. This ray intersects the tangent line to $\mathscr{C}$ at $\widehat{\mathbf{r}}$ at a point $\mathbf{q}$. The line segment $\overline{\mathbf{poq}}$ meets the two tangent lines perpendicularly, and its length equals the width of the oval measured at $\mathbf{p}$ or $\mathbf{q}$. Since the length of $\overline{\mathbf{oq}}$ is $p(\psi + \pi)$, we obtain (6.26). ∎

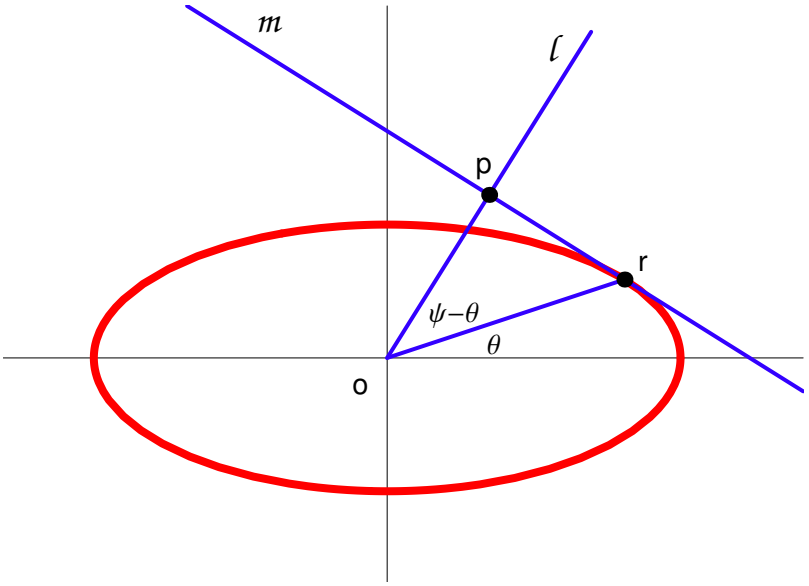


Figure 6.17: Parametrizing an oval by $p(\psi)$

6.9 Exercises

M **1.** Consider the ***piriform*** defined by

$$\mathsf{piriform}[a,b](t) = \big(a(1+\sin t),\ b\cos t(1+\sin t)\big).$$

Find a point where this curve fails to be regular, and plot $\mathsf{piriform}[1,1]$.

2. Finish the proof of Lemma 6.2.

M **3.** Show that the figure eight curve

$$t \mapsto (\sin t,\ \sin t\cos t), \qquad 0 \leqslant t \leqslant 2\pi$$

(from page 44) and the lemniscate

$$t \mapsto \left(\frac{a\cos t}{1+\sin^2 t},\ \frac{a\sin t\cos t}{1+\sin^2 t}\right), \qquad 0 \leqslant t \leqslant 2\pi$$

(from page 43), are both closed curves with total signed curvature and turning number equal to zero. Check the results by computer.

M **4.** Show that the limaçon

$$t \mapsto (2a\cos t + b)(\cos t,\ \sin t), \qquad 0 \leqslant t \leqslant 2\pi$$

(from page 58) has turning number equal to 2. Why is the turning number of a limaçon different from that of a figure eight curve or a lemniscate?

5. Verify that $\mathsf{limacon}[1, 1]$ is a nonsimple closed curve that has exactly two vertices, and plot its curvature.

6. Check that

$$\mathsf{epitrochoid}[a, b, h](t) = \mathsf{hypotrochoid}[a, -b, -h](t),$$

and explain this with reference to the definitions and Figure 6.18.

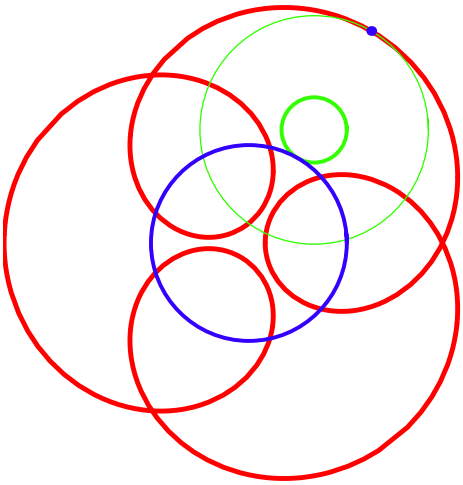


Figure 6.18: Definition of the epitrochoid

M **7.** Plot $\mathsf{hypotrochoid}[18, 2, 6]$ and compute its turning number.

M **8.** Find an explicit formula for the signed curvature of a general epitrochoid.

M **9.** Verify that a noncircular ellipse has exactly four vertices and plot its curvature.

M **10.** Prove (6.25) by computer.

11. Show that a parallel curve $\mathscr{P}$ to a closed curve $\mathscr{C}$ of constant width also has constant width, provided $\mathscr{C}$ is interior to $\mathscr{P}$.

12. Referring to Figure 6.17, suppose that $\mathbf{o} = (0, 0)$ is the center of an ellipse $\mathscr{E}$ parametrized as $(a\cos t,\, b\sin t)$. Find a relationship between $\tan t$ and $\tan\psi$, and deduce that the pedal parametrization of $\mathscr{E}$ can be obtained from the equation

$$p = \sqrt{(a^2 - b^2)\cos^2 t + b^2}.$$

13. Complete the determination of C_1, C_2 in the proof of Theorem 6.28. In the same notation, verify that the spread of an oval is given by

$$\lambda(\vartheta) = -\big(U(\vartheta) - \tfrac{1}{2}U(\pi)\big)\cos\vartheta - \big(V(\vartheta) - \tfrac{1}{2}V(\pi)\big)\sin\vartheta.$$

14. The straight lines in Figure 6.13 join opposite points on $R[1, 1]$, where "opposite" now means "half the total arc length apart." Investigate the deltoid-shaped curve formed as the envelope of this family.

Notebook 6

■ N6.0

This notebook starts, in N6.1, with a definition of total signed curvature and a computation of the turning number of a curve. In N6.2, particular attention is paid to epitrochoids and hypotrochoids, whose varied appearance is illustrated in a number of different ways.

We plot the curvature of trochoids in N6.2, and sine ovals in N6.5. A program is given to solve for (and then display) parallel tangent lines on curves, for the figure in Section 6.4. In N6.7, we plot Reuleaux polygons and deltoids. Finally, N6.8 illustrates the way in which an oval can be constructed from its support function.

At this juncture, we describe the technique mentioned in N0.1, which will be used extensively in the book's notebooks. Namely, we shall consistently suppress graphics displays unless they are called by the symbol S, which from now on is our customized version of *Mathematica*'s Show command.

This procedure is first used in Figure 6.1 in N6.1 below, and it will save us having to keep removing intermediate output manually by repeated use of the command DisplayFunction->Identity. The latter option is in fact set up as the default by the first line of the next cell, while the definition of S that follows overrides it. The use of this device will become more and more useful when plotting curves and surfaces in subsequent notebooks, at least using *Mathematica* Version 5.2 or below.

```
$DisplayFunction= Identity
  S[X__]:= Show[X, DisplayFunction->(Display[$Display,#]&)]
```

The following default options will provide different colors for multiple graphs (as in Figure 6.5). Similar options have already been used for parametric plots (as in Figure 6.3), and are included in N6.0.1.

```
SetOptions[Plot,
  Compiled->False,
  PlotStyle->Table[{Hue[.3n], Thickness[1/100]}, {n,2}]
]
```

Finally, we need the arrow package used in Notebook 1, and an abbreviation to draw Oxford blue arrows that are not too big:

```
<<Graphics`Arrow`
```

```
ar[v_,w_]:= {RGBColor[0,0,.7],
  Arrow[v,w, HeadLength->.04, HeadWidth->.6]}
```

Hidden below are definitions of the ellipse, circle, eight curve and deltoid, together with those of length, involute, J, L, $\kappa 2$. We also reload the text-placing command tx.

N6.0.1

N6.1

The definition of total signed curvature can be converted readily into *Mathematica*:

```
tsc[α_,{a_,b_}]:=
Module[{t},
  Integrate[κ2[α][t]L[α'[t]], {t,a,b}]
]
```

```
α:= ellipse[3,2]
κ2[α][t]//Simplify
tsc[α, {0,t}]
tsc[α, {0,3π/4}]//N
```

Here is a way to compute the turning number; the command Round is used to find the nearest integer to the result of a numerical integration.

```
turningnumber[α_,{a_,b_},opts___]:=
Module[{f,g,tt},
  f= κ2[α][tt]L[α'[tt]];
  g= NIntegrate[Evaluate[f], {tt,a,b}, AccuracyGoal->2, opts];
  Round[g/N[2π]]
]
```

```
turningnumber[ellipse[2,1], {0,2π}]
turningnumber[eightcurve[1], {0,2π}]
```

The following hypotrochoids are plotted using the definition of hypotrochoid from an initialization cell in N6.2. The command S is now used in place of Show to display the result. It can be applied in the usual way with square brackets, appended after //, or prepended before the symbol @. We prefer the latter method, wherever possible, in order to emphasize visually the presence of S and an important graphics display.

```
Module[{y,z},
  y[k_]= hypotrochoid[2k-1,1,k];
  z[k_]:= ParametricPlot[y[k][t], {t,0,2π}, Axes->False];
  S @ GraphicsArray[Table[z[k],{k,2,5}]];
  Table[turningnumber[y[k], {0,2π}], {k,2,5}]
]
```

Figure 6.1

N6.1.1

N6.2

Here are the generalizations of epicycloid and hypocycloid. Note that the hypotrochoid is easily defined in terms of the epitrochoid:

```
epitrochoid[a_,b_,h_]:=
   {(a+b)Cos[#]-h Cos[(a+b)#/b], (a+b)Sin[#]-h Sin[(a+b)#/b]} &
```

```
hypotrochoid[a_,b_,h_]:= epitrochoid[a,-b,-h]
```

An irrational parameter is used on purpose in the next figure.

```
α:= hypotrochoid[7, Sqrt[2], 8]
S @ ParametricPlot[α[t], {t,0,16π}, PlotStyle->Hue[0]]
```

Figure 6.3

The curvature of this curve has constant (negative) sign. Note the command to omit the x-axis (which is too far away) in the curvature plot:

```
Plot[κ2[α][t], {t,0,2π}, Axes->{False,True}]
```

We shall see shortly that the following curve has turning number 6:

```
β:= epitrochoid[5,1,3]
β[t]
```

The set val of parameter values where the tangent is parallel to the positive x-axis can be found by the numerical procedure

```
n:= 6
eq:= β'[t][[1]]/L[β'[t]]==1
fr[m_]:= t/.FindRoot[eq, {t,2π m/n}][[1]]
Off[FindRoot::cvnwt]
val:= Array[fr, n]
val
```

The arrow heads then indicate the points in question:

```
Module[{z,a},
  z[1]= ParametricPlot[β[t], {t,0,2π}];
  a[t_]= ar[β[t], β[t]+β'[t]/12];
  z[2]= Graphics[Map[a, val]];
  S @ Array[z,2]
]
```

Figure 6.4

If a/b is not rational, the curve will never close up. Nonetheless, the turning number can be computed over the range in question:

```
Module[{r,α,z,t},
  r= 16π;
  α[k_]:= epitrochoid[k,Sqrt[2],k];
  z[k_]:= ParametricPlot[α[k][t],{t,0,r}, PlotStyle->Hue[k/5]];
  t[k_]:= turningnumber[α[k],{0,r}]//Print;
  Do[{z[k], t[k]}, {k,8,10}]
]
```

▪ N6.2.1

■ N6.3

Constructing the homotopy square is an exercise in organizing graphics objects into lists. In the last line, we have added options for the command S, which is equivalent to Show except that the display is rendered visible. The presence of these options prevents us from using the symbol @ on this occasion.

```
Clear[v,z]
m:= 1
v[x_]:= {x, m - x}
z[x_]:= {Line[{{0,0}, v[x]}], Line[{v[x], {m,m}}]}
ta:= {Hue[.95], Thickness[1/70], z[0], z[1/2]}
tb:= {z[1]}
tc:= Table[z[x/16],{x,7}]
td:= {tx["A=(0,0)", {0, -.1}], tx["B=(0,L)", {0, 1.1}],
  tx["C=(L,L)", {1, 1.1}]}
gr:= Join[tb,tc,td,ta]//Graphics
S[gr, AspectRatio->Automatic, PlotRange->All, ImageSize->300]
```

Figure 6.6

▪ N6.3.1

■ N6.4

Here is an example of a closed nonconvex curve:

```
Clear[β]
β[t_]:= ellipse[5,3][t+π/2] + eightcurve[2][t]
```

```
ParametricPlot[β[t], {t,0,2π}]
```

The next program finds the values of t for which the tangent vector β'[t] has a given direction (cf. Figure 6.5). Several attempts were needed to get three values that remain distinct after adding integer multiples of 2π.

```
eq:= β'[t][[1]]/β'[t][[2]]==3.5
f[x_]:= t/.FindRoot[eq, {t, x}][[1]]
val= Array[f,3]
```

This work can now be used to construct the next figure. The program makes use of the commands Map and MapThread in place of Table to construct appropriate lists. The command β/@val is equivalent to Map[β,val], and lists the coordinates of the three points in question. The text is added by inserting subscripts directly into the input, though this necessitates the avoidance of quotation marks.

```
Clear[p]
Module[{z,li,po,t},
  z[1]= ParametricPlot[β[t], {t,0,2π}];
  li[t_]= Line[{β[t]-β'[t], β[t]+β'[t]}];
  po[t_]= Point[β[t]];
  t[3]= Map[po,val];
  t[4]= Map[li,val];
  t[2]= {Hue[.7], PointSize[.025], Thickness[.008]};
  t[1]= MapThread[tx,{{q1,"p",q2}, β/@val, Array[{0,2.2}&, 3]}];
  z[2]= Flatten[Array[t,4]]//Graphics;
  S[Array[z,2], ImageSize->400]
]
```

Figure 6.8

■ N6.5

The sine oval curve can be defined using the Nest command:

```
sinoval[n_,a_][t_]:={a Cos[t], a Nest[Sin, t, n]}
```

```
σ:= sinoval[3,1]
σ[t]
```

```
S @ ParametricPlot[σ[t], {t,0,2π}]
```

Figure 6.9

Curvature of the sine oval is represented by

```
S @ Plot[κ2[σ][t], {t,0,2π}, ImageSize->350]
```

Figure 6.10

We can list what appear to be the extreme values:

```
Table[κ2[σ][k π/4]//N, {k,0,7}]
```

```
ta:= Table[sinoval[n,1][t], {n,5}]
ParametricPlot[ta//Evaluate, {t,0,2π}]
```

■ N6.6

The next program uses the method of Figure 6.8, but the coordinates of p and q and the lengths of the arrows are determined by trial and error:

```
Module[{β,z,t,r},
  β[t_]= ellipse[5,4][t] + eightcurve[1][t];
  z[1]= ParametricPlot[β[t], {t,0,2π}];
  val:= {1, 4.75};
  li[t_]= {Thickness[.005], Line[{β[t]-β'[t], β[t]+β'[t]}]};
  po[t_]= Point[β[t]];
  t[3]= Map[po,val];
  t[4]= Map[li,val];
  t[2]= {Hue[.7], PointSize[.03], Thickness[.008]};
  t[1]= MapThread[tx, {{"p","q"}, β/@val, Array[{0,2.5}&, 2]}];
  r= β[1]+{1.5, 1};
  t[5]= {tx["l", r+{-.5,.7}], tx["m", r-{.4,5}]};
  t[6]= {Hue[.3], Arrow[r, r+.35β'[1]], Arrow[r, r+1.55J[β'[1]]]};
  z[2]= Flatten[Array[t,6]]//Graphics;
  S[Array[z,2], ImageSize->450]
]
```

Figure 6.11

■ N6.7

Reuleaux polygons can be generated by

```
reuleauxpolygon[n_,a_][t_]:=
Module[{h,f,ta},
  h= 1/(2n+1);
  f[k_]= Exp[2I π k h] + 2I Sin[π n h] Exp[π I(t+(k+n)h)];
  ta= Table[{k h ≤ t ≤ (k+1)h, a{Re[f[k]], Im[f[k]]}}, {k,0,2n}];
  Which@@Flatten[ta,1]
]
```

```
ρ[k_]:= reuleauxpolygon[k,k]
ParametricPlot[Table[ρ[k][t], {k,5}]//Evaluate, {t,0,1}, Axes->True]
```

The involute of a deltoid is computed next, and the result is saved as the expression ι[a,t] for use in Figure 6.13 in N6.7.1.

```
δ:= deltoid[1]
involute[δ,a][t]
%//PowerExpand//FullSimplify
```

```
ι[a_,t_]= {8Cos[3a/2]Cos[t/2]+2Cos[t]-Cos[2t],
  8(-Cos[3a/2]+Cos[t/2]^3)Sin[t/2]}/3
```

■ N6.7.1

■ N6.8

In this section, we shall base examples on a function p (of an angle ψ) which depends upon a parameter c, initially set to be less than one half:

```
c:= .45
p[ψ_]:= Csc[1+ c Sin[ψ]]
Plot[p[ψ], {ψ,0,2π}]
```

The next program displays the envelope of the corresponding family of lines. Each line is plotted by joining together its two points of intersection with the axes; this explains the "cross" effect.

```
Module[{li,t,gr},
  li[ψ_]= {Hue[.8], Line[{{p[ψ]/Cos[ψ],0}, {0, p[ψ]/Sin[ψ]}}]};
  t[a_,b_]:= Table[li[k 2π/100], {k,a,b}];
  gr= Join[t[1,24], t[26,49], t[51,74],t[76,99]]//Graphics;
  S[gr, AspectRatio->Automatic, Axes->True,
    PlotRange->{{-3,3},{-3,3}}, ImageSize->400]
]
```

Figure 6.14

In terms of a general function p, the pedal parametrization is

```
oval[p_][ψ_]:= {p[ψ]Cos[ψ] - p'[ψ]Sin[ψ], p[ψ]Sin[ψ] + p'[ψ]Cos[ψ]}
```

Its curvature is

```
Clear[p]
κ2[oval[p]][t]//Simplify//PowerExpand
```

A simple example gives rise to an astroid:

```
p[ψ_]:= Sin[2ψ]
γ[ψ_]= oval[p][ψ]//Simplify
ParametricPlot[γ[ψ], {ψ,0,2π}]
```

It is not so easy to find a function p that produces a closed convex curve, but our carefully chosen one gives an egg shape, already visible above. Below, we plot its curvature.

```
c:= .45;
p[ψ_]:= Csc[1+ c Sin[ψ]]
Module[{z},
  z[1]= ParametricPlot[oval[p][ψ], {ψ,0,2π}];
  z[2]= Plot[κ2[oval[p]][ψ], {ψ,0,2π}, PlotRange->{0,2.1}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->.3,
    ImageSize->500]
]
```

Figure 6.15

It is easy to verify that the curvature is everywhere positive. However, it becomes infinite if the parameter value c in the definition of p exceeds 1/2, and in this case the curve ceases to be an oval:

```
c:= .7
Module[{z},
  z[1]= ParametricPlot[oval[p][ψ], {ψ,0,2π}];
  z[2]= Plot[κ2[oval[p]][ψ], {ψ,0,2π}];
  GraphicsArray[Array[z,2]]
]
```

■ N6.8.1

■ N6.9

The next definition refers to Exercise 1:

```
piriform[a_,b_][t_]:= (1+Sin[t]){a, b Cos[t]}
```

```
ParametricPlot[piriform[1,1][t], {t,0,2π}]
```

A final computation validates the support function of an ellipse.

```
Clear[a,b]
pe[ψ_]:= Sqrt[(a^2-b^2)Cos[ψ]^2+b^2]
x= oval[pe][ψ][[1]]//Simplify
y= oval[pe][ψ][[2]]//Simplify
(x/a)^2 + (y/b)^2//Simplify
```

This can then be plotted:

```
a:= 2; b:= 1
ParametricPlot[oval[pe][ψ], {ψ,0,2π}]
```

■ N6.9.1

Chapter 7

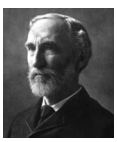
Curves in Space

In previous chapters, we have seen that the curvature $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ of a plane curve $\boldsymbol{\alpha}$ measures the failure of $\boldsymbol{\alpha}$ to be a straight line. In the present chapter, we define a similar curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ for a curve in $\mathbb{R}^n$; it measures the failure of the curve to be a straight line in n-dimensional space. The function $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ reduces to the absolute value of $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ when $n = 2$. For curves in $\mathbb{R}^3$, we can also measure the failure of the curve to lie in a plane by means of another function called the *torsion*, and denoted $\boldsymbol{\tau}[\boldsymbol{\alpha}]$.

We shall need the Gibbs[1] vector cross product to study curves in $\mathbb{R}^3$, just as we needed a complex structure to study curves in $\mathbb{R}^2$. We recall the definition and some of the properties of the vector cross product on $\mathbb{R}^3$ in Section 7.1. For completeness, and its use in Chapter 22, we take the opportunity to record analogous definitions for the complex vector spaces $\mathbb{C}^n$ and $\mathbb{C}^3$.

Curvature and torsion are defined in Section 7.2. We shall also define three orthogonal unit vector fields $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ along a space curve, that constitute the *Frenet frame field* of the curve. In Chapters 1–5, we made frequent use of the frame field $\{\mathbf{T}, J\mathbf{T}\}$ to study the geometry of a plane curve. The Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ plays the same role for space curves, but the algebra is somewhat different.

The twisting and turning of the Frenet frame field can be measured by curvature and torsion. In fact, the Frenet formulas use curvature and torsion to express the derivatives of the three vector fields of the Frenet frame field in terms



[1] Josiah Willard Gibbs (1839–1903). American physicist. When he tried to make use of Hamilton's quaternions (see Chapter 23), he found it more useful to split quaternion multiplication into a scalar part and a vector part, and to regard them as separate multiplications. Although this caused great consternation among some of Hamilton's followers, the formalism of Gibbs eventually prevailed; engineering and physics books using it started to appear in the early 1900s. Gibbs also played a large role in reviving Grassmann's vector calculus.

of the vector fields themselves. We establish the Frenet formulas for unit-speed space curves in Section 7.2, and for arbitrary-speed space curves in Section 7.4. The simplest space curves are discussed in the intervening Section 7.3. A number of other space curves, including Viviani's curve, the intersection of a sphere and a cylinder, are introduced and studied in Section 7.5.

Construction of the Frenet frame of a space curve leads one naturally to consider tubes about such curves. They are introduced in Section 7.6, and constitute our first examples of surfaces in $\mathbb{R}^3$. The torus is a special case, and we quickly generalize its definition to the case of elliptical cross-sections, though the detailed study of such surfaces does not begin until Chapter 10.

Knots form one of the most interesting classes of space curves. In the final section of this chapter we study *torus knots*, that is, knots which lie on the surface of a torus. Visualization of torus knots is considerably enhanced by making them thicker, and in practice considering the associated tube. We observe that such a tube can twist on itself, and explain how this phenomenon is influenced by the curvature and torsion of the torus knot.

7.1 The Vector Cross Product

Let us recall the notion of vector cross product on $\mathbb{R}^3$.

First, let us agree to use the notation

$$\mathbf{i} = (1,0,0), \qquad \mathbf{j} = (0,1,0), \qquad \mathbf{k} = (0,0,1).$$

For $\mathbf{a} \in \mathbb{R}^3$ we can write either $\mathbf{a} = (a_1, a_2, a_3)$ or $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$.

Definition 7.1. Let $\mathbf{a} = (a_1, a_2, a_3)$ *and* $\mathbf{b} = (b_1, b_2, b_3)$ *be vectors in* $\mathbb{R}^3$*. Then the* ***vector cross product*** *of* $\mathbf{a}$ *and* $\mathbf{b}$ *is formally given by*

$$\mathbf{a} \times \mathbf{b} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}.$$

More explicitly,

$$\mathbf{a} \times \mathbf{b} = \det \begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix} \mathbf{i} - \det \begin{pmatrix} a_1 & a_3 \\ b_1 & b_3 \end{pmatrix} \mathbf{j} + \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \mathbf{k}.$$

The vector cross product enjoys the following properties:

$$\mathbf{a} \times \mathbf{b} \;=\; -\mathbf{b} \times \mathbf{a},$$

$$(\mathbf{a} + \mathbf{b}) \times \mathbf{c} \;=\; \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c},$$

$$(\lambda\mathbf{a}) \times \mathbf{b} \;=\; \lambda(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (\lambda\mathbf{b}),$$

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) \;=\; (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}), \tag{7.1}$$

$$\|\mathbf{a} \times \mathbf{b}\|^2 \;=\; \|\mathbf{a}\|^2\|\mathbf{b}\|^2 - (\mathbf{a} \cdot \mathbf{b})^2, \tag{7.2}$$

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \;=\; (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}, \tag{7.3}$$

for $\lambda \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^3$. Frequently, (7.2), which is a special case of (7.1), is referred to as ***Lagrange's identity***.[2]

For $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ we define the ***vector triple product*** $[\mathbf{a}\,\mathbf{b}\,\mathbf{c}]$ by

$$[\mathbf{a}\,\mathbf{b}\,\mathbf{c}] = \det\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} = \det\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix},$$

where $\mathbf{a} = (a_1, a_2, a_3)$, and so forth. The vector triple product is related to the dot product and the cross product by the formulas

$$[\mathbf{a}\,\mathbf{b}\,\mathbf{c}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}), \tag{7.4}$$

reflecting properties of the determinant. In particular, $\mathbf{a} \times \mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$, and its direction is determined by the "right-hand rule."

Finally, note the following consequence of (7.2):

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\|\,\|\mathbf{b}\| \sin\theta, \tag{7.5}$$

where θ is the angle θ with $0 \leqslant \theta \leqslant \pi$ between $\mathbf{a}$ and $\mathbf{b}$. As a trivial consequence $\mathbf{a} \times \mathbf{a} = \mathbf{0}$ for any $\mathbf{a} \in \mathbb{R}^3$. Equation (7.5) can be interpreted geometrically; it says that $\|\mathbf{a} \times \mathbf{b}\|$ equals the area of the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$.

Complex Vector Algebra

For the sequel, it will be helpful to point out the differences between real and complex vector algebra. This subsection can be omitted on a first reading.

Definition 7.2. ***Complex Euclidean*** n***-space*** $\mathbb{C}^n$ *consists of the set of all complex* n*-tuples*:

$$\mathbb{C}^n = \big\{\, (p_1, \ldots, p_n) \mid p_j \text{ is a complex number for } j = 1, \ldots, n \,\big\}.$$

---

[2] 

Joseph Louis Lagrange (1736–1813). Born in Turin, Italy, Lagrange succeeded Euler as director of mathematics of the Berlin Academy of Science in 1766. In 1787 he left Berlin to become a member of the Paris Academy of Science, where he remained for the rest of his life. He made important contributions to mechanics, the calculus of variations and differential equations, in addition to differential geometry. During the 1790s he worked on the metric system and advocated a decimal base.

The elements of $\mathbb{C}^n$ are called ***complex vectors***, and one makes $\mathbb{C}^n$ into a complex vector space in the same way that we made $\mathbb{R}^n$ into a real vector space in Section 1.1. Thus if $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ are complex vectors, we define $\mathbf{p} + \mathbf{q}$ to be the element of $\mathbb{C}^n$ given by

$$\mathbf{p} + \mathbf{q} = (p_1 + q_1, \ldots, p_n + q_n).$$

Similarly, for $\lambda \in \mathbb{C}$ the vector $\lambda\mathbf{p}$ is defined by $\lambda\mathbf{p} = (\lambda p_1, \ldots, \lambda p_n)$. The ***complex dot product*** of $\mathbb{C}^n$ is given by the same formula as its real counterpart:

$$\mathbf{p} \cdot \mathbf{q} = \sum_{j=1}^{n} p_j q_j. \tag{7.6}$$

In addition to these operations, $\mathbb{C}^n$ also has a ***conjugation*** $\mathbf{p} \mapsto \overline{\mathbf{p}}$, defined by

$$\overline{\mathbf{p}} = (\overline{p}_1, \ldots, \overline{p}_n).$$

Conjugation has the following properties:

$$\overline{\mathbf{p} + \mathbf{q}} = \overline{\mathbf{p}} + \overline{\mathbf{q}}, \qquad \overline{\lambda\mathbf{p}} = \overline{\lambda}\,\overline{\mathbf{p}},$$
$$\overline{\mathbf{p} \cdot \mathbf{q}} = \overline{\mathbf{p}} \cdot \overline{\mathbf{q}}, \qquad \mathbf{p} \cdot \overline{\mathbf{p}} \geqslant 0.$$

The complex dot product should not be confused with the ***Hermitian product*** of $\mathbb{C}^n$, in which one of the two vectors in (7.6) is replaced by its conjugate. The latter is present in the definition of the ***norm***

$$\|\mathbf{p}\| = \sqrt{\mathbf{p} \cdot \overline{\mathbf{p}}} = \sqrt{|p_1|^2 + \cdots + |p_n|^2}$$

of a vector in $\mathbb{C}^n$, to ensure that this number is nonnegative. Whilst $\mathbf{p} \cdot \mathbf{p} = 0$ does not imply that $\mathbf{p} = 0$, the latter does follow if $\|\mathbf{p}\| = 0$.

Definition 7.1 extends to complex 3-tuples. Let $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ be vectors in $\mathbb{C}^3$. Written out in coordinates, the vector cross product of $\mathbf{a}$ and $\mathbf{b}$ is given by

$$\mathbf{a} \times \mathbf{b} = \big(a_2 b_3 - a_3 b_2,\ a_3 b_1 - a_1 b_3,\ a_1 b_2 - a_2 b_1\big).$$

It is easy to check that

$$\overline{\mathbf{a} \times \mathbf{b}} = \overline{\mathbf{a}} \times \overline{\mathbf{b}},$$

and that all of the identities for the real vector cross product given on page 193, with one exception, hold also for the complex vector cross product. The exception is the Lagrange identity (7.2), whose modification we include in the following lemma:

Lemma 7.3. *For* $\mathbf{a}, \mathbf{b} \in \mathbb{C}^3$ *we have*

$$\|\mathbf{a} \times \mathbf{b}\|^2 = \|\mathbf{a}\|^2\|\mathbf{b}\|^2 - |\mathbf{a} \cdot \overline{\mathbf{b}}|^2, \tag{7.7}$$

$$\|\mathbf{a} \times \overline{\mathbf{a}}\| = \|\mathbf{a}\|^2 \quad \text{if} \quad \mathbf{a} \cdot \mathbf{a} = 0, \tag{7.8}$$

$$\mathbf{a} \times \overline{\mathbf{a}} = 2i\big(\mathfrak{Im}(a_2\overline{a}_3),\, \mathfrak{Im}(a_3\overline{a}_1),\, \mathfrak{Im}(a_1\overline{a}_2)\big), \tag{7.9}$$

where $\mathbf{a} = (a_1, a_2, a_3)$ *and* $\mathfrak{Im}\, z$ *denotes the imaginary part of* z.

Proof. This is straightforward; for example, (7.8) follows directly because

$$\begin{aligned}\|\mathbf{a} \times \overline{\mathbf{a}}\|^2 = (\mathbf{a} \times \overline{\mathbf{a}}) \cdot (\overline{\mathbf{a} \times \overline{\mathbf{a}}}) &= -(\mathbf{a} \times \overline{\mathbf{a}}) \cdot (\mathbf{a} \times \overline{\mathbf{a}}) \\ &= -(\mathbf{a} \cdot \mathbf{a})(\overline{\mathbf{a}} \cdot \overline{\mathbf{a}}) + (\mathbf{a} \cdot \overline{\mathbf{a}})^2 = \|\mathbf{a}\|^4. \ \blacksquare\end{aligned}$$

7.2 Curvature and Torsion of Unit-Speed Curves

We first define the curvature of a *unit-speed* curve in $\mathbb{R}^n$. The definitions are more straightforward in this case, though arbitrary-speed curves in $\mathbb{R}^3$ will be considered in Section 7.4.

Definition 7.4. *Let* $\boldsymbol{\beta}\colon (c, d) \to \mathbb{R}^n$ *be a unit-speed curve. Write*

$$\boldsymbol{\kappa}[\boldsymbol{\beta}](s) = \|\boldsymbol{\beta}''(s)\|.$$

Then the function $\boldsymbol{\kappa}[\boldsymbol{\beta}]\colon (c, d) \to \mathbb{R}$ *is called the* ***curvature*** *of* $\boldsymbol{\beta}$.

We abbreviate $\boldsymbol{\kappa}[\boldsymbol{\beta}]$ to $\boldsymbol{\kappa}$ when there is no danger of confusion.

Intuitively, curvature measures the failure of a curve to be a straight line. More precisely, a straight line in $\mathbb{R}^n$ is characterized by the fact that its curvature vanishes, as we show in

Lemma 7.5. *Let* $\boldsymbol{\beta}\colon (c, d) \to \mathbb{R}^n$ *be a unit-speed curve. The following conditions are equivalent*:

(i) $\boldsymbol{\kappa} \equiv 0$;

(ii) $\boldsymbol{\beta}'' \equiv 0$;

(iii) $\boldsymbol{\beta}$ *is a straight line segment.*

Proof. It is clear from the definition that (i) and (ii) are equivalent. To show that (ii) and (iii) are equivalent, suppose that $\boldsymbol{\beta}'' \equiv 0$. Integration of this n-tuple of differential equations yields

$$\boldsymbol{\beta}(s) = \mathbf{u}\, s + \mathbf{v}, \tag{7.10}$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are constant vectors with $\|\mathbf{u}\| = 1$. Then (7.10) is a unit-speed parametrization of a straight line in $\mathbb{R}^n$. Conversely, any straight line in $\mathbb{R}^n$ has a parametrization of the form (7.10), which implies that $\boldsymbol{\beta}'' \equiv 0$. $\blacksquare$

In contrast to the signed curvature $\boldsymbol{\kappa}\mathbf{2}$ defined in Chapter 1, the curvature $\boldsymbol{\kappa}$ is manifestly nonnegative. For a curve in the plane, the two quantities are however related in the obvious way:

Lemma 7.6. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed curve. Then*

$$\boldsymbol{\kappa}[\boldsymbol{\beta}] = \big|\boldsymbol{\kappa}\mathbf{2}[\beta]\big|.$$

Proof. From Lemma 1.21 we have

$$\boldsymbol{\kappa}[\boldsymbol{\beta}] = \|\boldsymbol{\beta}''\| = \big\|\boldsymbol{\kappa}\mathbf{2}[\beta]J\boldsymbol{\beta}'\big\| = \big|\boldsymbol{\kappa}\mathbf{2}[\beta]\big|\,\|\boldsymbol{\beta}'\| = \big|\boldsymbol{\kappa}\mathbf{2}[\beta]\big|. \;\blacksquare$$

Definition 7.7. *Let $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^n$ be a unit-speed curve. Then*

$$\mathbf{T} = \boldsymbol{\beta}'$$

is called the ***unit tangent vector field*** *of $\boldsymbol{\beta}$.*

We need a general fact about a vector field along a curve. (The definition of vector field along a curve was given in Section 1.2.)

Lemma 7.8. *If $\mathbf{F}$ is a vector field of unit length along a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$, then $\mathbf{F}' \cdot \mathbf{F} = 0$.*

Proof. Differentiating the equation

$$\mathbf{F}(t) \cdot \mathbf{F}(t) = 1$$

gives $2\mathbf{F}(t) \cdot \mathbf{F}'(t) = 0$. $\blacksquare$

Taking $\mathbf{F} = \mathbf{T}$ in Lemma 7.8, we see that $\mathbf{T}' \cdot \mathbf{T} = 0$.

We now restrict our attention to unit-speed curves in $\mathbb{R}^3$ whose curvature is strictly positive. This implies that $\mathbf{T}' = \boldsymbol{\beta}''$ never vanishes. Now we can define the torsion, as well as the vector fields $\mathbf{N}$ and $\mathbf{B}$.

Definition 7.9. *Let $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^3$ be a unit-speed curve, and suppose that $\boldsymbol{\kappa}(s) > 0$ for $c < s < d$. The vector field*

$$\mathbf{N} = \frac{1}{\boldsymbol{\kappa}}\mathbf{T}'$$

is called the ***principal normal vector field*** *and $\mathbf{B} = \mathbf{T} \times \mathbf{N}$ is called the* ***binormal vector field****. The triple $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is called the* ***Frenet[3] frame field*** *on $\boldsymbol{\beta}$.*

---

[3] Jean Frédéric Frenet (1816–1900). French mathematician. Professor at Toulouse and Lyon.

The Frenet frame at a given point of a space curve is illustrated in Figure 7.1.

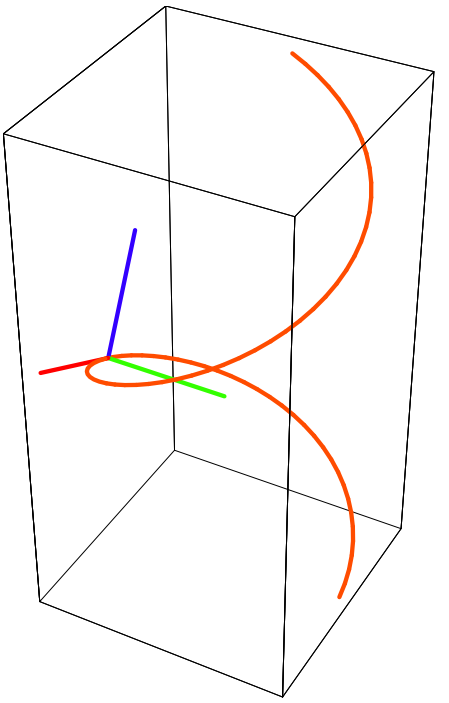


Figure 7.1: Frenet frame on the helix

We are now ready to establish the Frenet formulas; they form one of the basic tools for the differential geometry of space curves.

Theorem 7.10. *Let* $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^3$ *be a unit-speed curve with* $\boldsymbol{\kappa}(s) > 0$ *for* $c < s < d$. *Then*:

(i) $\|\mathbf{T}\| = \|\mathbf{N}\| = \|\mathbf{B}\| = 1$ *and* $\mathbf{T}\cdot\mathbf{N} = \mathbf{N}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{T} = 0$.

(ii) *Any vector field* $\mathbf{F}$ *along* $\boldsymbol{\beta}$ *can be expanded as*

$$\mathbf{F} = (\mathbf{F}\cdot\mathbf{T})\mathbf{T} + (\mathbf{F}\cdot\mathbf{N})\mathbf{N} + (\mathbf{F}\cdot\mathbf{B})\mathbf{B}. \tag{7.11}$$

(iii) *The* ***Frenet formulas*** *hold*:

$$\begin{cases} \mathbf{T}' = & \boldsymbol{\kappa}\mathbf{N}, \\ \mathbf{N}' = -\boldsymbol{\kappa}\mathbf{T} & +\boldsymbol{\tau}\mathbf{B}, \\ \mathbf{B}' = & -\boldsymbol{\tau}\mathbf{N}. \end{cases} \tag{7.12}$$

Here, the function $\boldsymbol{\tau} = \boldsymbol{\tau}[\boldsymbol{\beta}]$ *is called the* ***torsion*** *of the curve* $\boldsymbol{\beta}$.

Proof. By definition $\|\mathbf{T}\| = 1$; furthermore,

$$\|\mathbf{N}\| = \left\| \frac{1}{\boldsymbol{\kappa}} \mathbf{T}' \right\| = \frac{\|\boldsymbol{\beta}''\|}{|\boldsymbol{\kappa}|} = 1,$$

and by Lemma 7.8 we have $\mathbf{T} \cdot \mathbf{N} = 0$. Therefore, the Lagrange identity (7.2) implies that

$$\|\mathbf{B}\|^2 = \|\mathbf{T} \times \mathbf{N}\|^2 = \|\mathbf{T}\|^2 \|\mathbf{N}\|^2 - (\mathbf{T} \cdot \mathbf{N})^2 = \|\mathbf{T}\|^2 \|\mathbf{N}\|^2 = 1.$$

Finally, (7.4) implies that $\mathbf{B} \cdot \mathbf{T} = \mathbf{B} \cdot \mathbf{N} = 0$.

Since $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ are mutually orthogonal, they form a basis for the vector fields along $\boldsymbol{\beta}$. Hence there exist functions λ, μ, ν such that

$$\mathbf{F} = \lambda \mathbf{T} + \mu \mathbf{N} + \nu \mathbf{B}. \tag{7.13}$$

When we take the dot product of both sides of (7.13) with $\mathbf{T}$ and use (i), we find that $\lambda = \mathbf{F} \cdot \mathbf{T}$. Similarly, $\mu = \mathbf{F} \cdot \mathbf{N}$ and $\nu = \mathbf{F} \cdot \mathbf{B}$, proving (ii).

For (iii), we first observe that the first equation of (7.12) holds by definition of $\mathbf{N}$. To prove the third equation of (7.12), we differentiate $\mathbf{B} \cdot \mathbf{T} = 0$, obtaining $\mathbf{B}' \cdot \mathbf{T} + \mathbf{B} \cdot \mathbf{T}' = 0$; then

$$\mathbf{B}' \cdot \mathbf{T} = -\mathbf{B} \cdot \mathbf{T}' = -\mathbf{B} \cdot \boldsymbol{\kappa} \mathbf{N} = 0.$$

Lemma 7.8 implies $\mathbf{B}' \cdot \mathbf{B} = 0$. Since $\mathbf{B}'$ is perpendicular to $\mathbf{T}$ and $\mathbf{B}$, it must be a multiple of $\mathbf{N}$ by part (ii). This means that we can define the torsion $\boldsymbol{\tau}$ by the equation

$$\mathbf{B}' = -\boldsymbol{\tau} \mathbf{N}.$$

We have established the first and third of the Frenet formulas. To prove the second, we use the orthonormal expansion of $\mathbf{N}'$ in terms of $\mathbf{T}, \mathbf{N}, \mathbf{B}$ given by (ii), namely,

$$\mathbf{N}' = (\mathbf{N}' \cdot \mathbf{T})\mathbf{T} + (\mathbf{N}' \cdot \mathbf{N})\mathbf{N} + (\mathbf{N}' \cdot \mathbf{B})\mathbf{B}. \tag{7.14}$$

The coefficients in (7.14) are easy to find. First, differentiating $\mathbf{N} \cdot \mathbf{T} = 0$ we get

$$\mathbf{N}' \cdot \mathbf{T} = -\mathbf{N} \cdot \mathbf{T}' = -\mathbf{N} \cdot \boldsymbol{\kappa} \mathbf{N} = -\boldsymbol{\kappa}.$$

That $\mathbf{N}' \cdot \mathbf{N} = 0$ follows from Lemma 7.8. Finally,

$$\mathbf{N}' \cdot \mathbf{B} = -\mathbf{N} \cdot \mathbf{B}' = -\mathbf{N} \cdot (-\boldsymbol{\tau} \mathbf{N}) = \boldsymbol{\tau}.$$

Hence (7.14) reduces to the second Frenet formula. ∎

The first book containing a systematic treatment of space curves is Clairaut's *Recherches sur les courbes à double courbure*.[4] After that the term "courbe à double courbure" became a technical term for a space curve. The theory of space curves became much simpler after Frenet discovered the formulas named after him in 1847. Serret[5] found the formulas independently in 1851, and for this reason they are sometimes called the Frenet-Serret formulas (see [Frenet] and [Serret1]). In spite of their simplicity and usefulness, many years passed before they gained wide acceptance. The Frenet formulas were actually discovered for the first time in 1831 by Senff[6] and his teacher Bartels.[7] Needless to say, the isolation of Senff and Bartels in Dorpat, then a part of Russia, prevented their work from becoming widely known. (See [Reich] for details.)

It is impossible to define $\mathbf{N}$ for a straight line parametrized as a unit-speed curve $\boldsymbol{\beta}$, since $\mathbf{T}$ is a constant vector and $\boldsymbol{\kappa}$ vanishes. However, one is at liberty to take $\mathbf{N}$ and $\mathbf{B}$ to be arbitrary constant vector fields along $\boldsymbol{\beta}$, and to define the torsion of $\boldsymbol{\beta}$ to be zero. Then the formulas (7.12) remain valid (see Exercise 5 for details).

The following lemma shows that the torsion measures the failure of a curve to lie in a plane.

Lemma 7.11. *Let $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^3$ be a unit-speed curve with $\boldsymbol{\kappa}(s) > 0$ for $c < s < d$. The following conditions are equivalent:*

(i) *$\boldsymbol{\beta}$ is a plane curve;*

(ii) *$\boldsymbol{\tau} \equiv 0$.*

When (i) *and* (ii) *hold, $\mathbf{B}$ is perpendicular to the plane containing $\boldsymbol{\beta}$.*

---

4 

Alexis Claude Clairaut (1713–1765). French mathematician and astronomer, who at the age of 18 was elected to the French Academy of Sciences for his work on curve theory. In 1736–1737 he took part in an expedition to Lapland led by Maupertuis, the purpose of which was to verify Newton's theoretical proof that the earth is an oblate spheroid. Clairaut's precise calculations led to a near perfect prediction of the arrival of Halley's comet in 1759.

5 

Joseph Alfred Serret (1819–1885). French mathematician. Serret with other Paris mathematicians greatly advanced differential calculus in the period 1840–1865. Serret also worked in number theory and mechanics. Another Serret (Paul Joseph (1827–1898)) wrote a book *Théorie nouvelle géométrique et mécanique des lignes à double courbure* emphasizing space curves.

[6]Karl Eduard Senff (1810–1849). German professor at the University of Dorpat (now Tartu) in Estonia.

[7]Johann Martin Bartels (1769–1836). Another German professor at the University of Dorpat. Bartels was the first teacher of Gauss (in Brunswick); he and Gauss kept in contact over the years.

Proof. The condition that a curve $\boldsymbol{\beta}$ lie in a plane Π can be expressed analytically as

$$(\boldsymbol{\beta}(s) - \mathbf{p}) \cdot \mathbf{q} \equiv 0,$$

where $\mathbf{q}$ is a nonzero vector perpendicular to Π. Differentiation yields

$$\boldsymbol{\beta}'(s) \cdot \mathbf{q} \equiv 0 \equiv \boldsymbol{\beta}''(s) \cdot \mathbf{q}.$$

Thus both $\mathbf{T}$ and $\mathbf{N}$ are perpendicular to $\mathbf{q}$. Since $\mathbf{B}$ is also perpendicular to $\mathbf{T}$ and $\mathbf{N}$, it follows that $\mathbf{B}(s) = \pm\mathbf{q}/\|\mathbf{q}\|$ for all s. Therefore $\mathbf{B}' \equiv 0$, and from the definition of torsion, it follows that $\boldsymbol{\tau} \equiv 0$.

Conversely, suppose that $\boldsymbol{\tau} \equiv 0$; then (7.12) implies that $\mathbf{B}' \equiv -\boldsymbol{\tau}\mathbf{N} \equiv 0$. Thus $s \mapsto \mathbf{B}(s)$ is a constant curve; that is, there exists a unit vector $\mathbf{u} \in \mathbb{R}^3$ such that $\mathbf{B} \equiv \mathbf{u}$. Choose t_0 with $c < t_0 < d$ and consider the real-valued function $f\colon (c,d) \to \mathbb{R}^3$ given by

$$f(s) = \big(\boldsymbol{\beta}(s) - \boldsymbol{\beta}(t_0)\big) \cdot \mathbf{u}.$$

Then $f(t_0) = 0$ and $f'(s) \equiv \boldsymbol{\beta}'(s) \cdot \mathbf{u} \equiv \mathbf{T} \cdot \mathbf{B} \equiv 0$, so that f is identically zero. Thus $\big(\boldsymbol{\beta}(s) - \boldsymbol{\beta}(t_0)\big) \cdot \mathbf{u} \equiv 0$, and it follows that $\boldsymbol{\beta}$ lies in the plane perpendicular to $\mathbf{u}$ that passes through $\boldsymbol{\beta}(t_0)$. ∎

7.3 The Helix and Twisted Cubic

The ***circular helix*** is a curve in $\mathbb{R}^3$ that resembles a metal spring. Its usual parametrization is

$$\mathsf{helix}[a,b](t) = \big(a\cos t,\ a\sin t,\ bt\big), \tag{7.15}$$

where $a > 0$ is the radius and b is the ***incline*** of the helix. Observe that the projection of $\mathbb{R}^3$ onto the xy-plane maps the helix onto the circle $(a\cos t,\ a\sin t)$ of radius a. Figure 7.2 displays the helix $t \mapsto (\cos t,\ \sin t,\ \frac{1}{5}t)$.

The helix is one of the few curves for which a unit-speed parametrization is easy to find. In fact, a unit-speed parametrization of the helix is given by

$$\boldsymbol{\beta}(s) = \left(a\cos\frac{s}{\sqrt{a^2+b^2}},\ a\sin\frac{s}{\sqrt{a^2+b^2}},\ \frac{bs}{\sqrt{a^2+b^2}}\right).$$

We use Theorem 7.10 to compute $\boldsymbol{\kappa}, \boldsymbol{\tau}$ and $\mathbf{T}, \mathbf{N}, \mathbf{B}$ for $\boldsymbol{\beta}$. First,

$$\mathbf{T}(s) = \boldsymbol{\beta}'(s) = \frac{1}{\sqrt{a^2+b^2}}\left(-a\sin\frac{s}{\sqrt{a^2+b^2}},\ a\cos\frac{s}{\sqrt{a^2+b^2}},\ b\right);$$

thus $\|\mathbf{T}(s)\| = 1$ and $\boldsymbol{\beta}$ is indeed a unit-speed curve. Moreover,

$$\mathbf{T}'(s) = \frac{1}{a^2+b^2}\left(-a\cos\frac{s}{\sqrt{a^2+b^2}},\ -a\sin\frac{s}{\sqrt{a^2+b^2}},\ 0\right).$$

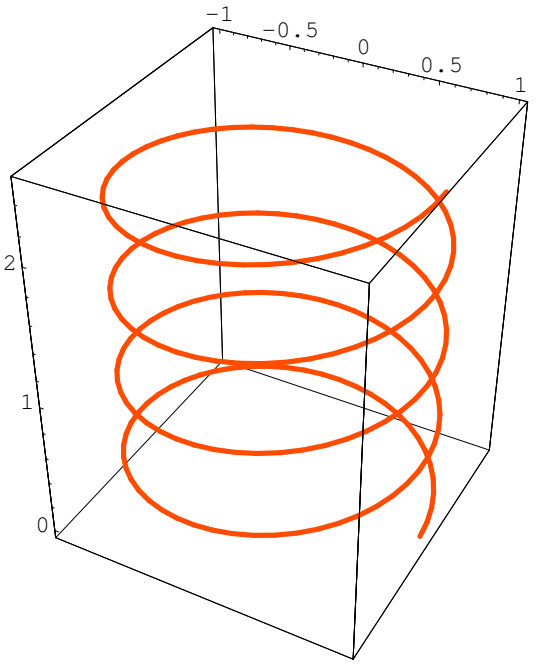


Figure 7.2: Part of the trace of helix$[1, 0.2]$

Since $a > 0$,

$$\boldsymbol{\kappa}(s) = \|\mathbf{T}'(s)\| = \frac{a}{a^2 + b^2}, \tag{7.16}$$

and

$$\mathbf{N}(s) = \frac{\mathbf{T}'(s)}{\|\mathbf{T}'(s)\|} = \left(-\cos\frac{s}{\sqrt{a^2 + b^2}},\ -\sin\frac{s}{\sqrt{a^2 + b^2}},\ 0 \right). \tag{7.17}$$

From the formula $\mathbf{B} = \mathbf{T} \times \mathbf{N}$ we get

$$\mathbf{B}(s) = \frac{1}{\sqrt{a^2 + b^2}} \left(b \sin\frac{s}{\sqrt{a^2 + b^2}},\ -b \cos\frac{s}{\sqrt{a^2 + b^2}},\ a \right).$$

Finally, to compute the torsion we note that

$$\mathbf{B}'(s) = \frac{1}{a^2 + b^2} \left(b \cos\frac{s}{\sqrt{a^2 + b^2}},\ b \sin\frac{s}{\sqrt{a^2 + b^2}},\ 0 \right).$$

When we compare this to (7.17), and use the Frenet formula $\mathbf{B}' = -\boldsymbol{\tau}\mathbf{N}$, we see that

$$\boldsymbol{\tau}(s) = \frac{b}{a^2 + b^2}.$$

In conclusion, both the curvature and torsion of a helix are constant, and there is no need to graph them!

Although we know that every regular curve has a unit-speed parametrization, in practice it is very difficult to find it. For example, consider the ***twisted cubic*** defined by

$$\mathsf{twicubic}(t) = (t,\, t^2,\, t^3). \tag{7.18}$$

One can readily check that this curve does not lie in a plane. For any such plane would have to be parallel to all three vectors

$$\boldsymbol{\gamma}'(0) = (1,0,0), \quad \boldsymbol{\gamma}''(0) = (0,2,0), \quad \boldsymbol{\gamma}'''(0) = (0,0,6),$$

where $\boldsymbol{\gamma} = \mathsf{twicubic}$. But this is clearly impossible.

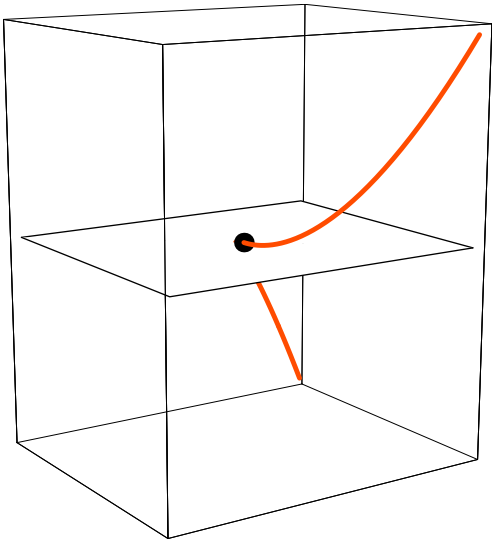


Figure 7.3: twicubic and the plane $z = 0$

Since $\boldsymbol{\gamma}'(0)$ is a unit vector, it coincides with $\mathbf{T}$ at $t = 0$. On the other hand, $\boldsymbol{\gamma}'(t)$ is not a unit vector unless $t = 0$, so $\boldsymbol{\gamma}''(0)$ is not parallel to $\mathbf{N}$. Nonetheless, $\boldsymbol{\gamma}''(0)$ is a linear combination of $\mathbf{T}$ and $\mathbf{N}$, and both pairs $\{\boldsymbol{\gamma}'(0),\, \boldsymbol{\gamma}''(0)\}$ and $\{\mathbf{T},\, \mathbf{N}\}$ (when applied to the origin) generate the xy-plane, shown in Figure 7.3.

The arc length function of the curve (7.18), starting at the origin, is

$$s(t) = \int_0^t \|\boldsymbol{\gamma}'(u)\|\, du = \int_0^t \sqrt{1 + 4u^2 + 9u^4}\, du.$$

The inverse of $s(t)$, which is needed to find the unit-speed parametrization, is an elliptic function that is too complicated to be of much use. We shall show how to compute the curvature and torsion of (7.18) more directly in the next section.

7.4 Arbitrary-Speed Curves in $\mathbb{R}^3$

For efficient computation of the curvature and torsion of an arbitrary-speed curve, we need formulas that bypass finding a unit-speed parametrization. Although we shall define the curvature and torsion of an arbitrary-speed curve in terms of the curvature and torsion of its unit-speed parametrization, ultimately we shall find formulas for these quantities that avoid finding a unit-speed parametrization explicitly.

The theoretical definition is

Definition 7.12. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular curve, and let $\widetilde{\boldsymbol{\alpha}}\colon (c,d) \to \mathbb{R}^3$ be a unit-speed reparametrization of $\boldsymbol{\alpha}$. Write $\boldsymbol{\alpha}(t) = \widetilde{\boldsymbol{\alpha}}(s(t))$, where $s(t)$ is the arc length function. Denote by $\widetilde{\boldsymbol{\kappa}}$ and $\widetilde{\boldsymbol{\tau}}$ the curvature and torsion of $\widetilde{\boldsymbol{\alpha}}$, respectively. Also, let $\{\widetilde{\mathbf{T}}, \widetilde{\mathbf{N}}, \widetilde{\mathbf{B}}\}$ be the Frenet frame field of $\widetilde{\boldsymbol{\alpha}}$. Then we define*

$$\boldsymbol{\kappa}(t) = \widetilde{\boldsymbol{\kappa}}\big(s(t)\big), \qquad \boldsymbol{\tau}(t) = \widetilde{\boldsymbol{\tau}}\big(s(t)\big),$$
$$\mathbf{T}(t) = \widetilde{\mathbf{T}}\big(s(t)\big), \quad \mathbf{N}(t) = \widetilde{\mathbf{N}}\big(s(t)\big), \quad \mathbf{B}(t) = \widetilde{\mathbf{B}}\big(s(t)\big).$$

Thus, the curvature, torsion and Frenet frame field of an arbitrary-speed curve $\boldsymbol{\alpha}$ are reparametrizations of those of a unit-speed parametrization of $\boldsymbol{\alpha}$.

Next, we generalize the Frenet formulas (7.12) to arbitrary-speed curves.

Theorem 7.13. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular curve with speed $v = \|\boldsymbol{\alpha}'\| = s'$. Then the following generalizations of the Frenet formulas hold:*

$$\text{(7.19)} \qquad \begin{cases} \mathbf{T}' = v\boldsymbol{\kappa}\mathbf{N}, \\ \mathbf{N}' = -v\boldsymbol{\kappa}\mathbf{T} + v\boldsymbol{\tau}\mathbf{B}, \\ \mathbf{B}' = -v\boldsymbol{\tau}\mathbf{N}. \end{cases}$$

Proof. By the chain rule we have

$$\text{(7.20)} \qquad \begin{cases} \mathbf{T}'(t) = s'(t)\widetilde{\mathbf{T}}'\big(s(t)\big) = v(t)\widetilde{\mathbf{T}}'\big(s(t)\big), \\ \mathbf{N}'(t) = s'(t)\widetilde{\mathbf{N}}'\big(s(t)\big) = v(t)\widetilde{\mathbf{N}}'\big(s(t)\big), \\ \mathbf{B}'(t) = s'(t)\widetilde{\mathbf{B}}'\big(s(t)\big) = v(t)\widetilde{\mathbf{B}}'\big(s(t)\big). \end{cases}$$

Thus from the Frenet formulas for $\widetilde{\boldsymbol{\alpha}}$, it follows that

$$\mathbf{T}'(t) = v(t)\widetilde{\boldsymbol{\kappa}}\big(s(t)\big)\widetilde{\mathbf{N}}\big(s(t)\big) = v(t)\boldsymbol{\kappa}(t)\mathbf{N}(t).$$

The other two formulas of (7.20) are proved similarly. ∎

Lemma 7.14. *The velocity $\boldsymbol{\alpha}'$ and acceleration $\boldsymbol{\alpha}''$ of a regular curve $\boldsymbol{\alpha}$ are given by*

$$\boldsymbol{\alpha}' = v\mathbf{T}, \tag{7.21}$$

$$\boldsymbol{\alpha}'' = \frac{dv}{dt}\mathbf{T} + \boldsymbol{\kappa} v^2 \mathbf{N}, \tag{7.22}$$

where v denotes the speed of $\boldsymbol{\alpha}$.

Proof. Write $\boldsymbol{\alpha}(t) = \widetilde{\boldsymbol{\alpha}}\big(s(t)\big)$, where $\widetilde{\boldsymbol{\alpha}}$ is a unit-speed parametrization of $\boldsymbol{\alpha}$. By the chain rule we have

$$\boldsymbol{\alpha}'(t) = \widetilde{\boldsymbol{\alpha}}'\big(s(t)\big)s'(t) = v(t)\widetilde{\mathbf{T}}\big(s(t)\big) = v(t)\mathbf{T}(t),$$

proving (7.21). Next, we take the derivative of (7.21) and use the first equation of (7.19) to get

$$\boldsymbol{\alpha}'' = v'\mathbf{T} + v\mathbf{T}' = v'\mathbf{T} + v^2\boldsymbol{\kappa}\mathbf{N},$$

and in stating the result we have chosen to write $v' = dv/dt$. ∎

Now we can derive useful formulas for the curvature and torsion for an arbitrary-speed curve. These formulas avoid finding a unit-speed reparametrization.

Theorem 7.15. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular curve with nonzero curvature. Then*

$$\mathbf{T} = \frac{\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|}, \tag{7.23}$$

$$\mathbf{N} = \mathbf{B} \times \mathbf{T}, \tag{7.24}$$

$$\mathbf{B} = \frac{\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''}{\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\|}, \tag{7.25}$$

$$\boldsymbol{\kappa} = \frac{\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\|}{\|\boldsymbol{\alpha}'\|^3}, \tag{7.26}$$

$$\boldsymbol{\tau} = \frac{[\boldsymbol{\alpha}'\ \boldsymbol{\alpha}''\ \boldsymbol{\alpha}''']}{\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\|^2}. \tag{7.27}$$

Proof. Clearly, (7.23) is equivalent to (7.21), and (7.24) is an algebraic consequence of the definition of vector cross product. Furthermore, it follows from (7.21) and (7.22) that

$$\begin{aligned} \boldsymbol{\alpha}' \times \boldsymbol{\alpha}'' &= v\mathbf{T} \times \left(\frac{dv}{dt}\mathbf{T} + \boldsymbol{\kappa} v^2 \mathbf{N}\right) \\ &= v\frac{dv}{dt}\mathbf{T} \times \mathbf{T} + \boldsymbol{\kappa} v^3 \mathbf{T} \times \mathbf{N} = \boldsymbol{\kappa} v^3 \mathbf{B}. \end{aligned} \tag{7.28}$$

Taking norms in (7.28), we get

$$\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\| = \|\boldsymbol{\kappa} v^3 \mathbf{B}\| = \boldsymbol{\kappa} v^3. \tag{7.29}$$

Then (7.29) implies (7.26). Furthermore, (7.28) and (7.29) imply that

$$\mathbf{B} = \frac{\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''}{\boldsymbol{\kappa} v^3} = \frac{\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''}{\|\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''\|},$$

and so (7.25) is proved.

To prove (7.27) we need a formula for $\boldsymbol{\alpha}'''$ analogous to (7.21) and (7.22). Actually, all we need is the component of $\boldsymbol{\alpha}'''$ in the $\mathbf{B}$ direction, because we want to take the dot product of $\boldsymbol{\alpha}'''$ with $\boldsymbol{\alpha}' \times \boldsymbol{\alpha}''$. So we compute

$$\begin{aligned} \boldsymbol{\alpha}''' &= \left(\frac{dv}{dt}\mathbf{T} + \boldsymbol{\kappa} v^2 \mathbf{N}\right)' = \boldsymbol{\kappa} v^2 \mathbf{N}' + \cdots \\ &= \boldsymbol{\kappa}\boldsymbol{\tau} v^3 \mathbf{B} + \cdots \end{aligned} \tag{7.30}$$

where the dots represent irrelevant terms. It follows from (7.28) and (7.30) that

$$(\boldsymbol{\alpha}' \times \boldsymbol{\alpha}'') \cdot \boldsymbol{\alpha}''' = \boldsymbol{\kappa}^2 \boldsymbol{\tau} v^6. \tag{7.31}$$

Now (7.27) follows from (7.31) and (7.29). ∎

To conclude this section, we compute the curvature and torsion of the curve $\boldsymbol{\gamma} = \mathsf{twicubic}$ defined by (7.18). Instead of finding a unit-speed reparametrization of $\boldsymbol{\gamma}$, we use Theorem 7.15. The computations are easy enough to do by hand:

$$\begin{aligned} \boldsymbol{\gamma}'(t) &= (1,\, 2t,\, 3t^2), & \boldsymbol{\gamma}''(t) &= (0,\, 2,\, 6t), \\ \boldsymbol{\gamma}'(t) \times \boldsymbol{\gamma}''(t) &= (6t^2,\, -6t,\, 2), & \boldsymbol{\gamma}'''(t) &= (0,\, 0,\, 6), \end{aligned}$$

so that

$$\begin{aligned} \|\boldsymbol{\gamma}'(t)\| &= (1 + 4t^2 + 9t^4)^{1/2}, \\ \|\boldsymbol{\gamma}'(t) \times \boldsymbol{\gamma}''(t)\| &= (4 + 36t^2 + 36t^4)^{1/2}. \end{aligned}$$

Therefore, the curvature and torsion of the twisted cubic are given by

$$\begin{aligned} \boldsymbol{\kappa}(t) &= \frac{(4 + 36t^2 + 36t^4)^{1/2}}{(1 + 4t^2 + 9t^4)^{3/2}}, \\ \boldsymbol{\tau}(t) &= \frac{3}{1 + 9t^2 + 9t^4}. \end{aligned}$$

Their respective maximum values are 2 and 3, and the graphs are plotted in Figure 7.4.

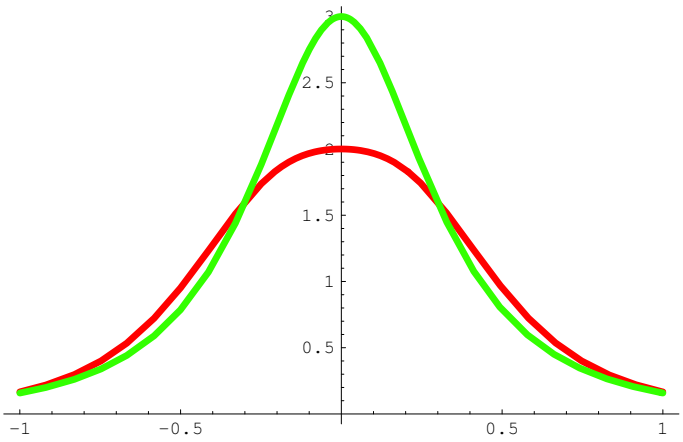


Figure 7.4: Curvature and (lighter) torsion of a cubic

The behavior of the Frenet frame of the twisted cubic for large values of $|t|$ is discussed on page 795 in Chapter 23.

7.5 More Constructions of Space Curves

Just as the helix sits over the circle, there are helical analogs of many other plane curves. Suppose that we are given a plane curve $t \mapsto \boldsymbol{\alpha}(t) = (x(t), y(t))$. A general formula for the helix with incline c over $\boldsymbol{\alpha}$ is

$$\mathsf{helical}[\boldsymbol{\alpha}, c](t) = \big(x(t),\ y(t),\ ct\big). \tag{7.32}$$

As an example, we may construct the helical curve over the logarithmic spiral. Using the paraemtrization on page 23, we obtain

$$\mathsf{helical}[\mathsf{logspiral}[a, b], c](t) = \big(ae^{bt}\cos t,\ ae^{bt}\sin t,\ ct\big).$$

A similar but more complicated example formed from the clothoid is shown in Figure 7.8, at the end of this section on page 209.

The formula (7.32) can of course be generalized by defining the third coordinate or "height" z to equal an arbitrary function of the parameter t. One can use this idea to graph the signed curvature of a plane curve. Suppose that $t \mapsto \boldsymbol{\alpha}(t) = (x(t),\ y(t))$ is a plane curve with signed curvature function $\boldsymbol{\kappa 2}(t)$, and define

$$\boldsymbol{\beta}(t) = \big(x(t),\ y(t),\ \boldsymbol{\kappa 2}(t)\big).$$

Figure 7.5 displays $\boldsymbol{\beta}$ when $\boldsymbol{\alpha}$ is one of the epitrochoids defined on page 157.

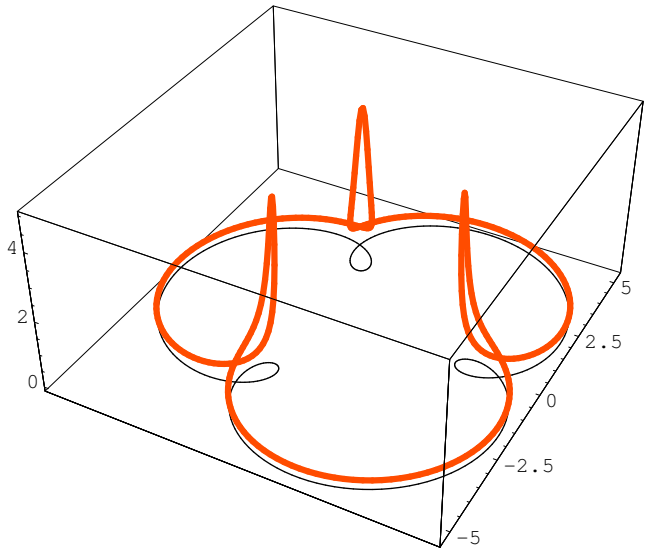


Figure 7.5: The curvature of epitrochoid$[3, 1, \frac{3}{2}]$

Let $a > 0$. Consider the sphere

$$x^2 + y^2 + z^2 = 4a^2 \tag{7.33}$$

of radius $2a$ and center the origin, and the cylinder

$$(x - a)^2 + y^2 = a^2 \tag{7.34}$$

of radius a containing the z axis and passing through the point $(2a, 0, 0)$. This point lies on ***Viviani's curve***, which is by definition the intersection of (7.33) and (7.34). Figure 7.6 displays the curve together with its projection onto the equatorial plane, and three great circles on the sphere to put the picture into perspective.

Using the identity $\cos t = 1 - 2\sin^2(t/2)$, one may verify that this intersection is parametrized by

$$\text{viviani}[a](t) = \left(a(1 + \cos t),\ a \sin t,\ 2a \sin\frac{t}{2}\right), \quad -2\pi \leqslant t \leqslant 2\pi. \tag{7.35}$$

Viviani[8] studied the curve in 1692. See [Stru2, pages 10–11] and [Gomes, volume 2, pages 311–320].

---

[8]Vincenzo Viviani (1622–1703). Student and biographer of Galileo. In 1660, together with Borelli, Viviani measured the velocity of sound by timing the difference between the flash and the sound of a cannon. In 1692, Viviani proposed the following problem which aroused much interest. How is it possible that a hemisphere has 4 equal windows of such a size that the remaining surface can be exactly squared? The answer involved the Viviani curve.

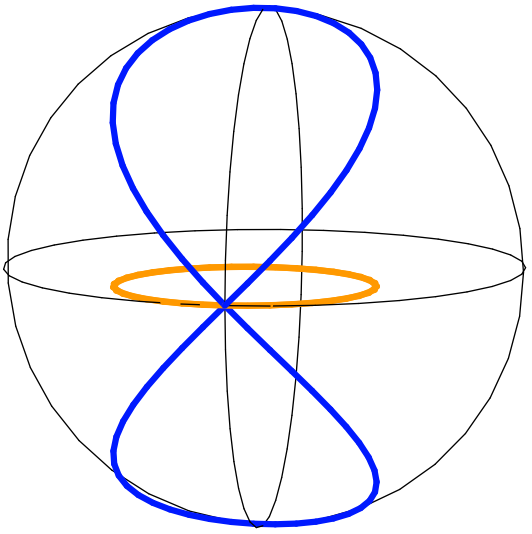


Figure 7.6: The trace of viviani[1]

Computation of the curvature and torsion of Viviani's curve in Notebook 7 yields the results

$$\begin{aligned} \boldsymbol{\kappa}(t) &= \frac{\sqrt{13+3\cos t}}{a(3+\cos t)^{3/2}}, \\ \boldsymbol{\tau}(t) &= \frac{6\cos(t/2)}{a(13+3\cos t)}. \end{aligned}$$

These functions are graphed simultaneously in Figure 7.7 for $a = 1$, $0 \leqslant t \leqslant 2\pi$. The torsion vanishes when $t = \pi$, and the most "planar" parts of Viviani's curve occur at the sphere's poles. Notebook 7 includes an animation of the Frenet frame on Viviani's curve; see also Figure 23.8 on page 795.

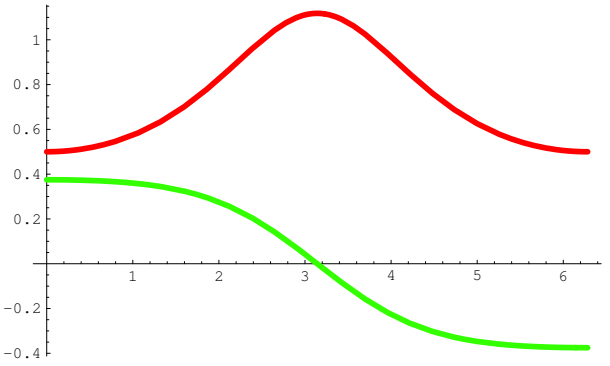


Figure 7.7: The curvature (above) and torsion (below) of viviani[1]

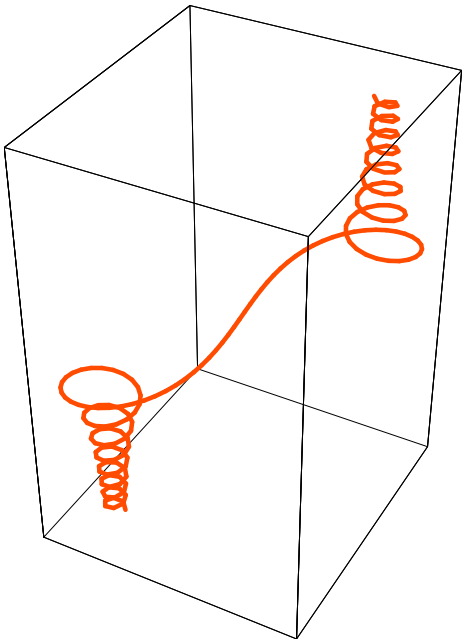


Figure 7.8: Helical curve over clothoid$[1, 1]$

7.6 Tubes and Tori

By definition, the ***tube*** of radius r around a set $\mathscr{C}$ is the set of points at a distance r from $\mathscr{C}$. In particular, let $\boldsymbol{\gamma}\colon (a, b) \to \mathbb{R}^3$ be a regular curve whose curvature does not vanish. Since the normal $\mathbf{N}$ and binormal $\mathbf{B}$ are perpendicular to $\boldsymbol{\gamma}$, the circle

$$\theta \mapsto r\big(-\cos\theta\,\mathbf{N}(t) + \sin\theta\,\mathbf{B}(t)\big)$$

is perpendicular to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$. As this circle moves along $\boldsymbol{\gamma}$ it traces out a surface about $\boldsymbol{\gamma}$, which will be the tube about $\boldsymbol{\gamma}$, provided r is not too large.

We can parametrize this surface by

$$\text{tubecurve}[\boldsymbol{\gamma}, r](t, \theta) = \boldsymbol{\gamma}(t) + r\big(-\cos\theta\,\mathbf{N}(t) + \sin\theta\,\mathbf{B}(t)\big), \tag{7.36}$$

where $a \leqslant t \leqslant b$ and $0 \leqslant \theta < 2\pi$. Figure 7.9 displays the tube of radius $1/2$ around the trace of helix$[2, \frac{1}{2}]$.

A tube about a curve $\boldsymbol{\gamma}$ in $\mathbb{R}^3$ has the following interesting property: the volume depends only on the length of $\boldsymbol{\gamma}$ and radius of the tube. In particular, the volume of the tube does not depend on the curvature or torsion of $\boldsymbol{\gamma}$. Thus, for example, tubes of the same radius about a circle and a helix of the same length will have the same volume. For the proofs of these facts and the study of tubes in higher dimensions, see [Gray].

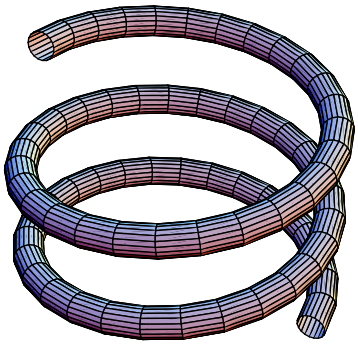


Figure 7.9: Tube around a helix

Figure 7.10 shows tubes around two ellipses. Actually, the "horizontal" tube is a circular torus, formed by rotating a circle around another circle, but both tubes have circular cross sections. More generally, an elliptical torus of revolution is parametrized by the function

$$\text{torus}[a,b,c](u,v) = \big((a+b\cos v)\cos u,\ (a+b\cos v)\sin u,\ c\sin v\big), \tag{7.37}$$

where u represents the angle of rotation about the z axis. Setting $u = 0$ gives the ellipse

$$(a+b\cos v,\ 0,\ c\sin v)$$

in the xz-plane centered at $(a,0,0)$, waiting to be rotated. Suppose for the time being that $a > b > 0$ and $c > 0$. In accordance with Figure 7.11, the ratio b/c determines how "flat" ($b > c$) or how "slim" ($b < c$) the torus is. On the other hand, the bigger $a - b$ is, the bigger the "hole" in the middle.

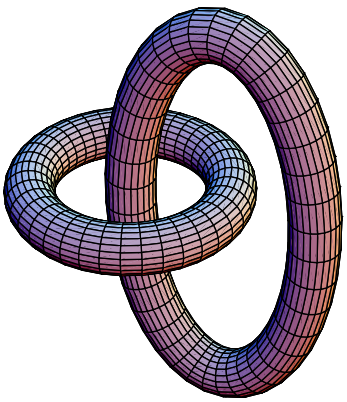


Figure 7.10: Tubes around linked ellipses

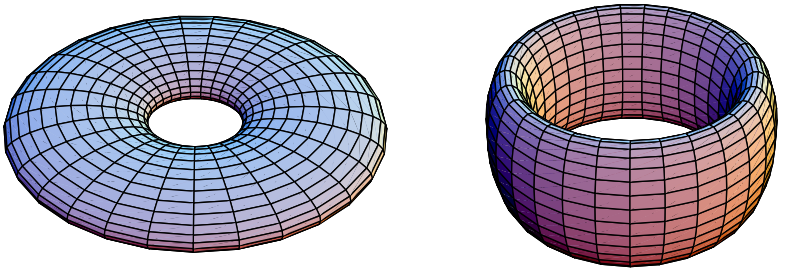


Figure 7.11: Tori formed by revolving ellipses

Elliptical tori will be investigated further in Section 10.4. Canal surfaces are generalizations of tubes that will be studied in Chapter 20.

7.7 Torus Knots

Curves that wind around a torus are frequently knotted. Let us define

$$\begin{aligned}\mathsf{torusknot}[a,b,c][p,q](t) &= \mathsf{torus}[a,b,c](pt,qt)\\ &= \Big(\big(a+b\cos(qt)\big)\cos(pt),\ \big(a+b\cos(qt)\big)\sin(pt),\ c\sin(qt)\Big)\end{aligned}$$

(refer to (7.37)). It follows that $\mathsf{torusknot}[a,b,c][p,q](t)$ lies on an elliptical torus and, assuming that p/q is rational, the curve is closed. For this reason, we have called it a ***torus knot***; it may or may not be truly knotted, depending on p and q. In fact, $\mathsf{torusknot}[a,b,c][1,q]$ is unknotted, whereas $\mathsf{torusknot}[a,b,c][2,3]$ is a ***trefoil knot***. Of the many books on the interesting subject of knot theory, we mention [BuZi], [Kauf1],[Kauf2] and [Rolf].

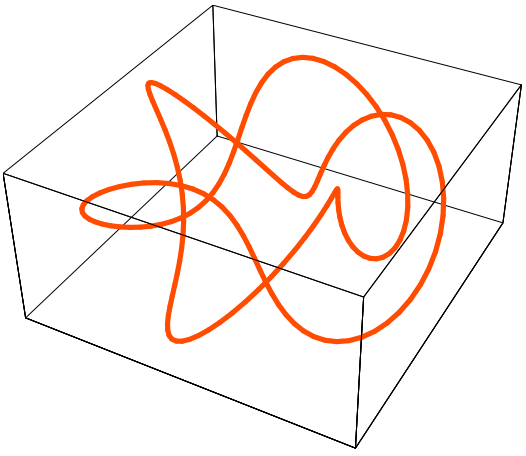


Figure 7.12: $\mathsf{torusknot}[8,3,5][2,5]$

For the remainder of this section, we shall study the curve

$$\boldsymbol{\alpha} = \mathsf{torusknot}[8, 3, 5][2, 5]$$

displayed in Figure 7.12. This representation is not as informative as we would like, because at an apparent crossing it is not clear which part of the curve is in front and which part is behind. To see what really is going on, let us draw a tube about this curve using the construction of the previous section. The crossings are now clear; furthermore, we can describe the winding a little more precisely by saying that $\boldsymbol{\alpha}$ spirals around the torus 2 times in the longitudinal sense and 5 times in the meridianal sense. (This is the terminology of [Costa2].) More generally, $\mathsf{torusknot}[a, b, c][p, q]$ will spiral around an elliptical torus p times in the longitudinal sense and q times in the meridianal sense.

Figure 7.13 raises a new issue. The tube itself seems to twist violently in several places. (This is even clearer when the graphics is rendered with fewer sample points in Notebook 7.) A careful look at the plot of the tube shows that this increased twisting occurs in five different places, in spite of the fact that other parts of the tube partially obscure two of these. Moreover, nearby to where the violent twisting occurs, the actual knot curve is straighter. This evidence leads us to suspect that there are five points on the original torus knot where simultaneously the curvature $\boldsymbol{\kappa}$ is small and the absolute value of the torsion $\boldsymbol{\tau}$ is large.

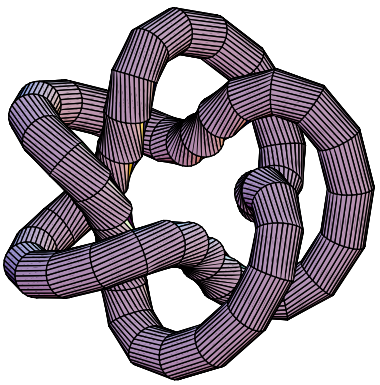


Figure 7.13: Tube around a torus knot

The graphs of the curvature and torsion of $\boldsymbol{\alpha}$ can be used to test this empirical evidence; they confirm that indeed there are five values of t at which the curvature of $\boldsymbol{\alpha}(t)$ is small when the absolute value of the torsion is large. Furthermore, the graph seems to indicate these points are $\pi/5$, $3\pi/5$, π, $7\pi/5$ and $9\pi/5$. In fact, this is the case; the numerical values at $\pi/5$ are

$$\boldsymbol{\kappa}[\boldsymbol{\alpha}](\pi/5) = 0.076, \qquad \boldsymbol{\tau}[\boldsymbol{\alpha}](\pi/5) = -0.414.$$

At $t = 2\pi/5$ the curvature is a little larger and the absolute value of the torsion is quite small:

$$\boldsymbol{\kappa}[\boldsymbol{\alpha}](2\pi/5) = 0.107, \qquad \boldsymbol{\tau}[\boldsymbol{\alpha}](2\pi/5) = -0.002.$$

See Figure 7.14.

The Frenet formulas (7.19), page 203, explain why small curvature and large absolute torsion produce so much twisting in the tube. The construction of the tube involves $\mathbf{N}$ and $\mathbf{B}$, but not $\mathbf{T}$. When the derivatives of $\mathbf{N}$ and $\mathbf{B}$ are large, the twisting in the tube will also be large. But a glance at the Frenet formulas shows that small curvature and large absolute torsion create exactly this effect.

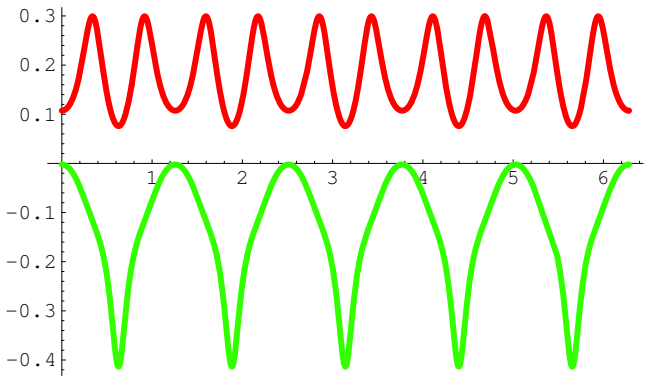


Figure 7.14: Curvature (positive) and torsion of torusknot$[8, 3, 5][2, 5]$

7.8 Exercises

1. Establish the following identities for $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^3$:

(a) $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = [\mathbf{a}\,\mathbf{c}\,\mathbf{d}]\,\mathbf{b} - [\mathbf{b}\,\mathbf{c}\,\mathbf{d}]\,\mathbf{a}$.

(b) $(\mathbf{d} \times (\mathbf{a} \times \mathbf{b})) \cdot (\mathbf{a} \times \mathbf{c}) = [\mathbf{a}\,\mathbf{b}\,\mathbf{c}](\mathbf{a} \cdot \mathbf{d})$.

(c) $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) = \mathbf{0}$.

(d) $(\mathbf{a} \times \mathbf{b}) \cdot \big((\mathbf{b} \times \mathbf{c}) \times (\mathbf{c} \times \mathbf{a})\big) = [\mathbf{a}\,\mathbf{b}\,\mathbf{c}]^2$.

M **2.** Graph helical curves over a cycloid, a cardioid and a figure eight curve, and plot the curvature and torsion of each.

3. Let $\boldsymbol{\alpha}$ be an arbitrary-speed curve in $\mathbb{R}^n$. Define

$$\boldsymbol{\kappa n}[\boldsymbol{\alpha}] = \frac{\Big\| \|\boldsymbol{\alpha}'\|\boldsymbol{\alpha}'' - \|\boldsymbol{\alpha}'\|'\boldsymbol{\alpha}' \Big\|}{\|\boldsymbol{\alpha}'\|^3}.$$

Use (7.26) to show that $\boldsymbol{\kappa n}[\boldsymbol{\alpha}]$ is the curvature of $\boldsymbol{\alpha}$ for $n = 3$.

M **4.** Plot the following space curves and graph their curvature and torsion:

(a) $s \mapsto \left(\dfrac{(1+s)^{3/2}}{3}, \dfrac{(1-s)^{3/2}}{3}, \dfrac{s}{\sqrt{2}} \right)$.

(b) $t \mapsto \big(f(t)\cos t,\ f(t)\sin t,\ a\cos 10t\big)$ where $f(t) = \sqrt{1 - a^2\cos^2 10t}$, and $a = 0.3$.

(c) $t \mapsto 0.5\big((1+\cos 10t)\cos t,\ (1+\cos 10t)\sin t,\ 1+\cos 10t\big)$.

(d) $t \mapsto (a\cosh t,\ a\sinh t,\ b\,t)$.

(e) $t \mapsto \big(te^t,\ e^{-t},\ \sqrt{2}\,t\big)$.

(f) $t \mapsto a\big(t - \sin t,\ 1 - \cos t,\ 4\cos(t/2)\big)$.

5. Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a straight line parametrized as $\boldsymbol{\beta}(s) = s\mathbf{e}_1 + \mathbf{q}$, where $\mathbf{e}_1, \mathbf{q} \in \mathbb{R}^3$ with $\mathbf{e}_1$ a unit vector. Let $\mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^3$ be such that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is an orthonormal basis of $\mathbb{R}^3$. Define constant vector fields by $\mathbf{T}(s) = \mathbf{e}_1$, $\mathbf{N}(s) = \mathbf{e}_2$ and $\mathbf{B}(s) = \mathbf{e}_3$ for $a < s < b$. Show that the Frenet formulas (7.12) hold provided we take $\boldsymbol{\kappa}(s) = \boldsymbol{\tau}(s) = 0$ for $a < s < b$.

6. Suppose that $a < b$ and choose ϵ to be 1 or -1. Show that the mapping

$$\mathsf{bicylinder}[a,b,\epsilon](t) = \left(a\cos t,\ a\sin t,\ \epsilon\sqrt{b^2 - a^2\sin^2 t}\ \right)$$

parametrizes one component of the intersection of a cylinder of radius a and an appropriately-positioned cylinder of radius b.

M **7.** A 3-dimensional version of the astroid is defined by

$$\mathsf{ast3d}[n,a,b](t) = (a\cos^n t,\ b\sin^n t,\ \cos 2t).$$

Compute the curvature and torsion of $\mathsf{ast3d}[n,1,1]$ and plot $\mathsf{ast3d}[n,1,1]$ for $n > 3$.

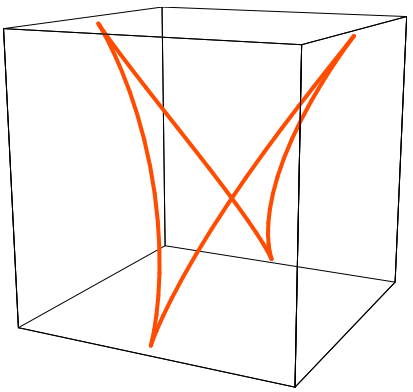


Figure 7.15: The 3-dimensional astroid $\mathsf{ast3d}[3,1,1]$

M **8.** An ***elliptical helix*** is given by

$$\mathsf{helix}[a,b,c](t) = \big(a\cos t,\ b\sin t,\ ct\big). \tag{7.38}$$

Plot an elliptical helix, its curvature and torsion. Show that an elliptical helix has constant curvature if and only if $a^2 = b^2$. [Hint: Compute the derivative of the curvature of an elliptical helix and evaluate it at $\pi/4$.]

M **9.** The ***Darboux***[9] ***vector field*** along a unit-speed curve $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ is defined by $\mathbf{D} = \boldsymbol{\tau}\mathbf{T} + \boldsymbol{\kappa}\mathbf{B}$. Show that

$$\begin{cases} \mathbf{T}' = \mathbf{D}\times\mathbf{T}, \\ \mathbf{N}' = \mathbf{D}\times\mathbf{N}, \\ \mathbf{B}' = \mathbf{D}\times\mathbf{B}. \end{cases}$$

Plot the curves traced out by the Darboux vectors of an elliptical helix, Viviani's curve, and a bicylinder parametrized in Exercise 6.

M **10.** A curve constructed from the Bessel functions J_n (mentioned on page 141) is defined by

$$\mathsf{besselcurve}[a,b,c](t) = \big(J_a(t), J_b(t), J_c(t)\big)$$

where a, b, c are integers. Plot $\mathsf{besselcurve}[0,1,2]$.

M **11.** A generalization of the twisted cubic (defined on page 202) is the curve $\mathsf{twistedn}[n,a]\colon \mathbb{R} \to \mathbb{R}^n$ given by

$$\mathsf{twistedn}[n,a](t) = \big(t,\ t^2,\ \ldots,\ t^n\big).$$

Compute the curvature of $\mathsf{twistedn}[n,a]$ for $1 \leqslant n \leqslant 6$.

M **12.** Plot the torus knot with $p = 3$ and $q = 2$ and graph its curvature and torsion. Plot a tubular surface surrounding it.

13. Find a parametrization for the tube about Viviani's curve illustrated in Figure 7.16.



[9] Jean Gaston Darboux (1842–1917). French mathematician, who is best known for his contributions to differential geometry. His four-volume work *Leçons sur la théorie general de surfaces* remains the bible of surface theory. In 1875 he provided new insight into the Riemann integral, first defining upper and lower sums and then defining a function to be integrable if the difference between the upper and lower sums tends to zero as the mesh size gets smaller.

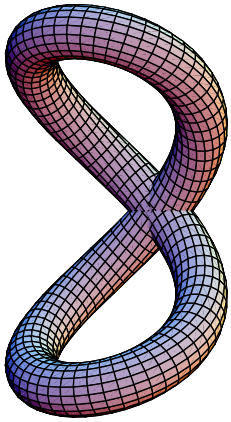


Figure 7.16: A tube of radius 1.5 around Viviani's curve

M **14.** The definition of a spherical spiral shown in Figure 7.17 is similar to that of a torus knot, namely

$$\text{sphericalspiral}[a][m,n](t) = a\Big(\cos(mt)\cos(nt),\ \sin(mt)\cos(nt),\ \sin(nt)\Big).$$

Find formulas for its curvature and torsion.

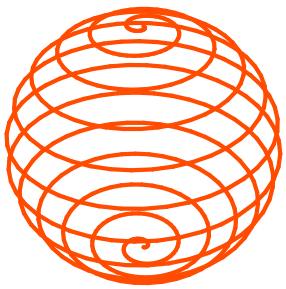


Figure 7.17: The spherical spiral with $m = 24$ and $n = 1$

M **15.** A ***figure eight knot*** can be parametrized by

$$\text{eightknot}(t) = \Big(10(\cos t + \cos 3t) + \cos 2t + \cos 4t,\\ 6\sin t + 10\sin 3t,\ 4\sin 3t \sin\frac{5t}{2} + 4\sin 4t - 2\sin 6t\Big).$$

Plot the tube about the figure eight knot.

Notebook 7

■ N7.0

Computer programs for constructing tangent, normal and binormal vectors, and for visualizing the resulting Frenet frame of a space curve, are described in N7.2. This is preceded by calculations that explain the definition of the vector cross product. In N7.3, we customize a command called highlight to enhance the plotting of space curves. This is first used to plot generalized helices and the twisted cubic, but will be used repeatedly in subsequent notebooks.

In N7.4, we test general formulas to compute the curvature and torsion of space curves. N7.5 includes an animation of the Frenet frame moving on Viviani's curve. Tubes around curves, elliptical tori, and torus knots are introduced in N7.7 and N7.8.

We first repeat the definition of the "length" or Euclidean norm of a vector, in a form given in N2.0.1. It is equally valid in 3 (or more) dimensions:

```
Unprotect[L]; L[v_]:= Simplify[Sqrt[v.v]]; Protect[L]
```

We shall also call upon a couple of definitions related to curves, using the slot symbol #, first explained in N5.0.

```
ellipse[a_,b_]:= {a Cos[#], b Sin[#]} &
logspiral[a_,b_]:= a{Exp[b #]Cos[#], Exp[b #]Sin[#]} &
epitrochoid[a_,b_,h_]:=
   {(a+b)Cos[#]-h Cos[(a+b)#/b], (a+b)Sin[#]-h Sin[(a+b)#/b]} &
```

Default options for plotting two functions simultaneously (typically the curvature and torsion) are given in N7.0.1, as well as the display command S discussed in N6.0.

■ N7.0.1

■ N7.1

To define the cross product in *Mathematica,* we can implement Definition 7.1:

```
id:= IdentityMatrix[3]
```

```
cross1[v_,w_]:=Table[Det[{id[[k]],v,w}],{k,3}]
```

This is the definition of vector cross product used in an early version of *Mathematica,* though (as explained in [De]) there is a more effective alternative that uses a cyclic permutation operator and component-wise multiplication (not of course the dot product). This is defined next:

```
cross2[v_,w_]:= RotateLeft[v RotateLeft[w] - RotateLeft[v] w]
```

```
v:= {a,b,c}
w:= {A,B,C}
RotateLeft[v]
v w
v.w
```

Needless to say, the built-in *Mathematica* command Cross, which we shall use, produces the same result:

```
cross1[v,w]
cross2[v,w]
Cross[v,w]
```

Using any of these definitions, we can verify some of the standard identities:

```
u:= {α,β,γ}
v:= {a,b,c}
w:= {A,B,C}
cross2[u, cross1[v,w]]//Expand
(u.w)v - (u.v)w//Expand
```

We can proceed to check Lagrange's identity:

```
L[Cross[v,w]]^2//Factor
L[v]^2 L[w]^2 - (v.w)^2//Expand
```

■ N7.2

In practice, we shall be computing the curvature of space curves via formulas in N7.4, based on Theorem 7.15. At this stage of the notebook, we merely construct the Frenet frame at each individual point of a curve. For implementation in *Mathematica*, the vectors T, N, B are referred to by their full names tangent, normal, binormal.

As we explained earlier, capital letters are used in internal commands and are best avoided; indeed N on its own computes numerical value and Normal is a built-in command used to convert a series:

```
Series[Sin[x], {x,0,9}]
Normal[%]
```

In the following definitions, tt is a dummy variable whose presence ensures that the differentiation is carried out before any value of t is actually substituted. The tangent vector T is obtained by dividing α'(t) by its norm. This having been done, it is more convenient to define the binormal vector B next, and the normal vector N is defined last as the cross product of B with T. The Off command suppresses the warning that normal may be a misspelling of Normal.

```
tangent[α_][t_]:=
Module[{A,tt},
  A= α'[tt];
  A/L[A] /. tt->t
]
```

```
binormal[α_][t_]:= binormal[α][t]=
Module[{A,tt},
  A= Cross[α'[tt], α''[tt]];
  A/L[A] /. tt->t
]
```

Repetition of the command string without a semicolon allows the program to record values it has already found, and this device can considerably reduce computation time.

```
Off[General::spell1]
normal[α_][t_]:= normal[α][t]= Cross[binormal[α][t], tangent[α][t]]
```

We may test these definitions on the cubic curve

```
γ[t_]:= {3t - t^3, 3t^2, 3t + t^3}
```

All three vectors have unit norm, by construction, though it does no harm to check:

```
tangent[γ][t]//PowerExpand
L[%]
```

```
normal[γ][3]//Simplify
L[%]
```

```
binormal[γ][5]//PowerExpand
L[%]
```

In the following computation, kv is not the curvature of γ because t is a nontrivial function of arclength s. Indeed, s'(t) is given by the function v in the next cell, and the theory of Section 7.4 tells us that the curvature is in fact kv/v. We confirm this by using the general definition of $\kappa[\gamma]$ given in an initialization cell at the start of N7.4.

```
Module[{kv,v},
  kv= tangent[γ]'[t].normal[γ][t]//Simplify//PowerExpand;
  v= L[γ'[t]]//PowerExpand;
  kv/v
]
κ[γ][t]
```

- N7.2.1

■ N7.3

One of the simplest space curves is

```
helix[a_,b_][t_]:= {a Cos[t], a Sin[t], b t}
```

The parameter t is a mere constant times arc length:

```
α:= helix[a,b]
L[α'[t]]
```

The following curve therefore has unit speed:

```
β[t_]:= helix[a,b][t]/Sqrt[a^2+b^2]
L[β'[t]]
```

The curvature of the helix is now given by

```
tangent[β]'[s].normal[β][s]//Simplify
%//PowerExpand
```

Its torsion is also constant:

```
-binormal[β]'[s].normal[β][s]//Simplify
```

Next, we turn our attention to plotting space curves. The symbols S@ must be added to the start of the third line to display the plot, as explained in N6.0.

```
Clear[α]
α:= helix[1,.2]
ParametricPlot3D[α[t], {t,0,4π}]
```

To color, thicken and otherwise improve the appearance of (we shall say "highlight") a space curve, it is not possible to use PlotStyle as in the case of 2 dimensional graphics. Instead, one needs to use the Append operation, which adds an element at the end of a list. We shall carry out this procedure so often that it is convenient to define the following command, that includes default color and thickness values.

```
highlight[α_, h_:1/20, k_:1/100][t_]:=
  Append[α[t], {Hue[h], Thickness[k]}]
```

Making *Mathematica* plot more points than the default is sometimes necessary to prevent the curve appearing "jagged"; this is used for the next figure. The ticks option eliminates labels on the y-axis.

```
αα:= highlight[helix[1,.1]]
αα[t]
S @ ParametricPlot3D[αα[t], {t,0,8π}, PlotPoints->200,
  Ticks->{Automatic, False, Automatic}, ImageSize->300]
```

Figure 7.2

The standard twisted cubic is defined next. Although its form is simpler than that of the cubic γ investigated in N7.2, we shall see in N7.4 that its curvature and torsion are a bit more complicated.

```
twicubic[t_]:= {t, t^2, t^3}
```

Two views of the twisted cubic are obtained in the electronic subsection that follows.

■ N7.3.1

■ N7.4

Computing curvature of mildly complicated curves by hand can be very tedious. However, the formulas of Section 7.4 are relatively straightforward to implement. The definition of the curvature relies upon the first and second derivatives of α:

```
κ[α_][t_]:=
Module[{tt,a,b},
  a= L[Cross[α'[tt], α''[tt]]];
  b= L[α'[tt]]^3;
  a/b /. tt->t
]
```

Torsion also involves the third derivative of α:

```
τ[α_][t_]:=
Module[{tt,a,C},
  a= Det[{α'[tt], α''[tt], α'''[tt]}];
  C= Cross[α'[tt], α''[tt]];
  a/(C.C) /.tt->t
]
```

As an application, we compute and plot the curvature and torsion of the standard twisted cubic curve. In such graphs, the curvature will always be in red, and the torsion in green, as a consequence of the default option from N7.0.1.

```
κ[twicubic][t]
τ[twicubic][t]
```

```
β:= twicubic
S @ Plot[{κ[β][t],τ[β][t]}, {t,-1,1}, ImageSize->500]
```

Figure 7.4

Here is confirmation of the constant curvature and torsion of the helix:

```
κ[helix[a,b]][t]
%//PowerExpand
```

```
τ[helix[a,b]][t]
```

■ N7.5

The code to generate a helical curve (depending upon a parameter c) over a given space curve α is given by

```
helical[α_,c_][t_]:= {α[t][[1]], α[t][[2]], c t}
```

We illustrate this by constructing the helical curve over a logarithmic spiral, and then computing the resulting curvature and torsion.

```
α:=helical[logspiral[1,.05], -.5]
α[t]
```

```
αα:=highlight[α]
ParametricPlot3D[αα[t], {t,-4π,16π}, PlotPoints->300,
  Boxed->False, Axes->False]
```

It is convenient to denote the pair of functions representing the curvature and torsion numerically by kt. Add S@ at the start of the second line to display it.

```
kt[t_]:= {κ[α][t], τ[α][t]}//N
Plot[kt[t]//Evaluate,{t,0,12π}]
```

For the next figure, we use the program κ2, for plane curvature, included in N7.0.1.

```
Module[{α,β,k,γ,γγ},
  α= epitrochoid[3,1,3/2];
  β[t_]:= Append[α[t], 0];
  k[t_]:= κ2[α][t]//N;
  γ[t_]:= {α[t][[1]], α[t][[2]], k[t]};
  γγ= highlight[γ];
  S @ ParametricPlot3D[{β[t], γγ[t]}//Evaluate, {t,0,2π},
    PlotPoints->300, PlotRange->All, ImageSize->400,
    Ticks->{False, Automatic, Automatic}]
]
```

Figure 7.5

```
α:= ellipse[5,3]
β[t_]:= Append[α[t],0]
k[t_]:= κ2[α][t]//N
ϕ[t_]:= {α[t][[1]], α[t][[2]], t k[t]/2}
ϕϕ:= highlight[ϕ]
ParametricPlot3D[{β[t], ϕϕ[t]}//Evaluate, {t,0,71π/8},
  PlotPoints->300, PlotRange->All]
```

Next, we give a parametrization of the celebrated Viviani curve, the intersection of a cylinder with axis x=a, y=0, and a sphere of radius 2a and center the origin.

```
viviani[a_][t_]:= a{1+Cos[t], Sin[t], 2Sin[t/2]}
```

When a=1, the curve lies on a sphere of radius 2.

```
L[viviani[1][t]]
```

A geometrical illustration of Viviani's curve is given in Figure 7.6. It displays three great circles of the sphere and the intersection of the cylinder with the equatorial plane.

```
vv:= highlight[viviani[1],.65]
f[t_]:= 2{{0,Cos[t],Sin[t]}, {Cos[t],0,Sin[t]}, {Cos[t],Sin[t],0}}
g[t_]:= {1+Cos[2t], Sin[2t], 0}
gg:= highlight[g,.1]
F[t_]:= Join[f[t], {vv[t],gg[t]}]
S @ ParametricPlot3D[F[t]//Evaluate, {t,-2π,2π},
  Boxed->False, Axes->False, ViewPoint->{20,3,3}]
```

Figure 7.6

The curvature and torsion of Viviani's curve are given by:

```
Module[{v},
  v= viviani[1];
  κ[v][t];
  τ[v][t]//Simplify;
  S @ Plot[{κ[v][t],τ[v][t]}//Evaluate, {t,0,2π}, ImageSize->500]
]
```

Figure 7.7

The following miniprogram enables us to plot and (if required) animate a Frenet frame. There is no danger in using the symbols T,N,B here, because their scope is bounded by the Module environment; highlighted versions are denoted by TT,NN,BB. Observe however that the colon in the definition of li is essential to avoid premature evaluation.

```
frenetframe[α_][t_]:=
Module[{A,T,N,B,li},
  A=α[t];
  T= tangent[α][t];
  N= normal[α][t];
  B= binormal[α][t];
  li[X_]:= Line[{A,A+X}];
  TT= {Hue[0], li[T]};
  NN:= {Hue[.3], li[N]};
  BB= {Hue[.7], li[B]};
  {TT,NN,BB}
]
```

```
frenetframe[viviani[1]][1]
```

The following program creates an animation of a labelled Frenet frame moving around the complete Viviani curve. Similar code was used in the plot of Figure 7.1 above. The

display is suppressed by the option for Show we have declared in N7.0.1, so to run the animation replace Show by S.

```
Module[{v,vv,z,g,s},
  v= viviani[1];
  vv= highlight[v];
  z= ParametricPlot3D[vv[t], {t,-2π,2π}];
  g[t_]:= Graphics3D[{Thickness[.01], Hue[.4], frenetframe[v][t]}];
  s[t_]:= Show[z, g[t], ViewPoint->{4,-4,1}, Axes->None,
    PlotRange->{{-1,3.5},{-2.2,2.2},{-3.5,3.5}}];
  Do[s[t], {t,-2π,2π, π/5}]
]
```

■ N7.5.1

■ N7.6

Following the text, we set

```
tubecurve[γ_][r_][t_,θ_]:=
  γ[t] + r N/@(-Cos[θ] normal[γ][t] + Sin[θ] binormal[γ][t])
```

The command N/@ has the effect of evaluating the vector components numerically, which speeds up the calculations. To explain the syntax, note that if g is a function and v a vector, the command g/@v is shorthand for applying g to each of the components of v.

```
v:= {a,b,c}
g/@v
```

The next figures show tubes around a helix, a circle, and an ellipse:

```
Γ:= tubecurve[helix[2,.2]][.2]
S @ ParametricPlot3D[Γ[t,θ], {t,0,5π}, {θ,0,2π},
  Axes->None, PlotPoints->{70,20}, Boxed->False]
```

Figure 7.9

```
Module[{α,β,A,B},
  α[t_]= 4{Cos[t], Sin[t], 0};
  β[t_]= 4{1+Cos[t], 0, 2Sin[t]};
  A= tubecurve[α][1];
  B= tubecurve[β][1];
  S @ ParametricPlot3D[{A[t,θ], B[t,θ]}//Evaluate, {t,0,2π}, {θ,0,2π},
    PlotPoints->{40,25}, Boxed->False, Axes->False]
]
```

Figure 7.10

The following definition of an elliptical torus of revolution allows one to alter three separate parameters describing it. This paramerization will be used again and again in future notebooks; see for example N10.4.

```
torus[a_,b_,c_][u_,v_]:=
  {(a+b Cos[v])Cos[u], (a+b Cos[v]) Sin[u], c Sin[v]}
```

The "torus" intersects itself if a<b:

```
zt[a_]:=ParametricPlot3D[torus[a,2,2][u,v], {u,0,π}, {v,0,2π},
  ViewPoint->{1,-2,1}]
Show/@Array[zt,3]
```

Note that if we display the Array rather than its constituents, we obtain all three tori simultaneously:

```
Show @ Array[zt,3]
```

■ N7.6.1

■ N7.7

We next define a torus knot in terms of torus:

```
torusknot[a_,b_,c_][p_,q_][t_]:= torus[a,b,c][p t, q t]
```

```
α:= torusknot[8,3,5][2,5]
S @ ParametricPlot3D[highlight[α][t], {t,0,2π}, PlotPoints->200,
  Ticks->None]
```

Figure 7.12

■ N7.7.1

■ N7.8

The next two cells refer to Exercises 7 and 12 respectively.

```
ast3d[n_,a_,b_][t_]:= {a Cos[t]^n, b Sin[t]^n, Cos[2t]}
αα:= highlight[ast3d[3,1,1]]
S @ ParametricPlot3D[αα[t]//Evaluate, {t,0,2π},
  ViewPoint->{1.5,-3,.8}, Axes->False]
```

Figure 7.15

A reduction of sample points is used to quicken the evaluation:

```
α:= torusknot[8,3,5][3,2]
αα:= highlight[α]
ParametricPlot3D[αα[t], {t,0,2π}, ViewPoint->{20,30,70}];
Γ:= tubecurve[torusknot[8,3,5][3,2]][1.3]
ParametricPlot3D[Γ[t,θ], {t,0,2π}, {θ,0,2π}, Axes->None,
  PlotPoints->{20,10}, ViewPoint->{20,30,70}]
```

A way of representing a torus as a "wire-frame" is provided by

```
Clear[u,v]
f[u_,v_]:= Append[torus[4,3,4][u,v], GrayLevel[.5]]
ta:= Table[f[u,v], {u,0,2π,π/20}]
ParametricPlot3D[ta//Evaluate, {v,0,2π}, Boxed->False,
  Axes->None]
```

■ N7.8.1

Chapter 8

Construction of Space Curves

The analog of the Fundamental Theorem of Plane Curves (Theorem 5.13) is the *Fundamental Theorem of Space Curves*. The uniqueness part of this theorem, proved in Section 8.1, states that two curves with the same torsion and positive curvature differ only by a Euclidean motion of $\mathbb{R}^3$. The torsion of a space curve is not defined at a point where the curvature is zero, so it is not unreasonable to require as a hypothesis that the curvature never vanish. But this requirement turns out to be essential, in the light of a counterexample that we present (see Figure 8.1).

In Section 8.2, we prove the existence part, namely that space curves exist with prescribed torsion and positive curvature. We then describe some of the examples with specified curvature and torsion that were generated by the associated computer programs in Notebook 8. Closed curves can be obtained by imposing constant curvature and periodic torsion (an example can be seen in Figure 8.14 on page 253).

The notion of contact between two curves or between a curve and a surface is defined in Section 8.3. This part of the chapter is also relevant to the theory of plane curves, although it leads eventually to the definition of the *evolute* of a space curve. The evolute is traced out by the centers of spheres having third order contact with the given space curve at each of its points.

Curves that lie on spheres are characterized in Section 8.4, and they furnish examples for the rest of the chapter. Curves of constant slope are generalizations of a helix, and such a curve $\boldsymbol{\gamma}$ can be constructed so as to project to a given plane curve $\boldsymbol{\beta}$. We see in Section 8.5 that the assumption that $\boldsymbol{\gamma}$ lie on a sphere requires $\boldsymbol{\beta}$ to be an epicycloid. Loxodromes on spheres are defined, discussed and plotted in Section 8.6.

8.1 The Fundamental Theorem of Space Curves

The curvature and torsion of a space curve determine the curve in very much the same way as the signed curvature $\boldsymbol{\kappa 2}$ determines a plane curve. First, we establish the invariance of curvature and torsion under Euclidean motions of $\mathbb{R}^3$. For that we need a fact about the vector triple product, defined in Section 7.1.

Lemma 8.1. *Let $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map and $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$. Then*

$$[A\mathbf{a}\ A\mathbf{b}\ A\mathbf{c}] = \det(A)\,[\mathbf{a}\ \mathbf{b}\ \mathbf{c}].$$

Proof. If we treat $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ as column vectors then we may regard A as a matrix premultiplying the vectors. Let $(\mathbf{a}|\mathbf{b}|\mathbf{c})$ denote the 3×3 matrix whose columns are $\mathbf{a}, \mathbf{b}, \mathbf{c}$. With this notation,

$$\begin{aligned}[A\mathbf{a}\ A\mathbf{b}\ A\mathbf{c}] = A\mathbf{a} \cdot (A\mathbf{b} \times A\mathbf{c}) &= \det\big(A\mathbf{a}|A\mathbf{b}|A\mathbf{c}\big) \\ &= \det\big(A(\mathbf{a}|\mathbf{b}|\mathbf{c})\big) = \det(A)\,[\mathbf{a}\ \mathbf{b}\ \mathbf{c}]. \ \blacksquare\end{aligned}$$

Now we can determine the effect of a Euclidean motion on arc length, curvature and torsion.

Theorem 8.2. *The curvature and the absolute value of the torsion are invariant under Euclidean motions of $\mathbb{R}^3$, though the torsion changes sign under an orientation-reversing Euclidean motion.*

Proof. Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^3$ be a curve, and let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a Euclidean motion. Denote by A the linear part of F, so that for all $\mathbf{p} \in \mathbb{R}^3$ we have $F(\mathbf{p}) = A\mathbf{p} + F(0)$. We define a curve $\boldsymbol{\gamma}\colon (a, b) \to \mathbb{R}^3$ by $\boldsymbol{\gamma} = F \circ \boldsymbol{\alpha}$; then for $a < t < b$ we have

$$\boldsymbol{\gamma}(t) = A\boldsymbol{\alpha}(t) + F(0).$$

Hence $\boldsymbol{\gamma}'(t) = A\boldsymbol{\alpha}'(t)$, $\boldsymbol{\gamma}''(t) = A\boldsymbol{\alpha}''(t)$ and $\boldsymbol{\gamma}'''(t) = A\boldsymbol{\alpha}'''(t)$. Let s_α and s_γ denote the arc length functions with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Since A is an orthogonal transformation, we have

$$s'_\gamma(t) = \big\|\boldsymbol{\gamma}'(t)\big\| = \big\|A\boldsymbol{\alpha}'(t)\big\| = \big\|\boldsymbol{\alpha}'(t)\big\| = s'_\alpha(t). \tag{8.1}$$

We compute the curvature of $\boldsymbol{\gamma}$, making use of the Lagrange identity (7.2) on page 193:

$$\begin{aligned}\boldsymbol{\kappa}[\gamma](t) &= \frac{\big\|\boldsymbol{\gamma}'(t) \times \boldsymbol{\gamma}''(t)\big\|}{\big\|\boldsymbol{\gamma}'(t)\big\|^3} = \frac{\big\|A\boldsymbol{\alpha}'(t) \times A\boldsymbol{\alpha}''(t)\big\|}{\big\|A\boldsymbol{\alpha}'(t)\big\|^3} \\ &= \frac{\sqrt{\big\|A\boldsymbol{\alpha}'(t)\big\|^2\big\|A\boldsymbol{\alpha}''(t)\big\|^2 - \big(A\boldsymbol{\alpha}'(t) \cdot A\boldsymbol{\alpha}''(t)\big)^2}}{\big\|A\boldsymbol{\alpha}'(t)\big\|^3}\end{aligned}$$

$$= \frac{\sqrt{\|\boldsymbol{\alpha}'(t)\|^2\|\boldsymbol{\alpha}''(t)\|^2 - \big(\boldsymbol{\alpha}'(t)\cdot\boldsymbol{\alpha}''(t)\big)^2}}{\|\boldsymbol{\alpha}'(t)\|^3} = \frac{\|\boldsymbol{\alpha}'(t)\times\boldsymbol{\alpha}''(t)\|}{\|\boldsymbol{\alpha}'(t)\|^3} = \boldsymbol{\kappa}[\boldsymbol{\alpha}](t).$$

Similarly, Lemma 8.1 and the fact that $\det(A) = \pm 1$ implies that

$$\boldsymbol{\tau}[\boldsymbol{\gamma}](t) = \frac{[\boldsymbol{\gamma}'(t)\,\boldsymbol{\gamma}''(t)\,\boldsymbol{\gamma}'''(t)]}{\|\boldsymbol{\gamma}'(t)\times\boldsymbol{\gamma}''(t)\|^2} = \frac{[A\boldsymbol{\alpha}'(t)\,A\boldsymbol{\alpha}''(t)\,A\boldsymbol{\alpha}'''(t)]}{\|A\boldsymbol{\alpha}'(t)\times A\boldsymbol{\alpha}''(t)\|^2}$$

$$= \frac{\det(A)[\boldsymbol{\alpha}'(t)\,\boldsymbol{\alpha}''(t)\,\boldsymbol{\alpha}'''(t)]}{\|\boldsymbol{\alpha}'(t)\times\boldsymbol{\alpha}''(t)\|^2} = \det(A)\boldsymbol{\tau}[\boldsymbol{\alpha}](t) = \pm\boldsymbol{\tau}[\boldsymbol{\alpha}](t). \blacksquare$$

Inherent in the above proof, and a consequence of (8.1), is the fact that the arc length of a curve is itself invariant under Euclidean motions.

In Theorem 5.13, page 136, we showed that a plane curve is determined up to a Euclidean motion of the plane by its signed curvature. This leads us to the conjecture that a space curve is determined up to a Euclidean motion of $\mathbb{R}^3$ by its curvature and torsion. We prove that this is true, with one additional (but essential) assumption.

Theorem 8.3. (Fundamental Theorem, Uniqueness) *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ be unit-speed curves in $\mathbb{R}^3$ defined on the same interval (a,b), and assume they have the same torsion and the same positive curvature. Then there is a Euclidean motion F of $\mathbb{R}^3$ that maps $\boldsymbol{\alpha}$ onto $\boldsymbol{\gamma}$.*

Proof. Since both $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have nonzero curvature, both of the Frenet frames $\{\mathbf{T}_\alpha, \mathbf{N}_\alpha, \mathbf{B}_\alpha\}$ and $\{\mathbf{T}_\gamma, \mathbf{N}_\gamma, \mathbf{B}_\gamma\}$ are well defined. Fix s_0 with $a < s_0 < b$. There exists a translation of $\mathbb{R}^3$ taking $\boldsymbol{\alpha}(s_0)$ into $\boldsymbol{\gamma}(s_0)$. Choose a rotation A that maps $\boldsymbol{\alpha}'(s_0)$ onto $\boldsymbol{\gamma}'(s_0)$. By rotating again in the plane perpendicular to this vector, we may assume that A also maps $\mathbf{N}_\alpha(s_0)$ onto $\mathbf{N}_\gamma(s_0)$. Lemma 8.1 then guarantees that A maps $\mathbf{B}_\alpha(s_0)$ onto $\mathbf{B}_\gamma(s_0)$. Thus there exists a Euclidean motion F of $\mathbb{R}^3$ such that

$$F\big(\boldsymbol{\alpha}(s_0)\big) = \boldsymbol{\gamma}(s_0)$$

$$A\big(\mathbf{T}_\alpha(s_0)\big) = \mathbf{T}_\gamma(s_0), \quad A\big(\mathbf{N}_\alpha(s_0)\big) = \mathbf{N}_\gamma(s_0), \quad A\big(\mathbf{B}_\alpha(s_0)\big) = \mathbf{B}_\gamma(s_0).$$

To show that $F\circ\boldsymbol{\alpha}$ coincides with $\boldsymbol{\gamma}$, we define a real-valued function f by

$$f(s) = \|(A\circ\mathbf{T}_\alpha)(s) - \mathbf{T}_\gamma(s)\|^2 + \|(A\circ\mathbf{N}_\alpha)(s) - \mathbf{N}_\gamma(s)\|^2 + \|(A\circ\mathbf{B}_\alpha)(s) - \mathbf{B}_\gamma(s)\|^2$$

for $a < s < b$. A computation similar to that of (5.5) on page 137 (using the assumptions that $\boldsymbol{\kappa}[\boldsymbol{\alpha}] = \boldsymbol{\kappa}[\boldsymbol{\gamma}]$ and $\boldsymbol{\tau}[\boldsymbol{\alpha}] = \boldsymbol{\tau}[\boldsymbol{\gamma}]$; see Exercise 1) shows that the derivative of f is 0. Since $f(s_0) = 0$, we conclude that $f(s) = 0$ for all s. Hence $(F\circ\boldsymbol{\alpha})'(s) = (A\circ\mathbf{T}_\alpha)(s) = \boldsymbol{\gamma}'(s)$ for all s, and so there exists $\mathbf{q}\in\mathbb{R}^3$ such that

$$(F\circ\boldsymbol{\alpha})(s) = \boldsymbol{\gamma}(s) + \mathbf{q}$$

for all s. By the choice of F we have $\mathbf{q} = \mathbf{0}$, and F maps $\boldsymbol{\alpha}$ onto $\boldsymbol{\gamma}$. $\blacksquare$

We now discuss what can go wrong when the assumption of nowhere-zero curvature is dropped. The first thing to note is that it is not possible to distinguish a unit normal vector $\mathbf{N}(t_0)$ at a point where $\boldsymbol{\kappa}(t_0) = 0$, and consequently $\boldsymbol{\tau}(t_0)$ is undefined. Nonetheless, if $\boldsymbol{\gamma}\colon (a, b) \to \mathbb{R}^3$ is a regular curve whose torsion vanishes everywhere except for a single t_0 with $a < t_0 < b$, it is reasonable to say that $\boldsymbol{\gamma}$ has zero torsion. With this convention, we shall explain that the assumption in Theorem 8.3 that $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have nonzero curvature is essential.

There are indeed pairs of space curves that have the same curvature and torsion for which it is impossible to find a Euclidean motion mapping one onto the other. For example, consider curves $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ defined by

$$\boldsymbol{\alpha}(t) = \begin{cases} 0, & \text{if } t = 0, \\ (t,\, 0,\, 5e^{-1/t^2}), & \text{if } t \neq 0, \end{cases}$$

and

$$\boldsymbol{\gamma}(t) = \begin{cases} (t,\, 5e^{-1/t^2},\, 0) & \text{if } t < 0, \\ 0, & \text{if } t = 0, \\ (t,\, 0,\, 5e^{-1/t^2}), & \text{if } t > 0. \end{cases}$$

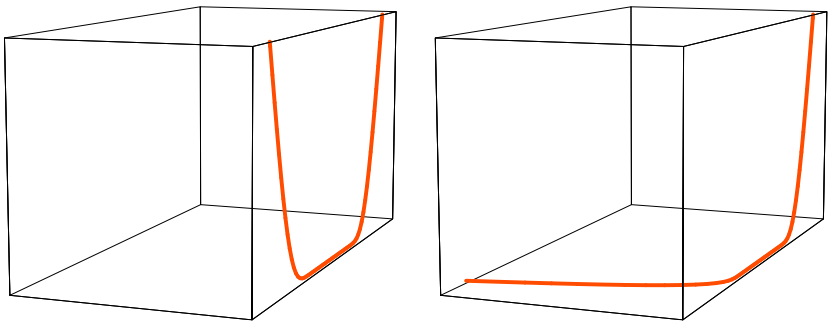


Figure 8.1: The "very flat" curves $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$

The function $f(t) = e^{-1/t^2}$ (with $f(0) = 0$) is infinitely differentiable, and $f^{(n)}(0) = 0$ for all $n \geqslant 0$. It follows that $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are regular, infinitely differentiable curves, which have identical curvature functions vanishing at $t = 0$. Furthermore, their torsion vanishes for both $t > 0$ and $t < 0$ since within these parameter ranges the curves are planar. Any Euclidean motion mapping $\boldsymbol{\alpha}$ onto $\boldsymbol{\gamma}$ would have to be the identity on one portion of $\mathbb{R}^3$ and a rotation on another portion of $\mathbb{R}^3$, which is impossible. Despite the nonexistence of an isometry between them, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ have the same curvature, and the same torsion function, wherever the latter can legitimately be defined.

8.2 Assigned Curvature and Torsion

Now we turn to the question of the existence of curves with prescribed curvature and torsion.

Theorem 8.4. (Fundamental Theorem of Space Curves, Existence) *Suppose that* $\boldsymbol{\kappa}\colon (a,b) \to \mathbb{R}$ *and* $\boldsymbol{\tau}\colon (a,b) \to \mathbb{R}$ *are differentiable functions with* $\boldsymbol{\kappa} > 0$. *Then there exists a unit-speed curve* $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ *whose curvature and torsion are* $\boldsymbol{\kappa}$ *and* $\boldsymbol{\tau}$. *For* $a < s_0 < b$ *the value* $\boldsymbol{\beta}(s_0)$ *can be prescribed arbitrarily. Also, the values of* $\mathbf{T}(s_0)$ *and* $\mathbf{N}(s_0)$ *can be prescribed subject to the conditions that* $\|\mathbf{T}(s_0)\| = \|\mathbf{N}(s_0)\| = 1$ *and* $\mathbf{T}(s_0) \cdot \mathbf{N}(s_0) = 0$.

Proof. Consider the following system of 12 differential equations

$$\left\{\begin{array}{rcl} x_i'(s) &=& t_i(s), \\ t_i'(s) &=& \boldsymbol{\kappa}(s) n_i(s), \\ n_i'(s) &=& -\boldsymbol{\kappa}(s) t_i(s) + \boldsymbol{\tau}(s) b_i(s), \\ b_i'(s) &=& -\boldsymbol{\tau}(s) n_i(s), \end{array}\right. \tag{8.2}$$

all for $1 \leqslant i \leqslant 3$, together with the initial conditions

$$\left\{\begin{array}{rcl} x_i(s_0) &=& p_i, \\ t_i(s_0) &=& q_i, \\ n_i(s_0) &=& r_i, \\ b_1(s_0) &=& q_2 r_3 - q_3 r_2, \\ b_2(s_0) &=& q_3 r_1 - q_1 r_3, \\ b_3(s_0) &=& q_1 r_2 - q_2 r_1. \end{array}\right. \tag{8.3}$$

In (8.3) we require that

$$\sum_{i=1}^{3} q_i^2 = 1 = \sum_{i=1}^{3} r_i^2 \quad \text{and} \quad \sum_{i=1}^{3} q_i r_i = 0.$$

From the theory of systems of differential equations, we know that this linear system has a unique solution [GrMePi]. If we put $\boldsymbol{\beta} = (x_1, x_2, x_3)$ and

$$\mathbf{T} = (t_1, t_2, t_3), \qquad \mathbf{N} = (n_1, n_2, n_3), \qquad \mathbf{B} = (b_1, b_2, b_3),$$

we see that (8.2) can be compressed into the equations $\boldsymbol{\beta}' = \mathbf{T}$ and

$$\left\{\begin{array}{rcl} \mathbf{T}' &=& \boldsymbol{\kappa}\, \mathbf{N}, \\ \mathbf{N}' &=& -\boldsymbol{\kappa}\, \mathbf{T} + \boldsymbol{\tau}\, \mathbf{B}, \\ \mathbf{B}' &=& -\boldsymbol{\tau}\, \mathbf{N}. \end{array}\right. \tag{8.4}$$

Comparison with the Frenet formulas (7.12) shows that $\boldsymbol{\kappa}$ is the curvature of the curve $\boldsymbol{\beta}$, and $\boldsymbol{\tau}$ is the torsion. ∎

In practice, Theorem 8.4 is implemented by a program in Notebook 8 which solves the system of differential equations numerically, and then plots the resulting space curve. The following initial examples of its use are designed to investigate cases in which one of $\boldsymbol{\kappa}, \boldsymbol{\tau}$ is constant.

If $\boldsymbol{\kappa} = 0$ then the definition of $\boldsymbol{\tau}$ is irrelevant, as a straight line results. If $\boldsymbol{\tau} = 0$ then the curve is planar. If both $\boldsymbol{\kappa}, \boldsymbol{\tau}$ are constant, then the curve is a helix, and this will be oriented according to the assignment of the initial Frenet frame $\{\mathbf{T}(s_0), \mathbf{N}(s_0), \mathbf{B}(s_0)\}$.

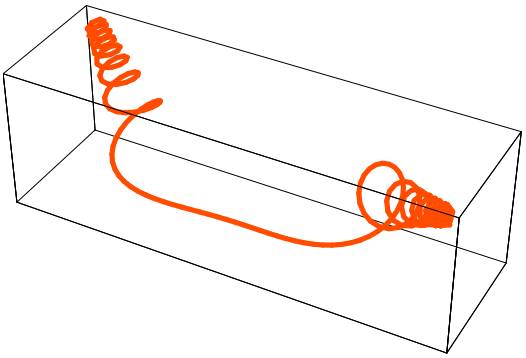


Figure 8.2: Curve with $\boldsymbol{\kappa}(s) = s$ and $\boldsymbol{\tau}(s) = 1$

Less familiar examples are formed by taking at least one of $\boldsymbol{\kappa}, \boldsymbol{\tau}$ to be the identity function $s \mapsto s$, or $s \mapsto cs$ with $c > 0$ a constant. Figure 8.2 is a 3-dimensional analogue of a clothoid, though a closer resemblance with a helix over a clothoid is obtained by setting $\boldsymbol{\kappa}(s) = |s|$ and $\boldsymbol{\tau}(s) = 1$. Figure 8.3 shows a "double corkscrew" characterized by both $\boldsymbol{\kappa}, \boldsymbol{\tau}$ being linear.

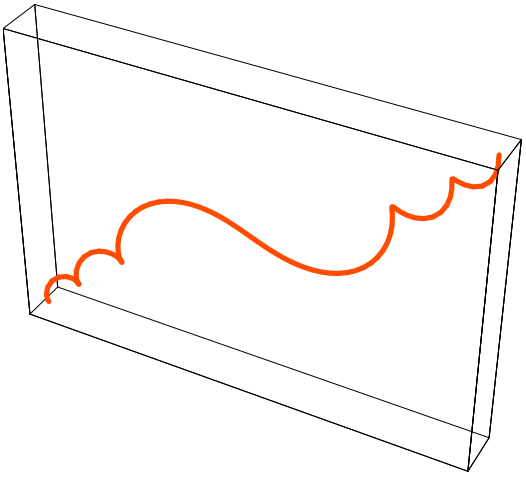


Figure 8.3: A curve with $\boldsymbol{\kappa}(s) = s$ and $\boldsymbol{\tau}(s) = s$

Figure 8.4 displays a straightforward example of a space curve with $\boldsymbol{\kappa}$ constant. It starts off at the origin resembling a semicircle, but becomes less planar as $|\boldsymbol{\tau}|$ increases. There are many other types of curves with constant curvature in space, characterized by different assignments of the torsion $\boldsymbol{\tau}(s)$ as a function of arc length.

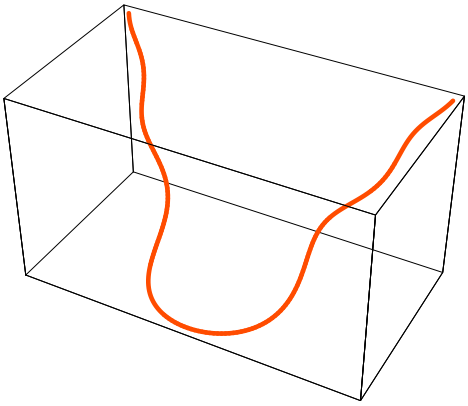


Figure 8.4: A curve with $\boldsymbol{\kappa}(s) = 1$ and $\boldsymbol{\tau}(s) = s$

Figure 8.5 consists of frames of an animation detailing the evolution of a space curve with constant curvature and periodic torsion proportional to $\sin s$. It begins with a circle traversed twice, which is then broken open and eventually joined up to become a simple closed curve admitting one point with $\boldsymbol{\tau}(s) = 1$. A similarly-formed closed curve is shown in Figure 8.14.

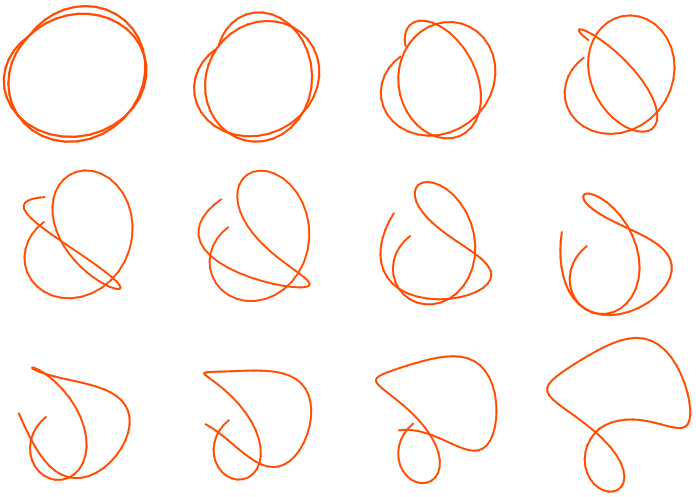


Figure 8.5: Curves with $\boldsymbol{\kappa}(s) = 1$ and $\boldsymbol{\tau}(s) = \frac{1}{12} n \sin s$, $n = 1, \ldots, 12$

8.3 Contact

In Section 4.1, we defined the evolute of a plane curve as the locus of its centers of curvature. In this section we show how to define the evolute of a space curve. This requires that we generalize the notion of center of curvature to space curves. For this purpose, we need the notion of *contact*. The latter is also important for plane curves, so we discuss it in a generality that will apply to both cases.

We begin by extending Definition 3.1.

Definition 8.5. *An* ***implicitly-defined hypersurface*** *in $\mathbb{R}^n$ is the set of zeros of a differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}$. We denote the set of zeros by $F^{-1}(0)$.*

An implicitly-defined hypersurface in $\mathbb{R}^2$ will be a (possibly singular) curve, and an implicitly-defined hypersurface in $\mathbb{R}^3$ will be a surface. However, we first consider the simplest hypersurfaces in $\mathbb{R}^n$, namely hyperplanes and hyperspheres. A ***hyperplane*** in $\mathbb{R}^n$ consists of the points represented by the set

$$\{\, \mathbf{p} \in \mathbb{R}^n \mid (\mathbf{p} - \mathbf{u}) \cdot \mathbf{v} = 0 \,\} \tag{8.5}$$

of vectors whose difference from a fixed one $\mathbf{u}$ is always perpendicular to another fixed vector $\mathbf{v}$ (normal to the plane). A hyperplane in $\mathbb{R}^2$ is a line (see the analogous Lemma 6.23), and a hyperplane in $\mathbb{R}^3$ is a plane.

A ***hypersphere*** in $\mathbb{R}^n$ of radius r centered at $\mathbf{q} \in \mathbb{R}^n$ is the set of points

$$\{\, \mathbf{p} \in \mathbb{R}^n \mid \|\mathbf{p} - \mathbf{q}\| = r \,\} \tag{8.6}$$

at a distance r from $\mathbf{q}$. In particular, a hypersphere in $\mathbb{R}^2$ is a circle and a hypersphere in $\mathbb{R}^3$ is a sphere. We leave it to the reader to find differentiable functions whose zero sets are precisely (8.5) and (8.6); recall (3.1) on page 75.

Among all lines passing through a point $\mathbf{p}$ on a curve $\boldsymbol{\alpha}$, the tangent line affords the best approximation, though other curves through $\mathbf{p}$ may approximate the curve even more closely. The mathematics that makes this idea precise arises from

Definition 8.6. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a regular curve, and let $F\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. We say that a parametrically-defined curve $\boldsymbol{\alpha}$ and an implicitly-defined hypersurface $F^{-1}(0)$ have* ***contact of order*** *k at $\boldsymbol{\alpha}(t_0)$ provided*

$$(F \circ \boldsymbol{\alpha})(t_0) = (F \circ \boldsymbol{\alpha})'(t_0) = \cdots = (F \circ \boldsymbol{\alpha})^{(k)}(t_0) = 0, \tag{8.7}$$

but

$$(F \circ \boldsymbol{\alpha})^{(k+1)}(t_0) \neq 0. \tag{8.8}$$

To check that contact is a geometric concept, we prove:

Lemma 8.7. *The definition of contact between* $\boldsymbol{\alpha}$ *and* $F^{-1}(0)$ *is independent of the parametrization of* $\boldsymbol{\alpha}$.

Proof. Let $\boldsymbol{\gamma}\colon (c,d) \to \mathbb{R}^n$ be a reparametrization of $\boldsymbol{\alpha}$; then there exists a differentiable function $h\colon (c,d) \to (a,b)$ such that $\boldsymbol{\gamma}(u) = \boldsymbol{\alpha}(h(u))$ for $c < u < d$. Let u_0 be such that $h(u_0) = h_0$. Then

$$\begin{aligned}
(F\circ\boldsymbol{\gamma})(u_0) &= (F\circ\boldsymbol{\alpha})(h_0) = 0,\\
(F\circ\boldsymbol{\gamma})'(u_0) &= (F\circ\boldsymbol{\alpha})'(h_0)h'(u_0) = 0,\\
(F\circ\boldsymbol{\gamma})''(u_0) &= (F\circ\boldsymbol{\alpha})''(h_0)h'(u_0)^2 + (F\circ\boldsymbol{\alpha})'(h_0)h''(u_0) = 0,\\
(F\circ\boldsymbol{\gamma})'''(u_0) &= (F\circ\boldsymbol{\alpha})'''(h_0)h'(u_0)^3 + 3(F\circ\boldsymbol{\alpha})''(h_0)h'(u_0)h''(u_0)\\
&\qquad +(F\circ\boldsymbol{\alpha})'(h_0)h'''(u_0) = 0,
\end{aligned}$$

and so forth. Such Leibniz formulas were computed in Notebook 1. In any case, we see that (8.7) implies that

$$(F\circ\boldsymbol{\gamma})(u_0) = (F\circ\boldsymbol{\gamma})'(u_0) = \cdots = (F\circ\boldsymbol{\gamma})^{(k)}(u_0) = 0,$$

and (8.8) implies that

$$(F\circ\boldsymbol{\gamma})^{(k+1)}(u_0) \neq 0. \ \blacksquare$$

For example, the x-axis is the set of zeros of the function $F\colon \mathbb{R}^2 \to \mathbb{R}$ given by $F(x,y) = y$. If n is a positive integer, the ***generalized parabola*** $t \mapsto (t,\, t^n)$ has contact of order $n-1$ with this horizontal line at $(0,0)$. A number of these curves are plotted in Figure 8.6.

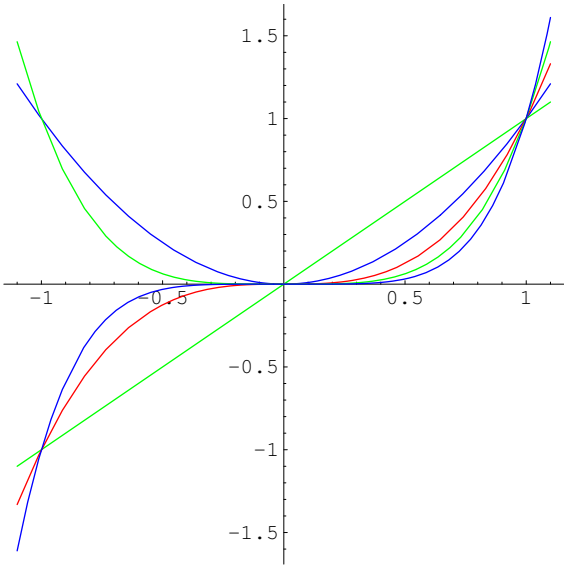


Figure 8.6: Generalized parabolas

Next, we discuss contact between hyperplanes and curves in $\mathbb{R}^n$.

Lemma 8.8. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^n$ be a unit-speed curve, let $a < s_0 < b$ and let $\mathbf{v} \in \mathbb{R}^n$. Then the hyperplane $\{\, \mathbf{q} \mid (\mathbf{q} - \boldsymbol{\beta}(s_0)) \cdot \mathbf{v} = 0 \,\}$ has at least order 1 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if $\mathbf{v}$ is perpendicular to $\boldsymbol{\beta}'(s_0)$.*

Proof. The hyperplane in question is the zero set of the function $F\colon \mathbb{R}^n \to \mathbb{R}$ given by $F(\mathbf{q}) = (\mathbf{q} - \boldsymbol{\beta}(s_0)) \cdot \mathbf{v}$, so we define $f(s) = (\boldsymbol{\beta}(s) - \boldsymbol{\beta}(s_0)) \cdot \mathbf{v}$. Then

$$f'(s) = \boldsymbol{\beta}'(s) \cdot \mathbf{v}.$$

Hence $f'(s_0) = 0$ if and only if $\boldsymbol{\beta}'(s_0) \cdot \mathbf{v} = 0$. ∎

Thus at each point on a regular curve, there are hyperplanes with at least order 1 contact with the curve. The tangent line to a *plane* curve is such a hypersurface, even though it only has dimension 1. As we have just seen, higher order contact with a hyperplane is possible for special curves; in this regard, see Exercise 11 and Figure 7.3.

We turn to contact between hyperspheres and curves. Intuitively, it should be the case that a hypersphere can be chosen to have higher order contact with a curve at a given point than is possible with a hyperplane. We first consider plane curves.

Lemma 8.9. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed curve, and choose s_0 such that $a < s_0 < b$. Let $\mathbf{v} \in \mathbb{R}^2$ be distinct from $\boldsymbol{\beta}(s_0)$ so that $r = \|\boldsymbol{\beta}(s_0) - \mathbf{v}\| > 0$. Let $\mathscr{C} = \{\, \mathbf{q} \in \mathbb{R}^2 \mid \|\mathbf{q} - \mathbf{v}\| = r \,\}$.*

(i) *The circle $\mathscr{C}$ has at least order 1 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if its center $\mathbf{v}$ lies on the line perpendicular to $\boldsymbol{\beta}'(s_0)$ at $\boldsymbol{\beta}(s_0)$.*

(ii) *Suppose that $\boldsymbol{\kappa 2}(s_0) \neq 0$. Then $\mathscr{C}$ has at least order 2 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if*

$$\mathbf{v} = \boldsymbol{\beta}(s_0) + \frac{1}{\boldsymbol{\kappa 2}(s_0)} J\mathbf{T}(s_0).$$

When this is the case, the radius r of $\mathscr{C}$ equals $1/|\boldsymbol{\kappa 2}(s_0)|$.

Proof. Define $g(s) = \|\boldsymbol{\beta}(s) - \mathbf{v}\|^2$. Then $g' = 2\boldsymbol{\beta}' \cdot (\boldsymbol{\beta} - \mathbf{v}) = 2\mathbf{T} \cdot (\boldsymbol{\beta} - \mathbf{v})$, so that $g'(s_0) = 0$ if and only if $\boldsymbol{\beta}'(s_0) \cdot (\boldsymbol{\beta}(s_0) - \mathbf{v}) = 0$. This proves (i).

It follows from (i) that there is a number λ such that $\boldsymbol{\beta}(s_0) - \mathbf{v} = \lambda J\mathbf{T}(s_0)$. Also, we have

$$g'' = 2\big(\mathbf{T} \cdot \mathbf{T} + \mathbf{T}' \cdot (\boldsymbol{\beta} - \mathbf{v})\big) = 2\big(1 + \boldsymbol{\kappa 2}\, J\mathbf{T} \cdot (\boldsymbol{\beta} - \mathbf{v})\big);$$

hence

$$g''(s_0) = 2\big(1 + \boldsymbol{\kappa 2}(s_0)\lambda\big).$$

Thus $g''(s_0) = 0$ if and only if $\lambda = -1/\boldsymbol{\kappa 2}(s_0)$, proving (ii). ∎

Part (ii) of Lemma 8.9 yields a characterization of the osculating circle defined in Section 4.4:

Corollary 8.10. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular plane curve, and let $a < t_0 < b$ be such that $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}](t_0) \neq 0$. Then the osculating circle of $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_0)$ is the unique circle which has at least order 2 contact with $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_0)$.*

Next, we consider contact between space curves and spheres. The next theorem shows in stages that contact of order 3 is generally possible.

Theorem 8.11. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a unit-speed curve, and choose s_0 such that $a < s_0 < b$. Let $\mathbf{v} \in \mathbb{R}^3$ be distinct from $\boldsymbol{\beta}(s_0)$ so that $r = \|\boldsymbol{\beta}(s_0) - \mathbf{v}\| > 0$. We abbreviate $\boldsymbol{\kappa}[\boldsymbol{\beta}]$ and $\boldsymbol{\tau}[\boldsymbol{\beta}]$ to $\boldsymbol{\kappa}$ and $\boldsymbol{\tau}$, and let $\mathscr{S} = \{\, \mathbf{q} \in \mathbb{R}^3 \mid \|\mathbf{q} - \mathbf{v}\| = r \,\}$.*

(i) *The sphere $\mathscr{S}$ has at least order 1 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if its center $\mathbf{v}$ lies on a line perpendicular to $\boldsymbol{\beta}'(s_0)$ at $\boldsymbol{\beta}(s_0)$.*

(ii) *Suppose that $\boldsymbol{\kappa}(s_0) \neq 0$. Then $\mathscr{S}$ has at least order 2 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if*

$$\mathbf{v} = \boldsymbol{\beta}(s_0) + \frac{1}{\boldsymbol{\kappa}(s_0)}\mathbf{N}(s_0) - b\mathbf{B}(s_0)$$

for some number b.

(iii) *Suppose that $\boldsymbol{\kappa}(s_0) \neq 0$ and $\boldsymbol{\tau}(s_0) \neq 0$. Then $\mathscr{S}$ has at least order 3 contact with $\boldsymbol{\beta}$ at s_0 if and only if*

$$\mathbf{v} = \boldsymbol{\beta}(s_0) + \frac{1}{\boldsymbol{\kappa}(s_0)}\mathbf{N}(s_0) - \frac{\boldsymbol{\kappa}'(s_0)}{\boldsymbol{\kappa}(s_0)^2\boldsymbol{\tau}(s_0)}\mathbf{B}(s_0).$$

Proof. Define $g(s) = \big\|\boldsymbol{\beta}(s) - \mathbf{v}\big\|^2$. We calculate as in part (i) of Lemma 8.9:

$$g' = 2\boldsymbol{\beta}' \cdot (\boldsymbol{\beta} - \mathbf{v}) = 2\mathbf{T} \cdot (\boldsymbol{\beta} - \mathbf{v}),$$

so that $g'(s_0) = 0$ if and only if $\boldsymbol{\beta}'(s_0) \cdot \big(\boldsymbol{\beta}(s_0) - \mathbf{v}\big) = 0$, proving (i). It follows from (i) that there are numbers a and b such that

$$\boldsymbol{\beta}(s_0) - \mathbf{v} = a\mathbf{N}(s_0) + b\mathbf{B}(s_0).$$

Also, computing as in part (ii) of Lemma 8.9, we find that

$$\tfrac{1}{2}g'' = \mathbf{T} \cdot \mathbf{T} + \mathbf{T}' \cdot (\boldsymbol{\beta} - \mathbf{v}) = 1 + \boldsymbol{\kappa}\mathbf{N} \cdot (\boldsymbol{\beta} - \mathbf{v}). \tag{8.9}$$

Hence

$$\tfrac{1}{2}g''(s_0) = 1 + \boldsymbol{\kappa}(s_0)\mathbf{N}(s_0) \cdot \big(a\mathbf{N}(s_0) + b\mathbf{B}(s_0)\big) = 1 + \boldsymbol{\kappa}(s_0)a.$$

Thus $g''(s_0) = 0$ if and only if $a = -1/\boldsymbol{\kappa}(s_0)$, proving (ii).

For (iii) we use (8.9) to compute

$$\tfrac{1}{2}g''' = \big(\boldsymbol{\kappa}'\mathbf{N} + \boldsymbol{\kappa}\mathbf{N}'\big)\cdot(\boldsymbol{\beta} - \mathbf{v}) + \boldsymbol{\kappa}\mathbf{N}\cdot\mathbf{T} = \big(\boldsymbol{\kappa}'\mathbf{N} - \boldsymbol{\kappa}^2\mathbf{T} + \boldsymbol{\kappa}\boldsymbol{\tau}\mathbf{B}\big)\cdot(\boldsymbol{\beta} - \mathbf{v}).$$

In particular, $g'''(s_0)/2$ equals

$$\begin{aligned}\big(\boldsymbol{\kappa}'(s_0)\mathbf{N}(s_0) - \boldsymbol{\kappa}(s_0)^2\mathbf{T}(s_0) + \boldsymbol{\kappa}(s_0)\boldsymbol{\tau}(s_0)\mathbf{B}(s_0)\big)\cdot\left(\frac{\mathbf{N}(s_0)}{\boldsymbol{\kappa}(s_0)} + b\mathbf{B}(s_0)\right)\\ = -\frac{\boldsymbol{\kappa}'(s_0)}{\boldsymbol{\kappa}(s_0)} + \boldsymbol{\kappa}(s_0)\boldsymbol{\tau}(s_0)b.\end{aligned}$$

Thus $g'''(s_0) = 0$ if and only if $b = \boldsymbol{\kappa}'(s_0)/\big(\boldsymbol{\kappa}(s_0)^2\boldsymbol{\tau}(s_0)\big)$. ∎

The space curve analog of an osculating circle is an osculating sphere, first considered by Fuss[1] in his paper [Fuss]. The analysis above allows us to define it precisely as follows:

Definition 8.12. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular space curve, and let $a < t_0 < b$ be such that $\boldsymbol{\kappa}[\boldsymbol{\alpha}](t_0) \neq 0$ and $\boldsymbol{\tau}[\boldsymbol{\alpha}](t_0) \neq 0$. Then the* ***osculating sphere*** *of $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_0)$ is the unique sphere which has at least order 3 contact with $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_0)$.*

Corollary 8.13. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular space curve whose curvature and torsion do not vanish at t_0. Then the osculating sphere at $\boldsymbol{\alpha}(t_0)$ is a sphere of radius*

$$\sqrt{\frac{1}{\boldsymbol{\kappa}[\boldsymbol{\alpha}](t_0)^2} + \frac{\boldsymbol{\kappa}[\boldsymbol{\alpha}]'(t_0)^2}{\boldsymbol{\tau}[\boldsymbol{\alpha}](t_0)^2\boldsymbol{\kappa}[\boldsymbol{\alpha}](t_0)^4}} \tag{8.10}$$

and center

$$\mathbf{v} = \boldsymbol{\alpha}(t_0) + \frac{1}{\boldsymbol{\kappa}[\boldsymbol{\alpha}](t_0)}\mathbf{N}(t_0) - \frac{\boldsymbol{\kappa}[\boldsymbol{\alpha}]'(t_0)}{\big\|\boldsymbol{\alpha}'(t_0)\big\|\boldsymbol{\kappa}[\boldsymbol{\alpha}](t_0)^2\boldsymbol{\tau}[\boldsymbol{\alpha}](t_0)}\mathbf{B}(t_0). \tag{8.11}$$

Proof. Let $\boldsymbol{\beta}$ be a unit-speed reparametrization of $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}(t) = \boldsymbol{\beta}(s(t))$; then by the chain rule

$$\boldsymbol{\kappa}[\boldsymbol{\beta}]'(s(t))s'(t) = \boldsymbol{\kappa}[\boldsymbol{\alpha}]'(t).$$

From this fact and part (iii) of Theorem 8.11 we get (8.11). Since the radius of the osculating sphere at $\boldsymbol{\alpha}(t_0)$ is $\|\mathbf{v} - \boldsymbol{\alpha}(t_0)\|$, equation (8.10) follows from (8.11). ∎

[1] Nicolaus Fuss (1755–1826). Born in Basel, Fuss went to St. Petersburg, where he became Euler's secretary. He edited Euler's works and wrote papers on insurance and many textbooks. From 1800 to his death he was permanent secretary of the St. Petersburg Academy.

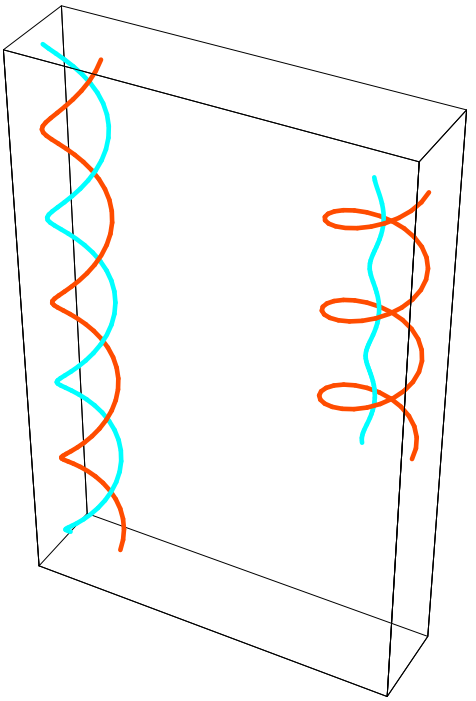


Figure 8.7: Dual pairs $\mathsf{helix}[1,1]$, $\mathsf{helix}[-1,1]$ (left) and $\mathsf{helix}[\frac{3}{2},\frac{1}{2}]$, $\mathsf{helix}[-\frac{1}{6},\frac{1}{2}]$ (right)

In Section 4.1, we defined the evolute of a plane curve to be the locus of the centers of the osculating circles to the curve. It is natural to define the evolute of a space curve to be the locus of the centers of the osculating spheres.

Definition 8.14. *The* ***evolute*** *of a regular space curve* $\boldsymbol{\alpha}$ *is the curve given by*

$$\mathsf{evolute3d}[\boldsymbol{\alpha}](t) = \boldsymbol{\alpha}(t) + \frac{1}{\boldsymbol{\kappa}[\boldsymbol{\alpha}](t)}\mathbf{N}(t) - \frac{\boldsymbol{\kappa}[\boldsymbol{\alpha}]'(t)}{\|\boldsymbol{\alpha}'(t)\|\boldsymbol{\kappa}[\boldsymbol{\alpha}](t)^2\,\boldsymbol{\tau}[\boldsymbol{\alpha}](t)}\mathbf{B}(t).$$

Consider for example $\boldsymbol{\alpha} = \mathsf{helix}[a,b]$. From (7.16), we have $\boldsymbol{\kappa}[\boldsymbol{\alpha}]' = 0$. Using (7.17), with t in place of $s/\sqrt{a^2+b^2}$, yields

$$\mathsf{evolute3d}[\boldsymbol{\alpha}](t) = \left(-\frac{b^2}{a}\cos t,\ -\frac{b^2}{a}\sin t,\ bt\right).$$

This is merely $\mathsf{helix}[-b^2/a,\ b]$ (it is now convenient to allow the radial parameter to be negative). The original helix can be recovered by applying taking the evolute again, so the two helices are "dual" to each other; see Figure 8.7. As a second exmple, the evolute of part of a twisted cubic is shown in Figure 8.8.

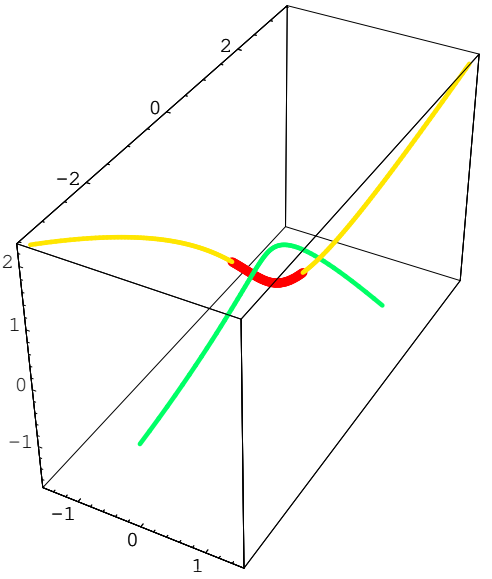


Figure 8.8: The evolute of a segment (thickened underneath) of the twisted cubic $t \mapsto (t, t^3, t^2)$

8.4 Space Curves That Lie on a Sphere

If a space curve (such as Viviani's curve) lies on a sphere, then it has contact of order n with that sphere for any positive integer n. In this section we show that certain relations between the curvature and torsion of a space curve are necessary and sufficient for the space curve to lie on a sphere.

Theorem 8.15. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a unit-speed curve, with curvature $\boldsymbol{\kappa}$ and torsion $\boldsymbol{\tau}$. Suppose that $\boldsymbol{\beta}$ lies on the sphere of radius $c > 0$ centered at $\mathbf{q} \in \mathbb{R}^3$. Then*:

(i) $\boldsymbol{\kappa} \geqslant 1/c$;

(ii) $\boldsymbol{\kappa}$ *and* $\boldsymbol{\tau}$ *are related by*

$$\boldsymbol{\tau}^2 \left(c^2 - \frac{1}{\boldsymbol{\kappa}^2} \right) = \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2} \right)^2 ; \tag{8.12}$$

(iii) $\boldsymbol{\kappa}'(t_0) = 0$ *implies* $\boldsymbol{\tau}(t_0) = 0$ *or* $\boldsymbol{\kappa}(t_0) = 1/c$;

(iv) $\dfrac{\boldsymbol{\tau}}{\boldsymbol{\kappa}} = \left(\dfrac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2} \right)'$;

(v) $\boldsymbol{\kappa}'(t_0) = \boldsymbol{\kappa}''(t_0) = 0$ *implies* $\boldsymbol{\tau}(t_0) = 0$;

(vi) $\boldsymbol{\kappa} \equiv 1/a$ *if and only if* $\boldsymbol{\beta}$ *is a circle of radius* a*, necessarily contained in some plane.*

Proof. Differentiation of the equation $\|\boldsymbol{\beta} - \mathbf{q}\|^2 = c^2$ yields $(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{T} = 0$. Another differentiation results in

$$1 + (\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{T}' = 0.$$

Hence $\mathbf{T}'$ is nonzero, which implies the Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is well defined on $\boldsymbol{\beta}$. Thus

$$1 + (\boldsymbol{\beta} - \mathbf{q}) \cdot (\boldsymbol{\kappa}\mathbf{N}) = 0. \tag{8.13}$$

The Cauchy–Schwarz inequality (from page 3) implies that

$$1 = |(\boldsymbol{\beta} - \mathbf{q}) \cdot (\boldsymbol{\kappa}\mathbf{N})| \leqslant \|\boldsymbol{\beta} - \mathbf{q}\| \, \|\boldsymbol{\kappa}\mathbf{N}\| = \boldsymbol{\kappa} c,$$

proving (i).

Differentiating (8.13), we obtain

$$\begin{aligned} 0 \;&=\; \mathbf{T} \cdot (\boldsymbol{\kappa}\mathbf{N}) + (\boldsymbol{\beta} - \mathbf{q}) \cdot \big(\boldsymbol{\kappa}'\mathbf{N} + \boldsymbol{\kappa}(-\boldsymbol{\kappa}\mathbf{T} + \boldsymbol{\tau}\mathbf{B})\big) \\ &=\; \boldsymbol{\kappa}'(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{N} + \boldsymbol{\kappa}\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{B} \qquad (8.14) \\ &=\; -\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}} + \boldsymbol{\kappa}\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{B}. \end{aligned}$$

Using (7.11) on page 197, we may now write

$$\begin{aligned} \boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \;&=\; \big(\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{T}\big)\mathbf{T} + \big(\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{N}\big)\mathbf{N} \\ &\qquad\qquad + \big(\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{B}\big)\mathbf{B} \qquad (8.15) \\ &=\; -\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\mathbf{N} + \frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}\mathbf{B}, \end{aligned}$$

the last line from (8.13) and (8.14). Taking norms,

$$\boldsymbol{\tau}^2 c^2 = \boldsymbol{\tau}^2 \|\boldsymbol{\beta} - \mathbf{q}\|^2 = \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\right)^2 + \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}\right)^2,$$

from which we obtain (ii). The latter implies (iii).

To prove (iv), we first rearrange (8.14) as

$$\boldsymbol{\tau}(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{B} = \frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}. \tag{8.16}$$

Differentiation of (8.16) together with (8.13) yields

$$\begin{aligned} \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}\right)' \;&=\; \boldsymbol{\tau}'(\boldsymbol{\beta} - \mathbf{q}) \cdot \mathbf{B} + \boldsymbol{\tau}\big(\mathbf{T} \cdot \mathbf{B} + (\boldsymbol{\beta} - \mathbf{q}) \cdot (-\boldsymbol{\tau}\mathbf{N})\big) \qquad (8.17) \\ &=\; \frac{\boldsymbol{\kappa}'\boldsymbol{\tau}'}{\boldsymbol{\kappa}^2\boldsymbol{\tau}} + \frac{\boldsymbol{\tau}^2}{\boldsymbol{\kappa}}. \end{aligned}$$

This implies (iv). Furthermore, (iv) implies (v).

Finally, if $\boldsymbol{\kappa} \equiv 1/a$, then (v) implies that $\boldsymbol{\tau} \equiv 0$. Hence by Lemma 7.11 we see that $\boldsymbol{\alpha}$ is a plane curve of constant curvature. From Theorem 1.22, page 16 it follows that $\boldsymbol{\alpha}$ is a circle. ∎

It is interesting to check that for Viviani's curve (which lies on a sphere of radius $2a$), $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}'$ have the same zeros. This is already suggested by Figure 7.7. Instead, Figure 8.9 shows the curve

$$\boldsymbol{\alpha}(t) = \left(\tfrac{3}{4}\cos t,\ \tfrac{1}{2}\sin t,\ \sqrt{1 - \tfrac{9}{16}\cos^2 t - \tfrac{1}{4}\sin^2 t} \right)$$

lying on a sphere of radius 1, projecting to an ellipse in the xy-plane. Figure 8.10 plots $\boldsymbol{\kappa}[\boldsymbol{\alpha}]'$ (with maximum value in excess of 4) and $\boldsymbol{\tau}[\boldsymbol{\alpha}]$.

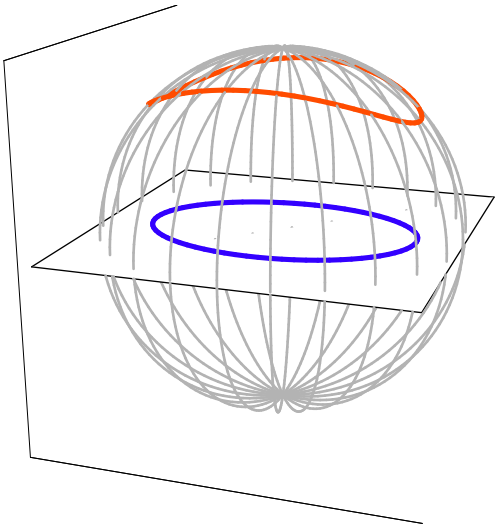


Figure 8.9: The elliptical space curve $\boldsymbol{\alpha}$

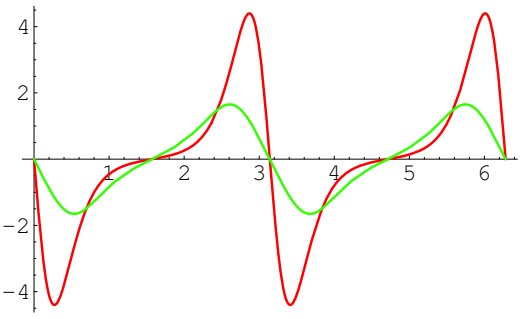


Figure 8.10: $\boldsymbol{\kappa}[\boldsymbol{\alpha}]'$ (darker) and $\boldsymbol{\tau}[\boldsymbol{\alpha}]$

Notice that equation (8.12) also holds for a helix of incline b over a circle of radius a where $ca = a^2 + b^2$, but that conditions (iv) and (v) fail. Indeed, a helix with nonzero torsion lies on no sphere, and we cannot therefore expect to characterize spherical curves by (8.12) alone. Nevertheless, we have the following result, which uses the technique of Theorem 1.22.

Theorem 8.16. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a unit-speed curve whose curvature $\boldsymbol{\kappa}$ and torsion $\boldsymbol{\tau}$ satisfy* (8.12). *Assume that $\boldsymbol{\kappa}$, $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}'$ vanish only at isolated points. Then $\boldsymbol{\beta}$ lies on a sphere of radius c centered at some $\mathbf{q} \in \mathbb{R}^3$.*

Proof. Let

$$\boldsymbol{\gamma} = \boldsymbol{\beta} + \frac{1}{\boldsymbol{\kappa}}\mathbf{N} - \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)\mathbf{B}.$$

The Frenet formulas (7.10) imply that

$$\begin{aligned} \boldsymbol{\gamma}' &= \mathbf{T} - \frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}\mathbf{N} + \frac{1}{\boldsymbol{\kappa}}(-\boldsymbol{\kappa}\mathbf{T} + \boldsymbol{\tau}\mathbf{B}) - \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)'\mathbf{B} + \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)\boldsymbol{\tau}\mathbf{N} \\ &= \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}} - \left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)'\right)\mathbf{B}. \end{aligned} \tag{8.18}$$

Differentiation of (8.12) yields

$$\left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)\left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)' = \frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^3}. \tag{8.19}$$

From (8.18) and (8.19), we get

$$\left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2}\right)\boldsymbol{\gamma}' = \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\frac{\boldsymbol{\kappa}'}{\boldsymbol{\tau}\boldsymbol{\kappa}^2} - \frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^3}\right)\mathbf{B} = 0$$

or $\boldsymbol{\kappa}'\boldsymbol{\gamma}' = 0$. Thus $\boldsymbol{\gamma}'(t) = 0$ whenever t is such that $\boldsymbol{\kappa}'(t) \neq 0$. Since $\boldsymbol{\gamma}'$ is continuous, $\boldsymbol{\gamma}'(t) = 0$ for all t. Hence there exists $\mathbf{q} \in \mathbb{R}^3$ such that $\boldsymbol{\gamma}(t) = \mathbf{q}$ for all t. In other words,

$$\boldsymbol{\beta}(t) - \mathbf{q} = -\frac{1}{\boldsymbol{\kappa}(t)}\mathbf{N}(t) + \frac{\boldsymbol{\kappa}'(t)}{\boldsymbol{\tau}(t)\boldsymbol{\kappa}(t)^2}\mathbf{B}(t) \tag{8.20}$$

for all t. Then (8.12) implies that

$$\|\boldsymbol{\beta}(t) - \mathbf{q}\|^2 = c^2.$$

Hence $\boldsymbol{\beta}$ lies on a sphere of radius c centered at $\mathbf{q}$. ∎

The conclusion of Theorem 8.16 can be expressed by saying that the osculating spheres of $\boldsymbol{\beta}$ all coincide.

8.5 Curves of Constant Slope

Since a helix has constant curvature $\boldsymbol{\kappa}$ and torsion $\boldsymbol{\tau}$, the ratio $\boldsymbol{\tau}/\boldsymbol{\kappa}$ is constant. Are there other curves with this property? In order to answer this question, we first show that the constancy of $\boldsymbol{\tau}/\boldsymbol{\kappa}$ for a curve is equivalent to the constancy of the angle between the curve and a fixed vector. More precisely:

Definition 8.17. *A curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ is said to have* ***constant slope*** *with respect to a unit vector $\mathbf{u} \in \mathbb{R}^3$, provided the angle ϕ between $\mathbf{u}$ and the unit tangent vector $\mathbf{T}$ of $\boldsymbol{\gamma}$ is constant. The analytical condition is*

$$\mathbf{T} \cdot \mathbf{u} = \cos\phi,$$

with $0 \leqslant \phi \leqslant \pi$, and we call $\cot\phi$ the ***slope*** *of $\boldsymbol{\gamma}$.*

With this convention, if the trace of $\boldsymbol{\gamma}$ is a straight line lying in the yz-plane $x = 0$ and $\mathbf{u} = (0,0,1)$, then the slope of $\boldsymbol{\gamma}$ with respect to $\mathbf{u}$ equals its gradient dz/dy.

Lemma 8.18. *If a curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ has nonzero curvature $\boldsymbol{\kappa}$ and constant slope with respect to some unit vector $\mathbf{u}$, then $\boldsymbol{\tau}/\boldsymbol{\kappa}$ has the constant value $\pm\cot\phi$.*

Proof. Without loss of generality, $\boldsymbol{\gamma}$ has unit speed. We differentiate $\mathbf{T} \cdot \mathbf{u} = \cos\phi$, obtaining $(\boldsymbol{\kappa}\mathbf{N}) \cdot \mathbf{u} = 0$. Therefore, $\mathbf{u}$ is perpendicular to $\mathbf{N}$, and so we can write

$$\mathbf{u} = \mathbf{T}\cos\phi \pm \mathbf{B}\sin\phi.$$

When we differentiate this equation, we obtain

$$0 = \mathbf{N}(\boldsymbol{\kappa}\cos\phi \mp \boldsymbol{\tau}\sin\phi),$$

and the lemma follows. ∎

The converse of Lemma 8.18 is true.

Lemma 8.19. *Let $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ be a curve for which the ratio $\boldsymbol{\tau}/\boldsymbol{\kappa}$ is constant. Write $\boldsymbol{\tau}/\boldsymbol{\kappa} = \cot\phi$. Then $\boldsymbol{\gamma}$ has constant slope $\cot\phi$ with respect to some unit vector $\mathbf{u} \in \mathbb{R}^3$.*

Proof. Without loss of generality, $\boldsymbol{\gamma}$ has unit speed. We have

$$0 = \mathbf{N}(\boldsymbol{\kappa}\cos\phi - \boldsymbol{\tau}\sin\phi) = \frac{d}{ds}(\mathbf{T}\cos\phi + \mathbf{B}\sin\phi).$$

It follows that there exists a constant unit vector $\mathbf{u} \in \mathbb{R}^3$ such that

$$\mathbf{u} = \mathbf{T}\cos\phi + \mathbf{B}\sin\phi.$$

In particular, $\mathbf{u} \cdot \mathbf{T}$ is constant. ∎

Returning to the helix, it is easily verified that (7.15) has constant slope b/a with respect to $(0,0,1)$ (Exercise 3). Next, we show that any unit-speed plane curve gives rise to a constant-slope space curve in the same way that a circle gives rise to a helix.

Lemma 8.20. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed plane curve, and write $\boldsymbol{\beta}(s) = \big(b_1(s), b_2(s)\big)$. Define a space curve $\boldsymbol{\gamma}$ by*

$$\boldsymbol{\gamma}(s) = \big(b_1(s),\ b_2(s),\ s\cos\psi\big),$$

where ψ is constant. Then $\boldsymbol{\gamma}$ has constant speed $\sqrt{1+\cos^2\psi}$ and constant slope $\cos\psi$ with respect to $(0,0,1)$.

Proof. The speed of $\boldsymbol{\gamma}$ is the norm of $\boldsymbol{\gamma}'$ and is obviously $\sqrt{1+\cos^2\psi}$. Using this fact, the angle ϕ between $\boldsymbol{\gamma}'(t)$ and $(0,0,1)$ is determined by

$$\cos\phi = \frac{\boldsymbol{\gamma}'(t)\cdot(0,0,1)}{\|\boldsymbol{\gamma}'(t)\|} = \frac{\cos\psi}{\sqrt{1+\cos^2\psi}},$$

whence

$$\sec^2\phi = 1 + \frac{1}{\cos^2\psi},$$

and $\cot\phi = \cos\psi$ as stated. ∎

The reverse construction also works: a space curve of constant slope projects to a plane curve in a natural way.

Lemma 8.21. *Let $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ be a unit-speed curve that has constant slope $\cot\phi$ with respect to a unit vector $\mathbf{u} \in \mathbb{R}^3$. Assume that $0 < \phi < \pi$. Let $\boldsymbol{\beta}$ be the projection of $\boldsymbol{\gamma}$ on a plane perpendicular to $\mathbf{u}$; that is,*

$$\boldsymbol{\beta}(s) = \boldsymbol{\gamma}(s) - (\boldsymbol{\gamma}(s)\cdot\mathbf{u})\mathbf{u}$$

for $a < s < b$. Then $\boldsymbol{\beta}$ has constant speed $|\sin\phi|$; furthermore, the curvature of $\boldsymbol{\gamma}$ and the signed curvature of $\boldsymbol{\beta}$ are related by

$$\boldsymbol{\kappa 2}[\boldsymbol{\beta}] = \pm\boldsymbol{\kappa}[\boldsymbol{\gamma}]\csc^2\phi. \tag{8.21}$$

Proof. Let v denote the speed of $\boldsymbol{\beta}$, and let $\mathbf{T}_\beta$ and $\mathbf{T}_\gamma$ denote the unit tangent vectors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. We have

$$v\mathbf{T}_\beta = \boldsymbol{\beta}'(s) = \mathbf{T}_\gamma - (\mathbf{T}_\gamma\cdot\mathbf{u})\mathbf{u} = \mathbf{T}_\gamma - (\cos\phi)\mathbf{u}.$$

Thus

$$1 = \|\mathbf{T}_\gamma\|^2 = v^2 + \cos^2\phi,$$

proving that $v = |\sin\phi|$. Also, we have $J\mathbf{T}_\beta = \pm\mathbf{N}_\gamma$. Hence

$$\boldsymbol{\kappa}[\boldsymbol{\gamma}]\mathbf{N}_\gamma = \mathbf{T}_\gamma' = v\mathbf{T}_\beta' = v^2\boldsymbol{\kappa 2}[\boldsymbol{\beta}]J\mathbf{T}_\beta = \pm(\sin^2\phi)\boldsymbol{\kappa 2}[\boldsymbol{\beta}]\,\mathbf{N}_\gamma.$$

Since $\sin\phi \neq 0$, we get (8.21). ∎

To find interesting curves of constant slope other than a helix or a plane curve, we look for curves of constant slope that lie on some sphere. Figure 8.11 represents a typical example of such a curve (viewed from above and from the sides). First, we determine what relations must exist between the curvature and torsion of such curves.

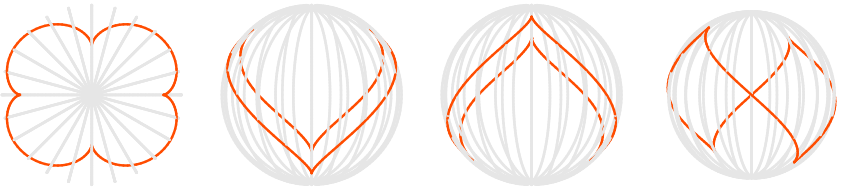


Figure 8.11: Four views of a curve on a sphere with constant slope

Lemma 8.22. *Let* $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ *be a unit-speed curve that has constant slope* $\cot\phi$ *with respect to a unit vector* $\mathbf{u} \in \mathbb{R}^3$*, where* $0 < \phi < \pi/2$*. Assume also that* $\boldsymbol{\gamma}$ *lies on a sphere of radius* $c > 0$*.*

(i) *The curvature and torsion of* $\boldsymbol{\gamma}$ *are given by*

$$\boldsymbol{\kappa}[\gamma](s)^2 = \frac{1}{c^2 - s^2\cot^2\phi} \quad \textit{and} \quad \boldsymbol{\tau}[\gamma](s)^2 = \frac{1}{c^2\tan^2\phi - s^2}. \tag{8.22}$$

(ii) *Let* $\boldsymbol{\beta}$ *be the projection of* $\boldsymbol{\gamma}$ *on a plane perpendicular to* $\mathbf{u}$*. Then the signed curvature of* $\boldsymbol{\beta}$ *satisfies*

$$\boldsymbol{\kappa 2}[\boldsymbol{\beta}](s_1)^2 = \frac{1}{c^2\,\sin^4\phi - {s_1}^2\,\cos^2\phi}\,, \tag{8.23}$$

where $s_1 = s\sin\phi$ *is the arc length function of* $\boldsymbol{\beta}$*.*

(iii) *Let* $c = a + 2b$ *and* $\cos\phi = a/(a+2b)$*. Then the natural equation* (8.23) *of* $\boldsymbol{\beta}$ *is the same as that of* epicycloid$[a,b]$ *defined on page* 144*.*

Proof. Since $\boldsymbol{\tau} = \boldsymbol{\kappa}\cot\phi$, from (8.12) we get

$$\left(\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2}\right)^2 = \boldsymbol{\tau}^2\left(c^2 - \frac{1}{\boldsymbol{\kappa}^2}\right) = (\boldsymbol{\kappa}^2c^2 - 1)\cot^2\phi,$$

or

$$\frac{\boldsymbol{\kappa}'}{\boldsymbol{\kappa}^2\sqrt{\boldsymbol{\kappa}^2c^2 - 1}} = \pm\cot\phi. \tag{8.24}$$

Integrating (8.24), we obtain

$$\frac{1}{\boldsymbol{\kappa}}\sqrt{\boldsymbol{\kappa}^2c^2 - 1} = \pm s\cot\phi. \tag{8.25}$$

When we solve (8.25) for $\boldsymbol{\kappa}$, we obtain (8.22), proving (i). Then (ii) follows from (8.21) and (i).

For (iii), we first compute

$$\sin^4\phi = \left(1 - \frac{a^2}{(a+2b)^2}\right)^2 = \frac{16b^2(a+b)^2}{(a+2b)^4}.$$

Hence (8.23) becomes

$$\frac{1}{\boldsymbol{\kappa}\mathbf{2}[\boldsymbol{\beta}](s_1)^2} = \frac{16b^2(a+b)^2}{(a+2b)^2} - \frac{a^2 s_1^2}{(a+2b)^2}. \tag{8.26}$$

Then (8.26) is the same as (5.13) with s replaced by s_1. ∎

Lemma 8.22 suggests the definition of a new kind of space curve.

Definition 8.23. *Write* $\mathsf{epicycloid}[a,b](t) = (x(t), y(t))$. *The* ***spherical helix*** *with parameters* a, b *is defined by*

$$\mathsf{sphericalhelix}[a,b](t) = \left(x(t),\ y(t),\ 2\sqrt{ab+b^2}\cos\frac{at}{2b}\right).$$

From Lemma 8.22 we have immediately

Lemma 8.24. *The space curve* $\mathsf{sphericalhelix}[a,b]$ *has constant slope with respect to* $(0,0,1)$ *and lies on a sphere of radius* $a+2b$ *centered at the origin.*

Figure 8.12 illustrates the following two examples. On the unit sphere $S^2(1)$, the curves of constant slope that project to a nephroid (see Exercise 6 on page 86) and a cardioid are respectively

$$\mathsf{sphericalhelix}[\tfrac{1}{2},\tfrac{1}{4}] \qquad \text{and} \qquad \mathsf{sphericalhelix}[\tfrac{1}{3},\tfrac{1}{3}].$$

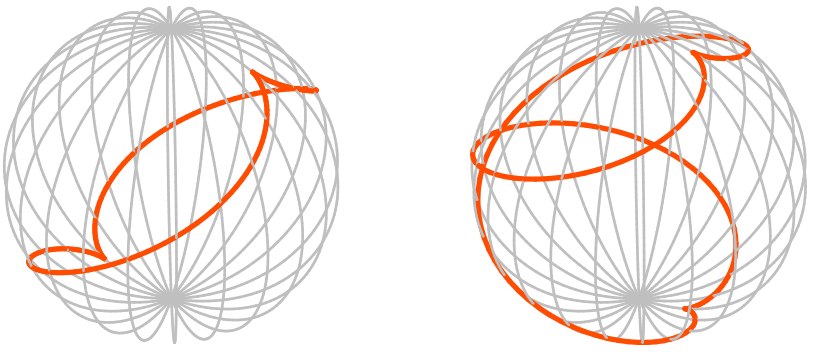


Figure 8.12: Spherical nephroid (left) and cardioid (right)

8.6 Loxodromes on Spheres

A ***meridian*** on a sphere is a great circle that passes through the north and south poles. A ***parallel*** on a sphere is a circle parallel to the equator. More generally:

Definition 8.25. *A **spherical loxodrome** or **rhumb line** is a curve on a sphere which meets each meridian of the sphere at the same angle, which we call the **pitch** of the loxodrome.*

A loxodrome has the north and south poles as asymptotic points. In the early years of terrestrial navigation, many sailors thought that a loxodrome was the same as a great circle, but Nuñez[2] distinguished the two.

To explain how to obtain the parametrization of a spherical loxodrome, we need the the so-called ***stereographic map*** $\boldsymbol{\Upsilon}\colon \mathbb{R}^2 \to S^2(1)$, defined by

$$\boldsymbol{\Upsilon}(p_1, p_2) = \frac{(2p_1,\, 2p_2,\, p_1^2 + p_2^2 - 1)}{p_1^2 + p_2^2 + 1}.$$

In fact, $\boldsymbol{\Upsilon}$ is the inverse of stereographic projection stereo, that we shall define in Chapter 22 using complex numbers (see page 730). In the meantime, the reader may check that the straight line joining (p_1, p_2) and the "north pole" $\mathbf{n} = (0, 0, 1)$ intersects the sphere in $\boldsymbol{\Upsilon}(p_1, p_2)$. It is easy to see that $\boldsymbol{\Upsilon}$ is differentiable and that $\|\boldsymbol{\Upsilon}(\mathbf{p})\| = 1$ for all $\mathbf{p} \in \mathbb{R}^2$.

Lemma 8.26. *The map $\boldsymbol{\Upsilon}$ has the following properties*:

(i) $\boldsymbol{\Upsilon}$ *maps circles in the plane onto circles on the sphere.*

(ii) $\boldsymbol{\Upsilon}$ *maps straight lines in the plane onto circles on the sphere that pass through* $(0, 0, 1)$.

(iii) $\boldsymbol{\Upsilon}$ *maps straight lines through the origin onto meridians.*

(iv) $\boldsymbol{\Upsilon}$ *preserves angles.*

Proof. Any circle or line in the plane is given implicitly by an equation of the form

$$a(x^2 + y^2) + bx + cy + d = 0. \tag{8.27}$$

Let $W = 1 + x^2 + y^2$, $X = 2x/W$, $Y = 2y/W$ and $Z = 1 - 2/W$. Under $\boldsymbol{\Upsilon}$, the equation (8.27) is transformed into

$$bX + cY + (a - d)Z + (d + a) = 0, \tag{8.28}$$

[2]Pedro Nuñez Salaciense (1502–1578). The first great Portuguese mathematician. Professor at Coimbra.

which is the equation of a plane in $\mathbb{R}^3$. This plane meets the sphere $S^2(1)$ in a great circle. This proves (i).

In the case of a straight line in the plane, $a = 0$ in (8.27). From (8.28), we see that $(0, 0, 1)$ is on the image curve, proving (ii). If in addition the straight line passes through the origin, then $a = d = 0$ in (8.27). Then from (8.28) we see that the plane containing the image curve also contains the Z-axis. Hence the curve is a meridian, proving (iii).

Statement (iv) will be proved in Lemma 12.14 on page 371. ▮

Lemma 8.27. *A spherical loxodrome is the image under the stereographic map of a logarithmic spiral.*

Proof. Lemma 1.29, page 24, implies that a logarithmic spiral meets every radial line from the origin at the same angle. The stereographic map $\boldsymbol{\Upsilon}$ maps each of these radial lines into a meridian of the sphere. Since $\boldsymbol{\Upsilon}$ preserves angles, it must map each logarithmic spiral onto a spherical loxodrome. ▮

This leads us to the definition

$$\mathsf{sphericalloxodrome}[a, b](t) = \left(\frac{2a\,e^{bt}\cos t}{1 + a^2 e^{2bt}}, \frac{2a\,e^{bt}\sin t}{1 + a^2 e^{2bt}}, \frac{a^2 e^{2bt} - 1}{1 + a^2 e^{2bt}} \right),$$

a particular case of which can be seen in Figure 8.13.

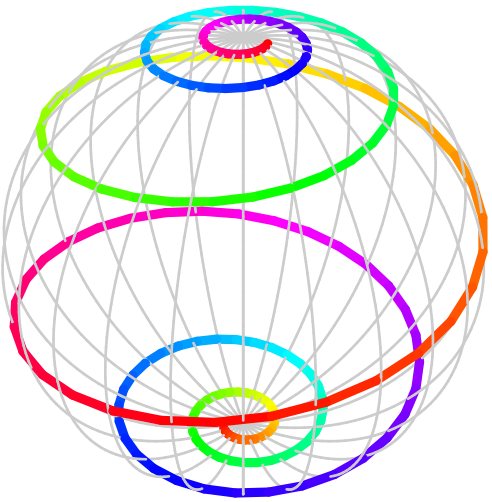


Figure 8.13: A spherical loxodrome of pitch 0.15

8.7 Exercises

1. Fill in the details of the proof of Theorem 8.3.

2. What is the evolute of Viviani's curve? No calculation is necessary!

3. Show that a helix has constant slope with respect to $(0, 0, 1)$.

4. Show that any curve in the xy-plane has zero slope with respect to $(0, 0, 1)$.

M **5.** Consider a space curve $\boldsymbol{\gamma}$ that projects to a curve $\boldsymbol{\alpha}$ in the xy-plane so that the third component of $\boldsymbol{\gamma}$ equals $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$, as explained in Section 7.5. Plot the result when $\boldsymbol{\gamma}$ is a figure eight, cardioid, astroid, cycloid.

M **6.** Another twisted generalization of the plane curve $\boldsymbol{\alpha}$ is the space curve

$$\mathsf{writhe}[\boldsymbol{\alpha}](t) = \big(a_1(t),\ a_2(t)\cos t,\ a_2(t)\sin t\big)$$

where $\boldsymbol{\alpha}(t) = (a_1(t), a_2(t))$. Compute the curvature and torsion of the curve $\mathsf{writhe}[\mathsf{circle}[a]]$. Verify that $\mathsf{writhe}[\mathsf{circle}[a]]$ and $\mathsf{viviani}[\frac{a}{2}](2t+\pi)$ have the same curvature and torsion. Conclude that there is a Euclidean motion that takes one curve into the other. Then plot the curves.

M **7.** Compute the curvature and torsion of $\mathsf{writhe}[\mathsf{lemniscate}[1]]$, and plot the curve.

8. Let $\boldsymbol{\theta}\colon (a, b) \to \mathbb{R}$ be an arbitrary differentiable function and define a curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^3$ by

$$\boldsymbol{\alpha}(t) = \left(\int_a^t \cos\boldsymbol{\theta}(u)\,du,\ \int_a^t \sin\boldsymbol{\theta}(u)\,du,\ ct\right),$$

where c is a constant. Compute the curvature and torsion of $\boldsymbol{\alpha}$ and show that $\boldsymbol{\alpha}$ has constant slope.

M **9.** Compute the curvature and torsion of $\mathsf{sphericalhelix}[a, b]$. Check that the ratio of the curvature to torsion is constant and that $\mathsf{sphericalhelix}[a, b]$ is a curve on a sphere. Plot several spherical helices.

10. A plane curve resembling a tear drop is defined by

$$\mathsf{teardrop}[a, b, c, d, n](t) = \big(a\sin t,\ d(b\sin t + c\sin 2t)^n\big).$$

Find the order of contact between $\mathsf{teardrop}[8, 1, 2, 1, n]$ and the x-axis at the origin $(0, 0)$.

11. Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^n$ be a unit-speed curve, let $a < s_0 < b$ and let $\mathbf{v} \in \mathbb{R}^n$. Prove that the hyperplane $\{\, \mathbf{q} \mid (\mathbf{q} - \boldsymbol{\beta}(s_0)) \cdot \mathbf{v} = 0 \,\}$ has at least order 2 contact with $\boldsymbol{\beta}$ at $\boldsymbol{\beta}(s_0)$ if and only if $\mathbf{v}$ is perpendicular to $\boldsymbol{\beta}'(s_0)$ and the curvature of $\boldsymbol{\beta}$ vanishes at the point of contact $\boldsymbol{\beta}(s_0)$. (This result is a strengthening of Lemma 8.8.)

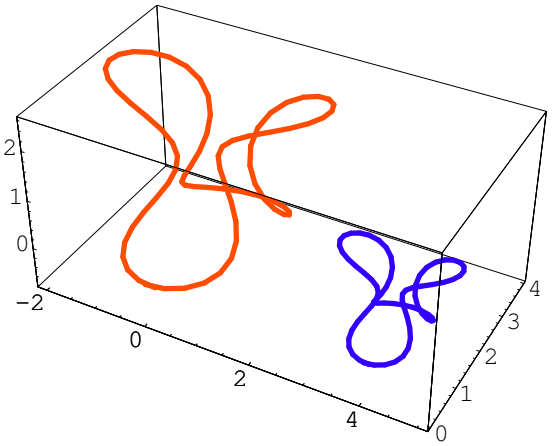


Figure 8.14: Space curves with $\boldsymbol{\kappa}(s) = 1$, $\boldsymbol{\tau}(s) = \sin\frac{s}{2}$ (large) and $\boldsymbol{\kappa}(s) = 2$, $\boldsymbol{\tau}(s) = 2\sin s$ (small)

12. Suppose that a space curve $\mathscr{C}$ has arc length s, curvature $\boldsymbol{\kappa}(s) > 0$ and torsion $\boldsymbol{\tau}(s)$. Let $c > 0$ be a constant, and set

$$\tilde{\boldsymbol{\kappa}}(s) = c\boldsymbol{\kappa}(cs), \quad \tilde{\boldsymbol{\tau}}(s) = c\boldsymbol{\tau}(cs).$$

Prove that a space curve with arc length s, curvature $\tilde{\boldsymbol{\kappa}}(s)$ and torsion $\tilde{\boldsymbol{\tau}}(s)$ is a rescaled version of $\mathscr{C}$. An example of this phenomenon can be seen in Figure 8.14.

Notebook 8

■ N8.0

This notebook incorporates the program to plot a space curve with assigned curvature and torsion. It shows how to plot spherical helices, and includes some animations of generalized helices spinning on spheres.

We begin by duplicating the definitions of the curvature and torsion, from Notebook 7.

```
κ[α_][t_]:=
  Module[{tt,a,b},
  a= L[Cross[α'[tt], α''[tt]]];
  b= L[α'[tt]]^3; a/b /. tt->t
]
```

```
τ[α_][t_]:=
  Module[{tt,a,C},
  a= Det[{α'[tt], α''[tt], α'''[tt]}];
  C= Cross[α'[tt], α''[tt]]; a/(C.C)/.tt->t
]
```

In this notebook, we shall need the packages

```
<<Graphics`Arrow`
<<Graphics`Shapes`
<<Graphics`Animation`
```

The definition of the Frenet frame, also from Notebook 7, is loaded next together with commands for the space curves helix and viviani.

■ N8.0.1

■ N8.1

Let us begin with a curve lying in the xy-plane, not strictly speaking defined at zero:

```
α[t_]:= {t, 5Exp[-1/t^2], 0}
α[1]//N
α[0]
```

Mathematica cannot evaluate the offending value of α, but it can compute the limit as `t` tends to zero.

```
Limit[α[t], t->0]
```

Curvature and torsion are given by

```
κ[α][t]//PowerExpand
κ[α][.01]
τ[α][t]
```

The cell for Figure 8.1 uses the following commands, which together provide a fancy way of permuting the components of α so as to define a similar curve `f[1]` in the xz-plane:

```
Reverse[α[t]]
RotateRight[%]
```

The `Which` command is then used to change the definition again for `t>0` to obtain a curve `f[2]` that no longer lies in the plane. Both curves are then ready for plotting. The presence of the `Off` command prevents an overflow message.

```
Module[{α,f,ff,z},
  α[t_]:= {t, 5Exp[-1/t^2], 0};
  f[1][t_]= RotateRight[Reverse[α[t]]];
  f[2][t_]= Which[t<0, α[t], t>0, f[1][t]];
  ff[k_]= highlight[f[k]];
  Off[General::unfl];
  z[k_]:= ParametricPlot3D[ff[k][t], {t,-1,1},
    PlotRange->{{-1,1},{-.01,1},{-.01,1}},
    Axes->False, ViewPoint->{-3.1,-1.2,.5}];
  S @ GraphicsArray[Array[z,2]]
]
```

Figure 8.1

■ N8.2

In this section, we introduce a program intrinsic3d that solves numerically a system of differential equations to determine the space curve with prescribed curvature and (positive) torsion. It is analogous to the one called intrinsic used in Notebook 5 for plane curves, but it is somewhat more complicated to set up the equations.

We begin by defining two commands V, Eq that enable one to quickly set up lists of variables and equations. Their definitions appear *inside* the relevant Module below, but we have placed and protected them externally for instructional purposes.

```
Unprotect[V, Eq]
  V[f_]:= Array[f,3]
  Eq[f_,g_]:= Thread[Equal[V[f],V[g]]]
Protect[V, Eq]
```

The first command V allows us to list three values of a given function f:

```
V[a]
f[k_]:=k^2
V[f]
V[#^2 &]
```

Observe that f may be treated as an explicit function in the definition of V, and in practice it is convenient to define it using #. We can then combine lists using Join, and get tables of equations using Eq:

```
Join[V[a],V[b]]
Eq[a, f]
Eq[a[#]'[t]&, p[#]&]
```

In the following program, initial values are given between the double asterisks for easy location. *Mathematica* uses (* and *) to open and close a comment that is not evaluated in the input, so one can use the sequence (**) as an innocuous marker.

```
intrinsic3d[κ_,τ_,c_,{min_:-10,max_:10}][u_]:=
Module[{V,Eq,x,t,n,b,s},
(**)    xa={0,0,0};  ta={1,0,0};  na={0,1,0};    (**)
  V[f_]:= Array[f,3];
  Eq[f_,g_]:= Thread[Equal[V[f], V[g]]];
  SF={Eq[x[#]'[s]&, t[#][s]&],
        Eq[t[#]'[s]&, κ[s] n[#][s]&],
        Eq[n[#]'[s]&, -κ[s] t[#][s] + τ[s] b[#][s]&],
        Eq[b[#]'[s]&, -τ[s] n[#][s]&]};
  IC={Eq[x[#][c]&, xa[[#]]&],
        Eq[t[#][c]&, ta[[#]]&],
        Eq[n[#][c]&, na[[#]]&],
        Eq[b[#][c]&, Cross[ta,na][[#]]&]};
  eqs= Flatten[{SF, IC}] ;
  vars= Join[V[x],V[t],V[n],V[b]];
  V[x[#][u]&] /. NDSolve[eqs, vars, {s,min,max}][[1]]
]
```

The plotting command is separated from the above program, for clarity:

```
plotintrinsic3d[κ_,τ_, a_,{min_:-10,max_:10},opts___]:=
   ParametricPlot3D[highlight[intrinsic3d[κ,τ,a,{min,max}]][u]
   //Evaluate, {u,min,max}, opts]
```

Here are examples to verify that zero torsion gives a plane curve, and that constant curvature and torsion give a helix.

```
plotintrinsic3d[10/(#+11)&, 0&, 0, {}, PlotPoints->200];
plotintrinsic3d[1&, 1&, 0, {-5,5}, PlotPoints->200];
```

The next example is less familiar. Three more chapter figures are included in the electronic subsection N8.2.1.

```
S @ plotintrinsic3d[#&, 1&, 0, {}, PlotPoints->200, Axes->None]
```

Figure 8.2

■ N8.2.1

■ N8.3

The Evaluate command is essential when plotting a Table of functions, in this case generalized parabolas:

```
ta:=Table[{t, t^n}, {n,1,5}]
S @ ParametricPlot[ta//Evaluate, {t,-1.1,1.1},
  PlotStyle->Table[Hue[n/3], {n,5}], AspectRatio->1,
  ImageSize->350]
```

Figure 8.6

The following Module paraphrases the definition of the evolute of a space curve. Two cases then illustrate that the evolute of the evolute of a helix is the original curve.

```
evolute3d[γ_][t_]:=
Module[{u, A, B},
  A=1/κ[γ][u];
  B= κ[γ]'[u]/(L[γ'[u]] κ[γ][u]^2 τ[γ][u]);
  γ[u] +A normal[γ][u] - B binormal[γ][u] /.u->t
]
```

```
γ:= helix[a,b]; γ[t]
ϵ:= evolute3d[γ]
ϵ[t]//Simplify
ξ:= evolute3d[ϵ]
ξ[t]//Simplify
```

```
α1:= helix[1,1]
αα1:= highlight[α1]
ϵ1:= evolute3d[α1]
ϵϵ1:= highlight[ϵ1, .5]
α2[t_]:= {10,0,8} + helix[3/2,1/2][t]
αα2:= highlight[α2]
ϵ2:= evolute3d[α2]
ϵϵ2:= highlight[ϵ2, .5]
S @ ParametricPlot3D[{αα1[t],ϵϵ1[t],αα2[t],ϵϵ2[t]}//Evaluate,
  {t,0,6π}, Axes->None]
```

Figure 8.7

■ N8.3.1

■ N8.4

The sphere of radius a can be parametrized in the same way as the torus in N7.6. Many similar such surfaces will be discussed in Notebook 10.

```
sphere[a_][u_,v_]:= a{Cos[u]Cos[v], Sin[u]Cos[v], Sin[v]}
```

The next program is a variant of the built-in WireFrame command.

```
meridians[gl_]:=
Module[{a,t},
  a[u_,v_]= Append[sphere[1][u,v], GrayLevel[gl]];
  t[v_]= Table[a[u,v], {u,0,2π,π/12}];
  ParametricPlot3D[t[v]//Evaluate, {v,-π/2,π/2}]
]
```

It outlines a unit sphere consisting of 24 meridians, that can then be added as a secondary plot to emphasize that a particular curve lies on a sphere. The argument in its definition controls the GrayLevel. Add S@ to display the graphics.

```
meridians[.5]
```

In the next picture, PlotPoints is used to construct a plane without the usual mesh.

```
Clear[v]
r:= 1
α[t_]:= {3Cos[t], 2Sin[t], 0}/4
αα:= highlight[α, .7]
δ[t_]:= {3Cos[t], 2Sin[t], Sqrt[16-9Cos[t]^2-4Sin[t]^2]}/4
δδ:= highlight[δ]
ParametricPlot3D[{αα[t], δδ[t]}//Evaluate, {t,0,2π}];
ParametricPlot3D[{u,v,0}, {u,-r,r}, {v,-r,r}, Shading->False,
  PlotPoints->2];
S[%, %%, meridians[.7], Boxed->False, Ticks->None,
  ViewPoint->{1,-3,1}]
```

Figure 8.9

```
S @ Plot[Evaluate[{κ[δ]'[t], τ[δ][t]}],{t,0,2π}]
```

Figure 8.10

```
v:= viviani[1]
Plot[Evaluate[{κ[v]'[t], τ[v][t]}], {t,-2π,2π}]
Clear[v]
```

■ N8.5

The meridians program of the preceding N8.4 is now used as part of a spherical plotting command. The latter adopts a default value for the GrayLevel.

```
splot[σ_, {a_,b_}, gl_:3/4]:= Module[{σσ,z},
  σσ= highlight[σ];
  z= ParametricPlot3D[σσ[t], {t,a,b},
    PlotPoints->200, Axes->None];
  Show[z, meridians[gl], Boxed->False]
]
```

Its input σ must be a curve whose trace lies on the unit sphere.

```
sphericalhelix[a_,b_][t_]:={(a+b)Cos[t] - b Cos[(a+b)t/b],
  (a+b)Sin[t] - b Sin[(a+b)t/b], 2Sqrt[a b+b^2]Cos[a t/(2b)]}
```

```
L[sphericalhelix[a,b][t]]//PowerExpand
```

```
Module[{s,z},
  σθ[t_]:= 1.01 sphericalhelix[2/3,1/6][t];
  z= splot[σθ, {0,2π}, .9];
  s[2]= Show[z, ViewPoint->{0,2,0}];
  s[1]= Show[z, ViewPoint->{0,0,2}];
  s[3]= Show[z, ViewPoint->{2,0,0}];
  s[4]= Show[z, ViewPoint->{2,2,0}];
  S @ GraphicsArray[Array[s,4], GraphicsSpacing->-.1]
]
```

Figure 8.11

```
sneph= splot[sphericalhelix[1/2,1/4], {0,2π}];
scard= splot[sphericalhelix[1/3,1/3], {0,4π}];
S @ GraphicsArray[{sneph, scard}, GraphicsSpacing->-.1]
```

Figure 8.12

One of these objects is rotated using the SpinShow command from the Animations package, loaded in N8.0. The number of frames can be increased at will.

```
SpinShow[scard, Frames->2, SpinOrigin->{0,0,0}]
```

■ N8.6

In this section, we shall work with the logarithmic spiral, loaded in N8.0.1. It will be inserted into the following parametrization of the unit sphere:

```
stereographicsphere[u_,v_]:= {2u, 2v, u^2+v^2-1}/(u^2+v^2+1)
```

```
stereographicsphere[u,v]//L
```

The next command now enables us to "lift" a plane curve up onto the sphere. The instruction @@ effectively strips away the curly brackets containing the components of the vector α[t] which get inserted into surf.

```
curveonsurf[surf_,α_][t_]:= surf@@α[t]
```

The result of applying this to the logarithmic spiral gives us the spherical loxodrome.

```
curveonsurf[stereographicsphere, logspiral[a,b]][t]//Simplify
L[logspiral[a,b][t]]//PowerExpand
L[stereographicsphere[u_,v_]]
```

```
sphericalloxodrome[a_,b_][t_]:=
{2a Exp[b t]Cos[t], 2a Exp[b t]Sin[t], -1+a^2 Exp[2b t]} /
  (1+a^2 Exp[2b t])
```

This is plotted using the command splot from N.8.5:

```
σλ:= sphericalloxodrome[1,.15]
slox= splot[σλ, {-20,20}, Boxed->False];
```

■ N8.6.1

■ N8.7

We illustrate two curves plotted from closely related curvature functions:

```
Module[{z,y,yy},
  z[1]= plotintrinsic3d[1&,  Sin[#/2]&, 0,{0,8π}];
  y= intrinsic3d[2&, 2Sin[#]&, 0, {0, 4π}][#]+{4,0,0}&;
  yy= highlight[y,.7];
  z[2]= ParametricPlot3D[yy[u]//Evaluate, {u,0,4π}];
  S @ Array[z,2]
]
```

Figure 8.14

These objects can be rotated using the SpinShow command, as in N8.5. Increase the number of frames below to 20.

```
plotintrinsic3d[1&, Sin[#/2]&, 0, {0,8π}, Axes->None,
  Boxed->False]
SpinShow[%, Frames->2]
```

Chapter 9

Calculus on Euclidean Space

More than the algebraic operations on $\mathbb{R}^n$ are needed for differential geometry. We need to know how to *differentiate* various geometric objects, and we need to know the relationship between differentiation and the algebraic operations. This chapter establishes the theoretical foundations of the theory of differentiation in $\mathbb{R}^n$ that will be the basis for subsequent chapters.

In Section 9.1, we define the notion of *tangent vector* to $\mathbb{R}^n$. The interpretation of a tangent vector to $\mathbb{R}^n$ as a directional derivative is given in Section 9.2. This allows us to define tangent maps as the derivatives of differentiable maps between Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ in Section 9.3.

It is important to understand how tangent vectors vary from point to point, so we begin the study of *vector fields* on $\mathbb{R}^n$ in Section 9.4. Derivatives of vector fields on $\mathbb{R}^n$ with respect to tangent vectors are considered in Section 9.5, in which we also prove some elementary facts about the linear map on vector fields induced by a diffeomorphism. The classical notions of gradient, divergence, and Laplacian are defined in Sections 9.4 and 9.5. Finally, in Section 9.6 we return to our study of curves, using tangent vectors to provide a different perspective and make more rigorous some previous definitions.

9.1 Tangent Vectors to $\mathbb{R}^n$

In Section 1.1, we defined the notion of a *vector* $\mathbf{v}$ in $\mathbb{R}^n$. So far, we have generally considered vectors in $\mathbb{R}^n$ to be *points* in $\mathbb{R}^n$; however, a vector $\mathbf{v}$ can be regarded as an arrow or *displacement* which starts at the origin and ends at the point $\mathbf{v}$. From this point of view, and with the theory of surfaces in mind, it makes sense to study displacements that start at points other than the origin, and to retain the starting point in the definition of the tangent vector.

Definition 9.1. *A* ***tangent vector*** $\mathbf{v_p}$ *to Euclidean space* $\mathbb{R}^n$ *consists of a pair of elements* $\mathbf{v}, \mathbf{p}$ *of* $\mathbb{R}^n$; $\mathbf{v}$ *is called the* ***vector part*** *and* $\mathbf{p}$ *is called the* ***point of application*** *of* $\mathbf{v_p}$.

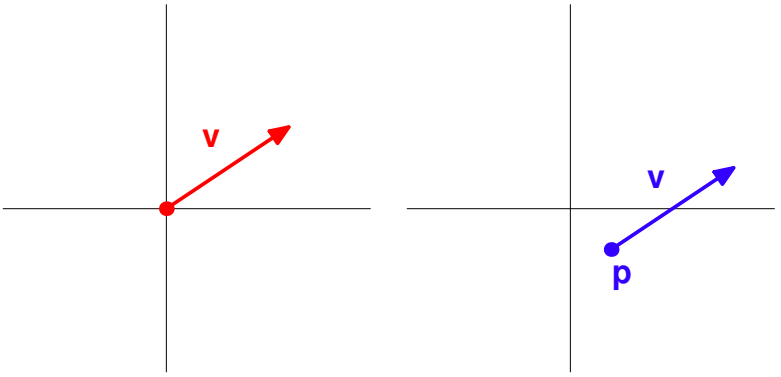


Figure 9.1: The same vector applied at different points

The tangent vector can be identified with the arrow from $\mathbf{p}$ to $\mathbf{p} + \mathbf{v}$. Furthermore, two tangent vectors $\mathbf{v_p}$ and $\mathbf{w_q}$ are declared to be equal if and only if *both* their vector parts and points of application coincide: $\mathbf{v} = \mathbf{w}$ and $\mathbf{p} = \mathbf{q}$.

Definition 9.2. *Let* $\mathbf{p} \in \mathbb{R}^n$. *The* ***tangent space*** *of* $\mathbb{R}^n$ *at* $\mathbf{p}$ *is the set*

$$\mathbb{R}^n_{\mathbf{p}} = \{\, \mathbf{v_p} \mid \mathbf{v} \in \mathbb{R}^n \,\}.$$

The tangent space $\mathbb{R}^n_{\mathbf{p}}$ is a carbon copy of the vector space $\mathbb{R}^n$; in fact, there is a canonical isomorphism between $\mathbb{R}^n_{\mathbf{p}}$ and $\mathbb{R}^n$ given by $\mathbf{v_p} \mapsto \mathbf{v}$. This isomorphism allows us to turn $\mathbb{R}^n_{\mathbf{p}}$ into a vector space; we can add elements in $\mathbb{R}^n_{\mathbf{p}}$ and multiply by scalars in the obvious way, retaining the common point of application:

$$\begin{aligned} \mathbf{v_p} + \mathbf{w_p} &= (\mathbf{v} + \mathbf{w})_{\mathbf{p}}, \\ \lambda \mathbf{v_p} &= (\lambda \mathbf{v})_{\mathbf{p}}. \end{aligned}$$

To sum up, $\mathbb{R}^n$ has a tangent space $\mathbb{R}^n_{\mathbf{p}}$ attached to each of its points, and the tangent space looks like $\mathbb{R}^n$ itself. For this reason, it is possible to transfer the dot product and norm to each tangent space. We simply put

$$\mathbf{v_p} \cdot \mathbf{w_p} = \mathbf{v} \cdot \mathbf{w}$$

for $\mathbf{v_p}, \mathbf{w_p} \in \mathbb{R}^n_{\mathbf{p}}$. It follows that there is a Cauchy–Schwarz inequality for tangent vectors, namely,

$$|\mathbf{v_p} \cdot \mathbf{w_p}| \leqslant \|\mathbf{v_p}\| \, \|\mathbf{w_p}\|$$

for $\mathbf{v_p}, \mathbf{w_p} \in \mathbb{R}^n_{\mathbf{p}}$. Angles between tangent vectors can now be defined in exactly the same way as angles between ordinary vectors (see page 3).

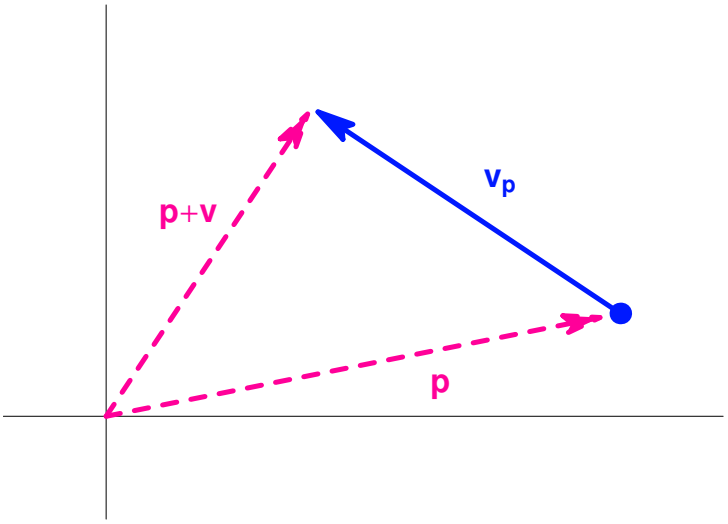


Figure 9.2: A tangent vector $\mathbf{v_p}$ at $\mathbf{p}$

To finish this introductory section, consider tangent vectors to $\mathbb{R}^2$ and $\mathbb{R}^3$. The complex structure of $\mathbb{R}^2$ that we defined on page 3 can be naturally extended to tangent vectors by putting

$$J\mathbf{v_p} = (J\mathbf{v})_{\mathbf{p}}$$

for a tangent vector $\mathbf{v_p}$ to $\mathbb{R}^2$. It is easy to see that $J\mathbf{v_p}$ is also a tangent vector to $\mathbb{R}^2$ and that $-J^2$ is the identity map on tangent vectors. Furthermore, $J\mathbf{v_p} \cdot J\mathbf{w_p} = \mathbf{v_p} \cdot \mathbf{w_p}$ for $\mathbf{v_p}, \mathbf{w_p} \in \mathbb{R}^2_{\mathbf{p}}$. Similarly, on $\mathbb{R}^3$ we extend the vector cross product to tangent vectors via

$$\mathbf{v_p} \times \mathbf{w_p} = (\mathbf{v} \times \mathbf{w})_{\mathbf{p}},$$

for $\mathbf{v_p}, \mathbf{w_p} \in \mathbb{R}^3_{\mathbf{p}}$. The usual identities, such as Lagrange's

$$\|\mathbf{v_p} \times \mathbf{w_p}\|^2 = \|\mathbf{v_p}\|^2\|\mathbf{w_p}\|^2 - (\mathbf{v_p} \cdot \mathbf{w_p})^2$$

(see page 193), will now hold for tangent vectors.

9.2 Tangent Vectors as Directional Derivatives

Frequently, it is important to know how a real-valued function $f\colon \mathbb{R}^n \to \mathbb{R}$ varies in different directions. For example, if we write the values of f as $f(u_1, \ldots, u_n)$,

the partial derivative $\partial f/\partial u_i$ measures how much f varies in the u_i direction, for $i = 1, \ldots, n$. However, it is possible to measure the variation of f in other directions as well. To measure the variation of f at $\mathbf{p} \in \mathbb{R}^n$ along the straight line $t \mapsto \mathbf{p} + t\mathbf{v}$, we need to make the tangent vector $\mathbf{v_p}$ operate on functions in a suitable way.

Throughout this section, we will use the words "f is a differentiable function" or "f is differentiable" to mean that f has partial derivatives of all orders on its domain. For details on why this is a reasonable definition, see the introduction to Chapter 24.

Definition 9.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and let $\mathbf{v_p}$ be a tangent vector to $\mathbb{R}^n$ at $\mathbf{p} \in \mathbb{R}^n$. We put*

$$\mathbf{v_p}[f] = \frac{d}{dt} f(\mathbf{p} + t\mathbf{v})\bigg|_{t=0}. \tag{9.1}$$

In elementary calculus, (9.1) goes under the name of *directional derivative in the $\mathbf{v_p}$ direction* (at least if $\|\mathbf{v_p}\| = 1$), but we prefer to consider (9.1) as defining how a tangent vector operates on functions.

A more explicit way to write (9.1) is

$$\mathbf{v_p}[f] = \lim_{t \to 0} \frac{f(\mathbf{p} + t\mathbf{v}) - f(\mathbf{p})}{t}.$$

For example, let $n = 2$, $f(x, y) = \sin(xy)$, $\mathbf{p} = (1, 2)$ and $\mathbf{v} = (2, 3)$. Then

$$\begin{aligned} f(\mathbf{p} + t\mathbf{v}) = f\big((1,2) + t(2,3)\big) &= f(2t+1, 3t+2) \\ &= \sin\big((2t+1)(3t+2)\big) \\ &= \sin\big(6t^2 + 7t + 2\big), \end{aligned}$$

so that

$$\frac{d}{dt} f(\mathbf{p} + t\mathbf{v}) = (12t + 7)\cos\big(6t^2 + 7t + 2\big).$$

Thus

$$\mathbf{v_p}[f] = (12t + 7)\cos\big(6t^2 + 7t + 2\big)\Big|_{t=0} = 7\cos 2 = -2.913.$$

Although it is possible to compute $\mathbf{v_p}[f]$ directly from the definition, it is usually easier to use a general formula, which we now derive.

Lemma 9.4. *Let $\mathbf{v_p} = (v_1, \ldots, v_n)_\mathbf{p}$ be a tangent vector to $\mathbb{R}^n$, and $f\colon \mathbb{R}^n \to \mathbb{R}$ a differentiable function. Then*

$$\mathbf{v_p}[f] = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial u_j}(\mathbf{p}). \tag{9.2}$$

Proof. Let $\mathbf{p} = (p_1, \dots, p_n)$, so that $\mathbf{p} + t\mathbf{v} = (p_1 + tv_1, \dots, p_n + tv_n)$. The chain rule implies that

$$\begin{aligned}\frac{d}{dt}\big(f(\mathbf{p}+t\mathbf{v})\big) &= \frac{d}{dt} f(p_1 + tv_1, \dots, p_n + tv_n) \\ &= \sum_{j=1}^{n} \frac{\partial f}{\partial u_j}(p_1 + tv_1, \dots, p_n + tv_n) \frac{d(p_j + tv_j)}{dt} \\ &= \sum_{j=1}^{n} \frac{\partial f}{\partial u_j}(p_1 + tv_1, \dots, p_n + tv_n) v_j.\end{aligned}$$

When we put $t = 0$ in this equation, we get (9.2). ∎

Next, we list some algebraic properties of the operation (9.1).

Lemma 9.5. *Let $f, g\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable functions, $\mathbf{v_p}, \mathbf{w_p}$ tangent vectors in $\mathbb{R}^n_{\mathbf{p}}$, and a, b real numbers. Then*

$$(a\mathbf{v_p} + b\mathbf{w_p})[f] = a\mathbf{v_p}[f] + b\mathbf{w_p}[f], \tag{9.3}$$

$$\mathbf{v_p}[af + bg] = a\mathbf{v_p}[f] + b\mathbf{v_p}[g], \tag{9.4}$$

$$\mathbf{v_p}[fg] = f(\mathbf{p})\mathbf{v_p}[g] + g(\mathbf{p})\mathbf{v_p}[f]. \tag{9.5}$$

Proof. For example, to prove (9.5):

$$\begin{aligned}\mathbf{v_p}[fg] &= \frac{d}{dt}(fg)(\mathbf{p}+t\mathbf{v})\bigg|_{t=0} \\ &= \left(f(\mathbf{p}+t\mathbf{v})\frac{d}{dt}g(\mathbf{p}+t\mathbf{v}) + g(\mathbf{p}+t\mathbf{v})\frac{d}{dt}f(\mathbf{p}+t\mathbf{v})\right)\bigg|_{t=0} \\ &= f(\mathbf{p})\mathbf{v_p}[g] + g(\mathbf{p})\mathbf{v_p}[f]. \ ∎\end{aligned}$$

There is a chain rule for tangent vectors:

Lemma 9.6. *Let $g_1, \dots, g_k\colon \mathbb{R}^n \to \mathbb{R}$ and $h\colon \mathbb{R}^k \to \mathbb{R}$ be differentiable functions and write $f = h \circ (g_1, \dots, g_k)$ for the composed function $\mathbb{R}^n \to \mathbb{R}$. Let $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ where $\mathbf{p} \in \mathbb{R}^n$. Then*

$$\mathbf{v_p}[f] = \sum_{j=1}^{k} \frac{\partial h}{\partial u_j}\big(g_1(\mathbf{p}), \dots, g_k(\mathbf{p})\big)\mathbf{v_p}[g_j].$$

Proof. Let $\mathbf{v_p} = (v_1, \dots, v_n)_{\mathbf{p}}$. We use the ordinary chain rule to compute

$$\begin{aligned}\mathbf{v_p}[f] = \sum_{i=1}^{n} v_i \frac{\partial f}{\partial u_i}(\mathbf{p}) &= \sum_{i=1}^{n}\sum_{j=1}^{k} v_i \frac{\partial h}{\partial u_j}\big(g_1(\mathbf{p}), \dots, g_k(\mathbf{p})\big)\frac{\partial g_j}{\partial u_i}(\mathbf{p}) \\ &= \sum_{j=1}^{k} \frac{\partial h}{\partial u_j}\big(g_1(\mathbf{p}), \dots, g_k(\mathbf{p})\big)\mathbf{v_p}[g_j]. \ ∎\end{aligned}$$

9.3 Tangent Maps or Differentials

We shall need to deal with functions $F\colon \mathbb{R}^n \to \mathbb{R}^m$ other than linear maps. For any such function, we can write

$$F(\mathbf{p}) = \big(f_1(\mathbf{p}), \ldots, f_m(\mathbf{p})\big), \tag{9.6}$$

where $f_j\colon \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, m$, and $\mathbf{p} \in \mathbb{R}^n$. It will be convenient to use the notation $F = (f_1, \ldots, f_m)$ as a shorthand for (9.6).

Clearly, (9.6) also makes sense for a function $F\colon \mathcal{U} \to \mathbb{R}^m$ if $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. Recall that a subset $\mathcal{U}$ of $\mathbb{R}^n$ is said to be ***open*** if for any $\mathbf{p} \in \mathcal{U}$ it is possible to "squeeze" a small ball with center $\mathbf{p}$ into $\mathcal{U}$, meaning that there exists $\varepsilon > 0$ such that $\{\mathbf{q} \in \mathbb{R}^m \mid \|\mathbf{q} - \mathbf{p}\| < \varepsilon\} \subseteq \mathcal{U}$.

Definition 9.7. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open. A function $F\colon \mathcal{U} \to \mathbb{R}^m$ is said to be* ***differentiable*** *provided that each f_j is differentiable. More generally, if $\mathcal{A}$ is any subset of $\mathbb{R}^n$, we say that $F\colon \mathcal{A} \to \mathbb{R}^m$ is differentiable if there exists an open set $\mathcal{U}$ containing $\mathcal{A}$ and a differentiable map $\widetilde{F}\colon \mathcal{U} \to \mathbb{R}^m$ such that the restriction of $\widetilde{F}$ to $\mathcal{A}$ is F. A* ***diffeomorphism*** *F between open subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathbb{R}^n$ is a differentiable map $F\colon \mathcal{U} \to \mathcal{V}$ possessing an inverse $F^{-1}\colon \mathcal{V} \to \mathcal{U}$ that is also differentiable.*

It is easy enough to prove that the composition of differentiable functions is differentiable:

Lemma 9.8. *Suppose that there are differentiable maps*

$$\begin{array}{ccccc} \mathcal{U} & \xrightarrow{F} & \mathcal{V} & \xrightarrow{G} & \mathbb{R}^n\,. \\ \cap & & \cap & & \\ \mathbb{R}^\ell & & \mathbb{R}^m & & \end{array}$$

Then the composition $G \circ F\colon \mathcal{U} \to \mathbb{R}^n$ is differentiable.

Any differentiable map $F\colon \mathcal{U} \to \mathbb{R}^m$ gives rise in a natural way to a linear map between tangent spaces. To establish this, first note that if $F\colon \mathcal{U} \to \mathbb{R}^m$ is differentiable, then so is $t \mapsto F(\mathbf{p} + t\mathbf{v})$ for $\mathbf{p} + t\mathbf{v} \in \mathcal{U}$, because it is the composition of differentiable functions. Therefore, the following makes sense:

Definition 9.9. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$, and let $\mathbf{p} \in \mathcal{U}$. For each tangent vector $\mathbf{v}_\mathbf{p}$ to $\mathbb{R}^n$ set*

$$F_{*\mathbf{p}}(\mathbf{v}_\mathbf{p}) = F(\mathbf{p} + t\mathbf{v})'(0).$$

Then $F_{*\mathbf{p}}\colon \mathbb{R}^n_\mathbf{p} \to \mathbb{R}^m_{F(\mathbf{p})}$ is called the ***tangent map*** *or* ***differential*** *of F at $\mathbf{p}$.*

Note that $F_{*\mathbf{p}}(\mathbf{v}_\mathbf{p})$ is the initial velocity of the curve $t \mapsto F(\mathbf{p} + t\mathbf{v})$.

One often abbreviates $F_{*\mathbf{p}}$ to F_* if the identity of the point in question is clear. For example, let $\mathbf{p}=(1,2,3)$, $\mathbf{v}=(1,0,-1)$ and $F(x,y,z) = (yz,\ zx,\ xy)$. Then

$$F(\mathbf{p}+t\mathbf{v}) = \big(6-2t,\ 3+2t-t^2,\ 2+2t\big),$$

so that

$$F_*(\mathbf{v}_\mathbf{p}) = \left(\frac{d}{dt}\big(6-2t,\ 3+2t-t^2,\ 2+2t\big)\Big|_{t=0}\right)_{F(\mathbf{p})} = (-2,2,2)_{(6,3,2)}.$$

The following lemma provides a useful way of computing tangent maps.

Lemma 9.10. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map defined on an open subset $\mathcal{U} \subseteq \mathbb{R}^n$, and write $F = (f_1, \dots, f_m)$. If $\mathbf{p} \in \mathcal{U}$ and $\mathbf{v}_\mathbf{p}$ is a tangent vector to $\mathbb{R}^n$ at $\mathbf{p}$, then*

$$F_*(\mathbf{v}_\mathbf{p}) = \big(\mathbf{v}_\mathbf{p}[f_1], \dots, \mathbf{v}_\mathbf{p}[f_m]\big)_{F(\mathbf{p})}. \tag{9.7}$$

Proof. Replace $\mathbf{p}$ by $\mathbf{p}+t\mathbf{v}$ in (9.6) and differentiate with respect to t:

$$F(\mathbf{p}+t\mathbf{v})' = \big(f_1(\mathbf{p}+t\mathbf{v})', \dots, f_m(\mathbf{p}+t\mathbf{v})'\big). \tag{9.8}$$

When we evaluate both sides of (9.8) at $t=0$ and use (9.1), we obtain (9.7). ∎

The next result is an easy consequence of (9.7) and (9.3):

Corollary 9.11. *If $F\colon \mathcal{U} \to \mathbb{R}^m$ is differentiable, then at each point $\mathbf{p} \in \mathcal{U}$ the tangent map $F_{*\mathbf{p}}\colon \mathbb{R}^n_\mathbf{p} \to \mathbb{R}^m_{F(\mathbf{p})}$ is* linear.

The tangent map of a differentiable map F can be thought of as the *best linear approximation* to F at a given point $\mathbf{p}$. We have already seen examples of tangent maps; for example, the velocity of a curve is a tangent map in disguise. To explain this, we let $1_\mathbf{p}$ be the tangent vector to $\mathbb{R}$ whose vector part is 1 and whose point of application is $\mathbf{p} \in \mathbb{R}$. This makes perfect logical sense, and it follows from the definitions that

$$1_\mathbf{p}[f] = \frac{d}{dt} f(\mathbf{p}+t1)\Big|_{t=0} = f'(\mathbf{p}) \tag{9.9}$$

for any differentiable function $f\colon \mathbb{R} \to \mathbb{R}$.

Lemma 9.12. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve. Then for $\mathbf{p} \in (a,b)$ we have*

$$\boldsymbol{\alpha}_*(1_\mathbf{p}) = \boldsymbol{\alpha}'(\mathbf{p}).$$

Proof. If we write $\boldsymbol{\alpha} = (a_1, \ldots, a_n)$, then from (9.9) and (9.7) we get

$$\boldsymbol{\alpha}_*(1_{\mathbf{p}}) = \big(1_{\mathbf{p}}[a_1], \ldots, 1_{\mathbf{p}}[a_n]\big) = \big(a_1'(\mathbf{p}), \ldots, a_n'(\mathbf{p})\big) = \boldsymbol{\alpha}'(\mathbf{p}). \blacksquare$$

If $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{U} \subseteq \mathbb{R}^n$ is a differentiable curve, and $F\colon \mathcal{U} \to \mathbb{R}^m$ a differentiable map, then $F \circ \boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^m$ is a differentiable curve by Lemma 9.8. It is reasonable to suspect that F_* maps the velocity of $\boldsymbol{\alpha}$ into the velocity of $F \circ \boldsymbol{\alpha}$. We show that this is indeed the case.

Lemma 9.13. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$, and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{U}$ be a curve. Then*

$$(F \circ \boldsymbol{\alpha})'(t) = F_*\big(\boldsymbol{\alpha}'(t)\big)$$

for $a < t < b$.

Proof. Write $F = (f_1, \ldots, f_m)$ and $\boldsymbol{\alpha} = (a_1, \ldots, a_n)$. Lemmas 9.4, 9.10 and the chain rule imply that

$$\begin{aligned} F_*\big(\boldsymbol{\alpha}'(t)\big) &= \big(\boldsymbol{\alpha}'(t)[f_1], \ldots, \boldsymbol{\alpha}'(t)[f_m]\big) \\ &= \left(\sum_{j=1}^n a_j'(t)\frac{\partial f_1}{\partial u_j}, \ldots, \sum_{j=1}^n a_j'(t)\frac{\partial f_m}{\partial u_j}\right)_{\boldsymbol{\alpha}(t)} \\ &= \big((f_1 \circ \boldsymbol{\alpha})'(t), \ldots, (f_m \circ \boldsymbol{\alpha})'(t)\big) \\ &= (F \circ \boldsymbol{\alpha})'(t). \blacksquare \end{aligned}$$

Corollary 9.14. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. Let $\mathbf{p} \in \mathcal{U}$, $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$, and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{U}$ be a curve such that $\boldsymbol{\alpha}'(0) = \mathbf{v}_{\mathbf{p}}$. Then*

$$F_{*\mathbf{p}}(\mathbf{v}_{\mathbf{p}}) = (F \circ \boldsymbol{\alpha})'(0).$$

We originally defined $F_*(\mathbf{v}_{\mathbf{p}}) = F_{*\mathbf{p}}(\mathbf{v}_{\mathbf{p}})$ to be the velocity vector at 0 of the image under F of the straight line $t \to \mathbf{p} + t\mathbf{v}$. Corollary 9.14 says that $F_*(\mathbf{v}_{\mathbf{p}})$ equals the velocity vector at 0 of the image under F of *any* curve $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \mathbf{v}_{\mathbf{p}}$.

Corollary 9.15. *Suppose that there are differentiable maps*

$$\begin{array}{ccccc} \mathcal{U} & \xrightarrow{F} & \mathcal{V} & \xrightarrow{G} & \mathbb{R}^n \\ \cap & & \cap & & \\ \mathbb{R}^\ell & & \mathbb{R}^m. & & \end{array}$$

Then $(G \circ F)_* = G_* \circ F_*$.

Proof. Let $\mathbf{v}_\mathbf{p}$ be a tangent vector to $\mathbb{R}^\ell$ at $\mathbf{p} \in \mathbb{R}^\ell$. Then $t \mapsto F(\mathbf{p} + t\mathbf{v})$ is a curve in $\mathbb{R}^m$ such that $F_*(\mathbf{v}_\mathbf{p}) = F(\mathbf{p} + t\mathbf{v})'(0)$. Hence

$$\begin{aligned}(G \circ F)_*(\mathbf{v}_\mathbf{p}) = \big((G \circ F)(\mathbf{p} + t\mathbf{v})\big)'(0) &= G\big(F(\mathbf{p} + t\mathbf{v})\big)'(0) \\ &= G_*\big(F(\mathbf{p} + t\mathbf{v})'(0)\big) = G_*\big(F_*(\mathbf{v}_\mathbf{p})\big). \ \blacksquare\end{aligned}$$

Recall from Chapter 1 that if $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then the matrix associated with L is the $m \times n$ matrix $A = (a_{ij})$ such that

$$L(\mathbf{v}) = A\mathbf{v}$$

for all $\mathbf{v} \in \mathbb{R}^n$. (Here, for clarity, we use different symbols for the map and the matrix.) Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbb{R}^n$, so $\mathbf{e}_i$ is to be thought of as the column vector with 1 in the ith row and zeros elsewhere; similarly, let $\{\mathbf{e}'_1, \ldots, \mathbf{e}'_m\}$ be the standard basis of $\mathbb{R}^m$. Then the formula we are adopting to relate A to L is

$$L(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{e}'_i, \qquad j = 1, \ldots, n.$$

It is the jth *column* of A that represents the image of the jth basis element.

If $F\colon \mathcal{U} \to \mathbb{R}^m$ is a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$, then for each $\mathbf{p} \in \mathcal{U}$ the tangent map $L = F_{*\mathbf{p}}$ is a linear map between the vector spaces $\mathbb{R}^n_\mathbf{p}$ and $\mathbb{R}^m_{F(\mathbf{p})}$. Since $\mathbb{R}^n_\mathbf{p}$ is canonically isomorphic to $\mathbb{R}^n$, and $\mathbb{R}^m_{F(\mathbf{p})}$ is $\mathbb{R}^m$, we can associate to $F_{*\mathbf{p}}$ an $m \times n$ matrix A, which has the property that $F_{*\mathbf{p}}(\mathbf{v}_\mathbf{p}) = A\,\mathbf{v}_\mathbf{p}$ for all $\mathbf{v}_\mathbf{p} \in \mathbb{R}^n_\mathbf{p}$, provided we interpret $\mathbf{v}_\mathbf{p}$ here as a column vector (despite the way we usually write it). More precisely,

Definition 9.16. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$, and write $F = (f_1, \ldots, f_m)$. The* ***Jacobian matrix***[1] *is the matrix-valued function $\mathscr{J}(F)$ of F given by*

$$\mathscr{J}(F)(\mathbf{p}) = \begin{pmatrix} \dfrac{\partial f_1}{\partial u_1}(\mathbf{p}) & \cdots & \dfrac{\partial f_1}{\partial u_n}(\mathbf{p}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial u_1}(\mathbf{p}) & \cdots & \dfrac{\partial f_m}{\partial u_n}(\mathbf{p}) \end{pmatrix}$$

for $\mathbf{p} \in \mathcal{U}$.

---

[1] 

Carl Gustav Jacob Jacobi (1804–1851). Professor of Mathematics at Königsberg and Berlin. He is remembered for his work on dynamics and elliptic functions. His *Fundamenta Nova Theoria Functionum Ellipticarum*, in which elliptic function theory is based on four theta functions, was published in 1829.

Recall that the ***rank*** of an $m \times n$ matrix A can be defined as the dimension ρ of the space spanned by the columns of A. A fundamental theorem of linear algebra asserts that it is also equal to the dimension of the space spanned by the rows of A, and is therefore no larger than the minimum of m and n. Accordingly, if ρ *equals* $\min(m, n)$, then A is said to have ***maximum rank***. The situation described by the next result can only occur when $m \geqslant n$.

Lemma 9.17. *Let $F\colon \mathcal{U} \to \mathbb{R}^m$ be a differentiable map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. Then each tangent map $F_{*\mathbf{p}}\colon \mathbb{R}^n_{\mathbf{p}} \to \mathbb{R}^m_{F(\mathbf{p})}$ is injective if and only if $\mathscr{J}(F)(\mathbf{p})$ has rank equal to n.*

Proof. Let $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ and write $\mathbf{v} = (v_1, \ldots, v_n)$; also put $F = (f_1, \ldots, f_m)$. Suppose that $F_*(\mathbf{v}_{\mathbf{p}}) = 0$ and $\mathbf{v} \neq 0$. Then Lemma 9.10 implies that

$$\mathbf{v}_{\mathbf{p}}[f_1] = \cdots = \mathbf{v}_{\mathbf{p}}[f_m] = 0,$$

and by Lemma 9.4 we have

$$\sum_{i=1}^{n} v_i \frac{\partial f_j}{\partial u_i}(\mathbf{p}) = 0$$

for $j = 1, \ldots, m$. It follows that

$$\sum_{i=1}^{n} v_i \left(\frac{\partial f_1}{\partial u_i}(\mathbf{p}), \ldots, \frac{\partial f_m}{\partial u_i}(\mathbf{p}) \right) = 0.$$

Therefore, the n columns of the Jacobian matrix $\mathscr{J}(F)(\mathbf{p})$ are linearly dependent, and consequently $\mathscr{J}(F)(\mathbf{p})$ has rank less than n.

Conversely, if the rank of $\mathscr{J}(F)(\mathbf{p})$ is less than n, the steps in the above proof can be reversed to conclude that $F_*(\mathbf{v}_{\mathbf{p}}) = 0$ for some nonzero $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$. ∎

9.4 Vector Fields on $\mathbb{R}^n$

It will be necessary to consider tangent vectors as they vary from point to point.

Definition 9.18. *A **vector field** $\mathbf{V}$ on an open subset $\mathcal{U}$ of $\mathbb{R}^n$ is a function that assigns to each $\mathbf{p} \in \mathcal{U}$ a tangent vector $\mathbf{V}(\mathbf{p}) \in \mathbb{R}^n_{\mathbf{p}}$. If $f : \mathcal{U} \to \mathbb{R}$ is differentiable, we let $\mathbf{V}$ act on f via*

$$\mathbf{V}[f](\mathbf{p}) = \mathbf{V}(\mathbf{p})[f].$$

*The vector field $\mathbf{V}$ is said to be **differentiable** provided that $\mathbf{V}[f]\colon \mathcal{U} \to \mathbb{R}$ is differentiable whenever f is.*

It is important to distinguish between the ith coordinate of a point $\mathbf{p} \in \mathbb{R}^n$ and the *function* that assigns to $\mathbf{p}$ its ith coordinate:

Definition 9.19. *The* ***natural coordinate functions*** *of* $\mathbb{R}^n$ *are the functions* u_i *defined by*

$$u_i(\mathbf{p}) = p_i, \qquad i = 1, \ldots, n,$$

for $\mathbf{p} = (p_1, \ldots, p_n)$.

In the special cases $\mathbb{R}$ or $\mathbb{R}^2$ or $\mathbb{R}^3$, we shall often denote the natural coordinate functions by symbols such as t or u, v or x, y, z, respectively.

Next, we need some "standard" vector fields on $\mathbb{R}^n$.

Definition 9.20. *The vector field* $\mathbf{U}_j$ *on* $\mathbb{R}^n$ *is defined by*

$$\mathbf{U}_j(\mathbf{p}) = (0, \ldots, 0, 1, 0, \ldots, 0)_{\mathbf{p}},$$

where 1 *occurs in the* j*th spot. We shall call* $\{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ *the* ***natural frame field*** *of* $\mathbb{R}^n$.

It is obvious from Lemma 9.4 that $\mathbf{U}_j[f] = \partial f/\partial u_j$ for any differentiable function f, and it is common practice to write $\partial/\partial u_j$ in place of $\mathbf{U}_j$. Note that the restrictions of $\mathbf{U}_1, \ldots, \mathbf{U}_n$ to any open subset $\mathcal{U} \subseteq \mathbb{R}^n$ are vector fields on $\mathcal{U}$.

The proofs of the following algebraic properties of vector fields are easy.

Lemma 9.21. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be vector fields on an open subset* $\mathcal{U} \subseteq \mathbb{R}^n$, *let* $a \in \mathbb{R}$ *and let* $g\colon \mathbb{R}^n \to \mathbb{R}$. *Define* $\mathbf{X} + \mathbf{Y}$, $a\mathbf{X}$, $g\mathbf{X}$ *and* $\mathbf{X} \cdot \mathbf{Y}$ *by*

$$(\mathbf{X} + \mathbf{Y})[f] = \mathbf{X}[f] + \mathbf{Y}[f], \qquad (a\mathbf{X})[f] = a\,\mathbf{X}[f],$$
$$(g\mathbf{X})[f] = g\,\mathbf{X}[f], \qquad (\mathbf{X} \cdot \mathbf{Y})(\mathbf{p}) = \mathbf{X}(\mathbf{p}) \cdot \mathbf{Y}(\mathbf{p}),$$

where $f\colon \mathcal{U} \to \mathbb{R}$ *is differentiable and* $\mathbf{p} \in \mathcal{U}$. *Then* $\mathbf{X} + \mathbf{Y}$, $a\mathbf{X}$, $g\mathbf{X}$ *are vector fields on* $\mathcal{U}$ *and* $\mathbf{X} \cdot \mathbf{Y}\colon \mathcal{U} \to \mathbb{R}$ *is a function. If* $\mathbf{X}$, $\mathbf{Y}$ *and* g *are differentiable, so are the vector fields* $\mathbf{X} + \mathbf{Y}$, $a\mathbf{X}$, $g\mathbf{X}$, *and the function* $\mathbf{X} \cdot \mathbf{Y}$.

It is not hard to express a vector field in terms of the $\mathbf{U}_j$'s:

Lemma 9.22. *If* $\mathbf{V}$ *is any vector field on an open subset* $\mathcal{U} \subseteq \mathbb{R}^n$, *there exist functions* $v_j\colon \mathcal{U} \to \mathbb{R}$ *for* $j = 1, \ldots, n$ *such that*

$$\mathbf{V} = \sum_{j=1}^{n} v_j \mathbf{U}_j. \tag{9.10}$$

Proof. Since $\mathbf{V}(\mathbf{p}) \in \mathbb{R}^n_{\mathbf{p}}$, we can write $\mathbf{V}(\mathbf{p}) = \big(v_1(\mathbf{p}), \ldots, v_n(\mathbf{p})\big)_{\mathbf{p}}$. In this way the functions $v_j\colon \mathcal{U} \to \mathbb{R}$ are defined. On the other hand,

$$\begin{aligned}\big(v_1(\mathbf{p}), \ldots, v_n(\mathbf{p})\big)_{\mathbf{p}} &= v_1(\mathbf{p})(1, 0, \ldots, 0)_{\mathbf{p}} + \cdots + v_n(\mathbf{p})(0, \ldots, 0, 1)_{\mathbf{p}} \\ &= \sum_{j=1}^{n} v_j(\mathbf{p})\mathbf{U}_j(\mathbf{p}).\end{aligned}$$

Since this equation holds for arbitrary $\mathbf{p} \in \mathcal{U}$, we get (9.10). ∎

The next lemma yields a important practical criterion for the differentiability of a vector field:

Corollary 9.23. *A vector field* $\mathbf{V}$ *on an open subset* $\mathcal{U} \subseteq \mathbb{R}^n$ *is differentiable if and only if the natural coordinate functions* $v_1, \ldots, v_n$ *of* $\mathbf{V}$ *given in* (9.10) *are differentiable.*

There is an important way to construct vector fields from functions on $\mathbb{R}^n$.

Definition 9.24. *Let* $g\colon \mathbb{R}^n \to \mathbb{R}$ *be a differentiable function. The* ***gradient*** *of* g *is the vector field* $\mathsf{grad}\, g$ *given by*

$$\mathsf{grad}\, g = \sum_{j=1}^{n} \frac{\partial g}{\partial u_j} \mathbf{U}_j.$$

This operator is characterized by

Lemma 9.25. *The gradient of a function* g *is the unique differentiable vector field* $\mathsf{grad}\, g$ *on* $\mathbb{R}^n$ *for which*

$$(\mathsf{grad}\, g) \cdot \mathbf{V} = \mathbf{V}[g]$$

for any differentiable vector field $\mathbf{V}$ *on* $\mathbb{R}^n$.

Proof. Write $\mathbf{V} = \sum_{j=1}^{n} v_j \mathbf{U}_j$. Then

$$(\mathsf{grad}\, g) \cdot \mathbf{V} = \left(\sum_{j=1}^{n} \mathbf{U}_j \frac{\partial g}{\partial u_j} \right) \cdot \mathbf{V} = \sum_{j=1}^{n} v_j \frac{\partial g}{\partial u_j}$$
$$= \sum_{j=1}^{n} v_j \mathbf{U}_j[g] = \mathbf{V}[g].$$

Conversely, let $\mathbf{Z}$ be a vector field for which $\mathbf{Z} \cdot \mathbf{V} = \mathbf{V}[g]$ for any vector field $\mathbf{V}$ on $\mathbb{R}^n$. We can write

$$\mathbf{Z} = \sum_{i=1}^{n} f_i \mathbf{U}_i,$$

where $f_i\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable for $i = 1, \ldots, n$. Then for $j = 1, \ldots, n$ we have

$$f_j = \mathbf{Z} \cdot \mathbf{U}_j = \mathbf{U}_j[g] = \frac{\partial g}{\partial u_j},$$

so that $\mathbf{Z} = \mathsf{grad}\, g$. ∎

Finally, we note some special operations on vector fields on $\mathbb{R}^2$ and $\mathbb{R}^3$, which extend the pointwise ones defined at the end of Section 9.1.

Lemma 9.26. *For a vector field* $\mathbf{Y}$ *on an open subset* $\mathcal{U} \subseteq \mathbb{R}^2$*, define* $J\mathbf{Y}$ *by*

$$(J\mathbf{Y})(\mathbf{p}) = J\big(\mathbf{Y}(\mathbf{p})\big)$$

for $\mathbf{p} \in \mathcal{U}$*. Similarly, for vector fields* $\mathbf{Y}, \mathbf{Z}$ *on* $\mathcal{U}$*, define their cross product in a pointwise fashion, namely*

$$(\mathbf{Y} \times \mathbf{Z})(\mathbf{p}) = \mathbf{Y}(\mathbf{p}) \times \mathbf{Z}(\mathbf{p})$$

for $\mathbf{p} \in \mathcal{U}$*.*

It is immediate that $J\mathbf{Y}$ and $\mathbf{Y} \times \mathbf{Z}$ are *differentiable* vector fields on $\mathcal{U}$ whenever $\mathbf{Y}$ and $\mathbf{Z}$ are differentiable.

9.5 Derivatives of Vector Fields

In Section 9.2, we learned how to differentiate *functions* by means of tangent vectors. A key concept in differential geometry is the notion of how to differentiate a *vector field*. For vector fields on open subsets of $\mathbb{R}^n$, the definition is particularly natural.

If $\mathbf{W}$ is a vector field on an open subset $\mathcal{U} \subseteq \mathbb{R}^n$, one frequently does not distinguish the element $\mathbf{W}(\mathbf{p})$ inside the tangent space $\mathbb{R}^n_{\mathbf{p}}$ from its vector part $\mathbf{v} \in \mathbb{R}^n$. Strictly speaking this is incorrect, since $\mathbf{W}(\mathbf{p}) = \mathbf{v}_{\mathbf{p}}$ and one loses knowledge of the point $\mathbf{p}$. It is however a convenient abuse of notation, which (as will be explained in Section 9.6) was carried out repeatedly in Chapter 1.

Restrict $\mathbf{W}$ to that portion of the straight line $t \mapsto \mathbf{p} + t\mathbf{v}$ contained in $\mathcal{U}$ so that, with the convention above, $t \mapsto \mathbf{W}(\mathbf{p} + t\mathbf{v})$ is now a curve in $\mathbb{R}^n$.

Definition 9.27. *Let* $\mathbf{W}$ *be a differentiable vector field on an open subset* $\mathcal{U}$ *of* $\mathbb{R}^n$*, and let* $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$*. The* derivative *of* $\mathbf{W}$ *with respect to* $\mathbf{v}_{\mathbf{p}}$ *is the element of* $\mathbb{R}^n_{\mathbf{p}}$ *given by*

$$\mathbf{W}(\mathbf{p} + t\mathbf{v})'(0)_{\mathbf{p}} = \left(\lim_{t \to 0} \frac{\mathbf{W}(\mathbf{p} + t\mathbf{v}) - \mathbf{W}(\mathbf{p})}{t}\right)_{\mathbf{p}}.$$

This is denoted by $\mathbf{D}_{\mathbf{v}_{\mathbf{p}}}\mathbf{W}$*, or* $\mathbf{D}_{\mathbf{v}}\mathbf{W}$ *if the identity of* $\mathbf{p}$ *is understood.*

It is possible to compute the tangent vector $\mathbf{D}_{\mathbf{v}}\mathbf{W}$ directly from its definition, but there is an alternative formula (similar to equation (9.7)), which is usually more convenient.

Lemma 9.28. *Let* $\mathbf{W}$ *be a differentiable vector field on an open subset* $\mathcal{U} \subseteq \mathbb{R}^n$*, and write*

$$\mathbf{W} = \sum_{i=1}^{n} w_i \mathbf{U}_i.$$

Then for $\mathbf{p} \in \mathcal{U}$ *and* $\mathbf{v}_\mathbf{p} \in \mathbb{R}^n_\mathbf{p}$,

$$\mathbf{D}_\mathbf{v}\mathbf{W} = \sum_{i=1}^{n} \mathbf{v}_\mathbf{p}[w_i]\mathbf{U}_i(\mathbf{p}) = \big(\mathbf{v}_\mathbf{p}[w_1], \ldots, \mathbf{v}_\mathbf{p}[w_n]\big)_\mathbf{p}. \tag{9.11}$$

Proof. We have

$$\begin{aligned} \mathbf{W}(\mathbf{p}+t\mathbf{v}) &= \sum_{i=1}^{n} w_i(\mathbf{p}+t\mathbf{v})\mathbf{U}_i(\mathbf{p}+t\mathbf{v}) \\ &= \big(w_1(\mathbf{p}+t\mathbf{v}), \ldots, w_n(\mathbf{p}+t\mathbf{v})\big)_{\mathbf{p}+t\mathbf{v}}. \end{aligned}$$

Hence $\mathbf{W}(\mathbf{p}+t\mathbf{v})' = \big(w_1(\mathbf{p}+t\mathbf{v})', \ldots, w_n(\mathbf{p}+t\mathbf{v})'\big)$. But by the definition of tangent vector, we have

$$w_j(\mathbf{p}+t\mathbf{v})'(0) = \mathbf{v}_\mathbf{p}[w_j]$$

for $j = 1, \ldots, n$, so

$$\mathbf{W}(\mathbf{p}+t\mathbf{v})'(0) = \big(\mathbf{v}_\mathbf{p}[w_1], \ldots, \mathbf{v}_\mathbf{p}[w_n]\big),$$

as required. ∎

It is an easy matter to compute the derivatives of the standard vector fields $\mathbf{U}_1, \ldots, \mathbf{U}_n$.

Corollary 9.29. *We have* $\mathbf{D}_\mathbf{p}\mathbf{U}_j = 0$ *for any tangent vector* $\mathbf{v}_\mathbf{p}$ *to* $\mathbb{R}^n$, *and* $j = 1, \ldots, n$.

Proof. If we use (9.10) to express $\mathbf{U}_j$ in terms of $\mathbf{U}_1, \ldots, \mathbf{U}_n$, then each of the coefficients $v_1, \ldots, v_n$ in (9.10) is either 0 or 1, and is in any case constant. The result follows from (9.11). ∎

Next, we show that the operator $\mathbf{D}$ behaves naturally with respect to the operations on vector fields given in Lemma 9.21.

Lemma 9.30. *Let* $\mathbf{Y}, \mathbf{Z}$ *be differentiable vector fields on an open subset* $\mathcal{U}$ *of* $\mathbb{R}^n$, *and* $\mathbf{v}_\mathbf{p}, \mathbf{w}_\mathbf{p}$ *tangent vectors to* $\mathbb{R}^n$ *at* $\mathbf{p} \in \mathcal{U}$. *Then*

$$\begin{aligned} \mathbf{D}_{a\mathbf{v}+b\mathbf{w}}\mathbf{Y} &= a\mathbf{D}_\mathbf{v}\mathbf{Y} + b\mathbf{D}_\mathbf{w}\mathbf{Y}, \\ \mathbf{D}_\mathbf{v}(a\mathbf{Y}+b\mathbf{Z}) &= a\mathbf{D}_\mathbf{v}\mathbf{Y} + b\mathbf{D}_\mathbf{v}\mathbf{Z}, \\ \mathbf{D}_\mathbf{v}(f\,\mathbf{Y}) &= \mathbf{v}_\mathbf{p}[f]\mathbf{Y}(\mathbf{p}) + f(\mathbf{p})\mathbf{D}_\mathbf{v}\mathbf{Y}, \\ \mathbf{v}_\mathbf{p}[\mathbf{Y}\cdot\mathbf{Z}] &= \mathbf{D}_\mathbf{v}\mathbf{Y}\cdot\mathbf{Z}(\mathbf{p}) + \mathbf{Y}(\mathbf{p})\cdot\mathbf{D}_\mathbf{v}\mathbf{Z}, \end{aligned} \tag{9.12}$$

where $a, b \in \mathbb{R}$ *and* $f\colon \mathcal{U} \to \mathbb{R}$ *is a differentiable function.*

Proof. We prove (9.12); the proofs of the other formulas are similar. Write

$$\mathbf{Y} = \sum_{i=1}^{n} y_i \mathbf{U}_i \quad \text{and} \quad \mathbf{Z} = \sum_{i=1}^{n} z_i \mathbf{U}_i.$$

From (9.4) and (9.5) it follows that

$$\begin{aligned}
\mathbf{v}_\mathbf{p}[\mathbf{Y} \cdot \mathbf{Z}] &= \sum_{i=1}^{n} \mathbf{v}_\mathbf{p}[y_i z_i] \\
&= \sum_{i=1}^{n} \Big\{ \mathbf{v}_\mathbf{p}[y_i] z_i(\mathbf{p}) + y_i(\mathbf{p}) \mathbf{v}_\mathbf{p}[z_i] \Big\} \\
&= \Big\{ \sum_{i=1}^{n} \mathbf{v}_\mathbf{p}[y_i] \mathbf{U}_i(\mathbf{p}) \Big\} \cdot \Big\{ \sum_{i=1}^{n} z_i(\mathbf{p}) \mathbf{U}_i(\mathbf{p}) \Big\} \\
&\qquad + \Big\{ \sum_{i=1}^{n} y_i(\mathbf{p}) \mathbf{U}_i(\mathbf{p}) \Big\} \cdot \Big\{ \sum_{i=1}^{n} \mathbf{v}_\mathbf{p}[z_i] \mathbf{U}_i(\mathbf{p}) \Big\} \\
&= \mathbf{D}_\mathbf{v}\mathbf{Y} \cdot \mathbf{Z}(\mathbf{p}) + \mathbf{Y}(\mathbf{p}) \cdot \mathbf{D}_\mathbf{v}\mathbf{Z}. \ \blacksquare
\end{aligned}$$

The operator $\mathbf{D}$ also behaves naturally with respect to the complex structure J of $\mathbb{R}^2$ and the vector cross product of $\mathbb{R}^3$:

Lemma 9.31. *Let $\mathbf{Y}$ and $\mathbf{Z}$ be differentiable vector fields on an open subset $\mathcal{U} \subseteq \mathbb{R}^n$, where $n = 2$ or 3, and let $\mathbf{v}_\mathbf{p}, \mathbf{w}_\mathbf{p}$ be tangent vectors to $\mathbb{R}^n$. Then*

$$\mathbf{D}_\mathbf{v}(J\mathbf{Y}) = J\big(\mathbf{D}_\mathbf{v}\mathbf{Y}\big) \qquad (\textit{for } n = 2), \tag{9.13}$$

$$\mathbf{D}_\mathbf{v}(\mathbf{Y} \times \mathbf{Z}) = \mathbf{D}_\mathbf{v}\mathbf{Y} \times \mathbf{Z}(\mathbf{p}) + \mathbf{Y}(\mathbf{p}) \times \mathbf{D}_\mathbf{v}\mathbf{Z} \quad (\textit{for } n = 3). \tag{9.14}$$

Proof. Given a vector field $\mathbf{Y}$ on $\mathbb{R}^2$, we can write $\mathbf{Y} = y_1\mathbf{U}_1 + y_2\mathbf{U}_2$; then $J\mathbf{Y} = -y_2\mathbf{U}_1 + y_1\mathbf{U}_2$. It follows from (9.11) that

$$\mathbf{D}_\mathbf{v}(J\mathbf{Y}) = \big(-\mathbf{v}_\mathbf{p}[y_2],\, \mathbf{v}_\mathbf{p}[y_1]\big)_\mathbf{p} = J\big(\mathbf{v}_\mathbf{p}[y_1],\, \mathbf{v}_\mathbf{p}[y_2]\big)_\mathbf{p} = J(\mathbf{D}_\mathbf{v}\mathbf{Y}),$$

giving (9.13). The proof of (9.14) is similar. ∎

It is also possible to define the derivative of one vector field by another.

Lemma 9.32. *Let $\mathbf{V}$ and $\mathbf{W}$ be vector fields defined on an open subset $\mathcal{U} \subseteq \mathbb{R}^n$, and assume that $\mathbf{W}$ is differentiable. Define $\mathbf{D}_\mathbf{V}\mathbf{W}$ by*

$$(\mathbf{D}_\mathbf{V}\mathbf{W})(\mathbf{p}) = \mathbf{D}_{\mathbf{V}(\mathbf{p})}\mathbf{W}$$

for $\mathbf{p} \in \mathcal{U}$. Then $\mathbf{D}_\mathbf{V}\mathbf{W}$ is a vector field on $\mathcal{U}$ which is differentiable if $\mathbf{V}$ and $\mathbf{W}$ are differentiable. Moreover,

$$\mathbf{D}_\mathbf{V}\Big(\sum_{i=1}^{n} w_i \mathbf{U}_i \Big) = \sum_{i=1}^{n} \mathbf{V}[w_i] \mathbf{U}_i. \tag{9.15}$$

Proof. Equation (9.15) is an immediate consequence of (9.11). ∎

Let $F\colon \mathcal{U} \to \mathcal{V}$ be a diffeomorphism between open subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathbb{R}^n$ (recall Definition 9.7 on page 268), and let $\mathbf{V}$ be a differentiable vector field on $\mathcal{U}$. We want to define the image $F_*(\mathbf{V})$ of $\mathbf{V}$ under F, in a manner consistent with (9.7). The following does the job:

Definition 9.33. *Let $F\colon \mathcal{U} \to \mathcal{V}$ be a diffeomorphism between open subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathbb{R}^n$ and let $\mathbf{V}$ be a differentiable vector field on $\mathcal{U}$. Write $F = (f_1, \ldots, f_n)$. Then $F_*(\mathbf{V})$ is the vector field on $\mathcal{V}$ given by*

$$F_*(\mathbf{V}) = \sum_{i=1}^{n} \big(\mathbf{V}[f_i] \circ F^{-1}\big)\mathbf{U}_i.$$

It is easy to prove

Lemma 9.34. *Suppose that $F\colon \mathcal{U} \to \mathcal{V}$ is a diffeomorphism between open subsets $\mathcal{U}$ and $\mathcal{V}$ of $\mathbb{R}^n$. Let $\mathbf{Y}$ and $\mathbf{Z}$ be differentiable vector fields on an open subset $\mathcal{U}$. Then*

$$F_*(a\mathbf{Y} + b\mathbf{Z}) = a F_*(\mathbf{Y}) + b F_*(\mathbf{Z}), \tag{9.16}$$

$$F_*(f\mathbf{Y}) = (f \circ F^{-1}) F_*(\mathbf{Y}), \tag{9.17}$$

$$F_*(\mathbf{Y})[f \circ F^{-1}] = \mathbf{Y}[f] \circ F^{-1},$$

where a, b are real numbers and $f\colon \mathcal{U} \to \mathbb{R}$ is a differentiable function.

Next, we show that certain maps of $\mathbb{R}^n$ into itself have a desirable effect on the derivatives of vector fields.

Lemma 9.35. *Suppose that $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is an affine transformation. Then*

$$\mathbf{D}_{F_*(\mathbf{V})}\big(F_*(\mathbf{W})\big) = F_*\big(\mathbf{D}_{\mathbf{V}}\mathbf{W}\big). \tag{9.18}$$

Proof. By assumption, $F(\mathbf{p}) = A\mathbf{p} + \mathbf{q}$, where $A\colon \mathbb{R}^n \to \mathbb{R}^n$ is linear. We can therefore write

$$F_*(\mathbf{U}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{U}_i, \qquad j = 1, \ldots, n \tag{9.19}$$

where $A = (a_{ij})$ and so each a_{ij} is constant. From (9.16) and (9.17), we get

$$\begin{aligned} \mathbf{D}_{F_*(\mathbf{V})}F_*(\mathbf{W}) &= \mathbf{D}_{F_*(\mathbf{V})}\bigg(F_*\Big(\sum_{j=1}^{n} w_j\mathbf{U}_j\Big)\bigg) \\ &= \mathbf{D}_{F_*(\mathbf{V})}\bigg(\sum_{j=1}^{n}(w_j \circ F^{-1})F_*(\mathbf{U}_j)\bigg) \\ &= \sum_{j=1}^{n} \mathbf{D}_{F_*(\mathbf{V})}\big((w_j \circ F^{-1})F_*(\mathbf{U}_j)\big). \end{aligned} \tag{9.20}$$

Then (9.15) and (9.19) imply that

$$\begin{aligned}(9.21)\quad \mathbf{D}_{F_*(\mathbf{V})}\big((w_j\circ F^{-1})F_*(\mathbf{U}_j)\big) &= \mathbf{D}_{F_*(\mathbf{V})}\bigg((w_j\circ F^{-1})\sum_{i=1}^n a_{ij}\mathbf{U}_i\bigg)\\ &= \sum_{i=1}^n F_*(\mathbf{V})\big[(w_j\circ F^{-1})a_{ij}\big]\mathbf{U}_i,\end{aligned}$$

for each fixed j. Since each a_{ij} is constant,

$$(9.22)\qquad F_*(\mathbf{V})\big[(w_j\circ F^{-1})a_{ij}\big] = F_*(\mathbf{V})\big[w_j\circ F^{-1}\big]a_{ij}.$$

From (9.20)–(9.22), we get

$$\begin{aligned}\mathbf{D}_{F_*(\mathbf{V})}F_*(\mathbf{W}) &= \sum_{i,j=1}^n F_*(\mathbf{V})\big[w_j\circ F^{-1}\big]a_{ij}\mathbf{U}_i\\ (9.23)\qquad &= \sum_{j=1}^n (\mathbf{V}[w_j]\circ F^{-1})F_*(\mathbf{U}_j)\\ &= \sum_{j=1}^n F_*\big(\mathbf{V}[w_j]\mathbf{U}_j\big).\end{aligned}$$

Now (9.18) follows from (9.23) and (9.15). ∎

We conclude this section by defining some well-known operators from vector analysis.

Definition 9.36. *Let $\mathbf{V}$ be a differentiable vector field defined on an open subset $\mathcal{U}\subseteq\mathbb{R}^n$. The **divergence** of $\mathbf{V}$ is the function*

$$\operatorname{div}\mathbf{V} = \sum_{i=1}^n \Big(\mathbf{D}_{\mathbf{U}_i}\mathbf{V}\cdot\mathbf{U}_i\Big).$$

*If $f\colon\mathcal{U}\to\mathbb{R}$ is differentiable, then the **Laplacian** of f is the function*

$$\boldsymbol{\Delta} f = \operatorname{div}(\operatorname{grad} f).$$

Lemma 9.37. *If $f\colon\mathcal{U}\to\mathbb{R}$ is a differentiable function, where $\mathcal{U}\subseteq\mathbb{R}^n$ is open, then*

$$\boldsymbol{\Delta} f = \sum_{i=1}^n \mathbf{U}_i^2 f = \sum_{i=1}^n \frac{\partial^2 f}{\partial u_i^2}.$$

Proof. We have

$$\begin{aligned}\boldsymbol{\Delta} f = \operatorname{div}(\operatorname{grad} f) &= \sum_{i=1}^n \mathbf{D}_{\mathbf{U}_i}(\operatorname{grad} f)\cdot\mathbf{U}_i\\ &= \sum_{i=1}^n \mathbf{U}_i\big[\operatorname{grad} f\cdot\mathbf{U}_i\big] = \sum_{i=1}^n \mathbf{U}_i\big[\mathbf{U}_i[f]\big].\end{aligned}$$

∎

A differentiable function $f\colon \mathcal{U} \to \mathbb{R}$ is called ***harmonic*** if it satisfes $\boldsymbol{\Delta} f = 0$. Such functions will be used in the theory of minimal surfaces in Sections 16.7 and 22.1.

9.6 Curves Revisited

In this section, we assume for simplicity that all curves are differentiable in the manner explained on page 266, so that derivatives of all orders exist.

In the study of curves and surfaces, it is common practice to make no distinction between a tangent vector $\mathbf{v_p}$ and its vector part. For example, let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve. On page 6, we defined its velocity $\boldsymbol{\alpha}'(t)$ to be a vector in $\mathbb{R}^n$. More properly, it should be a tangent vector at the point $\boldsymbol{\alpha}(t)$:

Definition 9.38. (*Revised*) *The **velocity** of a curve $\boldsymbol{\alpha}$ is the function*

$$t \mapsto \boldsymbol{\alpha}'(t)_{\alpha(t)},$$

*and the **acceleration** of $\boldsymbol{\alpha}$ is the function*

$$t \mapsto \boldsymbol{\alpha}''(t)_{\alpha(t)}.$$

Since $\boldsymbol{\alpha}'(t)_{\alpha(t)}$ is a tangent vector, it can be applied to a differentiable function f:

Lemma 9.39. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve and $f\colon \mathbb{R}^n \to \mathbb{R}$ a differentiable function. Then*

$$\boldsymbol{\alpha}'(t)_{\alpha(t)}[f] = (f \circ \boldsymbol{\alpha})'(t).$$

Proof. We write $\boldsymbol{\alpha}'(t) = \big(a_1'(t), \ldots, a_n'(t)\big)$. It follows from Lemma 9.4 that

$$\boldsymbol{\alpha}'(t)_{\alpha(t)}[f] = \sum_{j=1}^{n} a_j'(t) \frac{\partial f}{\partial u_j}\big(\boldsymbol{\alpha}(t)\big). \tag{9.24}$$

But the chain rule says that the right-hand side of (9.24) is $(f \circ \boldsymbol{\alpha})'(t)$. ∎

In Section 1.2 we defined the notion of vector field along a curve $\boldsymbol{\alpha}$ in $\mathbb{R}^n$. A vector field $\mathbf{W}$ on $\mathbb{R}^n$ gives rise to a vector field on $\boldsymbol{\alpha}$; it is simply $\mathbf{W} \circ \boldsymbol{\alpha}$. Note that for each t, $(\mathbf{W} \circ \boldsymbol{\alpha})(t)$ is a vector in $\mathbb{R}^n$. A closely related concept is the ***restriction*** of $\mathbf{W}$ to $\boldsymbol{\alpha}$, which is the function

$$t \mapsto \mathbf{W}\big(\boldsymbol{\alpha}(t)\big)_{\alpha(t)}.$$

Frequently, it is useful to make no distinction between the vector $(\mathbf{W} \circ \boldsymbol{\alpha})(t)$ in $\mathbb{R}^n$, and the tangent vector $\mathbf{W}\big(\boldsymbol{\alpha}(t)\big) = \mathbf{W}\big(\boldsymbol{\alpha}(t)\big)_{\alpha(t)}$ in $\mathbb{R}^n_{\alpha(t)}$, mainly because it is too much trouble to write the subscript. Similarly, unless complete clarity is required, we shall not distinguish between $\boldsymbol{\alpha}'(t)$ and $\boldsymbol{\alpha}'(t)_{\alpha(t)}$.

In Section 1.2, we also defined the derivative of a vector field along a curve. In particular, we know how to compute $(\mathbf{W} \circ \boldsymbol{\alpha})'$. Let us make precise the relation between $(\mathbf{W} \circ \boldsymbol{\alpha})'$ and the derivative of $\mathbf{W}$ with respect to $\boldsymbol{\alpha}'(t)$.

Lemma 9.40. *If* $\mathbf{W}$ *is a differentiable vector field on* $\mathbb{R}^n$ *and* $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ *is a curve, then*

$$\mathbf{D}_{\alpha'(t)}\mathbf{W} = (\mathbf{W} \circ \boldsymbol{\alpha})'(t)_{\alpha(t)}.$$

Proof. By Lemma 9.28 we have

$$\mathbf{D}_{\alpha'(t)}\mathbf{W} = \sum_{i=1}^{n} \boldsymbol{\alpha}'(t)[w_i]\mathbf{U}_i.$$

On the other hand, Lemma 9.39 implies that $\boldsymbol{\alpha}'(t)[w_i] = (w_i \circ \boldsymbol{\alpha})'(t)$. Thus,

$$\begin{aligned} \mathbf{D}_{\alpha'(t)}\mathbf{W} &= \sum_{i=1}^{n} \boldsymbol{\alpha}'(t)[w_i]\mathbf{U}_i \\ &= \sum_{i=1}^{n} (w_i \circ \boldsymbol{\alpha})'(t)\mathbf{U}_i \\ &= (\mathbf{W} \circ \boldsymbol{\alpha})'(t)_{\alpha(t)}. \blacksquare \end{aligned}$$

9.7 Exercises

1. Let $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{p} = (2, -1, 4)$. Define $f\colon \mathbb{R}^3 \to \mathbb{R}$ by

$$f(x,y,z) = xy^2z^4.$$

Compute $\mathbf{v}_\mathbf{p}[f]$.

2. Fill in the details of the proofs of Lemma 9.5, Lemma 9.8, Corollary 9.11 and Corollary 9.14.

3. Define $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ by $F(x,y,z) = (xy,\ yz,\ zx)$. Determine the following sets:

$$\begin{aligned} \mathcal{A} &= \{\, \mathbf{p} \in \mathbb{R}^3 \mid \|\mathbf{p}\| = 1,\ \ F(\mathbf{p}) = \mathbf{0} \,\}, \\ \mathcal{B} &= \{\, \mathbf{p} \in \mathbb{R}^3 \mid \|\mathbf{p}\| = 1,\ \ F_*(\mathbf{v}_\mathbf{p}) = \mathbf{0} \text{ for some } \mathbf{v}_\mathbf{p} \in \mathbb{R}^3_\mathbf{p} \,\}. \end{aligned}$$

4. Compute the gradients of the following functions:

(a) $(x,y,z) \mapsto e^{az}\cos ax - \cos ay$.

(b) $(x,y,z) \mapsto \sin z - \sinh x \sinh y$.

(c) $(x, y, z) \mapsto \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - c\,z.$

(d) $(x, y, z) \mapsto x^m + y^n + z^p.$

M **5.** Compute the gradient of

(a) $g_1(x, y, z) = x^2 + 2y^2 + 3z^2 - xz + yz - xy.$

(b) $g_2(x, y, z) = x^3 + y^3 + z^3 - 3axyz.$

(c) $g_3(x, y, z) = xyz\, e^{x+y+z}.$

(d) $g_4(x, y, z) = \arctan\left(x + y + z - \dfrac{xyz}{1 - xy - yz - xz}\right).$

6. Compute the divergence of each of the following vector fields:

(a) $(x, y, z) \mapsto \big(a\,yz\cos xy,\ byz\sin xy,\ cxy\big).$

(b) $(x, y, z) \mapsto \big(\sin(xy),\ \cos(yz),\ -3\cos^2(yz)\sin(xy) + \sin^3(xy)\big).$

(c) $(x, y, z) \mapsto \big((xyz)^a,\ (xyz)^b,\ (xyz)^c\big).$

(d) $(x, y, z) \mapsto \big(\log(x + y - z),\ e^{x-y+z},\ (-x + y + z)^5\big).$

M **7.** Compute the divergence of

(a) $\mathbf{V}_1(x, y, z) = (xyz,\ 2x + 3y + z,\ x^2 + z^2).$

(b) $\mathbf{V}_2(x, y, z) = (6x^2y^2 - z^3 + yz - 5,\ 4x^3y + xz + 2,\ xy - 3xz^2 - 3).$

8. Compute the Laplacian of the functions in Exercises 4 and 5.

9. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. A point $\mathbf{p} \in \mathbb{R}^n$ is said to be a ***critical point*** of f if $f_*\colon \mathbb{R}^n_{\mathbf{p}} \to \mathbb{R}_{f(\mathbf{p})}$ maps some nonzero tangent vector to zero. If $\mathbf{p}$ is a critical point of f, define the ***Hessian*** of f at $\mathbf{p}$ by

$$\mathsf{Hessian}[f]\big(\mathbf{v_p}, \mathbf{w_p}\big) = \mathbf{v_p}\big[\mathbf{W}[f]\big],$$

for $\mathbf{v_p}, \mathbf{w_p} \in \mathbb{R}^n_{\mathbf{p}}$. Here $\mathbf{W}$ is any vector field on $\mathbb{R}^n$ such that $\mathbf{W}(\mathbf{p}) = \mathbf{w_p}$.

(a) Show that $\mathsf{Hessian}[f]\big(\mathbf{v_p}, \mathbf{w_p}\big) = \displaystyle\sum_{i,j=1}^{n} v_i w_j \frac{\partial^2 f}{\partial u_i \partial u_j}.$

(b) Conclude that the definition of the Hessian is independent of the choice of the vector field $\mathbf{W}$.

10. Fill in the details of the proofs of Lemmas 9.30, 9.31 and 9.34.

Notebook 9

■ N9.0

This notebook is short, but it does include implementations of the various differential operators discussed in Chapter 9.

Loading the following package enables us to draw arrows on 2-dimensional plots:

```
<<Graphics`Arrow`
```

The next command is a bold variant of the function tx that we have used previously. It places text in the specified font at a point with position vector v.

```
txb[A_, v_]:= Text[FontForm[A,{"Helvetica-Bold",18}], v]
```

We shall continue to display a final plot using the command S, whose definition from N6.0 we repeat here with a protection against the same symbol being accidentally used for another purpose. Without the use of S or the DisplayFunction command each time, the programs for Figures 9.1 and 9.2 would display each arrow separately, before combining them into a single plot. This is acceptable for simple diagrams, but it can be confusing for more complicated ones.

```
$DisplayFunction= Identity
Unprotect[S]
  S[X__]:= Show[X, DisplayFunction->(Display[$Display,#1] &)]
Protect[S]
```

■ N9.1

The following objects are used in the chapter's first diagram. The vectors o,p,v are used to specify the arrows, whilst r is used to help offset the text. Data for the left-hand plot is given by z[1], and for the right-hand plot by z[2]. Both plots are then compiled into the function zz.

```
o:= {0,0}
v:= {3,2}
p:= {1,-1}
r:= {-.4,.8}
```

```
z[1]:= {Hue[0], Thickness[.01], Arrow[o,v,HeadWidth->.8],
  PointSize[.04], Point[o], txb["v",v/2+r]}
z[2]:= {Hue[.7], Thickness[.01],
  Arrow[p,v+p,HeadWidth->.8],
  PointSize[.04], Point[p], txb["v",v/2+p+r], txb["p",p-2r/3]}
```

The axes are added using the Show command:

```
zz[i_]:= Show[Graphics[z[i]], Axes->True, Ticks->None,
  AspectRatio->Automatic, PlotRange->{{-4,5},{-4,5}}]
S @ GraphicsArray[Array[zz,2], GraphicsSpacing->.1,
  ImageSize->450]
```

Figure 9.1

Similar commands are used in N9.1.1 in order to produce Figure 9.2.

■ N9.1.1

■ N9.4

The following program defines the gradient of a function f relative to an ordered set coor of coordinates:

```
grad[f_,coor_]:= Table[D[f,coor[[i]]], {i,Length[coor]}]
```

The theoretical definition can be inspected by inserting a generic function of three variables:

```
Clear[f]
grad[f[x,y,z], {x,y,z}]
```

More systematic notation is useful for four or more variables. The string @@ is shorthand for the command Apply, and has the practical effect of replacing curly brackets by square ones; it was already used in N8.6.

```
n:= 4
ay= Array[x,n]
grad[f@@ay, ay]
```

The Array command next is just a quicker way of using Table:

```
ay= Table[x[i], {i, n}]
grad[f@@ay, ay]
```

Here are some explicit examples:

```
grad[(x y z)^3, {x,y,z}]
```

```
grad[ArcTan[y/x], {x,y,z}]//Simplify
```

```
r:= Sqrt[x^2 + y^2 + z^2]
grad[r, {x,y,z}] //. r->d
```

```
s:= Sum[x[i], {i,n}]
p:= Product[x[i], {i,n}]
grad[s/p, ay]
Simplify[p %]
```

■ N9.5

We begin with the definitions of the divergence of a differentiable vector field and the Laplacian of a differentiable function.

```
divergence[vf_,coor_]:= Sum[D[vf[[i]],coor[[i]]], {i,Length[coor]}]
```

```
laplacian[f_,coor_]:= Sum[D[f,{coor[[i]],2}], {i,Length[coor]}]
```

An example is provided by

```
divergence[{x^a, y^b, z^c}, {x,y,z}]
```

The next computation will show that an appropriate power of r satisfies Laplace's equation in dimension n.

```
Do[ay= Array[x, n];
  r = Sqrt[Sum[x[i]^2, {i,n}]];
  Print[Simplify[laplacian[r^(-n+2), ay]]],
{n, 9}]
```

This confirms that the Laplacian has been defined correctly so as to equal the divergence of the gradient:

```
f:= (x y z)^3
divergence[grad[f, {x,y,z}], {x,y,z}]
laplacian[f, {x,y,z}]
```

■ N9.7

Exercise 1 is accomplished by

```
V[p_, f_]:= Sum[p[[i]] D[f,x[i]], {i, 3}]/.Table[x[i]->p[[i]], {i,3}]
V[{2,-1,4}, x[1] x[2]^2 x[3]^4]
```

In Exercise 3, we see that the zero set of F coincides with the coordinate axes, and therefore the first set $\mathcal{A}$ equals {(1,0,0), (0,1,0), (0,0,1)}. The condition on the differential of F is equivalent to the vanishing of the determinant

```
Det[{{y,0,z}, {x,z,0}, {0,y,x}}]
```

Since the result is twice the product of x,y,z, it follows that $\mathcal{B}$ consists of circles of radius 1 belonging to each of the three coordinate planes.

Chapter 10

Surfaces in Euclidean Space

The 2-dimensional analog of a curve is a surface. However, surfaces in general are much more complicated than curves. In this introductory chapter, we give basic definitions that will be used throughout the rest of the book.

The intuitive idea of a surface is a 2-dimensional set of points. Globally, the surface may be rather complicated and not look at all like a plane, but any sufficiently small piece of the surface should look like a "warped" portion of a plane. The most straightforward 2-dimensional generalization of a curve in $\mathbb{R}^n$ is a *patch* or local surface, which we define in Section 10.1. A given partial derivative of the patch function constitutes a column of the associated Jacobian matrix, which can then be used to characterize a *regular patch*.

Patches in $\mathbb{R}^3$ are discussed in Section 10.2, and the definition of an associated unit normal vector leads to a first encounter with the Gauss map of a surface, which is an analog of the mapping used in Chapters 1 and 6 to define the turning angle and turning number of a curve.

A second generalization of a curve in $\mathbb{R}^n$ is a *regular surface* in $\mathbb{R}^n$; this is a notion defined in Section 10.3 that adopts a more global perspective. A selection of surfaces in $\mathbb{R}^3$ is presented in Sections 10.4 and 10.6. As well as discussing familiar cases, we shall plot a few interesting but less well-known examples. In later chapters, we shall be showing how to compute associated geometric quantities, such as curvature, for all the surfaces that we have introduced.

In Section 10.5, we define a tangent vector to a regular surface, and a surface mapping; these are the surface analogs of a tangent vector to $\mathbb{R}^n$ and a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. As well as providing examples, Section 10.6 is devoted to the *nonparametric representation* of level surfaces as the zero sets of differentiable functions on $\mathbb{R}^3$. The Gauss map of such a surface is determined by the *gradient* of the defining function.

10.1 Patches in $\mathbb{R}^n$

Since a curve in $\mathbb{R}^n$ is a vector-valued function of one variable, it is reasonable to consider vector-valued functions of *two* variables. Such an object is called a patch. First, we give the precise definition.

Definition 10.1. *A **patch** or **local surface** is a differentiable mapping*

$$\mathbf{x}\colon \mathcal{U} \longrightarrow \mathbb{R}^n,$$

*where $\mathcal{U}$ is an open subset of $\mathbb{R}^2$. More generally, if $\mathcal{A}$ is any subset of $\mathbb{R}^2$, we say that a map $\mathbf{x}\colon \mathcal{A} \to \mathbb{R}^n$ is a patch provided that $\mathbf{x}$ can be extended to a differentiable mapping from $\mathcal{U}$ into $\mathbb{R}^n$, where $\mathcal{U}$ is an open set containing $\mathcal{A}$. We call $\mathbf{x}(\mathcal{U})$ (or more generally $\mathbf{x}(\mathcal{A})$) the **trace** of $\mathbf{x}$.*

Although the domain of definition of a patch can in theory be any set, more often than not it is an open or closed rectangle.

The standard parametrization of the ***sphere*** $S^2(a)$ of radius a is the mapping

$$\begin{aligned} \mathsf{sphere}[a]\colon [0,2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}] &\longrightarrow \mathbb{R}^3 \\ (u,v) &\mapsto \big(a\cos v\cos u,\ a\cos v\sin u,\ a\sin v\big). \end{aligned} \tag{10.1}$$

This is chosen so that u measures longitude and v measures latitude. Although $\mathsf{sphere}[a]$ is defined on the closed rectangle $\mathcal{R} = [0,2\pi] \times [-\pi/2, \pi/2]$, it has a differentiable extension to an open set containing $\mathcal{R}$. Any such extension will multiply cover substantial parts of $S^2(a)$, while the image of $\mathcal{R}$ multiply covers only a semicircle from the north pole $(0,0,a)$ to the south pole $(0,0,-a)$. By restricting the parametrization to an open *subset* of $\mathcal{R}$, it is possible to visualize the inside of the sphere.

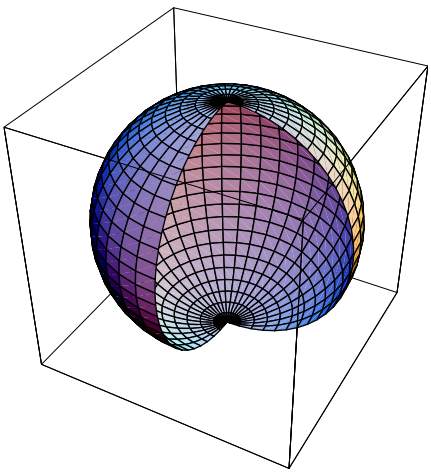


Figure 10.1: The image of $\mathsf{sphere}[1]$ for $0 \leqslant u \leqslant \dfrac{3\pi}{2}$

Since a patch can be written as an n-tuple of functions

$$\mathbf{x}(u,v) = \big(x_1(u,v), \ldots, x_n(u,v)\big), \tag{10.2}$$

we can define the partial derivative $\mathbf{x}_u$ of $\mathbf{x}$ with respect to u by

$$\mathbf{x}_u(u,v) = \left(\frac{\partial x_1}{\partial u}(u,v), \ldots, \frac{\partial x_n}{\partial u}(u,v)\right). \tag{10.3}$$

The other partial derivatives of $\mathbf{x}$, namely $\mathbf{x}_v, \mathbf{x}_{uu}, \mathbf{x}_{uv}, \ldots$, are defined similarly. Frequently, we abbreviate (10.2) and (10.3) to

$$\mathbf{x} = (x_1, \ldots, x_n) \qquad \text{and} \qquad \mathbf{x}_u = \left(\frac{\partial x_1}{\partial u}, \ldots, \frac{\partial x_n}{\partial u}\right).$$

The partial derivatives $\mathbf{x}_u$ and $\mathbf{x}_v$ can be expressed in terms of the tangent map of the patch $\mathbf{x}$.

Lemma 10.2. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a patch, and let $\mathbf{q} \in \mathcal{U}$. Then*

$$\mathbf{x}_*\big(\mathbf{e}_1(\mathbf{q})\big) = \mathbf{x}_u(\mathbf{q}) \qquad \textit{and} \qquad \mathbf{x}_*\big(\mathbf{e}_2(\mathbf{q})\big) = \mathbf{x}_v(\mathbf{q}),$$

where $\mathbf{x}_*$ denotes the tangent map of $\mathbf{x}$, and $\{\mathbf{e}_1, \mathbf{e}_2\}$ denotes the natural frame field of $\mathbb{R}^2$.

Proof. By Lemma 9.10, we have

$$\begin{aligned}\mathbf{x}_*\big(\,\mathbf{e}_1(\mathbf{q})\big) = \big(\mathbf{e}_1(\mathbf{q})[x_1], \ldots, \mathbf{e}_1(\mathbf{q})[x_n]\,\big)_{\mathbf{x}(\mathbf{q})} &= \left(\frac{\partial x_1}{\partial u}(\mathbf{q}), \ldots, \frac{\partial x_n}{\partial u}(\mathbf{q})\right)_{\mathbf{x}(\mathbf{q})} \\ &= \mathbf{x}_u(\mathbf{q}),\end{aligned}$$

and similarly for $\mathbf{x}_*\big(\mathbf{e}_2(\mathbf{q})\big)$. ∎

On page 271, we defined the Jacobian matrix of a differentiable map. We now specialize to the case of a patch.

Definition 10.3. *The **Jacobian matrix** of a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is the matrix-valued function $\mathscr{J}(\mathbf{x})$ given by*

$$\mathscr{J}(\mathbf{x})(u,v) = \begin{pmatrix} \frac{\partial x_1}{\partial u}(u,v) & \frac{\partial x_1}{\partial v}(u,v) \\ \vdots & \vdots \\ \vdots & \vdots \\ \frac{\partial x_n}{\partial u}(u,v) & \frac{\partial x_n}{\partial v}(u,v) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_u(u,v) \\ \mathbf{x}_v(u,v) \end{pmatrix}^T. \tag{10.4}$$

Observe that differentiation with respect to each of the two coordinates u, v corresponds to a *column* of (10.4), or a row of its transpose indicated with the superscript ^{T}. This is slightly inconvenient typographically, but consistent with the definition on page 271.

An equivalent definition of the ***rank*** of a matrix A is that it is the largest integer m such that A has an $m \times m$ submatrix whose determinant is nonzero. The following lemma is a consequence of well-known facts from linear algebra.

Lemma 10.4. *Let $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$. The following conditions are equivalent*:

(i) $\mathbf{p}$ *and* $\mathbf{q}$ *are linearly dependent*;

(ii) $\det\begin{pmatrix} \mathbf{p}\cdot\mathbf{p} & \mathbf{p}\cdot\mathbf{q} \\ \mathbf{q}\cdot\mathbf{p} & \mathbf{q}\cdot\mathbf{q} \end{pmatrix} = 0$;

(iii) *the $2 \times n$ matrix with rows $\mathbf{p}$, $\mathbf{q}$ has rank less than 2.*

Proof. If $\mathbf{p}$ and $\mathbf{q}$ are linearly dependent, then either $\mathbf{p}$ is a multiple of $\mathbf{q}$, or vice versa. For example, if $\mathbf{p} = \lambda\mathbf{q}$, then

$$\det\begin{pmatrix} \mathbf{p}\cdot\mathbf{p} & \mathbf{p}\cdot\mathbf{q} \\ \mathbf{q}\cdot\mathbf{p} & \mathbf{q}\cdot\mathbf{q} \end{pmatrix} = \det\begin{pmatrix} \lambda\mathbf{q}\cdot\lambda\mathbf{q} & \lambda\mathbf{q}\cdot\mathbf{q} \\ \lambda\mathbf{q}\cdot\mathbf{q} & \mathbf{q}\cdot\mathbf{q} \end{pmatrix} = \lambda^2 \det\begin{pmatrix} \mathbf{q}\cdot\mathbf{q} & \mathbf{q}\cdot\mathbf{q} \\ \mathbf{q}\cdot\mathbf{q} & \mathbf{q}\cdot\mathbf{q} \end{pmatrix} = 0.$$

Thus (i) implies (ii).

Next, suppose that condition (ii) holds. Let us write $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$. Then

$$\begin{aligned} 0 = \|\mathbf{p}\|^2 \|\mathbf{q}\|^2 - (\mathbf{p}\cdot\mathbf{q})^2 &= \left(\sum_{i=1}^n p_i^2\right)\left(\sum_{j=1}^n q_j^2\right) - \left(\sum_{i=1}^n p_i\, q_i\right)^2 \\ &= \sum_{1\leqslant i<j\leqslant n} (p_i q_j - p_j q_i)^2 . \end{aligned}$$

It follows that $p_i q_j = p_j q_i$ for all i and j. Hence (ii) implies (iii).

Finally, suppose (iii) holds. Then $p_i q_j = p_j q_i$ for all i and j. Without loss of generality, we can suppose that $q_i \neq 0$ for some i. Then $p_j = (p_i/q_i)q_j$ for all j, and

$$\mathbf{p} = (p_1, \ldots, p_n) = \frac{p_i}{q_i}(q_1, \ldots, q_n) = \frac{p_i}{q_i}\mathbf{q},$$

and $\mathbf{p}$ and $\mathbf{q}$ are linearly dependent. ∎

Corollary 10.5. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a patch. Then the following conditions are equivalent*:

(i) $\mathbf{x}_u(u_0, v_0)$ *and* $\mathbf{x}_v(u_0, v_0)$ *are linearly dependent*;

(ii) $\det\begin{pmatrix} \mathbf{x}_u \cdot \mathbf{x}_u & \mathbf{x}_u \cdot \mathbf{x}_v \\ \mathbf{x}_v \cdot \mathbf{x}_u & \mathbf{x}_v \cdot \mathbf{x}_v \end{pmatrix}$ *vanishes at* (u_0, v_0);

(iii) *the Jacobian matrix* $\mathscr{J}(\mathbf{x})$ *has rank less than* 2 *at* (u_0, v_0).

We shall need a stronger notion of patch in order that $\mathbf{x}(\mathcal{U})$ resemble the open set $\mathcal{U}$ more closely.

Definition 10.6. *A* ***regular patch*** *is a patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ *for which* $\mathscr{J}(\mathbf{x})(u, v)$ *has rank* 2 *for all* $(u, v) \in \mathcal{U}$. *An* ***injective patch*** *is a patch such that* $\mathbf{x}(u_1, v_1) = \mathbf{x}(u_2, v_2)$ *implies that* $u_1 = u_2$ *and* $v_1 = v_2$.

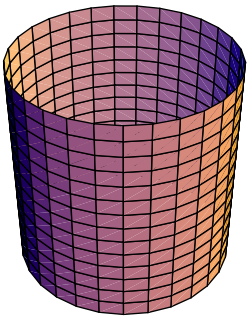


Figure 10.2: $(u, v) \mapsto (\cos u,\ \sin u,\ v)$

There are regular patches which are not injective. Consider, for example, the circular cylinder of Figure 10.2 defined by $\mathbf{x}(u, v) = (\cos u, \sin u, v)$, where $u \in \mathbb{R}$ and $-2 < v < 2$. There are also injective patches that are not regular. This is illustrated by the function $\mathbf{y}(u, v) = (u^3,\ v^3,\ uv)$ of Figure 10.3, for which $-1 < u, v < 1$.

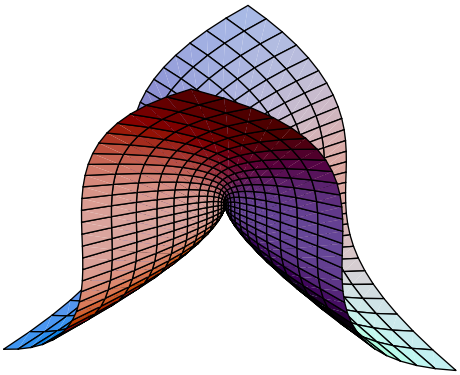


Figure 10.3: $(u, v) \mapsto (u^3,\ v^3,\ uv)$

The patch sphere$[a]$ fails to be regular when $v = \pm\pi/2$, that is, at the north and south poles. This is typical: often it is necessary to deal with patches that fail to be regular at just a few points. To handle such cases, we record the following obvious variant of Definition 10.6.

Definition 10.7. *A patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is **regular** at a point $(u_0, v_0) \in \mathcal{U}$ (or sometimes we say at $\mathbf{x}(u_0, v_0)$) provided $\mathscr{J}(\mathbf{x})(u_0, v_0)$ has rank 2.*

As a consequence of Lemma 10.2 and Corollary 10.5 we have

Lemma 10.8. *A patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is regular at $\mathbf{q} \in \mathcal{U}$ if and only if its tangent map $\mathbf{x}_*\colon \mathbb{R}^2_{\mathbf{q}} \to \mathbb{R}^n_{\mathbf{x}(\mathbf{q})}$ is injective.*

The following lemma is a consequence of the Inverse Function Theorem (as given, for example, in [Buck, page 276]), and refers to Definition 9.7 on page 268.

Lemma 10.9. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a regular patch and let $\mathbf{q} \in \mathcal{U}$. There exists a neighborhood $\mathcal{U}_{\mathbf{q}}$ of $\mathbf{q}$ such that $\mathbf{x}\colon \mathcal{U}_{\mathbf{q}} \to \mathbf{x}(\mathcal{U}_{\mathbf{q}})$ is the restriction of a diffeomorphism between open sets of $\mathbb{R}^n$.*

Proof. Write $\mathbf{x} = (x_1, \ldots, x_n)$. Since $\mathbf{x}$ is regular, its Jacobian matrix has a 2×2 submatrix with nonzero determinant. By renaming $x_1, \ldots, x_n$ if necessary, we can suppose that

$$\det \begin{pmatrix} \dfrac{\partial x_1}{\partial u}(u,v) & \dfrac{\partial x_1}{\partial v}(u,v) \\ \dfrac{\partial x_2}{\partial u}(u,v) & \dfrac{\partial x_2}{\partial v}(u,v) \end{pmatrix} \tag{10.5}$$

is nonzero for $(u, v) \in \mathcal{U}$. We extend $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ to a map

$$\widetilde{\mathbf{x}}\colon \mathcal{U} \times \mathbb{R}^{n-2} \longrightarrow \mathbb{R}^n \tag{10.6}$$

by defining

$$\widetilde{\mathbf{x}}(u, v, t_3, \ldots, t_n) = \big(x_1(u,v),\ x_2(u,v),\ x_3(u,v)+t_3, \ldots, x_n(u,v)+t_n\big).$$

It is clear that $\widetilde{\mathbf{x}}$ is differentiable; moreover, the determinant

$$\det\Big(\mathscr{J}(\widetilde{\mathbf{x}})^T\Big) = \det \begin{pmatrix} \dfrac{\partial x_1}{\partial u}(u,v) & \dfrac{\partial x_2}{\partial u}(u,v) & \dfrac{\partial x_3}{\partial u}(u,v) & \cdots & \dfrac{\partial x_n}{\partial u}(u,v) \\ \dfrac{\partial x_1}{\partial v}(u,v) & \dfrac{\partial x_2}{\partial v}(u,v) & \dfrac{\partial x_3}{\partial v}(u,v) & \cdots & \dfrac{\partial x_n}{\partial v}(u,v) \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

equals (10.5). Because $\det\big(\mathscr{J}(\widetilde{\mathbf{x}})\big)(\mathbf{q}) \neq 0$, the Inverse Function Theorem says that there is a neighborhood $\widetilde{\mathcal{U}}_{\mathbf{q}}$ of $(\mathbf{q}, \mathbf{0})$ upon which $\widetilde{\mathbf{x}}$ has a differentiable inverse. It follows that $\widetilde{\mathbf{x}}\colon \widetilde{\mathcal{U}}_{\mathbf{q}} \to \widetilde{\mathbf{x}}(\widetilde{\mathcal{U}}_{\mathbf{q}})$ is a diffeomorphism, and we may take $\mathcal{U}_{\mathbf{q}} = \widetilde{\mathcal{U}}_{\mathbf{q}} \cap \mathcal{U}$ to complete the proof. ▮

The conclusion of Lemma 10.9 can be expressed by saying that $\mathbf{x}\colon \mathcal{U}_{\mathbf{q}} \to \mathbf{x}(\mathcal{U}_{\mathbf{q}})$ is itself a diffeomorphism, anticipating Definition 10.42 on page 309. We may also state

Corollary 10.10. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be an injective regular patch. Then $\mathbf{x}$ maps $\mathcal{U}$ diffeomorphically onto $\mathbf{x}(\mathcal{U})$.*

Associated with any patch are some naturally-defined curves.

Definition 10.11. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a patch, and fix $(u_0, v_0) \in \mathcal{U}$. The curves*

$$u \mapsto \mathbf{x}(u, v_0) \qquad \text{and} \qquad v \mapsto \mathbf{x}(u_0, v)$$

*are called u- and v- **parameter curves** or **coordinate curves**.*

These are the curves that are usually displayed by computer graphics.

There is also a convenient way to represent a general curve whose trace is contained in the trace of a patch.

Lemma 10.12. *Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ be a curve whose trace lies on the image $\mathbf{x}(\mathcal{U})$ of a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ such that $\mathbf{x}\colon \mathcal{U} \to \mathbf{x}(\mathcal{U})$ is a homeomorphism. Then there exist unique differentiable functions $u, v\colon (a, b) \to \mathbb{R}$ such that*

$$\boldsymbol{\alpha}(t) = \mathbf{x}\big(u(t), v(t)\big) \tag{10.7}$$

for $a < t < b$.

Proof. For $a < t < b$ we can write

$$(\mathbf{x}^{-1} \circ \boldsymbol{\alpha})(t) = \big(u(t), v(t)\big);$$

this equation is equivalent to (10.7). It is clear that u and v are unique; by Lemma 10.9 they are also differentiable. ▮

Now we can define the notion of tangent vector to a patch.

Definition 10.13. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be an injective patch, and let $\mathbf{p} \in \mathbf{x}(\mathcal{U})$. A **tangent vector** to $\mathbf{x}$ at $\mathbf{p}$ is a tangent vector $\mathbf{v}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ for which there exists a curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ that can be written as*

$$\boldsymbol{\alpha}(t) = \mathbf{x}\big(u(t), v(t)\big) \qquad (a < t < b), \tag{10.8}$$

such that $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \mathbf{v}_{\mathbf{p}}$. We denote the set of tangent vectors to $\mathbf{x}$ at $\mathbf{p}$ by $\mathbf{x}(\mathcal{U})_{\mathbf{p}}$.

Lemma 10.14. *The set $\mathbf{x}(\mathcal{U})_{\mathbf{p}}$ of all tangent vectors to a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ at a regular point $\mathbf{p} = \mathbf{x}(u_0, v_0) \in \mathbf{x}(\mathcal{U})$ forms a vector space that is spanned by $\mathbf{x}_u(u_0, v_0)$ and $\mathbf{x}_v(u_0, v_0)$.*

Proof. By definition of tangent vector $\mathbf{v}_{\mathbf{p}}$ to $\mathbf{x}$, there exists a curve $\boldsymbol{\alpha}$ of the form (10.8) such that $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \mathbf{v}_{\mathbf{p}}$. The chain rule for curves (Lemma 1.9, page 7) implies that

$$\boldsymbol{\alpha}'(t) = u'(t)\mathbf{x}_u(u(t), v(t)) + v'(t)\mathbf{x}_v(u(t), v(t)).$$

In particular,

$$\mathbf{v}_{\mathbf{p}} = \boldsymbol{\alpha}'(0) = u'(0)\mathbf{x}_u(u_0, v_0) + v'(0)\mathbf{x}_v(u_0, v_0).$$

Conversely, if $\mathbf{v}_{\mathbf{p}} = c_1\mathbf{x}_u(u_0, v_0) + c_2\mathbf{x}_v(u_0, v_0)$, then $\mathbf{v}_{\mathbf{p}}$ is the velocity vector at $\mathbf{x}(u_0, v_0) = \mathbf{p}$ of the curve $t \mapsto \mathbf{x}(u_0 + tc_1,\ v_0 + tc_2)$. ∎

The proof of Lemma 10.14 yields

Corollary 10.15. *Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ be a curve whose trace lies on the image $\mathbf{x}(\mathcal{U})$ of a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ such that $\mathbf{x}\colon \mathcal{U} \to \mathbf{x}(\mathcal{U})$ is a homeomorphism. Then there exist unique differentiable functions $u, v\colon (a, b) \to \mathbb{R}$ such that*

$$\boldsymbol{\alpha}' = u'\mathbf{x}_u + v'\mathbf{x}_v. \tag{10.9}$$

Definition 10.16. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be an injective patch, and let $\mathbf{z}_{\mathbf{p}} \in \mathbb{R}^n_{\mathbf{p}}$ with $\mathbf{p} \in \mathbf{x}(\mathcal{U})$. We say that $\mathbf{z}_{\mathbf{p}}$ is **normal** or **perpendicular** to $\mathbf{x}$ at $\mathbf{p}$, provided $\mathbf{z}_{\mathbf{p}} \cdot \mathbf{v}_{\mathbf{p}} = 0$ for all vectors $\mathbf{v}_{\mathbf{p}}$ tangent to $\mathbf{x}$ at $\mathbf{p}$.*

Let $\mathbf{x}(\mathcal{U})^{\perp}_{\mathbf{p}}$ denote the ***orthogonal complement*** of $\mathbf{x}(\mathcal{U})_{\mathbf{p}}$, namely the space of vectors in $\mathbb{R}^n_{\mathbf{p}}$ perpendicular to $\mathbf{x}(\mathcal{U})_{\mathbf{p}}$. It is easy to verify that there is a direct sum

$$\mathbb{R}^n_{\mathbf{p}} = \mathbf{x}(\mathcal{U})_{\mathbf{p}} \oplus \mathbf{x}(\mathcal{U})^{\perp}_{\mathbf{p}} \tag{10.10}$$

of vector spaces.

We shall also need vector fields on patches:

Definition 10.17. *A **vector field** $\mathbf{V}$ on a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is a function that assigns to each $\mathbf{q} \in \mathcal{U}$ a tangent vector $\mathbf{V}(\mathbf{q}) \in \mathbb{R}^n_{\mathbf{p}}$, where $\mathbf{p} = \mathbf{x}(\mathbf{q})$. We say that $\mathbf{V}$ is **tangent** to $\mathbf{x}$ if $\mathbf{V}(\mathbf{q}) \in \mathbf{x}(\mathcal{U})_{\mathbf{p}}$ for all $\mathbf{q} \in \mathcal{U}$. Similarly, a vector field $\mathbf{W}$ on $\mathbf{x}$ is **normal** or **perpendicular** to $\mathbf{x}$ if $\mathbf{W}(\mathbf{q}) \cdot \mathbf{v}_{\mathbf{p}} = 0$ for all $\mathbf{v}_{\mathbf{p}} \in \mathbf{x}(\mathcal{U})_{\mathbf{p}}$ and $\mathbf{q} \in \mathcal{U}$.*

10.2 Patches in $\mathbb{R}^3$ and the Local Gauss Map

We now restrict our attention to patches in $\mathbb{R}^3$ because they are the ones that are easiest to visualize. Computations are simplified by using the vector cross product on $\mathbb{R}^3$. Here is a useful criterion for the regularity of a patch:

Lemma 10.18. *A patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ is regular at $(u_0, v_0) \in \mathcal{U}$ if and only if $\mathbf{x}_u \times \mathbf{x}_v$ is nonzero at (u_0, v_0).*

Proof. This follows immediately from the equation

$$\mathbf{x}_u \times \mathbf{x}_v = \det \begin{pmatrix} \dfrac{\partial x_1}{\partial u} & \dfrac{\partial x_1}{\partial v} & \mathbf{i} \\ \dfrac{\partial x_2}{\partial u} & \dfrac{\partial x_2}{\partial v} & \mathbf{j} \\ \dfrac{\partial x_3}{\partial u} & \dfrac{\partial x_3}{\partial v} & \mathbf{k} \end{pmatrix},$$

confirming that $(\mathbf{x}_u \times \mathbf{x}_v)(u_0, v_0) = \mathbf{0}$ if and only if the rank of $\mathscr{J}(\mathbf{x})(u_0, v_0)$ is less than 2. ∎

The vector cross product gives rise to a convenient way of finding a vector satisfying Definition 10.16 when $n = 3$.

Lemma 10.19. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be an injective regular patch. Then the vector field $(u, v) \mapsto \mathbf{x}_u \times \mathbf{x}_v$ is everywhere perpendicular to $\mathbf{x}(\mathcal{U})$.*

Proof. Certainly, $\mathbf{x}_u \times \mathbf{x}_v$ is perpendicular to both $\mathbf{x}_u$ and $\mathbf{x}_v$. Since any tangent vector $\mathbf{v_p}$ to $\mathbf{x}$ is a linear combination of $\mathbf{x}_u$ and $\mathbf{x}_v$ at $\mathbf{p}$, it follows from Lemma 10.14 that $(u, v) \mapsto \mathbf{x}_u \times \mathbf{x}_v$ is perpendicular to $\mathbf{x}(\mathcal{U})$. ∎

Now we can define a perpendicular vector field of unit length.

Definition 10.20. *For an injective patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ the* ***unit normal vector field*** *or* ***surface normal*** *$\mathbf{U}$ is defined by*

$$\mathbf{U}(u, v) = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}(u, v) \tag{10.11}$$

at those points $(u, v) \in \mathcal{U}$ at which $\mathbf{x}_u \times \mathbf{x}_v$ does not vanish.

The notion of regularity of a patch has the following geometric interpretation.

Corollary 10.21. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be an injective patch. Then $\mathbf{x}$ is regular if and only if the unit normal vector field $\mathbf{U}$ is everywhere well defined.*

The points for which $\mathbf{x}_u \times \mathbf{x}_v$ vanishes will be called ***singular*** in the next section.

One of the key elements necessary for the study of surfaces is the map that assigns to each point $\mathbf{p}$ on a surface $\mathcal{M} \subset \mathbb{R}^3$ the point on the unit sphere $S^2(1) \subset \mathbb{R}^3$ that is parallel to the unit normal $\mathbf{U}(\mathbf{p})$. It is called the ***Gauss***[1] ***map***. It is an analog of the mapping illustrated in Figure 6.2 on page 157, that associates to a point on a curve its unit tangent vector. It is not always possible to define this map at all points of a surface, for reasons that we take up in Chapter 11, but it *is* defined for any patch.

Definition 10.22. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be an injective patch. Then the unit normal $\mathbf{U}$ to $\mathbf{x}$ viewed as a mapping from $\mathcal{U}$ to the unit sphere $S^2(1) \subset \mathbb{R}^3$ is called the* ***local Gauss map*** *of $\mathbf{x}$.*

Note that $\mathbf{U}$ is undefined at singular points of $\mathbf{x}$.

We shall illustrate this concept for a well-known surface, that will be discussed again later in this chapter. The ***hyperbolic paraboloid*** is the trace of the patch

$$\mathbf{x}(u, v) = (u,\ v,\ uv). \tag{10.12}$$

Figure 10.4 shows the image of the rectangle $[-1, 1] \times [-1, 1]$, superimposed with the corresponding portion of the unit sphere $S^2(1)$ determined by the unit normal $\mathbf{U}(u, v)$. The patch (10.12) is closely related to the more general one (10.16) given below, and the hyperbolic paraboloid is an example of a ruled surface (see page 434 in Chapter 14).



[1] Carl Friedrich Gauss (1777–1855). Gauss's *Disquisitiones Generales Circa Superficies Curvas* (published in 1828) revolutionized differential geometry as his *Disquisitiones Arithmeticae* had revolutionized number theory. In particular, his approach depended only on the intrinsic properties of surfaces, not on their embedding in Euclidean 3-space. Gauss's genius extended into nearly every branch of mathematics. While a student at Göttingen he made his first major original discovery — the constructibility of the 17-sided regular polygon, closely followed by the first of his four proofs of the fundamental theorem of algebra (his doctoral dissertation) and his calculation of the orbit of the newly discovered asteroid Ceres. Gauss's interest in geodesy led to the invention of the heliotrope and the development of least squares approximation, a technique he also used in his investigation of the distribution of prime numbers. Gauss's work on terrestrial magnetism led to his construction, in collaboration with Wilhelm Weber, of the first operating electric telegraph. Other applied discoveries of significance were those in optics, potential theory, and astronomy. Gauss spent his entire career in Brunswick, under the patronage of the local royalty; for many years he directed the observatory there. Perhaps due to his deeply ingrained conservatism or to his distaste for controversy, Gauss did not publish his development of the theory of non-Euclidean geometry nor his work anticipating Hamilton's investigation of quaternions. Although he worked for the most part in mathematical isolation, his results laid the basis for new departures in number theory and statistics, as well as in differential geometry.

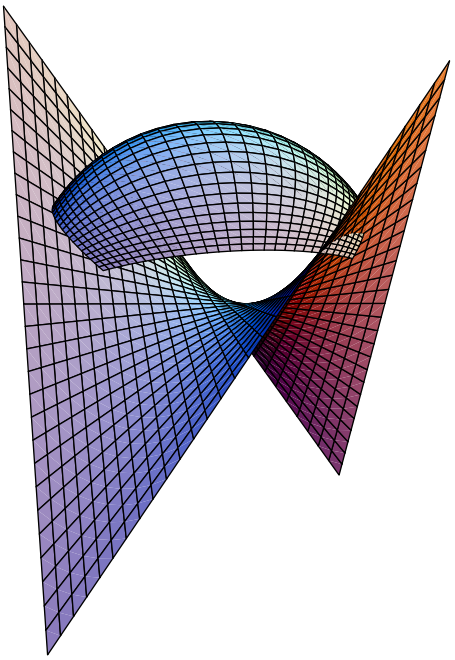


Figure 10.4: Part of the hyperboloid and its Gauss image

10.3 The Definition of a Regular Surface

There are some subsets of $\mathbb{R}^n$ that we would like to call surfaces, but they do not fit into the framework of Section 10.1. Such a subset cannot be expressed as the image of a single regular one-to-one patch defined on an open set; spheres and tori are examples. In this section we define the notion of regular surface; roughly speaking, what we shall do is to combine several patches. The resulting definition is more complicated, but it is what is needed in many situations.

Definition 10.23. *A subset $\mathcal{M} \subset \mathbb{R}^n$ is a* ***regular surface****, provided that for each point $\mathbf{p} \in \mathcal{M}$ there exist a neighborhood $\mathcal{V}$ of $\mathbf{p}$ in $\mathbb{R}^n$ and a map $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ of an open set $\mathcal{U} \subset \mathbb{R}^2$ onto $\mathcal{V} \cap \mathcal{M}$ such that*:

(i) $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *is a regular patch*;

(ii) $\mathbf{x}\colon \mathcal{U} \to \mathcal{V} \cap \mathcal{M}$ *is a homeomorphism; thus* $\mathbf{x}$ *has a continuous inverse* $\mathbf{x}^{-1}\colon \mathcal{V} \cap \mathcal{M} \to \mathcal{U}$ *such that* $\mathbf{x}^{-1}$ *is the restriction to* $\mathcal{V} \cap \mathcal{M}$ *of a continuous mapping* $F\colon \mathcal{W} \to \mathbb{R}^2$*, where* $\mathcal{W}$ *is an open subset of* $\mathbb{R}^n$ *that contains* $\mathcal{V} \cap \mathcal{M}$*.*

The map $\mathbf{x}^{-1}\colon \mathcal{V} \cap \mathcal{M} \to \mathbb{R}^2$ *is called a* ***chart*** *or a* ***system of local coordinates*** *on the regular surface* $\mathcal{M}$*.*

We shall need the following fact.

Lemma 10.24. *Let $\mathcal{W}$ be an open subset of $\mathbb{R}^n$, and suppose that $G\colon \mathcal{W} \to \mathbb{R}^m$ is a map such that $G\colon \mathcal{W} \to G(\mathcal{W})$ is a diffeomorphism. If $\mathcal{M} \subseteq \mathcal{W}$ is a regular surface, then $G(\mathcal{M})$ is also a regular surface.*

Proof. If $\mathbf{x}$ is a patch on $\mathcal{M}$ satisfying (i) and (ii) in the definition of regular surface, then $G \circ \mathbf{x}$ also satisfies (i) and (ii). ▌

There is an easy way to find new regular surfaces inside a given regular surface $\mathcal{M}$. A subset of a regular surface $\mathcal{M} \subset \mathbb{R}^n$ is said to be ***open*** provided it is the intersection of an open subset of $\mathbb{R}^n$ with $\mathcal{M}$.

Lemma 10.25. *An open subset $\mathcal{W}$ of a regular surface $\mathcal{M}$ is also a regular surface.*

Proof. Let $\mathbf{x}$ be a patch on $\mathcal{M}$ satisfying (i) and (ii) in the definition of regular surface. Then $\mathbf{x}$ is continuous, and the inverse image of an open set is open, so $\mathcal{U} = \mathbf{x}^{-1}(\mathcal{W})$ is open. If $\mathcal{U}$ is nonempty, the restriction $\mathbf{x}|\mathcal{U}$ becomes a patch on $\mathcal{W}$. It is clear that $\mathbf{x}|\mathcal{U}$ satisfies (i) and (ii) in the definition of regular surface. Thus $\mathcal{W}$ becomes a regular surface. ▌

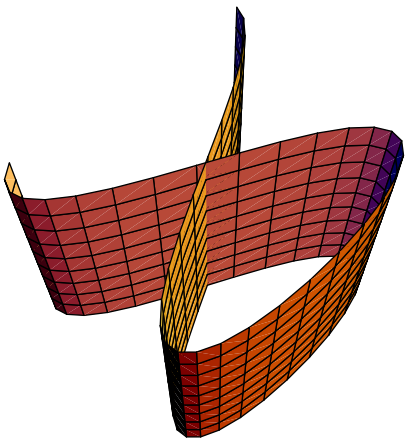


Figure 10.5: The patch $(u, v) \mapsto (\sin u, \sin 2u, v)$ with $-\dfrac{\pi}{3} \leqslant u \leqslant \dfrac{5\pi}{4}$

Let us clarify the relation between patches and regular surfaces. Certainly an arbitrary patch will fail to be a regular surface if its trace is 0- or 1-dimensional. Figure 10.5 shows that even a regular patch can fail to be a regular surface if it has self-intersections (which are precluded by condition (ii) in Definition 10.23). Nevertheless, let us prove that when we restrict the domain of definition of a regular patch we obtain a regular surface.

Lemma 10.26. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a regular patch. For any $\mathbf{q} \in \mathcal{U}$ there exists a neighborhood $\mathcal{U}_\mathbf{q}$ of $\mathbf{q}$ such that $\mathbf{x}(\mathcal{U}_\mathbf{q})$ is a regular surface.*

Proof. Lemma 10.9 states that $\mathbf{q}$ has a neighborhood $\mathcal{U}_\mathbf{q}$ such that $\mathbf{x}$ is a diffeomorphism between $\mathcal{U}_\mathbf{q}$ and $\mathbf{x}(\mathcal{U}_\mathbf{q})$. The neighborhood $\mathcal{U}_\mathbf{q}$ is an open set in the plane, and hence a regular surface by Lemma 10.25. Then Lemma 10.24 implies that $\mathbf{x}(\mathcal{U}_\mathbf{q})$ is a regular surface. ∎

Corollary 10.27. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a regular injective patch. Then $\mathbf{x}(\mathcal{U}_\mathbf{q})$ is a regular surface.*

Definition 10.28. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^n$. If there exists a single regular injective patch $\mathbf{x} : \mathcal{U} \to \mathbb{R}^n$ such that $\mathbf{x}(\mathcal{U}) = \mathcal{M}$, we say that $\mathcal{M}$ is* ***parametrized by*** $\mathbf{x}$.

This definition applies to the case of a Monge patch defined in Section 10.4 below. On the other hand, a sphere is an example of a regular surface which needs at least two patches to cover it. Figure 10.6 shows that the patches

$$\begin{aligned}(u,v) &\mapsto \big(\cos u\cos v,\ \sin u\cos v,\ \sin v\big)\\ (u,v) &\mapsto \big(\cos\big(\tfrac{3}{5}\pi - u\big)\cos v,\ \sin v,\ \sin\big(\tfrac{3}{5}\pi - u\big)\cos v\big),\end{aligned}$$

both defined for $0 < u < 7\pi/4$ and $-\pi/2 < v < \pi/2$, cover the sphere. For plotting purposes, the second patch is best regarded as the composition of the first with a rotation.

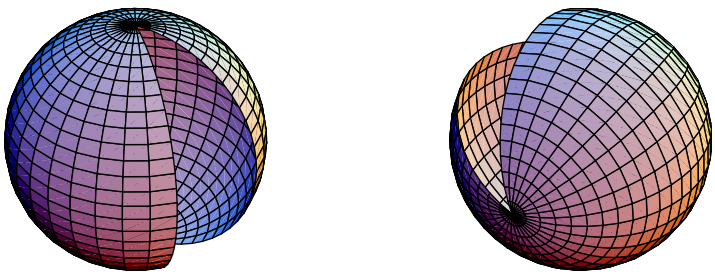


Figure 10.6: Two patches covering the sphere

Next, we establish a fact about differentiable maps into a regular surface.

Theorem 10.29. *Let $\mathcal{V} \subset \mathbb{R}^m$ be an open subset, and $\mathcal{M} \subset \mathbb{R}^n$ a regular surface. Suppose $F\colon \mathcal{V} \to \mathbb{R}^n$ is a differentiable mapping such that $F(\mathcal{V}) \subseteq \mathbf{x}(\mathcal{U})$, where $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a regular patch on $\mathcal{M}$. Then $\mathbf{x}^{-1}\circ F\colon \mathcal{V} \to \mathcal{U}$ is differentiable.*

Proof. It is tempting to say that since $\mathbf{x}^{-1}$ and F are differentiable, so is $\mathbf{x}^{-1} \circ F$. Unfortunately, this is not valid, because we do not know at this point what it means for a function defined on a regular surface to be differentiable. Therefore, we are forced to prove the theorem in a more roundabout fashion, modifying the proof of Lemma 10.9.

Write $\mathbf{x} = (x_1, \ldots, x_n)$. Since $\mathbf{x}$ is regular, its Jacobian matrix has a 2×2 submatrix with nonzero determinant. By renaming $x_1, \ldots, x_n$ if necessary, we can suppose that (10.5) holds, and we define (10.6) in exactly the same way. Once again, $\widetilde{\mathbf{x}}$ is differentiable, and

$$\det\big(\mathscr{J}(\widetilde{\mathbf{x}})(u,v)\big) \neq 0.$$

Let $\mathbf{p} \in \mathcal{V}$; then $F(\mathbf{p}) \in \mathbf{x}(\mathcal{U}) \subseteq \mathcal{M}$. Because $\det\big(\mathscr{J}(\widetilde{\mathbf{x}})\big) \neq 0$ on $\mathcal{U}$, the Inverse Function Theorem says that $\widetilde{\mathbf{x}}$ has an inverse $\widetilde{\mathbf{x}}^{-1}$ on a neighborhood $\widetilde{\mathcal{U}} \subseteq \mathbb{R}^n$ of $F(\mathbf{p})$, and that $\widetilde{\mathbf{x}}^{-1}$ is differentiable on $\widetilde{\mathcal{U}}$. Since F is continuous, there exists a neighborhood $\widetilde{\mathcal{V}}$ of $\mathbf{p}$ such that $F(\widetilde{\mathcal{V}}) \subset \widetilde{\mathcal{U}}$. Moreover,

$$\widetilde{\mathbf{x}}^{-1} \circ F\,|\,\widetilde{\mathcal{V}} = \mathbf{x}^{-1} \circ F\,|\,\widetilde{\mathcal{V}}.$$

Since $\widetilde{\mathbf{x}}^{-1} \circ F\,|\,\widetilde{\mathcal{V}}$ is differentiable at $\mathbf{p}$, so is $\mathbf{x}^{-1} \circ F\,|\,\widetilde{\mathcal{V}}$. Since $\mathbf{p} \in \mathcal{V}$ is arbitrary, it follows that $\mathbf{x}^{-1} \circ F$ is differentiable on all of $\mathcal{V}$. ∎

An important special case of Theorem 10.29 is:

Corollary 10.30. *Let $\mathcal{M}$ be a regular surface, and suppose we are given regular patches $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ such that $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V}) = \mathcal{W}$ is nonempty. Then the* ***change of coordinates***

$$\mathbf{x}^{-1} \circ \mathbf{y} :\ \mathbf{y}^{-1}(\mathcal{W}) \longrightarrow \mathbf{x}^{-1}(\mathcal{W}) \tag{10.13}$$

is a diffeomorphism between open subsets of $\mathbb{R}^2$.

Proof. That $\mathbf{x}^{-1} \circ \mathbf{y}$ and $\mathbf{y}^{-1} \circ \mathbf{x}$ are differentiable is a consequence of Theorem 10.29. Since both $\mathbf{x}$ and $\mathbf{y}$ are homeomorphisms, so are $\mathbf{x}^{-1} \circ \mathbf{y}$ and $\mathbf{y}^{-1} \circ \mathbf{x}$. Moreover, these maps are inverses of each other. Hence $\mathbf{x}^{-1} \circ \mathbf{y}$ and $\mathbf{y}^{-1} \circ \mathbf{x}$ are diffeomorphisms. ∎

We can write (10.13) more explicitly as follows: there exist differentiable functions $\bar{u}, \bar{v}$ such that

$$\mathbf{y}(u,v) = \mathbf{x}\big(\bar{u}(u,v),\ \bar{v}(u,v)\big).$$

Note that $(\bar{u}, \bar{v}) = (\mathbf{x}^{-1} \circ \mathbf{y})(u,v)$.

Lemma 10.31. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ be patches on a regular surface $\mathcal{M}$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ nonempty. Let $\mathbf{x}^{-1} \circ \mathbf{y} = (\bar{u}, \bar{v})\colon \mathcal{U} \cap \mathcal{V} \to \mathcal{U} \cap \mathcal{V}$ be the associated change of coordinates, so that*

$$\mathbf{y}(u,v) = \mathbf{x}\big(\bar{u}(u,v),\ \bar{v}(u,v)\big). \tag{10.14}$$

Then

$$\mathbf{y}_u = \frac{\partial \bar{u}}{\partial u}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_{\bar{v}} \quad \textit{and} \quad \mathbf{y}_v = \frac{\partial \bar{u}}{\partial v}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial v}\mathbf{x}_{\bar{v}}. \tag{10.15}$$

Proof. (10.15) is an immediate consequence of (10.14) and the chain rule. ∎

Usually there are only a few points on a general surface which cannot be in the image of a regular patch. We refer to these as ***nonregular points*** or ***singular points*** of the surface. All other points of the surface are called ***regular points***. Note that by Corollary 10.30, the definition of regular point and singular point does not depend on the choice of patch. Furthermore, if we remove all of the nonregular points from a general surface, we obtain a surface.

We need to extend the calculus that we developed in Chapter 9 for $\mathbb{R}^n$ to regular surfaces. Our first task is to define what it means for a real-valued function on a regular surface to be differentiable.

Definition 10.32. *Let $f\colon \mathcal{W} \to \mathbb{R}$ be a function defined on an open subset $\mathcal{W}$ of a regular surface $\mathcal{M}$. We say that f is **differentiable** at $\mathbf{p} \in \mathcal{W}$ provided that for some patch $\mathbf{x}\colon \mathcal{U} \subset \mathbb{R}^2 \to \mathcal{M}$ with $\mathbf{p} \in \mathbf{x}(\mathcal{U}) \subset \mathcal{W}$, the composition*

$$f \circ \mathbf{x}\colon \mathcal{U} \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$$

*is differentiable at $\mathbf{x}^{-1}(\mathbf{p})$. If f is differentiable at all points of $\mathcal{W}$, we say that f is **differentiable** on $\mathcal{W}$.*

It is important to realize that the definition of differentiability of a real-valued function on a regular surface does not depend on the choice of patch. If $\mathbf{x}$ and $\mathbf{y}$ are patches on a regular surface $\mathcal{M}$, then $\mathbf{y}^{-1} \circ \mathbf{x}$ is differentiable by Corollary 10.30, and the composition of differentiable functions is differentiable.

Lemma 10.33. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^n$. Then the restriction $f|\mathcal{M}$ of any differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ to $\mathcal{M}$ is differentiable.*

Proof. If $\mathbf{x}$ is any patch on $\mathcal{M}$, then $(f|\mathcal{M}) \circ \mathbf{x} = f \circ \mathbf{x}$ is differentiable, since it is the composition of differentiable functions. Thus by definition $f|\mathcal{M}$ is differentiable. ∎

As a simple consequence of Lemma 10.33, we see that the restrictions to any regular surface $\mathcal{M} \subset \mathbb{R}^n$ of the natural coordinate functions $u_1, \ldots, u_n$ of $\mathbb{R}^n$ are differentiable.

Here are two important examples of differentiable functions.

Definition 10.34. *Let $\mathcal{M}$ be a regular surface and let $\mathbf{v} \in \mathbb{R}^n$. The **height function** and the **square of the distance function** of $\mathcal{M}$ relative to $\mathbf{v}$ are the functions $h, f\colon \mathcal{M} \to \mathbb{R}$ defined by*

$$h(\mathbf{p}) = \mathbf{p} \cdot \mathbf{v} \qquad \textit{and} \qquad f(\mathbf{p}) = \|\mathbf{p} - \mathbf{v}\|^2.$$

Both the height and the square of the distance function are algebraic combinations of the natural coordinate functions of $\mathbb{R}^n$. Their differentiability therefore follows from Lemma 10.33.

We shall also need the notion of a curve on a regular surface.

Definition 10.35. *A* ***curve on a regular surface*** $\mathcal{M} \subset \mathbb{R}^n$ *is simply a curve* $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ *such that* $\boldsymbol{\alpha}(t) \in \mathcal{M}$ *for* $a < t < b$. *The curve* $\boldsymbol{\alpha}$ *is said to be* ***differentiable*** *provided that* $\mathbf{x}^{-1} \circ \boldsymbol{\alpha}\colon (a,b) \to \mathcal{U}$ *is differentiable for any patch* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$, *where* $\mathcal{U} \subseteq \mathbb{R}^2$.

10.4 Examples of Surfaces

An important class of surfaces in $\mathbb{R}^3$ consists of those that are graphs of a real-valued function of two variables.

Definition 10.36. *A* ***Monge[2] patch*** *is a patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *of the form*

$$\mathbf{x}(u,v) = \big(u,\ v,\ h(u,v)\big),$$

where $\mathcal{U}$ *is an open set in* $\mathbb{R}^2$ *and* $h\colon \mathcal{U} \to \mathbb{R}$ *is a differentiable function.*

This is a regular patch, because the Jacobian matrix

$$\mathscr{J}(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{\partial h}{\partial u} & \dfrac{\partial h}{\partial v} \end{pmatrix}$$

obviously has rank 2. Its trace is the graph $\mathcal{M}_h$ of h, whose nonparametric form is simply $z = h(x,y)$.

---

[2] 

Gaspard Monge (1746–1818). French mathematician, erstwhile secretary of the navy, founding director of the École Polytechnique, which played a leading role in the development and organization of scientific research and education in France. Monge's work on fortifications led him to descriptive geometry, and from there he went on to a broad exposition of the differential geometry of space curves. His major contribution to the development of differential geometry was the integration of geometrical facts and intuition with the use of partial differential equations. A fervent supporter first of the Revolution and then of Napoleon, Monge was expelled from the Institut de France following the Battle of Waterloo. Many consider Monge to be the father of French differential geometry.

Paraboloids and Monkey Saddles

The concept of Monge patch is well illustrated by taking $h(u, v)$ proportional to $au^2 + bv^2$. The trace of the parametrization

$$\text{paraboloid}[a, b](u, v) = \big(u,\ v,\ a\,u^2 + b\,v^2\big), \tag{10.16}$$

or equivalently the graph of $z = ax^2 + by^2$, is a ***paraboloid*** provided both a, b are nonzero. The paraboloid is ***circular*** if $a = b \neq 0$.

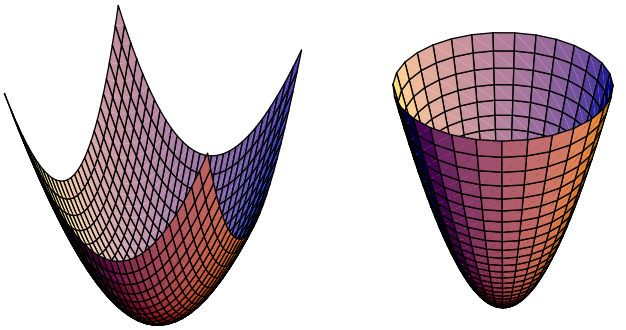


Figure 10.7: Subsets of the circular paraboloid

A clearer representation of a circular paraboloid is obtained by using polar coordinates (r, θ) instead of the rectangular coordinates (u, v). The relation between the two systems is of course

$$u = r\cos\theta, \qquad v = r\sin\theta,$$

and the polar parametric representation is merely

$$(r, \theta) \mapsto (r\cos\theta,\ r\sin\theta,\ ar^2).$$

Plotting this with a bound on the radius r provides the more symmetrical effect visible on the right of Figure 10.7.

Because of its shape, the hyperbolic paraboloid, defined by taking a, b to have different signs in (10.16), is a type of a ***saddle surface***. A person can sit comfortably on a hyperbolic paraboloid because there are indentations for two legs. The ***monkey saddle***, which also takes account of a tail, is represented by the patch

$$\text{monkeysaddle}(u, v) = (u,\ v,\ u^3 - 3uv^2). \tag{10.17}$$

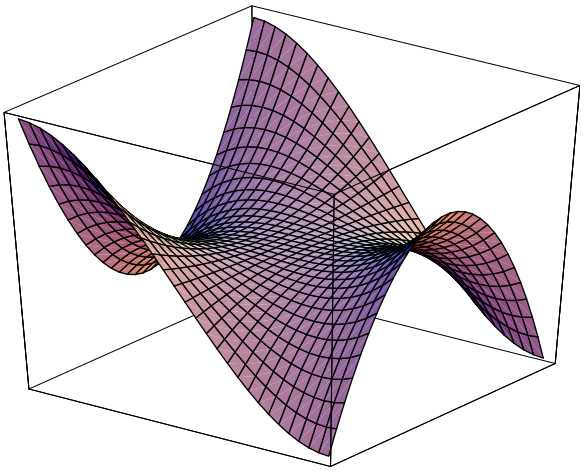


Figure 10.8: The monkey saddle

The plane, hyperbolic paraboloid ((10.16) with $a = 1$, $b = -1$) and monkey saddle are the first three members of a series of surfaces, namely the graphs of the real part of the function $z \mapsto z^n$, where z is a complex number and n a positive integer. Explicitly, we set

$$\mathsf{monkey}[n](u, v) = \big(u,\ v,\ \mathfrak{Re}\,[(u + iv)^n]\,\big),$$

so that $\mathsf{monkey}[2] = \mathsf{paraboloid}[1, -1]$. The imaginary part of z^n gives rise to a similar surface, namely the image of $\mathsf{monkey}[n]$ under a rotation by $\pi/(2n)$ radians.

The surface defined when $n = 5$ is illustrated on the left in Figure 10.9. On the right is the corresponding contour plot, in which the lighter the shading, the greater the value of the function $\mathsf{monkey}[5]$.

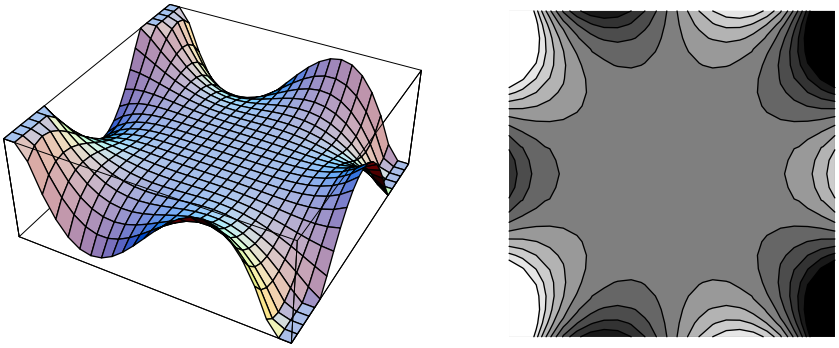


Figure 10.9: The $\mathsf{monkey}[5]$ surface

Elliptical Tori

The patch

$$\begin{aligned} &\mathsf{torus}[a,b,c]\colon [0,2\pi)\times[0,2\pi)\longrightarrow\mathbb{R}^3\\ &(u,v)\mapsto\big((a+b\cos v)\cos u,\ (a+b\cos v)\sin u,\ c\sin v\big)\end{aligned}\tag{10.18}$$

is a parametrization of an ***elliptical torus*** in $\mathbb{R}^3$. Examples were encountered in Section 7.6.

When $b=c<a$, we get a standard round torus, where a is the ***wheel radius*** and b the ***tube radius*** of the torus. A torus with $a>b$ is called a ***ring torus***, and is a regular surface in the sense of Section 10.3. The domain in (10.18) can be extended to any subset of $\mathbb{R}^2$, and including the rectangles

$$\big(0,\tfrac{3\pi}{2}\big)\times\big(0,\tfrac{3\pi}{2}\big),\qquad \big(\pi,\tfrac{5\pi}{2}\big)\times\big(0,\tfrac{3\pi}{2}\big),$$
$$\big(0,\tfrac{3\pi}{2}\big)\times\big(\pi,\tfrac{5\pi}{2}\big),\qquad \big(\pi,\tfrac{5\pi}{2}\big)\times\big(\pi,\tfrac{5\pi}{2}\big).$$

Their images are displayed in the same order in Figure 10.10 for an elliptical torus with $a=8$, $b=3$ and $c=7$.

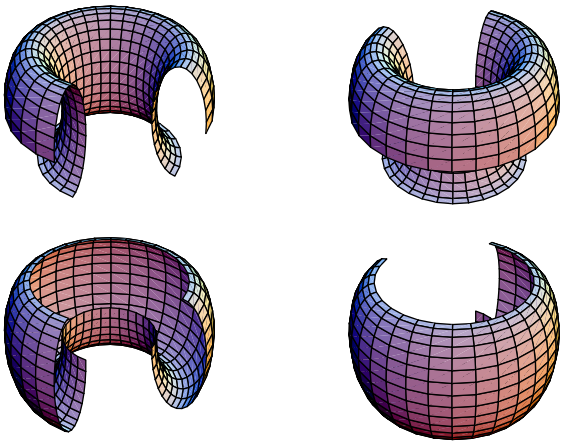


Figure 10.10: Four patches covering $\mathsf{torus}[8,3,7]$

If the wheel radius is less than the tube radius, the resulting torus becomes a self-intersecting surface and appears "inside-out." It is usually called a ***horn torus***, though the critical case when the wheel radius equals the tube radius is called a ***spindle torus***. Figure 10.11 shows both a horn and spindle torus. The number of self-intersections is 0 for a ring torus, 1 for a spindle torus and 2 for a horn torus. These facts are checked analytically in Notebook 10, using the determinant of the matrix in Exercise 1. See also [Fischer, page 28].

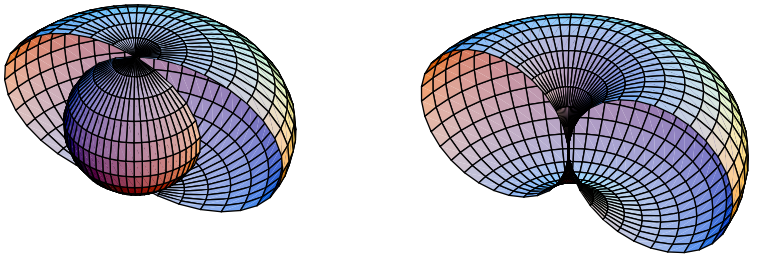


Figure 10.11: torus$[3, 8, 8]$ and torus$[4, 4, 4]$

Patches with Singularities

From the discussion in the previous section, we know that a surface fails to be regular at points of self-intersection, though the *patch* may or may not be singular at such points. Figure 10.11 shows isolated points of self-intersection where the patch is indeed singular, and we now give two more examples of this phenomenon.

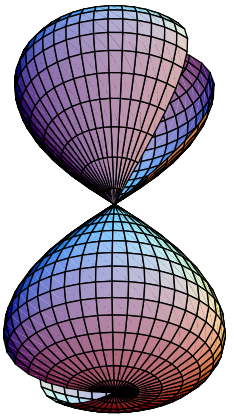


Figure 10.12: The eight surface

The ***eight surface*** consists of a figure eight revolved about the z-axis, and is defined as the trace of the patch

$$\text{eightsurface}(u, v) = \big(\cos u \cos v \sin v,\ \sin u \cos v \sin v,\ \sin v\big). \tag{10.19}$$

The eight surface becomes a regular surface only when the central "vertex" is excluded. In spite of the existence of a singular point, there is no difficulty in plotting the surface.

An example of a surface with a ***pinch point*** is the ***Whitney umbrella***, defined by

$$\text{whitneyumbrella}(u, v) = \big(uv,\ u,\ v^2\big).$$

The Jacobian matrix (10.4) of this patch has rank 2 unless $(u, v) = (0, 0)$, though whitneyumbrella maps $(0, \pm v)$ to the same point for all $v \neq 0$.

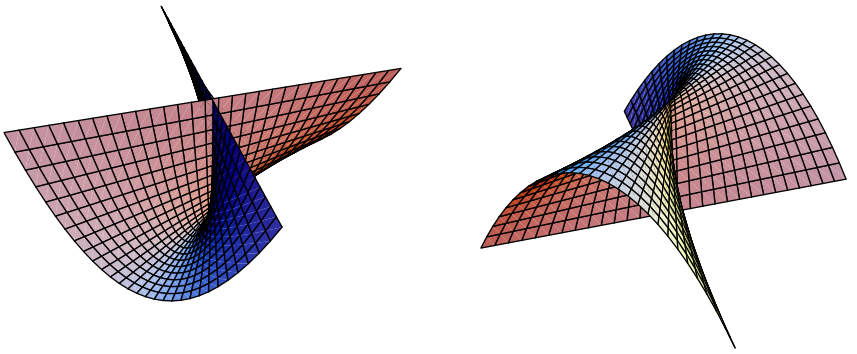


Figure 10.13: Two views of the Whitney umbrella

For a discussion of this surface see [Francis, pages 8–9].

10.5 Tangent Vectors and Surface Mappings

We are now ready to discuss the important notion of tangent vector to a regular surface. This is the next step in the extension to regular surfaces of the calculus of $\mathbb{R}^n$ that we developed in Chapter 9.

Definition 10.37. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^n$ and let $\mathbf{p} \in \mathcal{M}$. We say that $\mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}}$ is **tangent to** $\mathcal{M}$ at $\mathbf{p}$ provided there exists a curve $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ such that $\boldsymbol{\alpha}(0) = \mathbf{p}$, $\boldsymbol{\alpha}'(0) = \mathbf{v_p}$ and $\boldsymbol{\alpha}(t) \in \mathcal{M}$ for $a < t < b$. The **tangent space** to $\mathcal{M}$ at $\mathbf{p}$ is the set*

$$\mathcal{M}_{\mathbf{p}} = \big\{\, \mathbf{v_p} \in \mathbb{R}^n_{\mathbf{p}} \mid \mathbf{v_p} \text{ is tangent to } \mathcal{M} \text{ at } \mathbf{p} \,\big\}.$$

On page 289, we defined $\mathbf{x}_u(u, v)$ and $\mathbf{x}_v(u, v)$ to be vectors in $\mathbb{R}^n$. Sometimes it is useful to modify this definition so that $\mathbf{x}_u(u, v)$ and $\mathbf{x}_v(u, v)$ are tangent vectors at $\mathbf{x}(u, v)$. Thus if $\mathbf{x}(u, v) = \big(x_1(u, v), \ldots, x_n(u, v)\big)$ we can redefine

$$\mathbf{x}_u(u, v) = \left(\frac{\partial x_1}{\partial u}(u, v), \ldots, \frac{\partial x_n}{\partial u}(u, v)\right)_{\mathbf{x}(u,v)},$$

$$\mathbf{x}_v(u, v) = \left(\frac{\partial x_1}{\partial v}(u, v), \ldots, \frac{\partial x_n}{\partial v}(u, v)\right)_{\mathbf{x}(u,v)}.$$

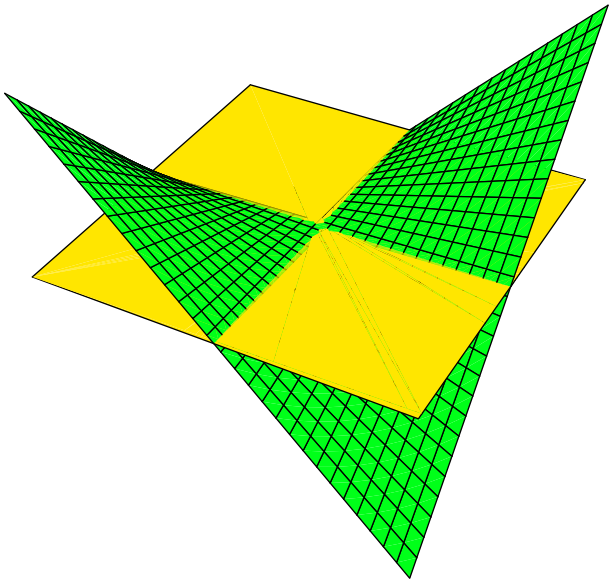


Figure 10.14: Tangent plane at the saddle point of a hyperbolic paraboloid

Similar remarks hold for $\mathbf{x}_{uu}$ and other higher partial derivatives of $\mathbf{x}$. However, such fine distinctions are unnecessary in practice, since when we calculate the partial derivatives of a patch explicitly (either by hand or computer) we consider them to be vectors in $\mathbb{R}^n$. In summary, we regard $\mathbf{x}_u$ and the other partial derivatives of $\mathbf{x}$ to be vectors in $\mathbb{R}^n$ unless for some theoretical reason we require them to be tangent vectors at the point of application.

Lemma 10.38. *Let $\mathbf{p}$ be a point on a regular surface $\mathcal{M} \subset \mathbb{R}^n$. Then the tangent space $\mathcal{M}_\mathbf{p}$ to $\mathcal{M}$ at $\mathbf{p}$ is a 2-dimensional vector subspace of $\mathbb{R}^n_\mathbf{p}$. If $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is any regular patch on $\mathcal{M}$ with $\mathbf{p} = \mathbf{x}(\mathbf{q})$, then*

$$\mathbf{x}_*(\mathbb{R}^2_\mathbf{q}) = \mathcal{M}_\mathbf{p}.$$

Proof. It follows from Lemma 10.2 that $\mathbf{x}_*(\mathbb{R}^2_\mathbf{q})$ is spanned by $\mathbf{x}_u(\mathbf{q})$ and $\mathbf{x}_v(\mathbf{q})$. The regularity of $\mathbf{x}$ at $\mathbf{q}$ implies that $\mathbf{x}_*\colon \mathbb{R}^2_\mathbf{q} \to \mathbb{R}^n_\mathbf{p}$ is injective; consequently, $\mathbf{x}_u(\mathbf{q})$ and $\mathbf{x}_v(\mathbf{q})$ are linearly independent and $\dim\bigl(\mathbf{x}_*(\mathbb{R}^2_\mathbf{q})\bigr) = 2$ by Lemma 10.8. On the other hand, Lemma 10.14 implies that any tangent vector to $\mathcal{M}$ at $\mathbf{q}$ is a linear combination of $\mathbf{x}_u(\mathbf{q})$ and $\mathbf{x}_v(\mathbf{q})$. ∎

Frequently, we shall need the notion of tangent vector perpendicular to a surface, in complete analogy to Definition 10.16.

Definition 10.39. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^n$ and let $\mathbf{z}_\mathbf{p} \in \mathbb{R}^n_\mathbf{p}$ with $\mathbf{p} \in \mathcal{M}$. We say that $\mathbf{z}_\mathbf{p}$ is **normal** or **perpendicular** to $\mathcal{M}$ at $\mathbf{p}$, provided $\mathbf{z}_\mathbf{p} \cdot \mathbf{v}_\mathbf{p} = 0$ for all tangent vectors $\mathbf{v}_\mathbf{p} \in \mathcal{M}_\mathbf{p}$.*

We denote the set of normal vectors to $\mathcal{M}$ at $\mathbf{p}$ by $\mathcal{M}^\perp_\mathbf{p}$, so that

$$\mathbb{R}^n_\mathbf{p} = \mathcal{M}_\mathbf{p} \oplus \mathcal{M}^\perp_\mathbf{p},$$

just as in (10.10).

The notions of tangent and normal vector fields also make sense.

Definition 10.40. *A **vector field** $\mathbf{V}$ on a regular surface $\mathcal{M}$ is a function which assigns to each $\mathbf{p} \in \mathcal{M}$ a tangent vector $\mathbf{V}(\mathbf{p}) \in \mathbb{R}^n_\mathbf{p}$. We say that $\mathbf{V}$ is **tangent** to $\mathcal{M}$ if $\mathbf{V}(\mathbf{p}) \in \mathcal{M}_\mathbf{p}$ for all $\mathbf{p} \in \mathcal{M}$ and that $\mathbf{V}$ is **normal** or **perpendicular** to $\mathcal{M}$ if $\mathbf{V}(\mathbf{p}) \in \mathcal{M}^\perp_\mathbf{p}$ for all $\mathbf{p} \in \mathcal{M}$.*

The definition of differentiability of a mapping between regular surfaces is similar to that of a real-valued function on a regular surface.

Definition 10.41. *A function $F\colon \mathcal{M} \to \mathcal{N}$ from one regular surface to another is **differentiable**, provided that, for any two regular injective patches $\mathbf{x}$ of $\mathcal{M}$ and $\mathbf{y}$ of $\mathcal{N}$, the composition $\mathbf{y}^{-1} \circ F \circ \mathbf{x}$ is differentiable. When this is the case, we call F a **surface mapping**.*

The simplest example of a surface mapping is the identity map $\mathbf{1}_\mathcal{M}\colon \mathcal{M} \to \mathcal{M}$, defined by $\mathbf{1}_\mathcal{M}(\mathbf{p}) = \mathbf{p}$ for $\mathbf{p} \in \mathcal{M}$.

Definition 10.42. *A **diffeomorphism** between regular surfaces $\mathcal{M}$ and $\mathcal{N}$ is a differentiable map $F\colon \mathcal{M} \to \mathcal{N}$ which has a differentiable inverse, that is, a surface mapping $G\colon \mathcal{N} \to \mathcal{M}$ such that*

$$G \circ F = \mathbf{1}_\mathcal{M} \qquad \text{and} \qquad F \circ G = \mathbf{1}_\mathcal{N},$$

where $\mathbf{1}_\mathcal{M}$ and $\mathbf{1}_\mathcal{N}$ denote the identity maps of $\mathcal{M}$ and $\mathcal{N}$.

Definition 10.43. *Let $\mathcal{M}$, $\mathcal{N}$ be regular surfaces in $\mathbb{R}^n$, and $\mathcal{W}$ an open subset of $\mathcal{M}$. We shall say that a mapping*

$$F\colon \mathcal{W} \longrightarrow \mathcal{N}$$

*is a **local diffeomorphism** at $\mathbf{p} \in \mathcal{W}$, if there exists a neighborhood $\mathcal{W}' \subset \mathcal{W}$ of $\mathbf{p}$ such that the restriction of F to $\mathcal{W}'$ is a diffeomorphism of $\mathcal{W}'$ onto an open subset $F(\mathcal{W}') \subseteq \mathcal{N}$.*

Just as a regular surface has a tangent *space*, a surface mapping has a tangent *map*.

Definition 10.44. *Let $\mathcal{M}, \mathcal{N}$ be regular surfaces in $\mathbb{R}^n$, and let $\mathbf{p} \in \mathcal{M}$. Let $F\colon \mathcal{M} \to \mathcal{N}$ be a surface mapping. Then the* ***tangent map*** *of F at $\mathbf{p}$ is the map*

$$F_*\colon \mathcal{M}_\mathbf{p} \longrightarrow \mathcal{N}_{F(\mathbf{p})}$$

given as follows. For $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$, choose a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ such that $\boldsymbol{\alpha}'(0) = \mathbf{v_p}$. Then we define $F_*(\mathbf{v_p})$ to be the initial velocity of the image curve $F \circ \boldsymbol{\alpha}\colon \to \mathcal{N}$; that is,

$$F_*(\mathbf{v_p}) = (F \circ \boldsymbol{\alpha})'(0).$$

Lemma 10.45. *Let $F\colon \mathcal{M} \to \mathcal{N}$ be a surface mapping, and let $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$. The definition of $F_*(\mathbf{v_p})$ is independent of the choice of curve $\boldsymbol{\alpha}$ with $\boldsymbol{\alpha}'(0) = \mathbf{v_p}$. Furthermore, $F_*\colon \mathcal{M}_\mathbf{p} \to \mathcal{N}_{F(\mathbf{p})}$ is a linear map.*

Proof. Let $\boldsymbol{\alpha}$ and $\widetilde{\boldsymbol{\alpha}}$ be curves in $\mathcal{M}$ with $\boldsymbol{\alpha}(0) = \widetilde{\boldsymbol{\alpha}}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \widetilde{\boldsymbol{\alpha}}'(0) = \mathbf{v_p}$. Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a regular patch on $\mathcal{M}$ such that $\mathbf{x}(u_0, v_0) = \mathbf{p}$. Write

$$\boldsymbol{\alpha}(t) = \mathbf{x}\big(u(t), v(t)\big) \qquad \text{and} \qquad \widetilde{\boldsymbol{\alpha}}(t) = \mathbf{x}\big(\widetilde{u}(t), \widetilde{v}(t)\big).$$

Just as in the proof of Lemma 10.14, the chain rule for curves implies that

$$(F \circ \boldsymbol{\alpha})'(t) = u'(t)\big(F \circ \mathbf{x}\big)_u\big(u(t), v(t)\big) + v'(t)\big(F \circ \mathbf{x}\big)_v\big(u(t), v(t)\big)$$

and

$$(F \circ \widetilde{\boldsymbol{\alpha}})'(t) = \widetilde{u}'(t)(F \circ \mathbf{x})_u\big(\widetilde{u}(t), \widetilde{v}(t)\big) + \widetilde{v}'(t)(F \circ \mathbf{x})_v\big(\widetilde{u}(t), \widetilde{v}(t)\big).$$

Since $u(0) = \widetilde{u}(0) = u_0$, $v(0) = \widetilde{v}(0) = v_0$, $u'(0) = \widetilde{u}'(0)$ and $v'(0) = \widetilde{v}'(0)$, it follows that

$$F_*(\mathbf{v_p}) = (F \circ \boldsymbol{\alpha})'(0) = (F \circ \widetilde{\boldsymbol{\alpha}})'(0).$$

This proves that $F_*(\mathbf{v_p})$ does not depend on the choice of $\boldsymbol{\alpha}$. To prove the linearity of F_*, we note that

$$F_*(\mathbf{v_p}) = u'(0)(F \circ \mathbf{x})_u\big(u(0), v(0)\big) + v'(0)(F \circ \mathbf{x})_v\big(u(0), v(0)\big).$$

Since $u'(0), v'(0)$ depend linearly on $\mathbf{v_p}$, it follows that F_* must be linear. ∎

We have the following consequence of the Inverse Function Theorem for $\mathbb{R}^2$.

Theorem 10.46. *If $\mathcal{M}, \mathcal{N}$ are regular surfaces and $F\colon \mathcal{W} \to \mathcal{N}$ is a differentiable mapping of an open subset $\mathcal{W} \subseteq \mathcal{M}$ such that the tangent map F_* of F at $\mathbf{p} \in \mathcal{W}$ is an isomorphism, then F is a local diffeomorphism at $\mathbf{p}$.*

Proof. Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{N}$ be injective regular patches on $\mathcal{M}$ and $\mathcal{N}$ such that $(F \circ \mathbf{x})(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ is nonempty. By restricting the domains of definition of $\mathbf{x}$ and $\mathbf{y}$ if necessary, we can assume that $(F \circ \mathbf{x})(\mathcal{U}) = \mathbf{y}(\mathcal{V})$. Then the tangent map of $\mathbf{y}^{-1} \circ F \circ \mathbf{x}$ is an isomorphism, and the Inverse Function Theorem for $\mathbb{R}^2$ implies that the map $\mathbf{y}^{-1} \circ F \circ \mathbf{x}$ possesses a local inverse. Hence F also has a local inverse. ∎

Examples of Surface Mappings

1. Let $\mathbf{y}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch on a regular surface $\mathcal{M}$. Theorem 10.29 implies that $\mathbf{y}$ is a diffeomorphism between $\mathcal{U}$ (itself, strictly speaking, a regular surface) and $\mathbf{y}(\mathcal{U})$.

2. Let $\mathbf{y}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch on a regular surface $\mathcal{M}$. Then the local Gauss map $\mathbf{U} \circ \mathbf{y}^{-1}$ is a surface mapping from $\mathbf{y}(\mathcal{U})$ to the unit sphere $S^2(1)$ (see Definition 10.22 and Exercise 5).

3. Let $S^2(a) = \{\, \mathbf{p} \mid \|\mathbf{p}\| = a \}$ be the sphere of radius a in $\mathbb{R}^3$. Then the ***antipodal map*** is the function $S^2(a) \to S^2(a)$ defined by $\mathbf{p} \mapsto -\mathbf{p}$. It is a diffeomorphism.

4. Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, and let $\mathcal{M} \subset \mathbb{R}^n$ be a regular surface. Then the restriction $F|\mathcal{M}\colon \mathcal{M} \to F(\mathcal{M})$ is a diffeomorphism of regular surfaces. In particular, any Euclidean motion of $\mathbb{R}^n$ gives rise to a surface mapping.

10.6 Level Surfaces in $\mathbb{R}^3$

So far we have been considering regular surfaces defined by patches. Such a description is called a ***parametric representation***. Another way to describe a regular surface $\mathcal{M}$ is by means of a ***nonparametric representation***. For a regular surface in $\mathbb{R}^3$, this means that $\mathcal{M}$ is the set of points mapped by a differentiable function $g\colon \mathbb{R}^3 \to \mathbb{R}$ into the same real number.

Definition 10.47. *Let $g\colon \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function and c a real number. Then the set*

$$\mathcal{M}(c) = \big\{\, \mathbf{p} \in \mathbb{R}^3 \mid g(\mathbf{p}) = c \,\big\}$$

is called the ***level surface*** *of g corresponding to c.*

Theorem 10.48. *Let $g\colon \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function and c a real number. Then the level surface $\mathcal{M}(c)$ of g is a regular surface if it is nonempty and the gradient* grad g *is nonzero at all points of $\mathcal{M}(c)$. When these conditions are satisfied,* grad g *is everywhere perpendicular to $\mathcal{M}(c)$.*

Proof. For each $\mathbf{p} \in \mathcal{M}(c)$, we must find a regular patch on a neighborhood of $\mathbf{p}$. The hypothesis that $(\text{grad}\, g)(\mathbf{p}) \neq 0$ is equivalent to saying that at least one of the partial derivatives $\partial g/\partial x$, $\partial g/\partial y$, $\partial g/\partial z$ does not vanish at $\mathbf{p}$. Let us suppose that $\big(\partial g/\partial z\big)(\mathbf{p}) \neq 0$. The Implicit Function Theorem states that the equation $g(x, y, z) = c$ can be solved for z. More precisely, there exists a function h such that

$$g\big(x, y, h(x, y)\big) = c.$$

The required patch is given by $\mathbf{x}(u, v) = (u, v, h(u, v))$, as in Definition 10.36.

Furthermore, let $\mathbf{v_p} = (v_1, v_2, v_3)_\mathbf{p} \in \mathcal{M}(c)_\mathbf{p}$. Then there exists a curve $\boldsymbol{\alpha}$ on $\mathcal{M}(c)$ with $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \mathbf{v_p}$. Write $\boldsymbol{\alpha}(t) = (a_1(t), a_2(t), a_3(t))$. Since $\boldsymbol{\alpha}$ lies on $\mathcal{M}(c)$, we have $g(a_1(t), a_2(t), a_3(t)) = c$ for all t. The chain rule implies that

$$\sum_{i=1}^{3} \left(\frac{\partial g}{\partial u_i} \circ \boldsymbol{\alpha} \right) \frac{da_i}{dt} = 0.$$

In particular,

$$0 = \sum_{i=1}^{3} \frac{\partial g}{\partial u_i}\big(\boldsymbol{\alpha}(0)\big) \frac{da_i}{dt}(0) = \sum_{i=1}^{3} \frac{\partial g}{\partial u_i}(\mathbf{p}) v_i = (\mathsf{grad}\, g)(\mathbf{p}) \cdot \mathbf{v_p}.$$

Hence $(\mathsf{grad}\, g)(\mathbf{p})$ is perpendicular to $\mathcal{M}(c)_\mathbf{p}$ for each $\mathbf{p} \in \mathcal{M}(c)$. ▌

To conclude this final section, we return to some examples.

Ellipsoids

The nonparametric equation that defines ***the ellipsoid*** is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{10.20}$$

Here, $a, b, c > 0$ are the lengths of the semi-axes of the ellipsoid. Ellipsoids with only two of a, b and c distinct are considerably simpler than general ellipsoids. Such an ellipsoid is called an ***ellipsoid of revolution*** and can be obtained by rotating an ellipse about one of its axes.

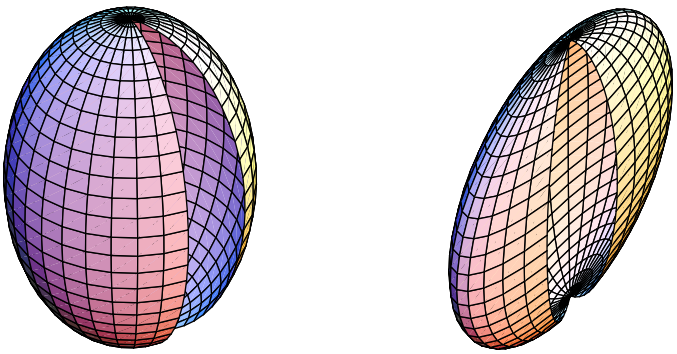


Figure 10.15: Regions of $\mathsf{ellipsoid}[1, 1, 2]$ and $\mathsf{ellipsoid}[1, 2, 3]$

We shall study surfaces of revolution in detail in Chapter 15. Figure 10.15 illustrates the difference between an ellipsoid of revolution and a general ellipsoid. A parametric form of (10.20) is

$$\mathsf{ellipsoid}[a,b,c](u,v) = \big(a\cos v\cos u,\ b\cos v\sin u,\ c\sin v\big).$$

The patch is regular except at the north and south poles, which are $(0,0,\pm c)$.

We next define a different parametrization of an ellipsoid, the ***stereographic ellipsoid***. It is a generalization of the stereographic projection of the sphere that will be defined on page 730. It is often useful to change parametrization in order to highlight a particular property of a surface, and we shall see several instances of this procedure in the sequel. The ellipsoid, for reasons of physics and cartography, is one of the surfaces most amenable to the procedure (see Exercise 6 and Section 19.6).

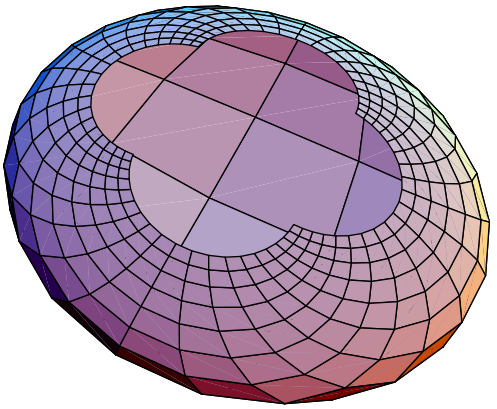


Figure 10.16: stereographicellipsoid$[5,3,1]$

For viewing convenience, the stereographic ellipse has been rotated so that the missing "northern" part is at the front of the picture. Its definition is:

$$\mathsf{stereographicellipsoid}[a,b,c](u,v) = \left(\frac{2a\,u}{1+u^2+v^2},\ \frac{2b\,v}{1+u^2+v^2},\ \frac{c(u^2+v^2-1)}{1+u^2+v^2}\right).$$

Hyperboloids

Consider the following patches:

$$\begin{aligned}\mathsf{hyperboloid1}[a,b,c](u,v) &= \big(a\cosh v\cos u,\ b\cosh v\sin u,\ c\sinh v\big),\\ \mathsf{hyperboloid2}[a,b,c](u,v) &= \big(a\sinh v\cos u,\ b\sinh v\sin u,\ c\cosh v\big).\end{aligned} \tag{10.21}$$

They describe surfaces whose shape is altered by varying the three parameters a, b, c. Assuming all three are positive, hyperboloid1 describes a hyperboloid of one sheet, whereas hyperboloid2 describes the upper half of a hyperboloid of two sheets (the lower half is obtained by taking $c < 0$). As $|v|$ becomes large, the surfaces are asymptotic to

$$\mathsf{ellipticalcone}[a, b, c](u, v) = \big(av\cos u,\ bv\cos u,\ cv\big),$$

and rotational symmetry is present if $a = b$. In this case, the surfaces are formed by revolving a branch of a hyperbola (or a line) about the z-axis, as is the case in Figure 10.17.

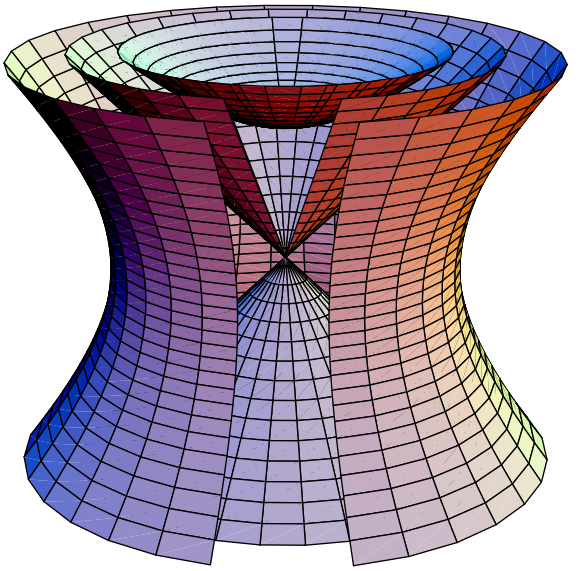


Figure 10.17: Hyperboloids in and outside of a cone

The nonparametric equations

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1,$$

describe a hyperboloid of one sheet and two sheets respectively. The equations can easily be distinguished, as the second has no solutions for $z = 0$, indicating that no points of the surface lie in the xy-plane.

Higher Order Surfaces

Consider the function $g_n\colon \mathbb{R}^3 \to \mathbb{R}$ defined by

$$g_n(x, y, z) = x^n + y^n + z^n,$$

where $n \geqslant 2$ is an even integer. Using Definition 9.24, the gradient of g_n is given by

$$(\operatorname{grad} g_n)(x, y, z) = \big(nx^{n-1},\, ny^{n-1},\, nz^{n-1}\big); \tag{10.22}$$

it vanishes if and only if $x = y = z = 0$, so the set

$$\{\, \mathbf{p} \in \mathbb{R}^3 \mid g_n(\mathbf{p}) = c \,\}$$

is a regular surface for any $c > 0$. It is a sphere for $n = 2$, but as n becomes larger and larger, (10.22) becomes more and more cube-like; see Figure 10.18.

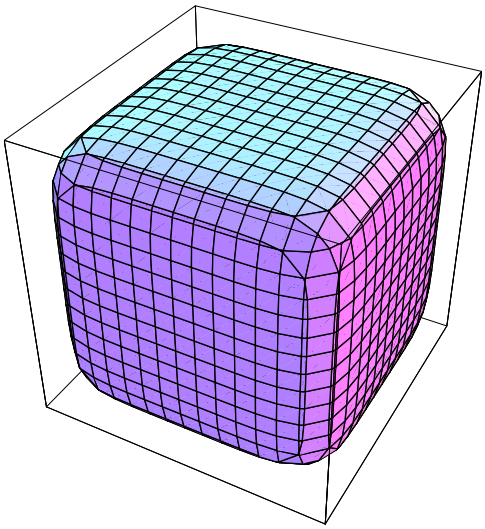


Figure 10.18: The regular surface $x^6 + y^6 + z^6 = 1$

There are functions g for which $\{\mathbf{p} \in \mathbb{R}^3 \mid g(\mathbf{p}) = c\}$ has several components. An artificial example is the regular surface defined by

$$(x^2 + y^2 + z^2 - 1)\big((x-3)^2 + y^2 + z^2 - 1\big) = 0,$$

which consists of two disjoint spheres.

Because the gradient of a function $g\colon \mathbb{R}^3 \to \mathbb{R}$ is always perpendicular to each of its level surfaces, there is an easy way to obtain the Gauss map for such surfaces.

Lemma 10.49. *Let $g\colon \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function and c a number such that* grad g *is nonzero on all points of the level surface $\mathcal{M}(c) = \{\, \mathbf{p} \in \mathbb{R}^3 \mid g(\mathbf{p}) = c \,\}$. Then the vector field $\mathbf{U}$ defined by*

$$\mathbf{U}(\mathbf{p}) = \frac{\operatorname{grad} g(\mathbf{p})}{\left\| \operatorname{grad} g(\mathbf{p}) \right\|}$$

is a globally-defined unit normal on $\mathcal{M}(c)$. Hence we can define the Gauss map of $\mathcal{M}(c)$ to be $\mathbf{U}$ considered as a mapping $\mathbf{U}\colon \mathcal{M}(c) \to S^2(1)$.

Steven Wilkinson's program ImplicitPlot3D is extremely effective in plotting surfaces such as those in Figures 10.18, 10.19 and 10.20. Details are given in Notebook 10. The reader is invited to test it out in Exercises 10–13.

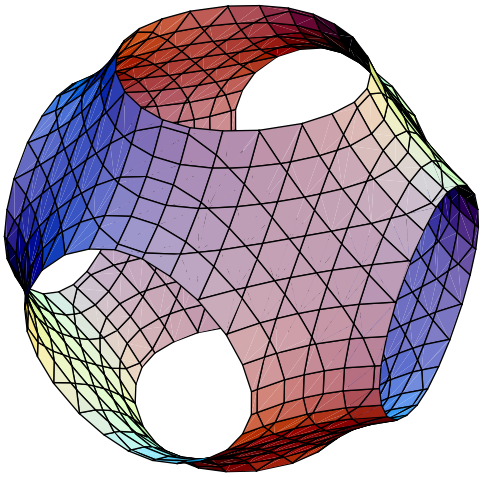


Figure 10.19: $(yz)^2 + (zx)^2 + (xy)^2 = 1$

10.7 Exercises

1. Show that the Jacobian matrix of a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is related to $\mathbf{x}_u$ and $\mathbf{x}_v$ by the formula

$$\mathscr{J}(\mathbf{x})^T \mathscr{J}(\mathbf{x}) = \begin{pmatrix} \mathbf{x}_u \cdot \mathbf{x}_u & \mathbf{x}_u \cdot \mathbf{x}_v \\ \mathbf{x}_v \cdot \mathbf{x}_u & \mathbf{x}_v \cdot \mathbf{x}_v \end{pmatrix},$$

where A^T denotes the transpose of a matrix A (see Corollary 10.5(ii)). Show that when $n = 3$ then $\det\big(\mathscr{J}(\mathbf{x})^T \mathscr{J}(\mathbf{x})\big) = \|\mathbf{x}_u \times \mathbf{x}_v\|^2$.

2. Determine where the patch

$$(u, v) \mapsto \big(\cos u \cos v \sin v,\ \sin u \cos v \sin v,\ \sin v\big)$$

is regular.

3. A torus can be defined as a level surface $\mathcal{M}(b^2)$ of the function $f\colon \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x, y, z) = z^2 + \big(\sqrt{x^2 + y^2} - a\big)^2,$$

where $a > b > 0$. Show that such a torus is a regular surface and compute its unit normal.

4. Show that the Gauss map of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is a surface mapping from $\mathcal{M}$ to the unit sphere $S^2(1) \subset \mathbb{R}^3$ in the sense of Definition 10.41.

5. Find unit normals to the monkey saddle, the sphere, the torus and the eight surface.

6. Define the ***Mercator*[3] *parametrization*** by

$$\mathsf{mercatorellipsoid}[a, b, c](u, v) = \big(a \operatorname{sech} v \cos u,\ b \operatorname{sech} v \sin u,\ c \tanh v\big).$$

Verify that its trace is an ellipsoid.

7. Determine appropriate domains for the mapping (10.19) so as to obtain two patches (satisfying Definition 10.23) that cover the eight surface illustrated in Figure 10.5.

M **8.** Given a patch $\mathbf{x}(u, v)$, let $\mathbf{x}^{[n]}(u, v)$ denote the surface whose x-, y- and z-entries are the nth powers of those of $\mathbf{x}(u, v)$. Plot the "cubed surface" $\mathsf{torus}^{[3]}[8, 3, 8]$.

M **9.** Verify that the following quartic factorizes into two quadratic polynomials, and hence plot the surface described by setting it equal to zero:

$$\frac{x^4}{9} + \frac{y^4}{9} + z^4 - \frac{82x^2y^2}{81} - \frac{10x^2z^2}{9} - \frac{10y^2z^2}{9} - \frac{10x^2}{9} + \frac{10y^2}{9} + 2z^2 + 1.$$



[3] Gerardus Mercator (Latinized name of Gerhard Kremer) (1512–1594). Flemish cartographer. In 1569 he first used the map projection which bears his name.

M **10.** Use ImplicitPlot3D to plot ***Kummer's surface***,[4] defined implicitly by

$$x^4 + y^4 + z^4 - (y^2z^2 + z^2x^2 + x^2y^2) - (x^2 + y^2 + z^2) + 1 = 0.$$

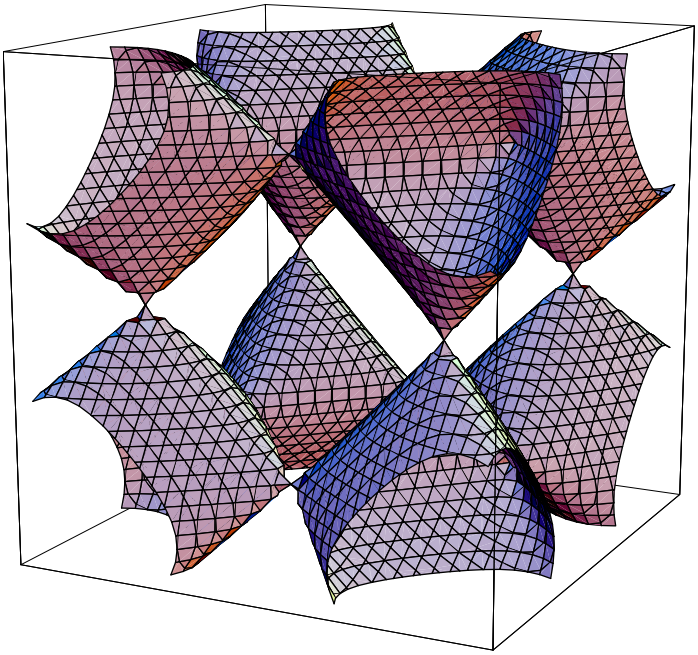


Figure 10.20: Part of Kummer's surface

M **11.** ***Goursat's surface***[5] is defined by

$$x^4 + y^4 + z^4 + a(x^2 + y^2 + z^2)^2 + b(x^2 + y^2 + z^2) + c = 0.$$

Plot it for $a = -1/5$, $b = 0$ and $c = -1$ in the range $-1 \leqslant x, y, z \leqslant 1$.

---

4 

Ernst Eduard Kummer(1810–1893). Kummer is remembered for his work in hypergeometric functions, number theory and algebraic geometry.

5 

Édouard Jean Babtiste Goursat (1858–1936). Goursat was a leading analyst of his day. His *Cours d'analyse mathématique* has long been a classic text in France and elsewhere.

M **12.** Plot the ***two-cusp surface***

$$(z-1)^2(x^2-z^2)-(x^2-z)^2-y^4-y^2(2x^2+z^2+2z-1)=0$$

and verify that it does indeed possess two singular points.

M **13.** Plot the ***sine surface*** defined by

$$\mathsf{sinsurface}(u,v)=\big(\sin u,\ \sin v,\ \sin(u+v)\big).$$

Find a nonparametric form of the surface. Describe the singular points of the sine surface.

14. Let $\mathcal{V}$ be an open subset of the monkey saddle containing its planar point $(0,0,0)$. Describe the image of $\mathcal{V}$ under the Gauss map.

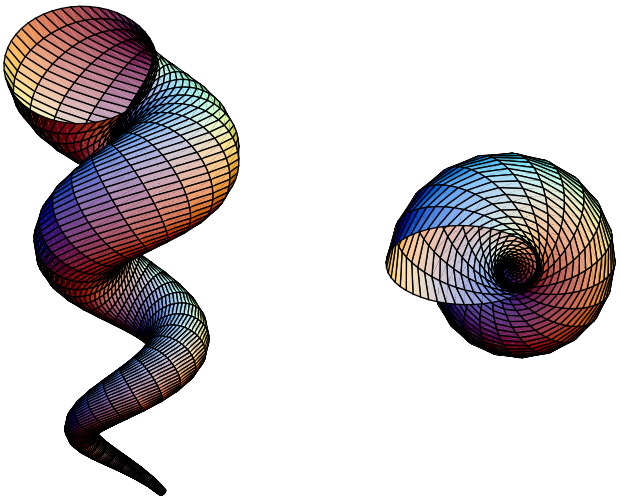


Figure 10.21: The $\mathsf{seashell}[\mathsf{helix}[1,0.6]][0.1]$

M **15.** In Section 7.6 we showed how to construct tubes about curves in $\mathbb{R}^3$. The construction can be easily modified to allow the radius of the tube to change from point to point. For example, given a curve $\boldsymbol{\gamma}\colon (a,b)\to\mathbb{R}^3$, let us define a ***sea shell***

$$\mathsf{seashell}[\boldsymbol{\gamma},r](t,\theta)=\boldsymbol{\gamma}(t)+rt\big(-\cos\theta\,\mathbf{N}(t)+\sin\theta\,\mathbf{B}(t)\big),$$

where $\mathbf{N}$ and $\mathbf{B}$ are the normal and binormal to $\boldsymbol{\gamma}$. Explain why the sea shell of a helix resembles Figure 10.21.

Notebook 10

■ N10.0

This notebook incorporates a first selection of surfaces. It investigates patches and associated normal vectors. It includes definitions of paraboloids, hyperboloids, and the elliptical tori already discussed in N7.6.

Towards the end of the notebook, in N10.8, we shall need Steven Wilkinson's package ImplicitPlot3D that is available from his website (see http://www.nku.edu). We have placed a copy inside the Graphics folder /AddOns/StandardPackages/Graphics of the *Mathematica* directory, to ensure that it can be loaded as follows:

```
<<Graphics`ImplicitPlot3D`
```

As explained in N6.0 and N9.0, use of the command S is necessary to display graphics, which are otherwise suppressed. This ensures that only the final picture appears each time a number of distinct plots are combined. Moreover, only the numbered figures will be displayed when the entire notebook is evaluated. As we shall point out again in N10.1, the string S@ needs to be added by hand to see the other plots. The user who prefers not to do this should replace the following cell by the command "S:=Show" so as to make S a mere abbreviation for Show, and then restart the notebook.

```
$DisplayFunction= Identity
Unprotect[S]
  S[X__]:= Show[X, DisplayFunction->(Display[$Display,#1] &)]
Protect[S]
```

We use the following options for ParametricPlot3D.

```
SetOptions[ParametricPlot3D,
  BoxRatios->Automatic, Compiled->False, Axes->None,
  Boxed->False
]
```

The definition of the gradient will be used again in N10.7.

```
grad[f_, coor_]:= Table[D[f, coor[[i]]], {i, Length[coor]}]
```

The hidden subsection contains the definition of the norm L, the Frenet frame commands tangent, normal, binormal, and the definition of helix, all from Notebook 7.

N10.0.1

N10.1

The sphere, first defined in N8.4, is an obvious surface with which to begin.

```
sphere[a_][u_,v_]:= a{Cos[v]Cos[u], Cos[v]Sin[u], Sin[v]}
```

Using this patch, we plot the image of a rectangle that does not cover the entire sphere:

```
x:= sphere[1]
S @ ParametricPlot3D[x[u,v], {u,0,3π/2}, {v,-π/2,π/2},
   Boxed->True]
```

Figure 10.1

The following figures illustrate noninjective patches:

```
S @ ParametricPlot3D[{Cos[u], Sin[u], v}, {u,0,2π}, {v,-1,1},
   PlotPoints->{30,15}]
```

Figure 10.2

```
S @ ParametricPlot3D[{u^3, v^3, u v}, {u,-1,1}, {v,-1,1},
   ViewPoint->{-2,-2.5,1}]
```

Figure 10.3

The PlotPoints command is used to decide how many units to divide the circumference and the length of the cylinder into. Taking smaller entries will produce a box shape. While this is not an accurate representation of the cylinder, this device is useful at times, as it would take much longer to plot the four sides separately. Add S@ to the first line to display the plot.

```
ParametricPlot3D[{Cos[u], Sin[u], v}, {u,0,2π}, {v,-1,1},
  PlotPoints->{5,5}]
```

The following miniprogram computes the Jacobian matrix of a patch. It begins by defining the partial derivatives that constitute the Jacobian's two columns. On the other hand, a matrix is represented in *Mathematica* by an ordered set of row vectors, so it is necessary to finish with the command Transpose:

```
jacobian[x_][u_,v_]:=
Module[{xu,xv,M},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  M= Simplify[{xu,xv}]/.{U->u,V->v};
  Transpose[M]
]
```

We display the Jacobian matrix of the standard patch to the sphere. It can be viewed in the more conventional way by using the MatrixForm command afterwards.

```
x:= sphere[a]
jacobian[x][u,v]
%//MatrixForm
```

The Jacobian matrix is not square, but we can manufacture a 2 by 2 matrix by multiplying it by its transpose. Its determinant is nonzero at points where the patch is regular:

```
surfacejacobian[x_][u_,v_]:=
Module[{M},
  M= jacobian[x][u,v];
  Det[Transpose[M].M]//Simplify
]
```

The next surface will provide a faithful companion for future constructions.

```
monkeysaddle[u_,v_]:= {u, v, u^3 - 3u v^2}
```

```
jacobian[monkeysaddle][u,v]
surfacejacobian[monkeysaddle][u,v]
```

■ N10.2

The following program constructs the unit normal of a patch.

```
unitnormal[x_][u_,v_]:=
Module[{U, V, xu, xv},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  Nx= Cross[xu,xv]/.{U->u,V->v};
  Simplify[Nx/L[Nx]]
]
```

The unit normal to a sphere is proportional to the patch describing the sphere itself:

```
Clear[a]
x:= sphere[a]
unitnormal[x][u,v]//PowerExpand
x[u,v]
```

The command unitnormal[x] is itself a function or patch whose image lies on the unit sphere. The patches obtained in this way have coordinate curves that represent the normal vector to the original surface and are therefore very different from those of the ordinary spherical patch:

```
y:= unitnormal[monkeysaddle]
ParametricPlot3D[y[u,v], {u,-1,1}, {v,-1,1}, ViewPoint->{1,3,-1}]
```

A Monge patch parametrizes the graph of a function h of two variables.

```
monge[h_][u_,v_]:={u, v, h[u,v]}
```

We compute its Jacobian matrix, and confirm that the patch is everywhere regular:

```
Clear[h]
jacobian[monge[h]][u,v]
surfacejacobian[monge[h]][u,v]
```

The program above for the unit normal also enables us to extract the normal vector Nx consisting of the cross product of the tangent vectors to the coordinate curves. The norm-squared of Nx coincides with the surfacejacobian output:

```
unitnormal[monge[h]][u,v]
L[Nx]^2
```

■ N10.2.1

■ N10.3

The definition of the next surface is obtained from that of an eight curve by adding the second parameter as the third coordinate. A more complicated surface, based on the same functions, can be seen in Figure 10.12 in N10.4.1.

```
eightcylinder[u_,v_]:= {Sin[u], Sin[2u], v}
```

```
jacobian[eightcylinder][u,v]
unitnormal[eightcylinder][u,v]
```

```
S @ ParametricPlot3D[eightcylinder[u,v], {u,-π/3,5π/4},
  {v,-1,1}, PlotPoints->{40,10}, ViewPoint->{1,-1,3}]
```

Figure 10.5

The easiest way of obtaining a second chart on the sphere is to cheat and rotate the first! To do this, we make use of the of the built-in package

```
<<Geometry`Rotations`
```

```
x:= sphere[1]
M:= RotationMatrix3D[.6π,.5π,0]
y[u_,v_]:= M . x[u,v]
pl[f_]:= ParametricPlot3D[f[u,v], {u,0,7π/4}, {v,-π/2,π/2}]
S @ GraphicsArray[{pl[x],pl[y]}]
```

Figure 10.6

■ N10.4

One can capture both types of paraboloid with the single definition

```
paraboloid[a_,b_][u_,v_]:= {u, v, a u^2 + b v^2}
```

The elliptic paraboloid arises when a and b are both positive or both negative. The hyperbolic paraboloid occurs when a and b have opposite signs:

```
r:= 1.5
z[a_,b_]:= ParametricPlot3D[paraboloid[a,b][u,v], {u,-r,r}, {v,-r,r}]
GraphicsArray[{z[1,1], z[1,-1]}]
```

Here is a simple command for conversion to polar coordinates:

```
polarize[f_][r_,θ_]:= f[r Cos[θ], r Sin[θ]]
```

```
x:= paraboloid[1,1]
xx:= polarize[x]
z1:= ParametricPlot3D[x[u,v], {u,-r,r}, {v,-r,r}]
z2:= ParametricPlot3D[xx[r,θ], {r,0,2.7}, {θ,0,2π}]
S @ GraphicsArray[{z1,z2}]
```

Figure 10.7

We plot the monkey saddle. Its parametrization in N10.1 corresponds to the case n=3 in the function f defined below using complex numbers.

```
r:= 1.3
S @ ParametricPlot3D[monkeysaddle[u,v], {u,-r,r}, {v,-r,r},
  ViewPoint->{3,-4,8}, AspectRatio->.8, Boxed->True]
```

Figure 10.8

```
f[n_,u_,v_]:= Re[(u + I v)^n]//ComplexExpand
f[3,u,v]
f[5,u,v]
```

```
r:= 1
z1:= Plot3D[f[5,u,v], {u,-r,r}, {v,-r,r}, Axes->False]
z2:= ContourPlot[f[5,u,v], {u,-r,r}, {v,-r,r}, Frame->False]
S @ GraphicsArray[{z1,z2}]
```

Figure 10.9

```
ParametricPlot3D[{u,v,f[5,u,v]}, {u,-r,r}, {v,-r,r},
  AspectRatio->1, Boxed->True]
```

The following definition was first given in Notebook 7:

```
torus[a_,b_,c_][u_,v_]:=
  {(a + b Cos[v])Cos[u], (a + b Cos[v])Sin[u], c Sin[v]}
```

```
ParametricPlot3D[torus[8,3,7][u,v], {u,0,π}, {v,0,2π},
  PlotPoints->{24,48}]
```

Four more chapter figures are included in an electronic subsection.

■ N10.4.1

■ N10.5

The following construction to combine the plot of a curved surface and plane is the basis for Figure 10.14 in N10.5.1. To display the plot, add S@ at the start of the first line, or replace Show by S in the penultimate line.

```
Module[{f,g,r,z},
  f[u_,v_]= {u, v, u v};
  g[u_,v_]= {u, v, 0};
  r= 1.5;
  z[x_]:= ParametricPlot3D[x[u,v], {u,-r,r}, {v,-r,r}];
  Show[z[f], z[g], PlotRange->All]
]
```

■ N10.5.1

■ N10.6

The next definition is similar to the function torus:

```
ellipsoid[a_,b_,c_][u_,v_]:=
  {a Cos[v]Cos[u], b Cos[v]Sin[u], c Sin[v]}
```

Changing the AmbientLight option gives a warmer coloring:

```
p[a_,b_,c_]:=
  ParametricPlot3D[ellipsoid[a,b,c][u,v], {u,0,9π/5}, {v,-π/2,π/2},
  AmbientLight->GrayLevel[.2]]
S @ GraphicsArray[{p[1,1,1.5], p[1,4,3]}]
```

Figure 10.15

We verify next that the ellipsoid patch is regular unless Cos[v] is zero:

```
Clear[a,b,c]
jacobian[ellipsoid[a,b,c]][u,v]
surfacejacobian[ellipsoid[a,b,c]][u,v]//FullSimplify
```

```
stereographicellipsoid[a_,b_,c_][u_,v_]:=
  {(2a u)/(u^2 + v^2 + 1), (2b v)/(u^2 + v^2 + 1),
   (c (u^2 + v^2 - 1))/(u^2 + v^2 + 1)}
```

```
x:= stereographicellipsoid[5,4,3]
S @ ParametricPlot3D[x[u,v], {u,-2.8,2.8}, {v,-2.8, 2.8},
  ViewPoint->{1,-2,4}, PlotPoints->20]
```

Figure 10.16

Here are more parametrizations of quadrics:

```
hyperboloid1[a_,b_,c_][u_,v_]:=
   {a Cosh[v]Cos[u], b Cosh[v]Sin[u], c Sinh[v]}
hyperboloid2[a_,b_,c_][u_,v_]:=
   {a Sinh[v]Cos[u], b Sinh[v]Sin[u], c Cosh[v]}
```

```
ellipticalcone[a_,b_,c_][u_, v_]:= {a v Cos[u], b v Sin[u], c v}
```

They are shown together using an Array of plots.

```
a = 1.3; b = 1.65; c = 2;
h[1]:= hyperboloid2[a,a,a]
h[2]:= hyperboloid2[a,a,-a]
h[3]:= hyperboloid1[b,b,b]
h[4]:= ellipticalcone[c,c,c]
z[i_]:= ParametricPlot3D[h[i][u,v], {u,1.8,7.6}, {v,-1,1},
  ViewPoint->{0,3,1}]
S @ Array[z,4]
```

Figure 10.17

An alternative parametrization of the hyperboloid of one sheet will show its straight-line rulings. This will be discussed in N14.1.

```
ellipticalhyperboloidplus[a_,b_,c_][u_,v_]:=
  {a Cos[u] - a v Sin[u], b v Cos[u] + b Sin[u], c v}
```

```
h[1]:= hyperboloid1[1,2,1]
h[2]:= ellipticalhyperboloidplus[1,2,1]
z[i_]:= ParametricPlot3D[h[i][u,v], {u,0,2π}, {v,-1,1}]
GraphicsArray[{z[1], z[2]}]
```

We next define a function g[n], find its gradient and use the ImplicitPlot3D program to approximate cubes:

```
Clear[x]
g[n_][x_,y_,z_]:= x^n + y^n + z^n
grad[g[n][x,y,z], {x,y,z}]
```

```
S @ ImplicitPlot3D[g[6][x,y,z]==1, {x,-2,2}, {y,-2,2}, {z,-2,2},
  PlotPoints->30, AmbientLight->GrayLevel[.5],
  LightSources->{{{2,0,0}, Hue[0]}, {{0,2,0}, Hue[.4]},
    {{0,0,2}, Hue[.7]}}]
```

Figure 10.18

```
ImplicitPlot3D[g[5][x,y,z]==1, {x,0,2}, {y,0,2}, {z,0,2},
  ViewPoint->{2,2,2}]
```

■ N10.7

The ImplicitPlot3D program used next was loaded at the start of N10.0.

```
Module[{f,r,z},
  r= 5;
  f= x^2 y^2 + y^2 z^2 + z^2 x^2;
  z[r_]:= ImplicitPlot3D[f==1, {x,-r,r}, {y,-r,r}, {z,-r,r},
  Boxed->False];
  S @ z[1.3]
]
```

Figure 10.19

The sine surface will be animated using quaternions in Notebook 23.

```
sinsurface[u_,v_]:={Sin[u], Sin[v], Sin[u+v]}
```

```
ParametricPlot3D[sinsurface[u,v], {u,0,2π}, {v,0,2π}, Axes->None,
  PlotPoints->{28,42}]
```

```
sinsurfaceimplicit[a_][x_,y_,z_]:=
  4x^2 y^2(a^2-z^2) - a^2(x^2+y^2-z^2)^2
```

ContourPlot3D is a built-in implicit plotting command that first divides the drawing in small cubes.

```
<<Graphics`ContourPlot3D`
```

Here we use it to display Kummer's surface, which is a quartic with 16 double points. Unlike the nonparametric definitions of curves given in Notebook 3, the implicit equations in this section are *polynomials* and have no equality sign in their definition.

```
kummer[x_,y_,z_]:=
  x^4 +y^4 +z^4 -x^2 -y^2 z^2 -y^2 -z^2 x^2 -z^2 -x^2 y^2 +1
```

```
Clear[x,y,z]
r:= 1.5
S @ ContourPlot3D[kummer[x,y,z], {x,-r,r}, {y,-r,r}, {z,-r,r},
  ViewPoint->{2,-4,1}, PlotPoints->{7,7}]
```

Figure 10.20

The next cell uses the commands normal and binormal, defined in N7.2 and N10.0.1. Using these, it defines a "tube" of varying radius around a curve, which is taken to be a helix in the subsequent figure.

```
seashell[γ_,r_][t_, θ_]:=
  γ[t] + r t (-(Cos[θ]normal[γ][t]) + Sin[θ]binormal[γ][t])
```

```
Module[{x},
  x= seashell[helix[1,.6], .1];
  z[v_]= ParametricPlot3D[x[t,θ]//Evaluate, {t,π/4,5π}, {θ,0,2π},
    Boxed->False, Axes->None, PlotPoints->50, ViewPoint->v];
  S@GraphicsArray[{z[{0,-2,0}],z[{0,0,2}]}, GraphicsSpacing->.4]
]
```

Figure 10.21

Here is the definition of another quartic surface:

```
goursat[a_,b_,c_][x_,y_,z_]:=
  x^4 +y^4 +z^4 + a(x^2+y^2+z^2)^2 + b(x^2+y^2+z^2) + c
```

```
r:= 1
ImplicitPlot3D[goursat[-1/5,0,-1][x,y,z]==0, {x,-r,r}, {y,-r,r},
  {z,-r,r}]
```

Chapter 11

Nonorientable Surfaces

An important discovery of the nineteenth century was that nonorientable surfaces exist. The goal of the present chapter is to gain an understanding of such surfaces. We begin in Section 11.1 by saying precisely what it means for a regular surface to be *orientable*, and defining the associated Gauss map. Examples of nonorientable surfaces are first described in Section 11.2 by means of various identifications of the edges of a square. We obtain topological descriptions of some of the simpler nonorientable surfaces, without reference to an underlying space like $\mathbb{R}^n$ with $n \geqslant 3$. The analytical theory of surfaces described in such an abstract way is developed in Chapter 26.

The rest of the chapter is devoted to realizing nonorientable surfaces inside 3-dimensional space. Section 11.3 is devoted to the Möbius strip, a building block for all nonorientable surfaces. Having parametrized the Möbius strip explicitly, we are able to depict its Gauss image in Figures 11.8 and 11.9. The Möbius strip has become an icon that has formed the basis of many artistic designs, including one on the Mall in Washington, DC.

Sections 11.4 and 11.5 describe two basic compact nonorientable surfaces, namely the Klein bottle and real projective plane. These necessarily have points of self-intersection in $\mathbb{R}^3$, and two different models of the Klein bottle are discussed. Different realizations of the projective plane are described using the notion of a map with the antipodal property (Definition 11.13).

11.1 Orientability of Surfaces

If V is a 2-dimensional vector space, we call a linear map $J\colon V \to V$ such that $J^2 = -\mathbf{1}$ a *complex structure* on V. We used this notion from the beginning of our study of curves, and observed on page 265 that each tangent space to $\mathbb{R}^2$ has a complex structure. More generally, because all 2-dimensional vector spaces are isomorphic, each tangent space $\mathcal{M}_{\mathbf{p}}$ to a regular surface $\mathcal{M}$ admits

a complex structure $J_{\mathbf{p}}\colon \mathcal{M}_{\mathbf{p}} \to \mathcal{M}_{\mathbf{p}}$. However, it may or may not be possible to make a *continuous* choice of $J_{\mathbf{p}}$. Continuity here means that the vector field $\mathbf{p} \mapsto J_{\mathbf{p}}\mathbf{X}$ is continuous for each continuous vector field $\mathbf{X}$ tangent to $\mathcal{M}$. This leads to

Definition 11.1. *A regular surface $\mathcal{M} \subset \mathbb{R}^3$ is called* ***orientable*** *provided each tangent space $\mathcal{M}_{\mathbf{p}}$ has a complex structure $J_{\mathbf{p}}\colon \mathcal{M}_{\mathbf{p}} \to \mathcal{M}_{\mathbf{p}}$ such that $\mathbf{p} \mapsto J_{\mathbf{p}}$ is a continuous function. An* ***oriented*** *regular surface $\mathcal{M} \subset \mathbb{R}^3$ is an orientable regular surface together with a choice of the complex structure $\mathbf{p} \mapsto J_{\mathbf{p}}$.*

We shall see in Chapter 26 that this definition works equally well for surfaces *not* contained in $\mathbb{R}^3$. For regular surfaces contained in $\mathbb{R}^3$, there is a more intuitive way to describe orientability using the vector cross product.

Theorem 11.2. *A regular surface $\mathcal{M} \subset \mathbb{R}^3$ is orientable if and only if there is a continuous map $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$ that assigns to each $\mathbf{p} \in \mathcal{M}$ a unit normal vector $\mathbf{U}(\mathbf{p}) \in \mathcal{M}_{\mathbf{p}}^{\perp}$.*

Proof. Suppose we are given $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$. Then for each $\mathbf{p} \in \mathcal{M}$ we define $J_{\mathbf{p}}\colon \mathcal{M}_{\mathbf{p}} \to \mathcal{M}_{\mathbf{p}}$ by

$$J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}} = \mathbf{U}(\mathbf{p}) \times \mathbf{v}_{\mathbf{p}}. \tag{11.1}$$

It is easy to check that $J_{\mathbf{p}}$ maps $\mathcal{M}_{\mathbf{p}}$ into $\mathcal{M}_{\mathbf{p}}$ and not merely into $\mathbb{R}^3_{\mathbf{p}}$, and that $\mathbf{p} \mapsto J_{\mathbf{p}}$ is continuous. From (7.3), page 193, it follows that

$$\begin{aligned} J_{\mathbf{p}}^{\,2}\mathbf{v}_{\mathbf{p}} &= \mathbf{U}(\mathbf{p}) \times \big(\mathbf{U}(\mathbf{p}) \times \mathbf{v}_{\mathbf{p}}\big) \\ &= \big(\mathbf{U}(\mathbf{p}) \cdot \mathbf{v}_{\mathbf{p}}\big)\mathbf{U}(\mathbf{p}) - \big(\mathbf{U}(\mathbf{p}) \cdot \mathbf{U}(\mathbf{p})\big)\mathbf{v}_{\mathbf{p}} = -\mathbf{v}_{\mathbf{p}}. \end{aligned}$$

Conversely, if we are given a regular surface $\mathcal{M} \subset \mathbb{R}^3$ with a globally-defined continuous complex structure $\mathbf{p} \mapsto J_{\mathbf{p}}$, we define $\mathbf{U}(\mathbf{p}) \in \mathbb{R}^3_{\mathbf{p}}$ by

$$\mathbf{U}(\mathbf{p}) = \frac{\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}}{\|\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}\|} \tag{11.2}$$

for any nonzero $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$. Then $\mathbf{U}(\mathbf{p})$ is perpendicular to both $\mathbf{v}_{\mathbf{p}}$ and $J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}$. Since $\mathbf{v}_{\mathbf{p}}$ and $J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}$ form a basis for $\mathcal{M}_{\mathbf{p}}$, it follows that $\mathbf{U}(\mathbf{p})$ is perpendicular to $\mathcal{M}_{\mathbf{p}}$. To check that the $\mathbf{U}(\mathbf{p})$ defined by (11.2) is independent of the choice of $\mathbf{v}_{\mathbf{p}}$, let $\mathbf{w}_{\mathbf{p}}$ be another nonzero tangent vector in $\mathcal{M}_{\mathbf{p}}$. Then $\mathbf{w}_{\mathbf{p}} = a\mathbf{v}_{\mathbf{p}} + bJ_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}$ is a linear combination of $\mathbf{v}_{\mathbf{p}}$, and

$$\mathbf{w}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{w}_{\mathbf{p}} = (a\,\mathbf{v}_{\mathbf{p}} + b\,J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}) \times (-b\,\mathbf{v}_{\mathbf{p}} + a\,J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}) = (a^2 + b^2)(\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}).$$

Hence

$$\frac{\mathbf{w}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{w}_{\mathbf{p}}}{\|\mathbf{w}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{w}_{\mathbf{p}}\|} = \frac{(a^2 + b^2)(\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}})}{\|(a^2 + b^2)(\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}})\|} = \frac{\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}}{\|\mathbf{v}_{\mathbf{p}} \times J_{\mathbf{p}}\mathbf{v}_{\mathbf{p}}\|},$$

and (11.2) is unambiguous. Since $\mathbf{p} \mapsto J_{\mathbf{p}}$ is continuous, so is $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$. ∎

Theorem 11.2 permits us to define the Gauss map of an arbitrary orientable regular surface in $\mathbb{R}^3$, rather than just a patch. Definition 10.22 on page 296 extends in an obvious fashion to

Definition 11.3. *Let $\mathcal{M}$ be an oriented regular surface in $\mathbb{R}^3$, and let $\mathbf{U}$ be a globally-defined unit normal vector field on $\mathcal{M}$ that defines the orientation of $\mathcal{M}$. Let $S^2(1)$ denote the unit sphere in $\mathbb{R}^3$. Then $\mathbf{U}$, viewed as a map*

$$\mathbf{U}\colon \mathcal{M} \longrightarrow S^2(1),$$

is called the ***Gauss map*** *of $\mathcal{M}$.*

It is easy to find examples of orientable regular surfaces in $\mathbb{R}^3$.

Lemma 11.4. *Let $\mathcal{U} \subseteq \mathbb{R}^2$. The graph $\mathcal{M}_h$ of a function $h\colon \mathcal{U} \to \mathbb{R}$ is an orientable regular surface.*

Proof. As in Section 10.4, we have the Monge patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}_h$, given by

$$\mathbf{x}(u, v) = \big(u,\ v,\ h(u, v)\big).$$

Then $\mathbf{x}$ covers all of $\mathcal{M}_h$; that is, $\mathbf{x}(\mathcal{U}) = \mathcal{M}_h$. Furthermore, $\mathbf{x}$ is regular and injective. The surface normal $\mathbf{U}$ to $\mathcal{M}_h$ is given by

$$\mathbf{U} \circ \mathbf{x} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = \frac{(-h_u,\ -h_v,\ 1)}{\sqrt{1 + h_u^2 + h_v^2}}.$$

The unit vector $\mathbf{U}$ is everywhere nonzero, and it follows from Theorem 11.2 that $\mathcal{M}_h$ is orientable. ∎

More generally, the method of proof of Lemma 11.4 yields:

Lemma 11.5. *Any surface $\mathcal{M} \subset \mathbb{R}^3$ which is the trace of a single injective regular patch $\mathbf{x}$ is orientable.*

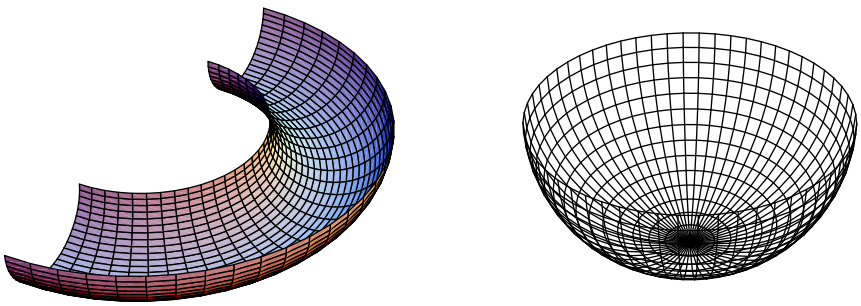


Figure 11.1: Gauss image of a quarter torus

As an example, Figure 11.1 shows that the Gauss map of a *quarter* of a torus covers *half* the unit sphere. Here is another generalization of Lemma 11.4.

Lemma 11.6. *Let $g\colon \mathbb{R}^3 \to \mathbb{R}$ be a differentiable function and c a number such that $\operatorname{grad} g$ is nonzero on all points of $\mathcal{M}(c) = \{\, \mathbf{p} \in \mathbb{R}^3 \mid g(\mathbf{p}) = c \,\}$. Then $\mathcal{M}(c)$, as well as every component of $\mathcal{M}(c)$, is orientable.*

Proof. Theorem 10.48, page 311, implies that $\mathcal{M}(c)$ is a regular surface and that $\operatorname{grad} g$ is everywhere perpendicular to $\mathcal{M}(c)$. We are assuming that $\operatorname{grad} g$ never vanishes on $\mathcal{M}(c)$. Putting these facts together we see that

$$\frac{\operatorname{grad} g}{\| \operatorname{grad} g \|}$$

is a unit vector field that is well defined at all points of $\mathcal{M}(c)$ and everywhere perpendicular to $\mathcal{M}(c)$. Then Theorem 11.2 says that $\mathcal{M}(c)$ and each of its components are orientable. ∎

Recall that a subset X of $\mathbb{R}^n$ is said to be ***closed*** if it contains all its limit points. Equivalently, X is closed if and only if its complement $\mathbb{R}^n \setminus X$ is open (see page 268). A subset X is said to be ***connected*** if any decomposition $X = X_1 \cup X_2$ of X into closed subsets X_1, X_2 with $X_1 \cap X_2$ empty must be trivial; that is, either $X = X_1$ or $X = X_2$.

Lemma 11.7. *Let $\mathcal{M}$ be a connected orientable regular surface in $\mathbb{R}^3$. Then $\mathcal{M}$ has exactly two globally-defined unit normal vector fields.*

Proof. Since $\mathcal{M}$ is orientable, it has at least one globally-defined unit normal vector field $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$. Let $\mathbf{p} \mapsto \mathbf{v}(\mathbf{p})$ be any other globally-defined unit normal vector field on $\mathcal{M}$. The sets $\mathcal{W}_\pm = \{\, \mathbf{p} \in \mathcal{M} \mid \mathbf{U}(\mathbf{p}) = \pm \mathbf{v}(\mathbf{p}) \,\}$ are closed, because $\mathbf{U}$ and $\mathbf{v}$ are continuous. Also $\mathcal{M} = \mathcal{W}_+ \cup \mathcal{W}_-$. The connectedness of $\mathcal{M}$ implies that $\mathcal{M}$ coincides with either $\mathcal{W}_+$ or $\mathcal{W}_-$. Consequently, the globally-defined unit vector fields on $\mathcal{M}$ are $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$ and $\mathbf{p} \mapsto -\mathbf{U}(\mathbf{p})$. ∎

Definition 11.8. *Let $\mathcal{M}$ be an orientable regular surface in $\mathbb{R}^3$. An **orientation** of $\mathcal{M}$ is the choice of a globally-defined unit normal vector field on $\mathcal{M}$.*

In general, it may not be possible to cover a regular surface with a single patch. If we have a family of patches, we must know how their orientations are related in order to define the orientation of a regular surface.

Lemma 11.9. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface. If $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ are patches such that $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ is nonempty, then*

$$\mathbf{y}_u \times \mathbf{y}_v = \det\big(\mathscr{J}(\mathbf{x}^{-1} \circ \mathbf{y})\big)\, \mathbf{x}_{\bar{u}} \times \mathbf{x}_{\bar{v}},$$

where $\mathscr{J}(\mathbf{x}^{-1} \circ \mathbf{y})$ denotes the Jacobian matrix of $\mathbf{x}^{-1} \circ \mathbf{y}$.

Proof. It follows from Lemma 10.31, page 300, that

$$\begin{aligned}\mathbf{y}_u \times \mathbf{y}_v &= \left(\frac{\partial \bar{u}}{\partial u}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_{\bar{v}}\right) \times \left(\frac{\partial \bar{u}}{\partial v}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial v}\mathbf{x}_{\bar{v}}\right) \\ &= \det \begin{pmatrix} \dfrac{\partial \bar{u}}{\partial u} & \dfrac{\partial \bar{u}}{\partial v} \\ \dfrac{\partial \bar{v}}{\partial u} & \dfrac{\partial \bar{v}}{\partial v} \end{pmatrix} \mathbf{x}_{\bar{u}} \times \mathbf{x}_{\bar{v}} \\ &= \det\big(\mathscr{J}(\mathbf{x}^{-1} \circ \mathbf{y})\big)\, \mathbf{x}_{\bar{u}} \times \mathbf{x}_{\bar{v}}. \ \blacksquare\end{aligned}$$

Definition 11.10. *We say that patches $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ on a regular surface $\mathcal{M}$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ nonempty are* ***coherently oriented****, provided the determinant of the Jacobian matrix $\mathscr{J}(\mathbf{x}^{-1} \circ \mathbf{y})$ is positive on $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$.*

Theorem 11.11. *A regular surface $\mathcal{M} \subset \mathbb{R}^3$ is orientable if and only if it is possible to cover $\mathcal{M}$ with a family $\mathfrak{B}$ of regular injective patches such that any two patches $(\mathbf{x},\mathcal{U})$, $(\mathbf{y},\mathcal{V})$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V}) \neq \emptyset$ are coherently oriented.*

Proof. Suppose that $\mathcal{M}$ is orientable. Theorem 11.2 implies that $\mathcal{M}$ has a globally-defined surface normal $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$. Let $\mathfrak{A}$ be a family of regular injective patches whose union covers $\mathcal{M}$. Without loss of generality, we may suppose that the domain of definition of each patch in $\mathfrak{A}$ is connected. We must construct from $\mathfrak{A}$ a family $\mathfrak{B}$ of coherently-oriented patches whose union covers $\mathcal{M}$.

To this end, we first note that if $\mathbf{x}$ is any regular injective patch on $\mathcal{M}$, then $\widetilde{\mathbf{x}}$ defined by

$$\widetilde{\mathbf{x}}(u, v) = \mathbf{x}(v, u)$$

is also a regular injective patch on $\mathcal{M}$. Moreover, $\widetilde{\mathbf{x}}$ and $\mathbf{x}$ are oppositely oriented, because $\widetilde{\mathbf{x}}_u \times \widetilde{\mathbf{x}}_v = -\mathbf{x}_u \times \mathbf{x}_v$. Now it is clear how to choose the patches that are to be members of the family $\mathfrak{B}$. If $\mathbf{x}$ is a patch in $\mathfrak{A}$ and

$$\frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = \mathbf{U}$$

we put $\mathbf{x}$ into $\mathfrak{B}$, but if

$$\frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = -\mathbf{U}$$

we put $\widetilde{\mathbf{x}}$ into $\mathfrak{B}$. Then $\mathfrak{B}$ is a family of coherently-oriented patches that covers $\mathcal{M}$.

Conversely, suppose that $\mathcal{M}$ is covered by a family $\mathfrak{B}$ of coherently-oriented patches. If $\mathbf{x}$ is a patch in $\mathfrak{B}$ defined on $\mathcal{U} \subset \mathbb{R}^2$, we define $\mathbf{U}_{\mathbf{x}}$ on $\mathbf{x}(\mathcal{U})$ by

$$\mathbf{U}_{\mathbf{x}}\big(\mathbf{x}(u, v)\big) = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}(u, v).$$

Let $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ be another patch in $\mathfrak{B}$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V}) \neq \emptyset$. Lemma 11.9 implies that on $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ we have

$$\mathbf{U}_{\mathbf{y}}\big(\mathbf{y}(u,v)\big) = \frac{\mathbf{y}_u \times \mathbf{y}_v}{\|\mathbf{y}_u \times \mathbf{y}_v\|}(u,v) \;=\; \left(\frac{\det \mathscr{J}(\mathbf{x}^{-1}\circ \mathbf{y})}{|\det \mathscr{J}(\mathbf{x}^{-1}\circ \mathbf{y})|}\frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}\right)(u,v)$$

$$= \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}(u,v) \;=\; \mathbf{U}_{\mathbf{x}}\big(\mathbf{x}(u,v)\big).$$

Thus we get a well-defined surface normal $\mathbf{U}$ on $\mathcal{M}$ by putting $\mathbf{U} = \mathbf{U}_{\mathbf{x}}$ on $\mathbf{x}(\mathcal{U})$ for any patch $\mathbf{x}$ in $\mathfrak{B}$. By Theorem 11.2, $\mathcal{M}$ is orientable. ▌

The proof of Theorem 11.11 establishes

Corollary 11.12. A family of coherently-oriented regular injective patches on a regular surface in $\mathbb{R}^3$ defines globally a unit normal vector field on a surface $\mathcal{M}$ in $\mathbb{R}^3$, that is, an orientation of $\mathcal{M}$.

The image under the Gauss map of the equatorial region

$$\big\{\mathsf{hyperboloid1}[1,1,1](u,v) \;\mid\; 0 \leqslant u \leqslant 2\pi,\; -1 \leqslant v \leqslant 1\big\}$$

of a hyperboloid of one sheet is illustrated in Figure 11.2. The image of the whole hyperboloid is also a bounded equatorial region, because the normals to the surface approach those of the asymptotic cone (shown in Figure 10.17). The image of the whole hyperboloid of two sheets under the Gauss map consists of two antipodal disks (see Exercise 2).

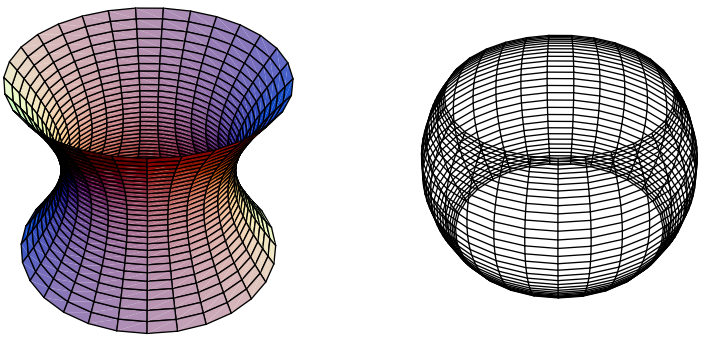


Figure 11.2: Part of a hyperboloid (left) and its Gauss image

11.2 Surfaces by Identification

There are useful topological descriptions of some elementary surfaces that are obtained from identifying edges of a square. For example, if we identify the

top and bottom edges, we obtain a cylinder. It is conventional to describe this identification by means of an arrow along the top edge and an arrow pointing in the same direction along the bottom edge, as in Figure 11.3 (left).

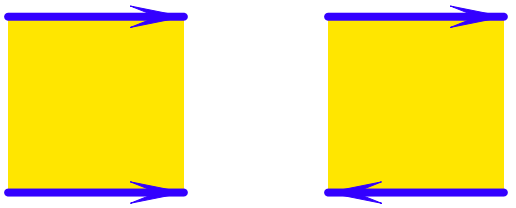


Figure 11.3: Models of the cylinder and Möbius strip

Now consider a square with the top and bottom edges identified, but in reverse order. This identification is indicated by means of an arrow along the top edge pointing in one direction and an arrow along the bottom edge pointing in the opposite direction. The resulting surface is called a ***Möbius***[1] ***strip*** or a ***Möbius band***. It is easy to put a cylinder into Euclidean space $\mathbb{R}^3$, but to find the actual parametrization of a Möbius strip in $\mathbb{R}^3$ is more difficult. We shall return to this problem shortly.

Now let us see what happens when we identify the vertical as well as the horizontal edges of a square. There are three possibilities. If the vertical arrows point in the same direction, and if the direction of the horizontal arrows is also the same, we obtain a torus. We have already seen on page 305 how to parametrize a torus in $\mathbb{R}^3$ using the function $\mathsf{torus}[a, b, c]$. For different values of a, b, c, we get tori which are different in shape, but provided $a > b$ they are topologically of the same type.

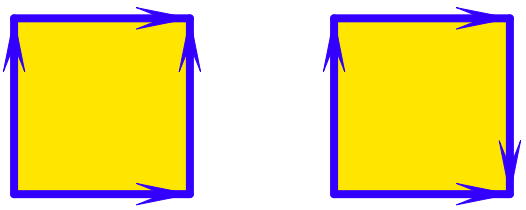
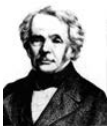


Figure 11.4: Models of the torus and Klein bottle

---

[1] 

August Ferdinand Möbius (1790–1868). Professor at the University of Leipzig. Möbius's name is attached to several important mathematical objects such as the Möbius function, the Möbius inversion formula, and Möbius transformations. He discovered the Möbius strip at age 71.

The ***Klein***[2] ***bottle*** is the surface that results when the edges of the square are identified with the vertical arrows pointing in the same direction, but the horizontal arrows pointing in opposite directions. Clearly, the same surface results if we interchange horizontal with vertical, to obtain the right-hand side of Figure 11.4.

Finally, the surface that results when we identify the edges of the square with the two vertical arrows pointing in different directions *and* the two horizontal arrows pointing in different directions is called the ***real projective plane***, and sometimes denoted $\mathbb{RP}^2$. This surface can also be thought of as a sphere with antipodal points identified. It is one of a series of important topological objects, namely the real projective spaces $\mathbb{RP}^n$, that are defined in Exercise 12 on page 799. The superscript indicates dimension, a concept that will be defined rigorously in Chapter 24. While $\mathbb{RP}^1$ is equivalent to a circle, it turns out that $\mathbb{RP}^3$ can be identified with the set of rotations of $\mathbb{R}^3$ discussed in Chapter 23.

We have described the Möbius strip, the Klein bottle and the real projective plane topologically. All of these surfaces turn out to be nonorientable. It is quite another matter, however, to find effective parametrizations of these surfaces in $\mathbb{R}^3$. That is the subject of the rest of this chapter.

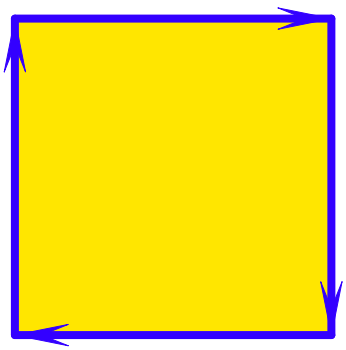


Figure 11.5: Model of the projective plane



[2] Christian Felix Klein (1849–1925). Klein made fruitful contributions to many branches of mathematics, including applied mathematics and mathematical physics. His Erlanger Programm (1872) instituted research directions in geometry for a half century; his subsequent work on Riemann surfaces established their essential role in function theory. In his writings, Klein concerned himself with what he saw as a developing gap between the increasing abstraction of mathematics and applied fields whose practitioners did not appreciate the fundamental rôle of mathematics, as well as with mathematics instruction at the secondary level. Klein discussed nonorientable surfaces in 1874 in [Klein].

11.3 The Möbius Strip

Another useful description of the Möbius strip is as the surface resulting from revolving a line segment around an axis, putting one twist in the line segment as it goes around the axis.

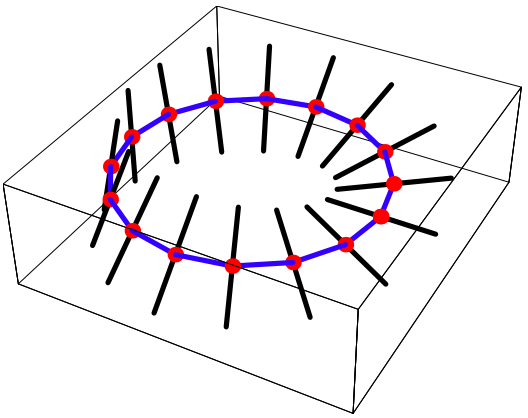


Figure 11.6: Lines twisting to form a Möbius strip

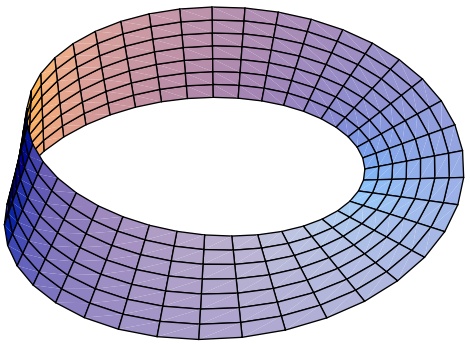


Figure 11.7: Möbius strip

We can use this model to get a parametrization of the strip:

$$\text{moebiusstrip}[a](u,v) = \Big(a\cos u + v\cos\frac{u}{2}\cos u,\ a\sin u + v\cos\frac{u}{2}\sin u,\ v\sin\frac{u}{2}\Big). \tag{11.3}$$

We see that the points

$$\text{moebiusstrip}[a](u,0) = \big(a\cos u,\ a\sin u,\ 0\big), \quad 0 \leqslant u < 2\pi,$$

constitute the central circle which therefore has radius a. It is convenient to retain this parameter a on the left-hand side in order to plot Möbius strips of different shape, and our notation reflects that of Notebook 11.

The points moebiusstrip$[a](u_0, v)$, as v varies for each fixed u_0, form a line segment meeting the central circle. As u_0 increases from 0 to 2π, the angle between this line and the xy-plane changes from 0 to π.

The nonorientability of the Möbius strip means that the Gauss map is not well defined on the whole surface: any attempt to define a unit normal vector on the *entire* Möbius strip is doomed to failure. However, the Gauss map is defined on any orientable portion, for example, on a Möbius strip minus a line orthogonal to the central circle.

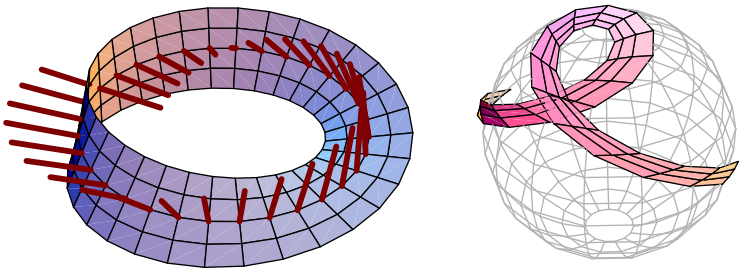


Figure 11.8: Gauss image of a Möbius strip traversed once

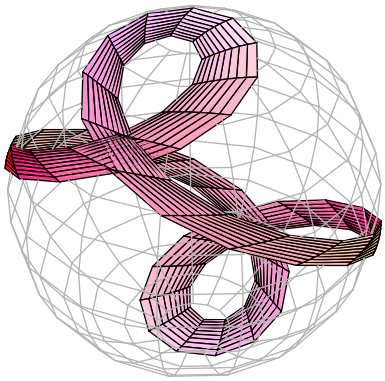


Figure 11.9: Gauss image of a Möbius strip traversed twice

It is possible to compute a normal vector field to the patch moebiusstrip[1]; the result is

$$\Big(\big(-v\cos\frac{u}{2}+2\cos u+v\cos\frac{3u}{2}\big)\sin\frac{u}{2},\ \cos\frac{u}{2}-\cos\frac{3u}{2}+v(\cos u+\sin^2 u),\ -2\cos\frac{u}{2}\big(1+v\cos\frac{u}{2}\big)\Big). \tag{11.4}$$

The Gauss map is determined by normalizing this vector. Although (11.4) is complicated, it is used in Notebook 11 to plot the accompanying figures. If one tries to extend the definition of the unit normal so that it is defined on all of the Möbius strip by going around the center circle, then the unit normal comes back on the other side of the surface (Figure 11.8). If we repeat this operation to return to where we started on the sphere, we have effectively associated *two* unit normal vectors to each point of the strip (Figure 11.9).

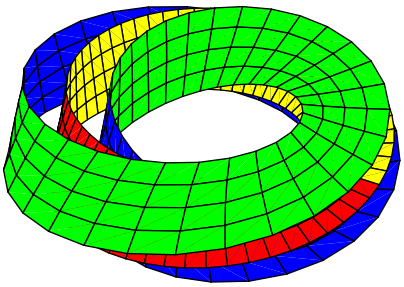


Figure 11.10: Parallel surface to the Möbius strip

Figure 11.10 displays a Möbius strip together with the surface formed by moving a small fixed distance along both of the unit normals emanating from each point of the strip. This construction is considered further in Notebook 11 and Section 19.8; here we merely observe that the resulting surface parallel to the Möbius strip is orientable.

11.4 The Klein Bottle

The Klein bottle is also a nonorientable surface, but in contrast to the Möbius strip it is ***compact***, equivalently closed without boundary, like the torus. An elementary account of the theory of compact surfaces can be found in [FiGa].

One way to define the Klein bottle is as follows. It is the surface that results from rotating a figure eight about an axis, but putting a twist in it. The rotation and twisting are the same as a Möbius strip but, instead of a line segment, one uses a figure eight.

Here is a parametrization of the Klein bottle that uses this construction:

$$\begin{aligned}\mathsf{kleinbottle}[a](u,v) \;=\; \Big(&\big(a+\cos\tfrac{u}{2}\sin v-\sin\tfrac{u}{2}\sin 2v\big)\cos u,\\ &\big(a+\cos\tfrac{u}{2}\sin v-\sin\tfrac{u}{2}\sin 2v\big)\sin u,\quad \sin\tfrac{u}{2}\sin v+\cos\tfrac{u}{2}\sin 2v\Big).\end{aligned}$$

The central circle of the Klein bottle is traversed twice by the curve

$$u \mapsto \mathsf{kleinbottle}[a](u,0), \quad 0 \leqslant u < 4\pi.$$

Each of the curves $v \mapsto \mathsf{kleinbottle}[a](u_0, v)$ is a figure eight. As u varies from 0 to 2π, the figure eights twist from 0 to π; this is the same twisting that we encountered in the parametrization (11.3) of the Möbius strip.

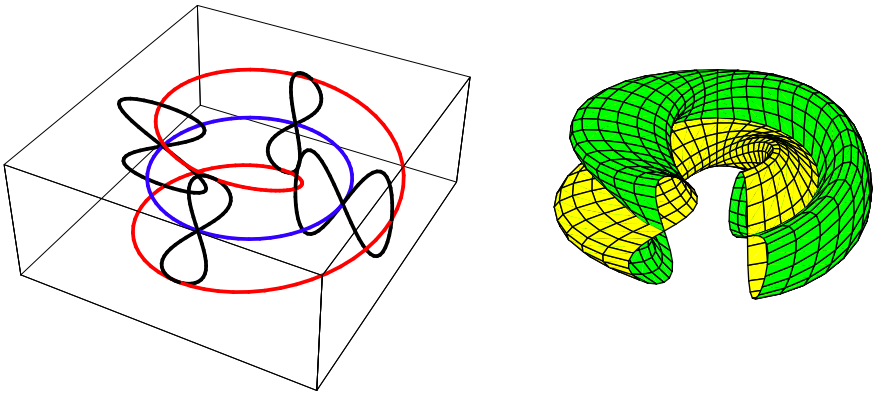


Figure 11.11: Figure eights twisting to form a Klein bottle

The Klein bottle is not a regular surface in $\mathbb{R}^3$ because it has self-intersections. However, the Klein bottle can be shown to be an abstract surface, a concept to be defined in Chapter 26. There are two kinds of Klein bottles in $\mathbb{R}^3$. The first one, $\mathcal{K}_1$ illustrated in Figure 11.11, has the feature that a neighborhood $\mathcal{V}_1$ of the self-intersection curve is nonorientable. In fact, $\mathcal{V}_1$ is formed from an "X" that rotates and twists about an axis in the same way that the figure eights move when they form the surface.

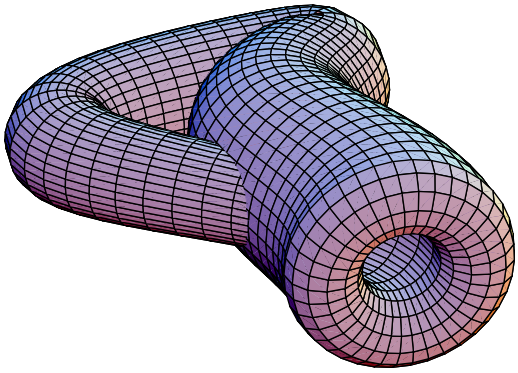


Figure 11.12: Klein bottle with an orientable neighborhood of the self-intersection curve

The original description of Klein (see [HC-V, pages 308–311]) of his surface was much different, and is described as follows. Consider a tube $\mathscr{T}$ of variable radius about a line. Topologically, a torus is formed from $\mathscr{T}$ by bending the tube until the ends meet and then gluing the boundary circles together. Another way to glue the ends is as follows. Let one end of $\mathscr{T}$ be a little smaller than the other. We bend the smaller end, then push it through the surface of the tube, and move it so that it is a concentric circle with the larger end, lying in the same plane. We complete the surface by adjoining half a torus on the other side of the plane. The result is shown in Figures 11.12 and 11.13.

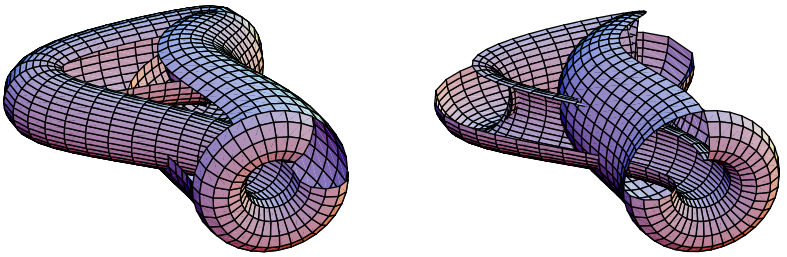


Figure 11.13: Open views displaying the self-intersection

To see that this new Klein bottle $\mathcal{K}_2$ is (as a surface in $\mathbb{R}^3$) distinct from the Klein bottle $\mathcal{K}_1$ formed by twisting figure eights, consider a neighborhood $\mathcal{V}_2$ of the self-intersection curve of $\mathcal{K}_2$. This neighborhood is again formed by rotating an "X," but in an orientable manner. Thus the difference between $\mathcal{K}_1$ and $\mathcal{K}_2$ is that $\mathcal{V}_1$ is nonorientable, but $\mathcal{V}_2$ is orientable. Nevertheless, $\mathcal{K}_1$ and $\mathcal{K}_2$ are topologically the same surface, because each can be formed by identifying sides of squares, as described in Section 11.2. A parametrization of $\mathcal{K}_2$ is given in Notebook 11.

11.5 Realizations of the Real Projective Plane

Let

$$S^2 = S^2(1) = \{\, \mathbf{p} \mid \|\mathbf{p}\| = 1\}$$

be the sphere of unit radius in $\mathbb{R}^3$. Recall that the ***antipodal map*** of the sphere is the diffeomorphism

$$\begin{aligned} S^2 &\longrightarrow S^2 \\ \mathbf{p} &\mapsto -\mathbf{p}. \end{aligned}$$

The real projective plane $\mathbb{RP}^2$ can be defined as the set that results when antipodal points of S^2 are identified; thus

$$\mathbb{RP}^2 = \big\{\, \{\mathbf{p}, -\mathbf{p}\} \mid \|\mathbf{p}\| = 1 \,\big\}.$$

To realize the real projective plane as a surface in $\mathbb{R}^3$ let us look for a map of $\mathbb{R}^3$ into itself with a special property.

Definition 11.13. *A map $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ such that*

$$F(-\mathbf{p}) = F(\mathbf{p}) \tag{11.5}$$

is said to have the ***antipodal property****.*

That we can use a map with the antipodal property to realize the real projective plane is a consequence of the following easily-proven lemma:

Lemma 11.14. *A map $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ with the antipodal property gives rise to a map $\widetilde{F}\colon \mathbb{RP}^2 \to F\big(S^2\big) \subset \mathbb{R}^3$ defined by*

$$\widetilde{F}(\{\mathbf{p}, -\mathbf{p}\}) = F(\mathbf{p}).$$

Thus we can realize $\mathbb{RP}^2$ as the image of S^2 under a map F which has the antipodal property. Moreover, any patch $\mathbf{x}\colon \mathcal{U} \to S^2$ will give rise to a patch $\widetilde{\mathbf{x}}\colon \mathcal{U} \to F\big(S^2\big)$ defined by

$$\widetilde{\mathbf{x}}(u, v) = F\big(\mathbf{x}(u, v)\big).$$

Ideally, one should choose F so that its Jacobian matrix is never zero. It turns out that this can be accomplished with quartic polynomials (see page 347), but the zeros can be kept to a minimum with well-chosen quadratic polynomials. We present three examples of maps with the antipodal property and associated realizations of $\mathbb{RP}^2$. The first is illustrated in Figure 11.14.

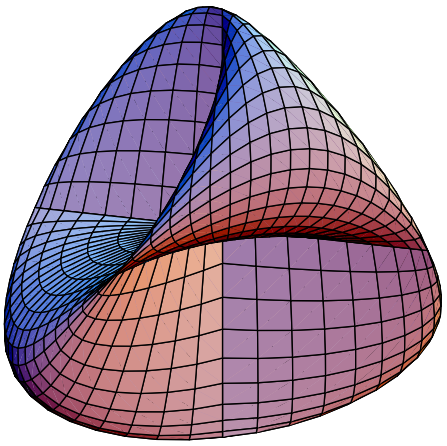


Figure 11.14: Steiner's Roman surface

Steiner's Roman Surface

When Jakob Steiner[3] visited Rome in 1844 he developed the concept of a surface that now carries his name (see [Apéry, page 37]). It is a realization of the real projective plane. To describe it, we first define the map

$$\mathsf{romanmap}(x, y, z) = (yz,\ zx,\ xy) \tag{11.6}$$

from $\mathbb{R}^3$ to itself. It is obvious that romanmap has the antipodal property; it therefore induces a map $\mathbb{RP}^2 \to \mathsf{romanmap}\big(S^2\big)$. We call this image ***Steiner's Roman surface*** of radius 1. Moreover, we can plot a portion of $\mathsf{romanmap}(S^2)$ by composing romanmap with any patch on S^2, for example, the standard parametrization

$$\mathsf{sphere}[1]\colon (u, v) \mapsto \big(\cos v \cos u,\ \cos v \sin u,\ \sin v\big).$$

Then the composition $\mathsf{romanmap} \circ \mathsf{sphere}[1]$ parametrizes all of Steiner's Roman surface.

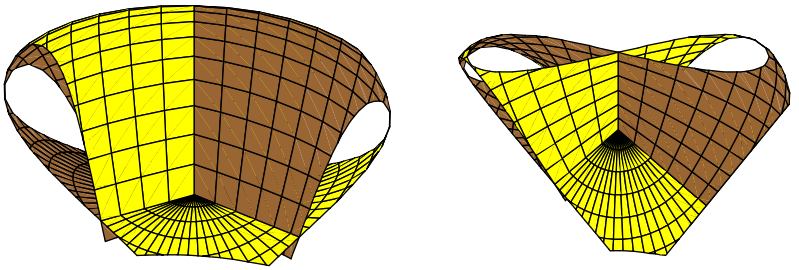


Figure 11.15: Two cut views of the Roman surface

Therefore, we define

$$\mathsf{roman}(u, v) = \tfrac{1}{2}\big(\sin u \sin 2v,\ \cos u \sin 2v,\ \sin 2u \cos^2 v\big).$$

Figure 11.15 cuts open Steiner's surface and displays the inside and outside in different colors to help visualize the intersections. Notice that the origin $(0, 0, 0)$ of $\mathbb{R}^3$ can be represented in one of the equivalent ways

$$\mathsf{romanmap}(\pm 1, 0, 0) = \mathsf{romanmap}(0, \pm 1, 0) = \mathsf{romanmap}(0, 0, \pm 1).$$



[3] Jakob Steiner (1796–1863). Swiss mathematician who was professor at the University of Berlin. Steiner did not learn to read and write until he was 14 and only went to school at the age of 18, against the wishes of his parents. Synthetic geometry was revolutionized by Steiner. He hated analysis as thoroughly as Lagrange hated geometry, according to [Cajori]. He believed that calculation replaces thinking while geometry stimulates thinking.

This confirms that the origin is a ***triple point*** of Steiner's surface, meaning that three separate branches intersect there. In addition, there are six singularities of the umbrella type illustrated in Figure 10.14. These occur at either end of each of the three "axes" visible in Figure 11.15, and one is clearly visible on the left front in Figure 11.14.

The Cross Cap

A mapping with the antipodal property formed from homogeneous quadratic polynomials, and similar to (11.6), is given by

$$\mathsf{crosscapmap}(x, y, z) = \big(yz,\ 2xy,\ x^2 - y^2\big).$$

We can easily get a parametrization of the cross cap, just as we did for Steiner's Roman surface. An explicit parametrization is

$$\mathsf{crosscap}(u, v) = \big(\tfrac{1}{2}\sin u \sin 2v,\ \sin 2u \cos^2 v,\ \cos 2u \cos^2 v\big). \tag{11.7}$$

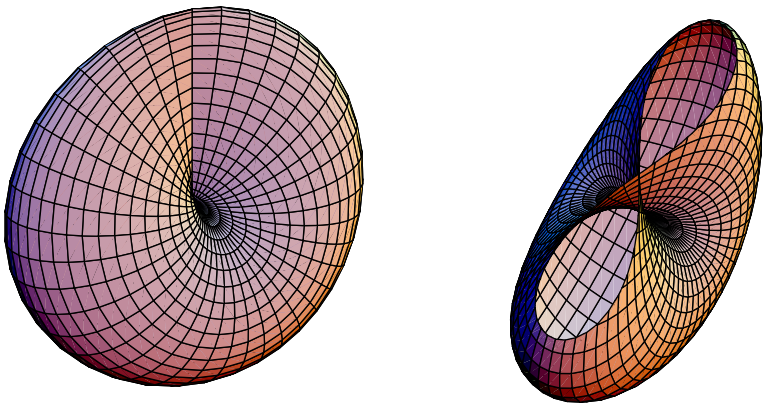


Figure 11.16: A cross cap and a cut view

The cross cap is perhaps the easiest realization of the sphere with antipodal points identified. Given a pair of points $\pm\mathbf{p}$ in S^2, either (i) both points belong to the equator (meaning $z = 0$), or (ii) they correspond to a unique point in the "southern hemisphere" (for which $z < 0$). To obtain $\mathbb{RP}^2$ from S^2, it therefore suffices to remove the open northern hemisphere (for which $z > 0$), and then deform the equator upwards towards where the north pole was, and "sew" it to itself so that opposite points on (what was) the equator are placed next to each other. This requires a segment in which the surface intersects itself, but is the idea behind Figure 11.16.

Boy's Surface

A more complicated mapping,

$$F = \big(F_1(x,y,z),\ F_2(x,y,z),\ F_3(x,y,z)\big),$$

with the antipodal property is one defining the remarkable surface discovered in 1901 by Boy,[4] whose components are homogeneous quartic polynomials. This description of the surface was found by Apéry [Apéry], and is obtained by taking

$$\begin{aligned} 4F_1(x,y,z) &= (x+y+z)\big((x+y+z)^3 + 4(y-x)(z-y)(x-z)\big), \\ F_2(x,y,z) &= (2x^2-y^2-z^2)(x^2+y^2+x^2) + 2yz(y^2-z^2) \\ &\qquad + zx(x^2-z^2) + xy(y^2-x^2), \\ \tfrac{1}{\sqrt{3}}F_3(x,y,z) &= (y^2-z^2)(x^2+y^2+z^2) + zx(z^2-x^2) + xy(y^2-x^2). \end{aligned}$$

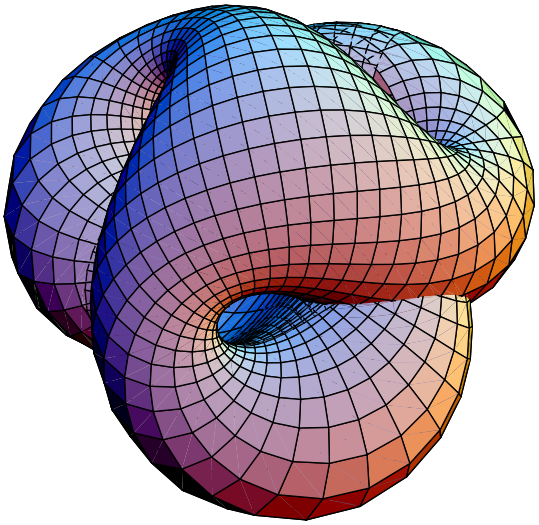


Figure 11.17: A view of Boy's surface

Figure 11.17 displays another realization of the surface, described in Notebook 11, using the concept of inversion from Section 20.4. Boy's surface can be covered by patches without singularities, in contrast to the two previous examples. Figure 11.18 helps to understand that curves of self-intersection meet in a triple point, at which the surface has a 3-fold symmetry, but there are no pinch points.

[4] Werner Boy, a student of David Hilbert (see page 602).

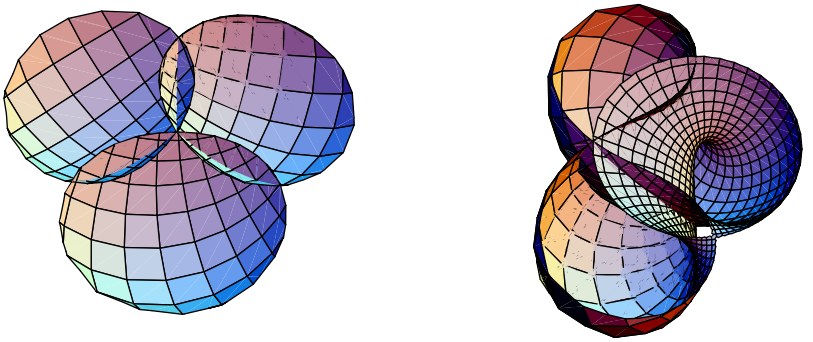


Figure 11.18: Two cut views of Boy's surface, revealing a triple self-intersection point

11.6 Twisted Surfaces

In this section we define a class of "twisted" surfaces that generalize the Klein bottle and Möbius strip.

Definition 11.15. *Let $\boldsymbol{\alpha}$ be a plane curve with the property that*

$$\boldsymbol{\alpha}(-t) = -\boldsymbol{\alpha}(t) \tag{11.8}$$

*and write $\boldsymbol{\alpha}(t) = (\varphi(t), \psi(t))$. The **twisted surface** with **profile curve** $\boldsymbol{\alpha}$ and parameters a and b is defined by*

$$\begin{aligned}\mathsf{twist}[\boldsymbol{\alpha}, a, b](u, v) \;=\; &\Big(a + \cos(bu)\varphi(v) - \sin(bu)\psi(v)\Big)(\cos u,\ \sin u,\ 0) \\ &+\Big(\sin(bu)\varphi(v) + \cos(bu)\psi(v)\Big)(0, 0, 1).\end{aligned}$$

To understand the significance of this definition, take $a = 0$ and $b = 1/2$. The coordinate curve of the surface defined for each fixed value of u consists of a copy of $\boldsymbol{\alpha}$ mapping to the plane Π_u generated by the vectors $(\cos u, \sin u, 0)$ and $(0, 0, 1)$. This is the image of the xz-plane Π_0 under a rotation by u, and the curve $\boldsymbol{\alpha}$ in Π_u is rotated by the same angle u. Note that the curve $\boldsymbol{\alpha}$ starts from the xz-plane (corresponding to $u = 0$), and by the time it returns to the same plane ($u = 2\pi$) it has been rotated by only 180°, though the two traces coincide by (11.8). The resulting surface may or may not be orientable, depending on the choice of $\boldsymbol{\alpha}$.

We shall now show that both the Möbius strip and Klein bottle can be constructed in this way.

The Möbius Strip

Define $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}(t) = (at,\ 0)$. Then

$$\begin{aligned}\mathsf{twist}\big[\boldsymbol{\alpha}, a, \tfrac{1}{2}\big](u,v) &= a\Big(\cos u + v\cos\frac{u}{2}\cos u,\ \sin u + v\cos\frac{u}{2}\sin u,\ v\sin\frac{u}{2}\Big)\\ &= \mathsf{moebiusstrip}[a](u,v).\end{aligned}$$

The Klein Bottle

Define $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}(t) = (\sin t,\ \sin 2t)$. Then

$$\begin{aligned}\mathsf{twist}\big[\boldsymbol{\gamma}, a, \tfrac{1}{2}\big](u,v) &= \Big((a + \cos\frac{u}{2}\sin v - \sin\frac{u}{2}\sin 2v)\cos u,\\ &\qquad (a + \cos\frac{u}{2}\sin v - \sin\frac{u}{2}\sin 2v)\sin u,\quad \sin\frac{u}{2}\sin v + \cos\frac{u}{2}\sin 2v\Big)\\ &= \mathsf{kleinbottle}[a](u,v).\end{aligned}$$

Twisted Surfaces of Lissajous Curves

Instead of twisting a figure eight around a circle we can twist a Lissajous curve. The latter was defined on page 58 by the parametrization

$$\mathsf{lissajous}[n,d,a,b](t) = \big(a\sin(nt+d),\ b\sin t\big).$$

For example, $\mathsf{lissajous}[2,0,1,1]$ resembles a figure eight, and the associated twisted surface is a Klein bottle. Figure 11.19 (left) displays the trace of the curve $\boldsymbol{\beta} = \mathsf{lissajous}[4,0,1,1]$, whilst the right-hand side consists of the points

$$\mathsf{twist}[\boldsymbol{\beta}, 2, \tfrac{1}{2}](u,v) \qquad 0 \leqslant u, v < 2\pi,$$

describing another self-intersecting surface.

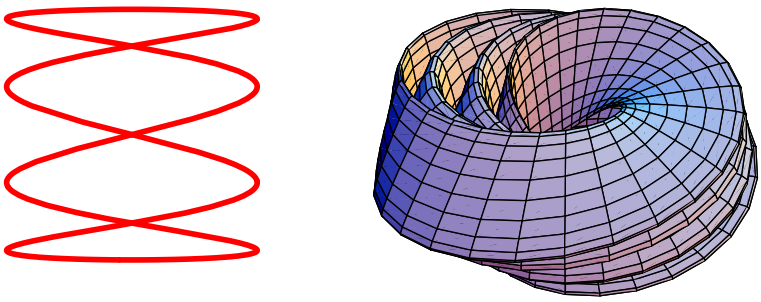


Figure 11.19: A Lissajous curve and twisted surface

11.7 Exercises

1. Prove Corollary 11.12.

M **2.** Sketch the image under the Gauss map of the region

$$\left\{ \mathsf{hyperboloid2}[1,1,1](u,v) \mid 0 \leqslant u \leqslant 2\pi,\ 1 \leqslant v \right\}$$

of the 2-sheeted hyperboloid. Show that the image of the whole hyperboloid of two sheets under the Gauss map consists of two antipodal disks, and is not therefore the entire sphere.

3. The image of the Gauss map of an ellipsoid is obviously the whole unit sphere. Nonetheless, plotting the result can be of interest since one can visualize the image of coordinate curves. Explain how the curves visible in Figure 11.20, left, describe the orientation of the unit normal to the patch $\mathsf{ellipsoid}[3,2,1]$. The associated approximation, right, is therefore an "elliptical" polyhedral representation of the sphere.

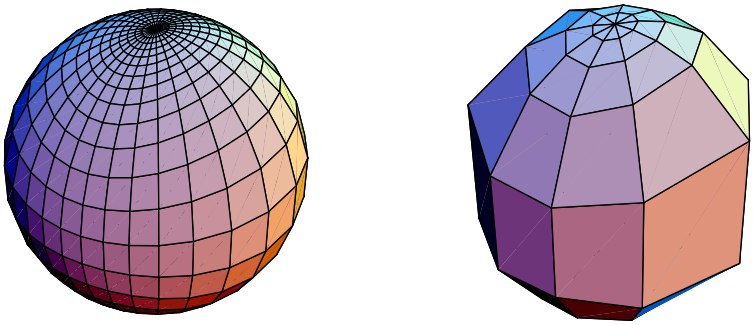


Figure 11.20: Gauss images of $\mathsf{ellipsoid}[3,2,1]$

M **4.** Let $\mathbf{x} = \mathsf{moebiusstrip}[1]$, and consider the two patches

$$\mathbf{x}(u,v), \qquad 0 < u < \tfrac{3}{2}\pi,\ -1 < v < 1;$$
$$\mathbf{x}(u,v), \qquad -\pi < u < \tfrac{1}{2}\pi,\ -1 < v < 1.$$

Compute the unit normal (see (11.4)), and prove that the two normals have equal and opposite signs on their two respective regions of intersection. Deduce that the Möbius strip is nonorientable.

5. Verify that Equation (11.7) is indeed the composition of sphere$[a]$ with crosscapmap.

M **6.** A version of the cross cap can be defined implicitly by the equation

$$(ax^2 + by^2)(x^2 + y^2 + z^2) - 2z(x^2 + y^2) = 0.$$

Plot this surface using ImplicitPlot3D.

7. Let us put a family of figure eight curves into $\mathbb{R}^3$ in such a way that the first and last figure eights reduce to points. This defines a surface, which we call a ***pseudo cross cap***, shown in Figure 11.21. It can be parametrized as follows:

$$\text{pseudocrosscap}(u, v) = \big((1 - u^2)\sin v,\ (1 - u^2)\sin 2v,\ u\big).$$

Show that pseudocrosscap is not regular at the points $(0, 0, \pm 1)$. Why should the pseudo cross cap be considered orientable, even though it is not a regular surface?

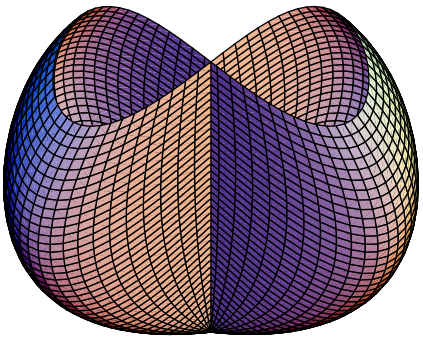


Figure 11.21: A pseudo cross cap

8. Show that the image under crosscapmap of a torus centered at the origin is topologically a Klein bottle. To do this, explain how points of the image can be made to correspond to those in Figure 11.4 with the right-hand identifications.

Notebook 11

■ N11.0

This notebook is primarily concerned with the visualization of nonorientable surfaces. However, it initially makes extensive use of the Jacobian matrix, to understand the role of the unit normal associated to a patch, and the Gauss map is illustrated for regions of a torus and a hyperboloid.

From a practical point of view, the Arrows package plays an important role in drawing the diagrams used to model nonorientable surfaces, the topic of N11.2. The Möbius strip is studied in N11.3, and Figure 11.8 illustrates its nonorientability by displaying the behavior of a unit normal vector. A similar analysis is reserved for the Klein bottle in N11.4. The unified construction of these surfaces via *twisting* is verified in N11.6. Steiner's Roman surface and the cross cap are investigated in N11.5.

The electronic version of this notebook includes a miniprogram for producing parallel surfaces, illustrated with reference to the Möbius strip, in N11.3.1. (This topic is discussed in Section 19.8 of the text.) Boy's surface is plotted in N11.5.1, and animations of other nonorientable surfaces are to be found in N11.7.1.

The operator J represents rotation by $\pi/2$, and was introduced in N1.1.

```
Unprotect[J]
  J[{x_, y_}]:= {-y, x}
Protect[J]
```

We shall need the following packages. The first will be used to provide the names of two distinct colors in plotting surfaces, the second for its WireFrame command, and the third in N11.2.

```
<< Graphics`Colors`
<< Graphics`Shapes`
<< Graphics`Arrow`
```

The programs jacobian and unitnormal from the previous notebook are reloaded in N11.0.1, as are standard definitions of the torus, ellipsoid and 1-sheeted hyperboloid, all used in Chapter 11 to illustrate the Gauss map.

■ N11.0.1

■ N11.1

Take the example in which the patch x is a torus of revolution:

```
x:= torus[3,1,1]
x[u, v]
```

By definition, the Jacobian matrix has size 3 by 2, since each column represents a partial derivative of the patch.

```
jacobian[x][u,v]
MatrixForm[%]
```

Once simplified, the unit normal to the patch is evidently a point of the unit sphere:

```
unitnormal[x][u,v]
%//PowerExpand
```

The above computations are combined in the following program, in which x and y are used as abbreviations for the surface and its unit normal to make the subsequent commands clearer. The chart y is plotted in the usual way prior to applying the WireFrame command. To view the Gauss map of another surface, it suffices to change the definition of x and adjust the plotting ranges.

```
Module[{x,y,z},
  x= torus[3,1,1];
  y= unitnormal[x];
  z[1]:= ParametricPlot3D[x[u,v], {u,-π/2,π/2}, {v,-π,0}];
  z[2]:= ParametricPlot3D[y[u,v], {u,-π/2,π/2}, {v,-π,0}];
  z[3]:= Show[WireFrame[z[2]]];
  S @ GraphicsArray[{z[1], z[3]}, GraphicsSpacing->-.1]
]
```

Figure 11.1

■ N11.1.1

N11.2

We shall use the Arrow package loaded in N11.0 and ensuing macro to construct arrows of exactly the desired shape and size. Unlike the definition in N6.0, this version of ar accepts a single argument consisting of a *pair* of endpoints. It is adapted for use in the command sq to construct arrows on specified edges of a square.

```
ar[{v_, w_}]:= {Hue[.7], AbsoluteThickness[3],
  Arrow[v, w, HeadScaling->Absolute, HeadWidth->.4,
  HeadLength->20, HeadCenter->.5]}
```

The Module command is avoided in the next cell in order to leave the definition of sq available for two more graphics objects z2,z3 corresponding to Figures 11.4 and 11.5 in N11.2.1.

```
a:= .15;
o:= {0,0}
p:= {1,0}
q:= {0,1}
r:= {1,1}
bg:= {Hue[0.15],Polygon[{o,p,r,q}]}
sq[A__]:= Show[Graphics[{bg, ar/@ {A}}],
  AspectRatio->Automatic, PlotRange->{{-a,1+a},{-a,1+a}}]
z1:= {sq[{o,p},{q,r}], sq[{p,o},{q,r}]}
S @ GraphicsArray[z1, GraphicsSpacing->.4]
```

Figure 11.3

N11.2.1

N11.3

A standard parametrization of the Möbius strip or band is

```
moebiusstrip[a_][u_,v_]:= {a Cos[u] + v Cos[u/2]Cos[u],
  a Sin[u] + v Cos[u/2]Sin[u], v Sin[u/2]}
```

```
Module[{m,z},
  m= moebiusstrip[1];
  z[u_]= {Thickness[.01], Line[{m[u,-.4], m[u,.4]}],
    Hue[.7], Line[{m[u,0], m[u+π/8,0]}],
    Hue[0], PointSize[.03], Point[m[u,0]]};
  S @ Graphics3D[Table[z[u], {u,0,2π,π/8}]]
]
```

Figure 11.6

```
S @ ParametricPlot3D[moebiusstrip[4][u,v], {u,0,2π}, {v,-1,1},
  PlotPoints->{40, 8}]
```

Figure 11.7

We point out that a Möbius strip can be plotted using the built-in command from the Shapes package, loaded in N11.0. The following produces comparable output to the previous cell; it can be seen by adding S@ at the front.

```
Graphics3D[MoebiusStrip[4,1,50]]
```

Next we compute the Jacobian of our parametrization.

```
jacobian[moebiusstrip[1]][u,v]//MatrixForm
```

Having run the command unitnormal below, we can refer to an unnormalized normal vector Nx, the definition of which was not bound to the Module. The last line then produces output that was used in the chapter text.

```
unitnormal[moebiusstrip[1]][u,v]
FullSimplify[Nx]
InputForm[%]
```

The next figures represent the unit normal vectors to a Möbius strip in various ways. In the first cell, z[1] combines the strip with a family of normal lines and z[2] places a unit normal vector of the same strip on a spherical background.

```
Module[{m,y,pm,z,py},
  m= moebiusstrip[1];
  y= unitnormal[m];
  col = RGBColor[.5,0,0];
  li[u_]= {col, Thickness[.01],
    Line[{m[u,0], m[u,0]+.5y[u,0]}]};
  pm= ParametricPlot3D[m[u,v], {u,0,2π}, {v,-.3,.3},
    PlotPoints->{33,5}];
  z[1]= Show[pm, Graphics3D[Table[li[u], {u,π,3π,π/16}]]];
  py:= ParametricPlot3D[y[u,v], {u,π,3π}, {v,-.3,.3},
    PlotPoints->{20,4}, AmbientLight->col];
  wf= WireFrame[{GrayLevel[.7], Sphere[.95,15,15]}];
  z[2]= Show[py, wf];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->-.3]
]
```

Figure 11.8

```
Module[{m,y,py},
  m= moebiusstrip[1];
  y= unitnormal[m];
  py:= ParametricPlot3D[y[u,v], {u,0,4π}, {v,-.3,.3},
    PlotPoints->{40,10}, AmbientLight->Hue[0]];
  wf= WireFrame[{GrayLevel[.7], Sphere[.98,15,15]}];
  S[py, wf]
]
```

Figure 11.9

N11.3.1

N11.4

The Klein bottle can be defined in an analogous way to the Möbius strip:

```
kleinbottle[a_][u_,v_]:=
  {(a + Cos[u/2]Sin[v] - Sin[u/2]Sin[2 v])Cos[u],
   (a + Cos[u/2]Sin[v] - Sin[u/2]Sin[2 v])Sin[u],
   Sin[u/2]Sin[v] + Cos[u/2]Sin[2 v]}
```

The next figure displays its definition in terms of a twisted figure eight.

```
klb:= kleinbottle[2]
z0:= {Thickness[.008],
  Table[Line[Table[klb[u,v], {v,0,2π, π/32}]], {u,0,2π,π/2}],
  Hue[.7], Line[Table[klb[u,0], {u,0,2π, π/32}]], Hue[0],
  Line[Table[klb[u, π/2], {u,0,4π,π/32}]]}
z1:= Graphics3D[z0]
z2:= ParametricPlot3D[Append[klb[u,v], FaceForm[Green, Yellow]],
   {v,0,2π}, {u,-π/4,3π/2}, PlotPoints->{32,32},
   Lighting->False]
S @ GraphicsArray[{z1,z2}, GraphicsSpacing->0]
```

Figure 11.11

N11.4.1

N11.5

Steiner's Roman surface is the subject of the following cells.

```
romanmap[{x_,y_,z_}]:= {y z, z x, x y}
Composition[Sin,Cos][x]
Simplify[Composition[romanmap,sphere[1]][u, v]]
```

```
roman[u_,v_]:=
  {Sin[u]Sin[2v], Cos[u]Sin[2v], Sin[2u]Cos[v]^2}/2
```

ViewPoint coordinates can easily be changed in the next figure:

```
Clear[y]
y[opts___]:= ParametricPlot3D[roman[u,v], {u,0,π}, {v,-π/2,π/2},
  PlotPoints->{39,39}, opts]
S[y[ViewPoint->{-4,-2,1}]]
```

Figure 11.14

The parameter ranges in Figure 11.15 were selected by trial and error to best illustrate the inner structure of the Roman Surface.

```
rr[u_,v_]:= Append[roman[u,v], FaceForm[Green,Yellow]]
y[X__]:= ParametricPlot3D[rr[u,v], {u,0,π}, {v,0,π},
  Lighting->False, PlotPoints->39, X]
z1:= y[PlotRange->{{-.4,.4},{-.3,.3},{-.05,.5}},
  ViewPoint->5{1.2, -1, .5}]
z2:= y[PlotRange->{{-.4,.3}, Automatic, Automatic}]
z2:= y[PlotRange->{{-.4,.4},{-.3,.3},{-.1,.4}},
 ViewPoint->{1,-1,2}/2]
z3:= y[ViewPoint -> {-2.8,-1.9,.3}]
S @ GraphicsArray[{z1,z2}, GraphicsSpacing->.1]
```

Figure 11.15

Next we compute the Jacobian and regularity of the patch defining the Roman surface.

```
JM= jacobian[roman][u,v]
Det[Transpose[JM].JM]
%//Simplify
```

The cross cap formulas from the text are given next, and then used for Figure 11.16.

```
crosscapmap[{x_,y_,z_}]:= {y z, 2 x y, x^2 - y^2}
```

```
crosscap[u_,v_]:=
  {Sin[u]Sin[2 v]/2, Sin[2 u]Cos[v]^2, Cos[2 u]Cos[v]^2}
```

The relationship between these two definitions is verified by a computation in N11.7 below.

```
Module[{z,y},
  cc= crosscap;
  y[opts__]:= ParametricPlot3D[cc[u,v], {u,0,π}, {v,-π/2,π/2},
    PlotPoints->{40, 40}, opts];
  z[1]= y[ViewPoint -> {2.7,-1.5,1.4}];
  z[2]= y[ViewPoint->{.5,-1,3.2},
    PlotRange->{Automatic, Automatic, {-1,.5}}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->-.1]
]
```

Figure 11.16

Excellent images of nonorientable surfaces can be found within Richard Palais's 3D-Xplor-Math gallery (see http://vmm.math.uci.edu), which was the source for Figure 11.17.

■ N11.5.1

■ N11.6

Twisted surfaces can be plotted by means of the following program. The curve determined by fixing u lies in the plane generated by the unit vector F[u] in the xy-plane and the unit vector k in the direction of the z-axis.

```
twist[α_,a_,b_][u_,v_]:=
Module[{f,F,k},
  f= {Sin[#], Cos[#]} & ;
  F= {Cos[#], Sin[#], 0} & ;
  k= {0,0,1};
  (a + f[b u].J[α[v]])F[u] + f[b u].α[v] k
]
```

A Möbius strip is obtained by twisting a line segment:

```
moebiusstrip[a][u,v]
twist[{#1,0}&, a, 1/2][u,v]
Simplify[% - %%]
```

A Klein bottle is obtained by twisting an eight curve:

```
kleinbottle[a][u,v]
twist[{Sin[#],Sin[2#]}&, a, 1/2][u,v]
% - %%
```

The lemnsicate curve (first defined in N2.2) provides a more complicated example, and we include instructions to plot its twisted surface. The latter can be seen in Figure 13.11 at the end of Chapter 13.

```
lemniscate[t_]:= {Cos[t], Sin[t]Cos[t]}/(1 + Sin[t]^2)
```

```
ParametricPlot3D[twist[lemniscate, 2, 1/2][u,v], {u,-π/8,3π/2},
  {v,0,2π}, PlotPoints->{40,40}]
```

■ N11.6.1

■ N11.7

The following cells help with some of the exercises. For Exercise 2, we simplify the unit normal of a hyperboloid of one sheet.

```
x:= hyperboloid1[1,1,1];
jacobian[x][u,v]
Simplify[PowerExpand[unitnormal[x][u,v]]]
```

Similar calculations are needed for Exercise 3:

```
x:= ellipsoid[3,2,1];
jacobian[x][u,v];
y[u_,v_]:= unitnormal[x][u,v]
y[u,v]//FullSimplify
```

```
Module[{z},
  z[n_]:= ParametricPlot3D[y[u,v], {u,0,2π}, {v,-π/2, π/2},
    PlotPoints->n];
  S @ GraphicsArray[{z[30],z[10]}, GraphicsSpacing->0]
]
```

Figure 11.20

For Exercise 5, we use

```
Composition[crosscapmap, sphere[a]][u,v]//Simplify
```

■ N11.7.1

Chapter 12

Metrics on Surfaces

In this chapter, we begin the study of the geometry of surfaces from the point of view of distance and area. When mathematicians began to study surfaces at the end of the eighteenth century, they did so in terms of *infinitesimal* distance and area. We explain these notions intuitively in Sections 12.1 and 12.4, but from a modern standpoint.

We first define the coefficients E, F, G of the first fundamental form and investigate their transformation under a change of coordinates (Lemma 12.4). Lengths and areas are defined by integrals involving the quantities E, F, G.

The concept of an *isometry* between surfaces is defined and illustrated by means of a circular cone in Section 12.2. The extension of this concept to *conformal maps* is introduced in Section 12.3. The latter also includes a discussion of the *distance function* defined by a metric on a surface.

In Section 12.5, we compute metrics and areas for a selection of simple surfaces, as an application of the theory.

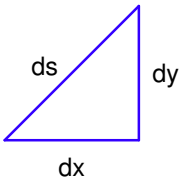
12.1 The Intuitive Idea of Distance

So far we have not discussed how to measure distances on a surface. One of the key facts about distance in Euclidean space $\mathbb{R}^n$ is that the Pythagorean Theorem holds. This means that if $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ are points in $\mathbb{R}^n$, then the distance s from $\mathbf{p}$ to $\mathbf{q}$ is given by

$$s^2 = (p_1 - q_1)^2 + \cdots + (p_n - q_n)^2. \tag{12.1}$$

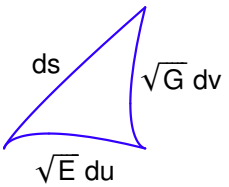
How is this notion different for a surface? Because a general surface is curved, distance on it is not the same as in Euclidean space; in particular, (12.1) is in general false however we interpret the coordinates. To describe how to measure distance on a surface, we need the mathematically imprecise concept of an "infinitesimal." The infinitesimal version of (12.1) for $n = 2$ is

$$ds^2 = dx^2 + dy^2. \tag{12.2}$$



We can think of dx and dy as small quantities in the x and y directions. The formula (12.2) is valid for $\mathbb{R}^2$. For a surface, or more precisely a patch, the corresponding equation is

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2, \tag{12.3}$$



where E, F, G are functions that we define below. This is the classical notation[1] for a metric on a surface. We may consider (12.3) as an infinitesimal warped version of the Pythagorean Theorem.

Among the many ways to define metrics on surfaces, one is especially simple and important. Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^n$. The scalar or dot product of $\mathbb{R}^n$ gives rise to a scalar product on $\mathcal{M}$ by restriction. If $\mathbf{v_p}$ and $\mathbf{w_p}$ are tangent vectors to $\mathcal{M}$ at $\mathbf{p} \in \mathcal{M}$, we can take the scalar product $\mathbf{v_p} \cdot \mathbf{w_p}$ because $\mathbf{v_p}$ and $\mathbf{w_p}$ are also tangent vectors to $\mathbb{R}^n$ at $\mathbf{p}$.

In this chapter we deal only with local properties of distance, so without loss of generality, we can assume that $\mathcal{M}$ is the image of an injective regular patch. However, the following definition makes sense for any patch in $\mathbb{R}^n$.

Definition 12.1. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a patch. Define functions $E, F, G\colon \mathcal{U} \to \mathbb{R}$ by*

$$E = \|\mathbf{x}_u\|^2, \qquad F = \mathbf{x}_u \cdot \mathbf{x}_v, \qquad G = \|\mathbf{x}_v\|^2.$$

Then $ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$ is the ***Riemannian metric*** *or* ***first fundamental form*** *of the patch $\mathbf{x}$. Furthermore, E, F, G are called the* ***coefficients of the first fundamental form*** *of $\mathbf{x}$.*

The equivalence of the formal definition of E, F, G with the infinitesimal version (12.3) can be better understood if we consider curves on a patch.

---

[1] This notation (including the choice of letters E, F and G) was already in use in the early part of the 19th century; it can be found, for example, in the works of Gauss (see [Gauss2] and [Dom]). Gauss had the idea to study properties of a surface that are independent of the way the surface sits in space, that is, properties of a surface that are expressible in terms of E, F, G alone. One such property is Gauss's *Theorema Egregium*, which we shall prove in Section 17.2.

Lemma 12.2. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^n$ be a curve that lies on a regular injective patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$, and fix c with $a < c < b$. The arc length function s of $\boldsymbol{\alpha}$ starting at $\boldsymbol{\alpha}(c)$ is given by*

$$s(t) = \int_c^t \sqrt{E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}\, dt. \tag{12.4}$$

Proof. Write $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ for $a < t < b$. By the definition of the arc length function and Corollary 10.15 on page 294, we have

$$\begin{aligned} s(t) = \int_c^t \|\boldsymbol{\alpha}'(t)\| dt &= \int_c^t \big\|u'(t)\mathbf{x}_u\big(u(t),v(t)\big) + v'(t)\mathbf{x}_v\big(u(t),v(t)\big)\big\|\, dt \\ &= \int_c^t \sqrt{u'(t)^2\big\|\mathbf{x}_u\big\|^2 + 2u'(t)v'(t)\,\mathbf{x}_u\cdot\mathbf{x}_v + v'(t)^2\big\|\mathbf{x}_v\big\|^2}\, dt \\ &= \int_c^t \sqrt{E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}\, dt. \end{aligned}$$

∎

It follows from (12.4) that

$$\frac{ds}{dt} = \sqrt{E\left(\frac{du}{dt}\right)^2 + 2F\frac{du}{dt}\frac{dv}{dt} + G\left(\frac{dv}{dt}\right)^2}. \tag{12.5}$$

Now we can make sense of (12.3). We square both sides of (12.5) and multiply through by dt^2. Although strictly speaking multiplication by the infinitesimal dt^2 is not permitted, at least formally we obtain (12.3).

Notice that the right-hand side of (12.3) does not involve the parameter t, except insofar as u and v depend on t. We may think of ds as the ***infinitesimal arc length***, because it gives the arc length function when integrated over any curve. Geometrically, ds can be interpreted as the infinitesimal distance from a point $\mathbf{x}(u,v)$ to a point $\mathbf{x}(u+du, v+dv)$ measured along the surface. Indeed, to first order,

$$\boldsymbol{\alpha}(t+dt) \;\approx\; \boldsymbol{\alpha}(t) + \boldsymbol{\alpha}'(t)dt \;=\; \boldsymbol{\alpha}(t) + \mathbf{x}_u u'(t)dt + \mathbf{x}_v v'(t)dt,$$

so that

$$\begin{aligned} \big\|\boldsymbol{\alpha}(t+dt) - \boldsymbol{\alpha}(t)\big\| &\;\approx\; \big\|\mathbf{x}_u u'(t) + \mathbf{x}_v v'(t)\big\|\, dt \\ &= \sqrt{E\,u'^2 + 2F\,u'v' + G\,v'^2}\, dt \;=\; ds. \end{aligned}$$

A standard notion from calculus of several variables is that of the *differential* of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$; it is given by

$$df = \frac{\partial f}{\partial u}du + \frac{\partial f}{\partial v}dv.$$

More generally, if $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a regular injective patch and $f\colon \mathcal{M} \to \mathbb{R}$ is a differentiable function, we put

$$df = \frac{\partial(f \circ \mathbf{x})}{\partial u} du + \frac{\partial(f \circ \mathbf{x})}{\partial v} dv.$$

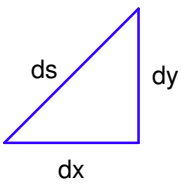
We call df the ***differential*** of f. The differentials of the functions $\mathbf{x}(u, v) \mapsto u$ and $\mathbf{x}(u, v) \mapsto v$ are denoted by du and dv. In spite of its appearance, ds will hardly ever be the differential of a function on a surface, since ds^2 represents a *nondegenerate* quadratic form. But formally, $d\mathbf{x} = \mathbf{x}_u du + \mathbf{x}_v dv$, so that

$$\begin{aligned} d\mathbf{x} \cdot d\mathbf{x} &= \big(\mathbf{x}_u du + \mathbf{x}_v dv\big) \cdot \big(\mathbf{x}_u du + \mathbf{x}_v dv\big) \\ &= \|\mathbf{x}_u\|^2 du^2 + 2\mathbf{x}_u \cdot \mathbf{x}_v du dv + \|\mathbf{x}_v\|^2 dv^2 \\ &= E\, du^2 + 2F\, du dv + G\, dv^2 \\ &= ds^2. \end{aligned}$$

Equation (12.2) represents the square of the infinitesimal distance on $\mathbb{R}^2$ written in terms of the Cartesian coordinates x and y. There is a different, but equally familiar, expression for ds^2 in polar coordinates.

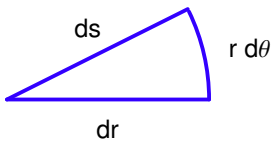
Lemma 12.3. *The metric ds^2 on $\mathbb{R}^2$, given in Cartesian coordinates as*

$$ds^2 = dx^2 + dy^2,$$



in polar coordinates becomes

$$ds^2 = dr^2 + r^2 d\theta^2.$$



Proof. We have the standard change of variable formulas from rectangular to polar coordinates:

$$\begin{cases} x = r\cos\theta, \\ y = r\sin\theta. \end{cases}$$

Therefore,

$$\begin{cases} dx &= -r\sin\theta\, d\theta + \cos\theta\, dr, \\ dy &= r\cos\theta\, d\theta + \sin\theta\, dr. \end{cases} \tag{12.6}$$

Hence

$$\begin{aligned} dx^2 + dy^2 &= (-r\sin\theta\, d\theta + \cos\theta\, dr)^2 + (r\cos\theta\, d\theta + \sin\theta\, dr)^2 \\ &= dr^2 + r^2 d\theta^2. \end{aligned}$$ ∎

We need to know how the expression for a metric changes under a change of coordinates.

Lemma 12.4. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ be patches on a regular surface $\mathcal{M}$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ nonempty. Let $\mathbf{x}^{-1} \circ \mathbf{y} = (\bar{u}, \bar{v})\colon \mathcal{U} \cap \mathcal{V} \to \mathcal{U} \cap \mathcal{V}$ be the associated change of coordinates, so that*

$$\mathbf{y}(u, v) = \mathbf{x}\big(\bar{u}(u, v), \bar{v}(u, v)\big).$$

Suppose that $\mathcal{M}$ has a metric and denote the induced metrics on $\mathbf{x}$ and $\mathbf{y}$ by

$$ds^2_{\mathbf{x}} = E_{\mathbf{x}} d\bar{u}^2 + 2F_{\mathbf{x}} d\bar{u} d\bar{v} + G_{\mathbf{x}} d\bar{v}^2 \quad \text{and} \quad ds^2_{\mathbf{y}} = E_{\mathbf{y}} du^2 + 2F_{\mathbf{y}} du dv + G_{\mathbf{y}} dv^2.$$

Then

$$\left\{ \begin{aligned} E_{\mathbf{y}} &= E_{\mathbf{x}} \left(\frac{\partial \bar{u}}{\partial u}\right)^2 + 2F_{\mathbf{x}} \frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial u} + G_{\mathbf{x}} \left(\frac{\partial \bar{v}}{\partial u}\right)^2, \\ F_{\mathbf{y}} &= E_{\mathbf{x}} \frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{u}}{\partial v} + F_{\mathbf{x}} \left(\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial v} + \frac{\partial \bar{u}}{\partial v}\frac{\partial \bar{v}}{\partial u}\right) + G_{\mathbf{x}} \frac{\partial \bar{v}}{\partial u}\frac{\partial \bar{v}}{\partial v}, \\ G_{\mathbf{y}} &= E_{\mathbf{x}} \left(\frac{\partial \bar{u}}{\partial v}\right)^2 + 2F_{\mathbf{x}} \frac{\partial \bar{u}}{\partial v}\frac{\partial \bar{v}}{\partial v} + G_{\mathbf{x}} \left(\frac{\partial \bar{v}}{\partial v}\right)^2. \end{aligned} \right. \tag{12.7}$$

Proof. We use Lemma 10.31 on page 300, to compute

$$\begin{aligned} E_{\mathbf{y}} = \mathbf{y}_u \cdot \mathbf{y}_u &= \left(\frac{\partial \bar{u}}{\partial u}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_{\bar{v}}\right) \cdot \left(\frac{\partial \bar{u}}{\partial u}\mathbf{x}_{\bar{u}} + \frac{\partial \bar{v}}{\partial u}\mathbf{x}_{\bar{v}}\right) \\ &= \left(\frac{\partial \bar{u}}{\partial u}\right)^2 \mathbf{x}_{\bar{u}} \cdot \mathbf{x}_{\bar{u}} + 2\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial u}\mathbf{x}_{\bar{u}} \cdot \mathbf{x}_{\bar{v}} + \left(\frac{\partial \bar{v}}{\partial u}\right)^2 \mathbf{x}_{\bar{v}} \cdot \mathbf{x}_{\bar{v}} \\ &= \left(\frac{\partial \bar{u}}{\partial u}\right)^2 E_{\mathbf{x}} + 2\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial u}F_{\mathbf{x}} + \left(\frac{\partial \bar{v}}{\partial u}\right)^2 G_{\mathbf{x}}. \end{aligned}$$

The other equations are proved similarly. ∎

12.2 Isometries between Surfaces

In Section 10.5, we defined the notion of a mapping F between surfaces in $\mathbb{R}^n$. Even if F is a local diffeomorphism (see page 309), $F(\mathcal{M})$ can be quite different from $\mathcal{M}$. Good examples are the maps romanmap and crosscapmap of Section 11.5 that map a sphere onto Steiner's Roman surface and a cross cap. Differentiable maps that preserve infinitesimal distances, on the other hand, distort much less. The search for such maps leads to the following definition.

Definition 12.5. *Let $\mathcal{M}_1, \mathcal{M}_2$ be regular surfaces in $\mathbb{R}^n$. A map $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ is called a* ***local isometry*** *provided its tangent map satisfies*

$$\|\Phi_*(\mathbf{v_p})\| = \|\mathbf{v_p}\| \tag{12.8}$$

for all tangent vectors $\mathbf{v_p}$ to $\mathcal{M}_1$. An ***isometry*** *is a surface mapping which is simultaneously a local isometry and a diffeomorphism.*

Lemma 12.6. *A local isometry is a local diffeomorphism.*

Proof. It is easy to see that (12.8) implies that each tangent map of a local isometry Φ is injective. Then the Inverse Function Theorem implies that Φ is a local diffeomorphism. ∎

Every isometry of $\mathbb{R}^n$ which maps a regular surface $\mathcal{M}_1$ onto a regular surface $\mathcal{M}_2$ obviously restricts to an isometry between $\mathcal{M}_1$ and $\mathcal{M}_2$. When we study minimal surfaces in Chapter 16, we shall see that there are isometries between surfaces that do not however arise in this fashion.

Let $\mathbf{v_p}, \mathbf{w_p}$ be tangent vectors to $\mathcal{M}_1$. Suppose that (12.8) holds. Since Φ_* is linear,

$$\begin{aligned}
\Phi_*(\mathbf{v_p}) \cdot \Phi_*(\mathbf{w_p}) &= \frac{1}{2}\Big(\|\Phi_*(\mathbf{v_p} + \mathbf{w_p})\|^2 - \|\Phi_*(\mathbf{v_p})\|^2 - \|\Phi_*(\mathbf{w_p})\|^2\Big) \\
&= \frac{1}{2}\Big(\|(\mathbf{v_p} + \mathbf{w_p})\|^2 - \|\mathbf{v_p}\|^2 - \|\mathbf{w_p}\|^2\Big) \\
&= \mathbf{v_p} \cdot \mathbf{w_p}.
\end{aligned}$$

It follows that (12.8) is equivalent to

$$\Phi_*(\mathbf{v_p}) \cdot \Phi_*(\mathbf{w_p}) = \mathbf{v_p} \cdot \mathbf{w_p}. \tag{12.9}$$

Next, we show that a surface mapping is an isometry if and only if it preserves Riemannian metrics.

Lemma 12.7. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a regular injective patch and let $\mathbf{y}\colon \mathcal{U} \to \mathbb{R}^n$ be any patch. Let*

$$ds^2_{\mathbf{x}} = E_{\mathbf{x}}du^2 + 2F_{\mathbf{x}}dudv + G_{\mathbf{x}}dv^2 \quad \text{and} \quad ds^2_{\mathbf{y}} = E_{\mathbf{y}}du^2 + 2F_{\mathbf{y}}dudv + G_{\mathbf{y}}dv^2$$

denote the induced Riemannian metrics on $\mathbf{x}$ and $\mathbf{y}$. Then the map

$$\Phi = \mathbf{y} \circ \mathbf{x}^{-1}\colon \mathbf{x}(\mathcal{U}) \longrightarrow \mathbf{y}(\mathcal{U})$$

is a local isometry if and only if

$$ds^2_{\mathbf{x}} = ds^2_{\mathbf{y}}. \tag{12.10}$$

Proof. First, note that

$$\Phi_*(\mathbf{x}_u) = \big((\mathbf{y}\circ\mathbf{x}^{-1})_*\circ\mathbf{x}_*\big)\left(\frac{\partial}{\partial u}\right) = \mathbf{y}_*\left(\frac{\partial}{\partial u}\right) = \mathbf{y}_u;$$

similarly, $\Phi_*(\mathbf{x}_v) = \mathbf{y}_v$. If Φ is a local isometry, then

$$E_{\mathbf{y}} = \|\mathbf{y}_u\|^2 = \|\Phi_*(\mathbf{x}_u)\|^2 = \|\mathbf{x}_u\|^2 = E_{\mathbf{x}}.$$

In the same way, $F_{\mathbf{y}} = F_{\mathbf{x}}$ and $G_{\mathbf{y}} = G_{\mathbf{x}}$. Thus (12.10) holds.

To prove the converse, consider a curve $\boldsymbol{\alpha}$ of the form

$$\boldsymbol{\alpha}(t) = \mathbf{x}\big(u(t), v(t)\big);$$

then $(\Phi\circ\boldsymbol{\alpha})(t) = \mathbf{y}(u(t), v(t))$. From Corollary 10.15 we know that

$$\boldsymbol{\alpha}' = u'\mathbf{x}_u + v'\mathbf{x}_v \qquad \text{and} \qquad (\Phi\circ\boldsymbol{\alpha})' = u'\mathbf{y}_u + v'\mathbf{y}_v.$$

Hence

$$\text{(12.11)} \qquad \begin{cases} \|\boldsymbol{\alpha}'\|^2 \;=\; E_{\mathbf{x}}u'^2 + 2F_{\mathbf{x}}u'v' + G_{\mathbf{x}}v'^2, \\ \|(\Phi\circ\boldsymbol{\alpha})'\|^2 \;=\; E_{\mathbf{y}}u'^2 + 2F_{\mathbf{y}}u'v' + G_{\mathbf{y}}v'^2. \end{cases}$$

Then (12.10) and (12.11) imply that

$$\text{(12.12)} \qquad \|\boldsymbol{\alpha}'(t)\| = \|(\Phi\circ\boldsymbol{\alpha})'(t)\|.$$

Since every tangent vector to $\mathbf{x}(\mathcal{U})$ can be represented as $\boldsymbol{\alpha}'(0)$ for some curve $\boldsymbol{\alpha}$, it follows that Φ is an isometry. ∎

Corollary 12.8. *Let $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a surface mapping. Given a patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}_1$, set $\mathbf{y} = \Phi\circ\mathbf{x}$. Then Φ is a local isometry if and only if for each regular injective patch $\mathbf{x}$ on $\mathcal{M}_1$ we have $ds^2_{\mathbf{x}} = ds^2_{\mathbf{y}}$.*

We shall illustrate these results with a vivid example by constructing a patch on a circular cone in terms of Euclidean coordinates (u, v) in the plane. We begin with the region in the plane surrounded by a circular arc AB of radius 1 subtending an angle AOB of α radians, shown in Figure 12.1.

We shall attach the edge OA to OB, in the fashion of the previous chapter, but make this process explicit by identifying the result with a circular (half) cone in 3-dimensional space. The top circle of the cone has circumference α (this being the length of the existing arc AB) and thus radius $\alpha/2\pi$. It follows that if β denotes the angle (between the axis and a generator) of the cone, then

$$\sin\beta = \frac{\alpha}{2\pi}.$$

For example, if $\alpha = \pi$, the initial region is semicircular and the cone has angle $\beta = \pi/6$ or 30°.

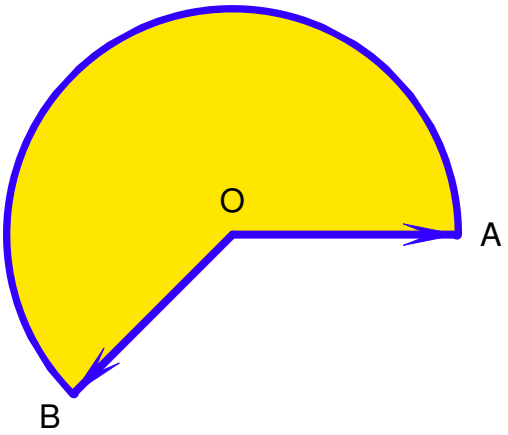


Figure 12.1: A planar region

The planar region may be realized as a surface in $\mathbb{R}^3$ by merely mapping it into the xy-plane. Using polar coordinates, it is then the image of the patch

$$\mathbf{x}(r,\theta) = (r\cos\theta,\ r\sin\theta,\ 0)$$

defined for $0 \leqslant r \leqslant 1$ and $0 \leqslant \theta \leqslant \alpha$. We can parametrize the resulting conical surface by realizing that a radial line in the preceding plot becomes a generator of the cone. A distance r along this generator contributes a horizontal distance of $r\sin\beta$ and a vertical distance $r\cos\beta$. The angle about the z-axis has to run through a full turn, so a patch of the conical surface is

$$\mathbf{y}(r,\theta) = \Big(r\cos\frac{2\pi\theta}{\alpha}\sin\beta,\ r\sin\frac{2\pi\theta}{\alpha}\sin\beta,\ r\cos\beta\Big), \tag{12.13}$$

for $0 \leqslant r \leqslant 1$ and $0 \leqslant \theta \leqslant \alpha$. The final result is close to Figure 12.2 (right).

Let $\mathcal{V}$ denote the open interior of the region in $\mathbb{R}^2$ shown in Figure 12.1. The above procedure determines a mapping $\Phi\colon \mathcal{V} \to \mathbb{R}^3$ for which

$$\mathbf{y}(u,v) = \Phi\big(\mathbf{x}(u,v)\big),$$

in accord with Corollary 12.8. Indeed, Φ represents the "rigid" folding of a piece of paper representing Figure 12.1 into the cone, and is therefore an isometry. This is confirmed by a computation of the respective first fundamental forms. The metric $ds^2_{\mathbf{x}}$ is given in polar coordinates (r,θ) by Lemma 12.3. But an easy computation of E, F, G for $\mathbf{y}$ (carried out in Notebook 12) shows that $ds^2_{\mathbf{y}}$ has the identical form. This means that the same norms and angles are induced on the corresponding tangent vectors on the surface of the cone.

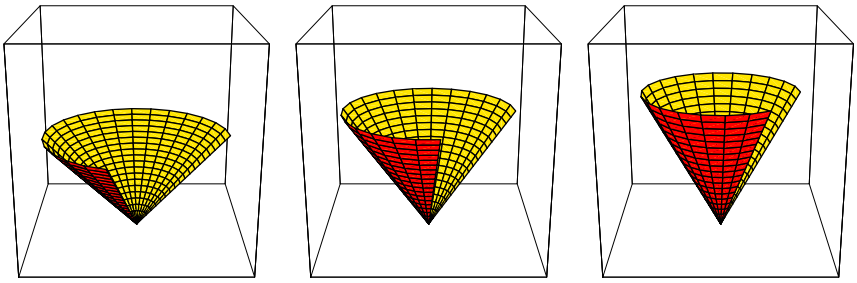


Figure 12.2: Isometries between conical surfaces

The following patch interpolates between $\mathbf{x}$ and $\mathbf{y}$ by representing an intermediate conical surface

$$\mathbf{x}_t(r,\theta) = \Big(r\cos\frac{\theta}{\lambda_t}\sin\beta_t,\ r\sin\frac{\theta}{\lambda_t}\sin\beta_t,\ r\cos\beta_t\Big),$$

where

$$\lambda_t = 1 - t + \frac{\alpha t}{2\pi}, \qquad \beta_t = \arcsin\lambda_t.$$

For each fixed t with $0 \leqslant t \leqslant 1$, this process determines an isometry from the planar region to the intermediate surface, for example Figure 12.2 (middle). In Notebook 12, it is used to create an animation of surfaces passing from $\mathbf{x}$ to the closed cone $\mathbf{y}$, representing the act of folding the piece of paper.

12.3 Distance and Conformal Maps

A Riemannian metric determines a function which measures distances on a regular surface.

Definition 12.9. *Let $\mathcal{M} \subset \mathbb{R}^n$ be a regular surface, and let $\mathbf{p}, \mathbf{q} \in \mathcal{M}$. Then the* ***intrinsic distance*** $\boldsymbol{\rho}(\mathbf{p},\mathbf{q})$ *is the greatest lower bound of the lengths of all piecewise-differentiable curves $\boldsymbol{\alpha}$ that lie entirely on $\mathcal{M}$ and connect $\mathbf{p}$ to $\mathbf{q}$. We call $\boldsymbol{\rho}$ the* ***distance function*** *of $\mathcal{M}$.*

In general, the intrinsic distance $\boldsymbol{\rho}(\mathbf{p},\mathbf{q})$ will be greater than the Euclidean distance $\|\mathbf{p}-\mathbf{q}\|$, since the surface will not contain the straight line joining $\mathbf{p}$ to $\mathbf{q}$. This fact is evident in Figure 12.3.

An isometry also preserves the intrinsic distance:

Lemma 12.10. *Let $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ be an isometry. Then Φ identifies the intrinsic distances $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2$ of $\mathcal{M}_1$ and $\mathcal{M}_2$ in the sense that*

$$\boldsymbol{\rho}_2\big(\Phi(\mathbf{p}),\Phi(\mathbf{q})\big) = \boldsymbol{\rho}_1(\mathbf{p},\mathbf{q}), \tag{12.14}$$

for $\mathbf{p}, \mathbf{q} \in \mathcal{M}_1$.

Proof. Let $\boldsymbol{\alpha}\colon (c,d) \to \mathcal{M}$ be a piecewise-differentiable curve with $\boldsymbol{\alpha}(a) = \mathbf{p}$ and $\boldsymbol{\alpha}(b) = \mathbf{q}$, where $c < a < b < d$. From the definition of local isometry, it follows that

$$\mathsf{length}[\boldsymbol{\alpha}] = \int_a^b \|\boldsymbol{\alpha}'(t)\|\,dt = \int_a^b \|(\Phi \circ \boldsymbol{\alpha})'(t)\|\,dt = \mathsf{length}[\Phi \circ \boldsymbol{\alpha}]$$

(see page 9). Since Φ is a diffeomorphism, there is a one-to-one correspondence between the piecewise-differentiable curves on $\mathcal{M}_1$ connecting $\mathbf{p}$ to $\mathbf{q}$ and the piecewise-differentiable curves on $\mathcal{M}_2$ connecting $\Phi(\mathbf{p})$ to $\Phi(\mathbf{q})$. Since corresponding curves have equal lengths, we obtain (12.14). ▮

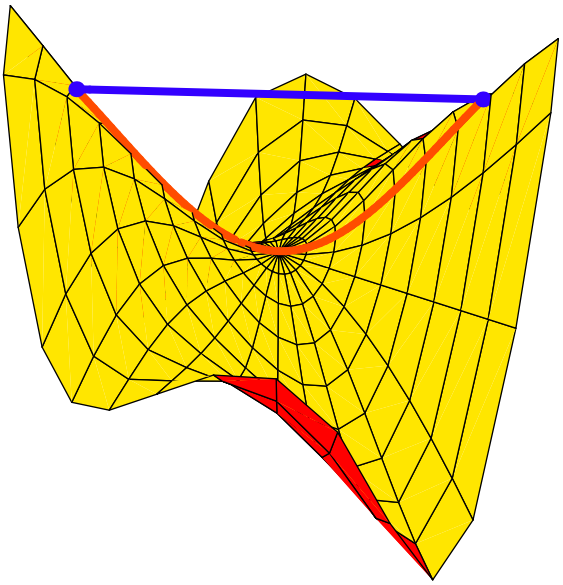


Figure 12.3: Distance on a surface

Next, let us consider a generalization of the notion of local isometry.

Definition 12.11. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be regular surfaces in $\mathbb{R}^n$. Then a map $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ is called a* ***conformal map*** *provided there is a differentiable everywhere positive function $\lambda\colon \mathcal{M}_1 \to \mathbb{R}$ such that*

$$\|\Phi_*(\mathbf{v}_\mathbf{p})\| = \lambda(\mathbf{p})\|\mathbf{v}_\mathbf{p}\| \tag{12.15}$$

for all $\mathbf{p} \in \mathcal{M}_1$ and all tangent vectors $\mathbf{v}_\mathbf{p}$ to $\mathcal{M}_1$ at $\mathbf{p}$. We call λ the ***scale factor****. A* ***conformal diffeomorphism*** *is a surface mapping which is simultaneously a conformal map and a diffeomorphism.*

Since the tangent map Φ_* of an isometry preserves inner products, it also preserves both lengths and angles. The tangent map of a conformal diffeomorphism in general will change the *lengths* of tangent vectors, but we do have

Lemma 12.12. *A conformal map $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ preserves the angles between nonzero tangent vectors.*

Proof. The proof that (12.9) is equivalent to (12.8) can be easily extended to show that (12.15) is equivalent to

$$\Phi_*(\mathbf{v_p}) \cdot \Phi_*(\mathbf{w_p}) = \lambda(\mathbf{p})^2\, \mathbf{v_p} \cdot \mathbf{w_p} \tag{12.16}$$

for all nonzero tangent vectors $\mathbf{v_p}$ and $\mathbf{w_p}$ to $\mathcal{M}_1$. Then (12.16) implies that

$$\frac{\Phi_*(\mathbf{v_p}) \cdot \Phi_*(\mathbf{w_p})}{\|\Phi_*(\mathbf{v_p})\|\, \|\Phi_*(\mathbf{w_p})\|} = \frac{\mathbf{v_p} \cdot \mathbf{w_p}}{\|\mathbf{v_p}\|\, \|\mathbf{w_p}\|}.$$

From the definition of angle on page 3, it follows that Φ_* preserves angles between tangent vectors. ∎

Here is an important example of a conformal map, which was in fact introduced in Section 8.6.

Definition 12.13. *Let $S^2(1)$ denote the unit sphere in $\mathbb{R}^3$. The* ***stereographic map*** *$\boldsymbol{\Upsilon}\colon \mathbb{R}^2 \to S^2(1)$ is defined by*

$$\boldsymbol{\Upsilon}(p_1, p_2) = \frac{(2p_1,\ 2p_2,\ p_1^2 + p_2^2 - 1)}{p_1^2 + p_2^2 + 1}.$$

We abbreviate the last equation to

$$\boldsymbol{\Upsilon}(\mathbf{p}) = \frac{(2\mathbf{p};\ \|\mathbf{p}\|^2 - 1)}{1 + \|\mathbf{p}\|^2}, \tag{12.17}$$

for $\mathbf{p} \in \mathbb{R}^2$ (the semicolon reminds us that the numerator is a vector, not an inner product). This makes it easy to prove that $\|\boldsymbol{\Upsilon}(\mathbf{p})\| = 1$ and that, as remarked on page 250, $\boldsymbol{\Upsilon}$ is differentiable. Moreover,

Lemma 12.14. *$\boldsymbol{\Upsilon}$ is a conformal map.*

Proof. Let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ be a curve. It follows from (12.17) that the image of $\boldsymbol{\alpha}$ by $\boldsymbol{\Upsilon}$ is the curve

$$\boldsymbol{\Upsilon} \circ \boldsymbol{\alpha} = \frac{(2\boldsymbol{\alpha};\ \|\boldsymbol{\alpha}\|^2 - 1)}{1 + \|\boldsymbol{\alpha}\|^2},$$

and this equation can be rewritten as

$$(1 + \|\boldsymbol{\alpha}\|^2)(\boldsymbol{\Upsilon} \circ \boldsymbol{\alpha}) = (2\boldsymbol{\alpha};\ \|\boldsymbol{\alpha}\|^2 - 1). \tag{12.18}$$

Differentiating (12.18), we obtain

$$2(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}')(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})+(1+\|\boldsymbol{\alpha}\|^2)(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})'=2(\boldsymbol{\alpha}';\,\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}'),$$

and taking norms,

$$\big\|2(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}')(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})+(1+\|\boldsymbol{\alpha}\|^2)(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})'\big\|^2=4\|(\boldsymbol{\alpha}';\,\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}')\|^2. \tag{12.19}$$

Since $\boldsymbol{\Upsilon}(\mathbb{R}^2)\subset S^2(1)$, we have

$$\|\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha}\|^2=1 \qquad\text{and}\qquad (\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})\cdot(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})'=0.$$

It therefore follows from (12.19) that

$$4(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}')^2+(1+\|\boldsymbol{\alpha}\|^2)^2\big\|(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})'\big\|^2=4\|\boldsymbol{\alpha}'\|^2+4(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}')^2,$$

whence

$$\big\|(\boldsymbol{\Upsilon}\circ\boldsymbol{\alpha})'\big\|=\frac{2\|\boldsymbol{\alpha}'\|}{1+\|\boldsymbol{\alpha}\|^2}.$$

We conclude that

$$\big\|\boldsymbol{\Upsilon}_*(\mathbf{v_p})\big\|=\frac{2\|\mathbf{v_p}\|}{1+\|\mathbf{p}\|^2} \tag{12.20}$$

for all $\mathbf{p}$ and all tangent vectors $\mathbf{v_p}$ to $\mathbb{R}^2$ at $\mathbf{p}$. Hence $\boldsymbol{\Upsilon}$ is conformal with scale factor $\lambda(\mathbf{p})=2/(1+\|\mathbf{p}\|^2)$. ∎

12.4 The Intuitive Idea of Area

In $\mathbb{R}^n$, the infinitesimal hypercube bounded by $dx_1,\ldots,dx_n$ has as its volume the product

$$dV=dx_1dx_2\cdots dx_n.$$

We call dV the ***volume element*** of $\mathbb{R}^n$. The corresponding concept for a surface with metric (12.3), page 362, is the ***element of area*** dA, given by

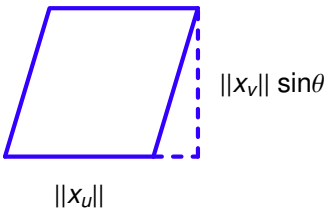
$$dA=\sqrt{EG-F^2}\,dudv \tag{12.21}$$



For a patch $\mathbf{x}\colon\mathcal{U}\to\mathbb{R}^3$, it is easy to see why this is the appropriate definition. By (7.2) on page 193, we have

$$\sqrt{EG-F^2}=\|\mathbf{x}_u\times\mathbf{x}_v\|=\|\mathbf{x}_u\|\,\|\mathbf{x}_v\|\sin\theta, \tag{12.22}$$

where θ is the oriented angle from the vector $\mathbf{x}_u$ to the vector $\mathbf{x}_v$. On the other hand, the quantity $\|\mathbf{x}_u\|\,\|\mathbf{x}_v\|\sin\theta$ represents the area of the parallelogram with sides $\mathbf{x}_u, \mathbf{x}_v$ and angle θ between the sides:



It is important to realize that the expression "$du dv$" occurring in (12.21) has a very different meaning to that in the middle term $2F\,du\,dv$ of (12.3). The latter arises from the *symmetric* product $\mathbf{x}_u \cdot \mathbf{x}_v$, while equation (12.22) shows that dA is determined by an *antisymmetric* product which logically should be written $du \times dv$ or, as is customary, $du \wedge dv$, satisfying

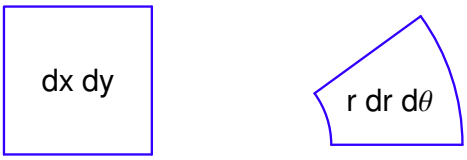
$$du \wedge dv = -dv \wedge du. \tag{12.23}$$

In fact, (12.23) is an example of a ***differential form***, an object which is subject to transformation rules to reflect (12.25) below and the associated change of variable formula in double integrals. To keep the presentation simple, we shall avoid this notation, though the theory can be developed using the approach of Section 24.6.

Let us compute the element of area for the usual metric on the plane, but in polar coordinates $u = r > 0$ and $v = \theta$. Lemma 12.3 tells us that in this case, $E = 1$, $F = 0$ and $G = r^2$. Thus,

Lemma 12.15. *The metric ds^2 on $\mathbb{R}^2$ given by* (12.2) *has as its element of area*

$$dA = dx\,dy = r\,dr\,d\theta.$$



Equation (12.21) motivates the following definition of the area of a closed subset of the trace of a patch. Recall that a subset $\mathcal{S}$ of $\mathbb{R}^n$ is ***bounded***, provided there exists a number M such that $\|\mathbf{p}\| \leqslant M$ for all $\mathbf{p} \in \mathcal{S}$. A ***compact*** subset of $\mathbb{R}^n$ is a subset which is closed and bounded.

Definition 12.16. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be an injective regular patch, and let $\mathcal{R}$ be a compact subset of $\mathbf{x}(\mathcal{U})$. Then the* ***area*** *of $\mathcal{R}$ is*

$$\text{area}(\mathcal{R}) = \iint_{\mathbf{x}^{-1}(\mathcal{R})} \sqrt{EG - F^2}\, du dv. \tag{12.24}$$

That this definition is geometric is a consequence of

Lemma 12.17. *The definition of area is independent of the choice of patch.*

Proof. Let $\mathbf{x}\colon \mathcal{V} \to \mathbb{R}^n$ and $\mathbf{y}\colon \mathcal{W} \to \mathbb{R}^n$ be injective regular patches, and assume that $\mathcal{R} \subseteq \mathbf{x}(\mathcal{V}) \cap \mathbf{y}(\mathcal{W})$. A long but straightforward computation using Lemma 12.4 (and its notation) shows that

$$\sqrt{E_{\mathbf{y}} G_{\mathbf{y}} - F_{\mathbf{y}}^2} \;=\; \sqrt{E_{\mathbf{x}} G_{\mathbf{x}} - F_{\mathbf{x}}^2} \left| \frac{\partial \bar{u}}{\partial u} \frac{\partial \bar{v}}{\partial v} - \frac{\partial \bar{u}}{\partial v} \frac{\partial \bar{v}}{\partial u} \right|. \tag{12.25}$$

But the last factor is the determinant of the Jacobian matrix of $\mathbf{x}^{-1} \circ \mathbf{y}$. By the change of variables formulas for multiple integrals, we have

$$\begin{aligned} \iint_{\mathbf{y}^{-1}(\mathcal{R})} \sqrt{E_{\mathbf{y}} G_{\mathbf{y}} - F_{\mathbf{y}}^2}\, du dv \;&=\; \iint_{\mathbf{y}^{-1}(\mathcal{R})} \sqrt{E_{\mathbf{x}} G_{\mathbf{x}} - F_{\mathbf{x}}^2} \left| \frac{\partial \bar{u}}{\partial u} \frac{\partial \bar{v}}{\partial v} - \frac{\partial \bar{u}}{\partial v} \frac{\partial \bar{v}}{\partial u} \right| du dv \\ &=\; \iint_{\mathbf{x}^{-1}(\mathcal{R})} \sqrt{E_{\mathbf{x}} G_{\mathbf{x}} - F_{\mathbf{x}}^2}\, d\bar{u}\, d\bar{v}. \;\blacksquare \end{aligned}$$

Surface area cannot, in general, be computed by taking the limit of the area of approximating polyhedra. For example, see [Krey1, pages 115–117]. Thus the analog for surfaces of Theorem 1.14, page 10, is false.

12.5 Examples of Metrics

The Sphere

The components of the metric and the infinitesimal area of a sphere $S^2(a)$ of radius a are easily computed. For the standard parametrization on page 288, we obtain from Notebook 12 that

$$E = a^2 \cos^2 v, \qquad F = 0, \qquad G = a^2.$$

Hence the Riemannian metric of $S^2(a)$ can be written as

$$ds^2 = a^2\big(\cos^2 v\, du^2 + dv^2\big). \tag{12.26}$$

The element of area is

$$dA = \sqrt{EG - F^2}\, du dv = a^2 \cos v\, du dv.$$

If we note that, except for one missing meridian, $\mathsf{sphere}[a]$ covers the sphere exactly once when $0 < u < 2\pi$ and $-\pi/2 < v < \pi/2$, we can compute the total area of a sphere by computer. As expected, the result is $4\pi a^2$.

Paraboloids

The two types of paraboloid are captured by the single definition

$$\mathsf{paraboloid}[a,b](u,v) = (u,\ v,\ au^2 + bv^2),$$

copied from page 303, where both a and b are nonzero. If these two parameters have the same sign, we obtain the ***elliptical paraboloid***, and otherwise the ***hyperbolic paraboloid***; both types are visible in Figure 12.4.

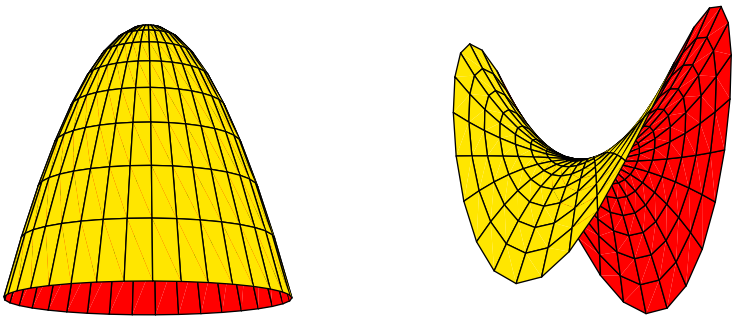


Figure 12.4: Elliptical and hyperbolic paraboloids

The metric can be computed as a function of a, b, and the result is

$$ds^2 = (1 + 4a^2u^2)du^2 + 8abuv\,du\,dv + (1 + 4b^2v^2)dv^2.$$

Only the sign of $F = 4abuv$ changes in passing from "elliptical" to "hyperbolic," as E and G are invariant. The element of area

$$dA = \sqrt{1 + 4a^2u^2 + 4b^2v^2}\,du\,dv$$

is identical in the two cases.

Cylinders

Perhaps an even simpler example is provided by the two cylinders illustrated in Figure 12.5. Whilst the first is a conventional circular cylinder, the leaning one on the right is formed by continuously translating an ellipse in a constant direction. Indeed, by a "cylinder," we understand the surface parametrized by

$$\mathsf{cylinder}[\mathbf{d},\boldsymbol{\gamma}](u,v) = \boldsymbol{\gamma}(u) + v\,\mathbf{d},$$

where $\mathbf{d}$ is some fixed nonzero vector (representing the direction of the axis), and $\boldsymbol{\gamma}$ is a given curve. Our two cylinders are determined by the choices

$$\begin{aligned} \mathbf{d} &= (0,0,1), & \boldsymbol{\gamma}(u) &= (\cos u,\ \sin u,\ 0),\\ \mathbf{d} &= (0,1,1), & \boldsymbol{\gamma}(u) &= (\sin u,\ \tfrac{1}{4}\cos u,\ 0). \end{aligned}$$

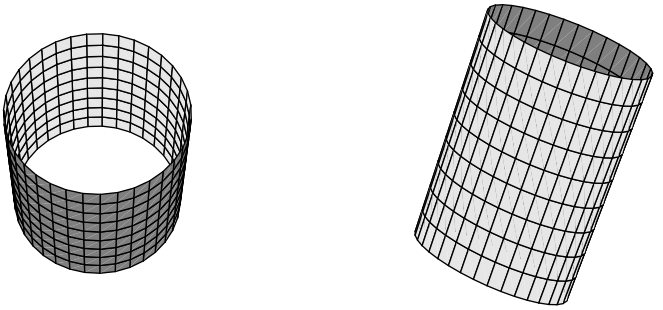


Figure 12.5: Cylinders

While the circular cylinder has the Pythagorean metric $ds^2 = du^2 + dv^2$ (as in (12.2) apart from renaming the variables), the elliptical cylinder has

$$ds^2 = (1 - \frac{15}{16}\sin^2 u)du^2 - \frac{1}{2}\sin u\, du dv + 2\, dv^2.$$

Nonetheless, there exist local isometries from the plane onto both cylinders.

The Helicoid

The geometric definition of the *helicoid* is the surface generated by a line ℓ attached orthogonally to an axis m such that ℓ moves along m and also rotates, both at constant speed. The effect depends whether or not ℓ extends to both sides of m (see Figure 12.6); any point of ℓ not on m describes a circular helix.

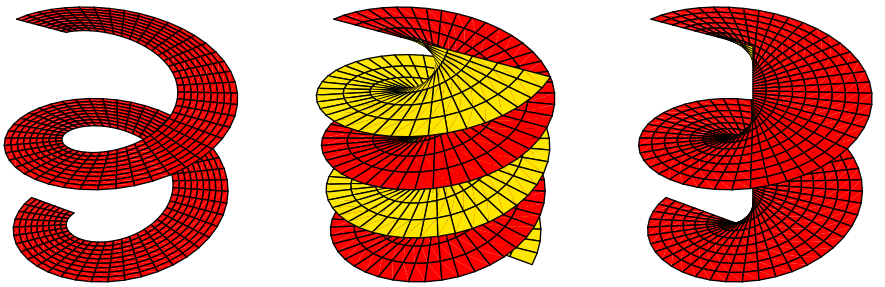


Figure 12.6: Helicoids

The helicoid surface is explicitly parametrized by

$$\mathsf{helicoid}[a,b](u,v) = \big(a\,v\cos u,\ a\,v\sin u,\ b\,u\big). \tag{12.27}$$

See Exercise 3, and the description on page 432. Both cylinders and helicoids are special cases of *ruled surfaces*, the subject of Chapter 14.

The Monkey Saddle

We can use Notebook 12 to find the metric and infinitesimal area of the monkey saddle defined on page 304. The Riemannian metric of the monkey saddle is given by

$$ds^2 = \big(1 + (3u^2 - 3v^2)^2\big)du^2 - 36uv(u^2 - v^2)dudv + \big(1 + 36u^2v^2\big)dv^2,$$

and a computation gives

$$dA = \sqrt{1 + 9u^4 + 18u^2v^2 + 9v^4}\,dudv.$$

Numerical integration can be used to find the area of (for instance) the square-like portion

$$\mathcal{S} = \{\mathsf{monkeysaddle}(p, q) \mid -1 < p, q < 1\},$$

which is approximately 2.33.

Twisted Surfaces

Recall Definition 11.15 from the previous chapter.

Lemma 12.18. *Suppose* $\mathbf{x}$ *is a twisted surface in* $\mathbb{R}^3$ *whose profile curve is* $\boldsymbol{\alpha} = (\varphi, \psi)$. *Then*

$$\begin{aligned} E &= a^2 + 2a\varphi(v)\cos bu + \varphi(v)^2\cos^2 bu + \psi(v)^2\sin^2 bu \\ &\qquad + b^2(\varphi(v)^2 + \psi(v)^2) - 2a\psi(v)\sin bu - \varphi(v)\psi(v)\sin(2bu), \\ F &= b(-\psi\varphi' + \varphi\psi'), \\ G &= \varphi'^2 + \psi'^2. \end{aligned}$$

This can be proved computationally (see Exercise 4).

12.6 Exercises

1. Find the metric, infinitesimal area, and total area of a circular torus defined by

$$\mathsf{torus}[a, b](u, v) = \big((a + b\cos v)\cos u,\ (a + b\cos v)\sin u,\ b\sin v\big).$$

2. Fill in the details of the proof of Lemma 12.12.

M **3.** Find the metric and infinitesimal area of the helicoid parametrized by (12.27). Compute the area of the region

$$\big\{\,\mathsf{helicoid}[a, b](p, q) \mid 0 < p < 2\pi,\ c < q < d\,\big\},$$

and plot the helicoid with $a = b = 1$.

M **4.** Prove Lemma 12.18 using the program metric from Notebook 12.

5. Describe an explicit isometry between an open rectangle in the plane and a circular cylinder, in analogy to (12.13) and the analysis carried out in Section 12.2 for the circular cone.

M **6.** Determine the metric and infinitesimal area of ***Enneper's minimal surface***,[2] defined by

$$\mathsf{enneper}(u,v) = \left(u - \frac{u^3}{3} + u\,v^2,\ -v + \frac{v^3}{3} - v\,u^2,\ u^2 - v^2\right).$$

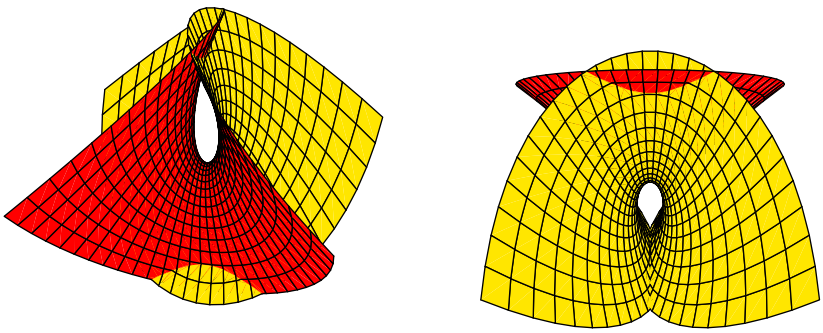


Figure 12.7: Two views of Enneper's surface

Plot Enneper's surface with viewpoints different from the ones shown in Figure 12.7. Finally, compute the area of the image under enneper of the set $[-1,1]\times[-1,1]$. (See also page 509 and [Enn1].)

M **7.** Find formulas for the components E, F, G of the metrics of the following surfaces parametrized in Chapter 10: the circular paraboloid, the eight surface, and the Whitney umbrella.

M **8.** Find E, F, G for the following surfaces parametrized in Chapter 11: the Möbius strip, the Klein bottle, Steiner's Roman surface, the cross cap, and the pseudo cross cap.

[2] Alfred Enneper (1830–1885). Professor at the University of Göttingen. Enneper also studied minimal surfaces and surfaces of constant negative curvature.

Notebook 12

■ N12.0

The main task of this notebook is the computation of the metric or first fundamental form of various surfaces. The end of N12.1 includes a program that accomplishes this. It is used in N12.2 in the construction of a mapping between a planar region and a circular cone that is designed to illustrate the notion of isometry. An animation of this process is included in N12.2.1. This notebook also incorporates *Mathematica* code used to draw a number of simple diagrams used in the text to represent the concept of metric and area. A command for plotting surfaces in polar coordinates can be found in N12.3.

The next instruction for the two-sided coloring of surfaces will be used repeatedly inside a ParametricPlot3D command in this notebook.

```
col[x_][u_,v_]:= Append[x[u,v], FaceForm[Hue[.15], Hue[0]]]
```

It is analogous to the command highlight in N7.3. To make the two colors visible, the Lighting option must be set to False, and we prefer to make this the default in order to compile the figures in Chapter 12. (Splitting the book's programs into notebooks enables us to choose the most appropriate defaults for a particular notebook to save time in carrying out the plotting.) For Notebook 12, we accordingly set

```
SetOptions[ParametricPlot3D,
  Compiled->False, Boxed->False, Axes->None,
  BoxRatios->{1,1,1}, Lighting->False
]
```

In N12.0.1, we reload the arrows command ar, as defined in N11.2, based on

```
<<Graphics`Arrow`
```

Options for ParametricPlot, the definitions of J, L, unitnormal, polarize, the display commands tx, highlight, S, are given in N12.0.1, as well as standard parametrizations of the sphere and paraboloids.

■ N12.0.1

■ N12.1

To help plot the uncaptioned figures of Chapter 12, we shall use abbreviations for color and thickness, which we combine inside an initialization cell:

```
hu:= Hue[.7]
th:= Thickness[.015]
```

```
Module[{o,p,q,z,ga},
  o= {0,0}; p= {1,0}; q= {1,1};
  z[1]= tx["dx", {.5,-.2}];
  z[2]= tx["dy", {1.2,.5}];
  z[3]= tx["ds", {.3,.55}];
  z[4]= {hu, th, Line[{o,p}], Line[{o,q}], Line[{p,q}]};
  ga= Graphics/@Array[z,4];
  S[ga, AspectRatio->1, PlotRange->All, ImageSize->130]
]
```

The following miniprogram works in the same way as the jacobian command introduced in N10.1. Its input consists of a patch x and two coordinates. The output displays the metric or first fundamental form as a symmetric matrix with entries ee,ff,gg (our notation for E,F,G, partly to avoid a clash with E=exp(1)).

```
metric[x_][u_,v_]:=
Module[{xu,xv,U,V,si},
  si= Simplify[#]/.{U->u, V->v} &;
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  ee= xu.xu//si;
  ff=  xu.xv//si;
  gg= xv.xv//si;
  {{ee,ff},{ff,gg}}
]
```

The coefficients ee,ff,gg can also be listed upon request, though these quantities depend upon the chosen coordinates, and are not formally functions of either the coordinates or the patch x. This saves excessive typing, but one needs to be careful to

ensure that the current values of ee,ff,gg refer to the required surface. We shall use the program metric to compute surface area in N12.4.

N12.1.1

N12.2

We set a default value for the angle α, and consider a circular region surrounded by a circular arc AB of radius 1 subtending an angle AOB of α radians:

```
α:= 5π/4
```

```
Module[{o,f,z},
  o= {0,0};
  f[t_]= {Cos[t], Sin[t]};
  jt= Join[Table[.99f[k α/64], {k,0,64}], {o}];
  z[1]= ParametricPlot[f[t], {t,0,α},
    PlotStyle->{Hue[.7],Thickness[.02]}];
  z[2]= {Hue[.15], Polygon[jt]}//Graphics;
  z[3]= {ar[{o,f[0]}], ar[{o,f[α]}]}//Graphics;
  z[4]= {tx["O", {0,.15}], tx["A", 1.15f[0]],
    tx["B", 1.15f[α]]}//Graphics;
  S @ Array[z,4]
]
```

Figure 12.1

This region defines a flat surface in 3-dimensional space by regarding it as part of the xy-plane. As such, it is given by the patch

```
x[r_,θ_]:= {r Cos[θ], r Sin[θ], 0}
ParametricPlot3D[col[x][r,θ], {r,0,1}, {θ,0,α}]
```

The metric of this planar surface is the Euclidean one expressed in polar coordinates:

```
Clear[r]
metric[x][r,θ]//MatrixForm
```

The planar surface is to be mapped into a circular cone in 3-dimensional space by attaching the edge OA to OB. The resulting angle of the cone is

```
β:= ArcSin[α/(2π)]
```

We can parametrize the conical surface by the patch

```
reg:= {{-1,1},{-1,1},{0,1}}
y[r_,θ_]:= {r Cos[2π θ/α]Sin[β], r Sin[2π θ/α]Sin[β], r Cos[β]}
ParametricPlot3D[col[y][r,θ], {r,0,1}, {θ,0,α}, PlotRange->reg];
```

```
metric[y][r,θ]
```

The hidden subsection below contains both Figure 12.2 and an animation of the above construction.

N12.2.1

N12.3

The surface in Figure 12.3 approximates a hyperbolic paraboloid near the origin. The command polarize from N10.4 and N12.0.1 is used to provide a "symmetrical" effect.

```
Module[{f,g,a,b,α,k,v,w,z},
  f[u_,v_]:= {u, v, u Sin[2v]};
  g:= polarize[f];
  a= 1.5;
  b= π/7;
  α[r_]:= g[r,b] + k;
  k= {0,0,.01};
  v= g[a,b] + k;
  w= g[-a,b] +k;
  z[1]= ParametricPlot3D[col[g][r,θ], {r,0,2}, {θ,0,2π},
    PlotPoints->{10,30}];
  z[2]= ParametricPlot3D[highlight[α][r], {r,-a,a}];
  z[3]= Graphics3D[{hu, Thickness[.01], Line[{v,w}]}];
  z[4]= Graphics3D[{hu, PointSize[.02], Point[v], Point[w]}];
  S @ Array[z,4]
]
```

Figure 12.3

■ N12.4

The rotation operator J is used to produce two of the sides of the curvilinear rectangle in the next diagram. Instructions for the remaining two uncaptioned diagrams of Section 12.4 are contained in subsection N12.4.1 of the electronic version of this notebook.

```
Module[{f,ff,ep},
  f[t_]= {t,-t}^3{t,t-1};
  ff[t_]= {f[t], J[f[t]], J[f[t]]+{1,0}, f[t]+{0,1}};
  ep= tx["du dv", {.4,.6}];
  S @ ParametricPlot[ff[t]//Evaluate, {t,0,1},
    PlotStyle->{{hu, th}}, Axes->None, AspectRatio->1,
    PlotRange->All, ImageSize->120, Epilog->ep]
]
```

The following program computes the integrand for surface area:

```
area[x_][u_,v_]:=
Module[{M},
  M= metric[x][u,v];
  Sqrt[Simplify[Det[M]]]
]
```

We test it on the paraboloid, and then show how it is possible to integrate the result to obtain the area of a region on the surface.

```
Clear[a,b]
x:= paraboloid[a,b]
metric[x][u,v]//MatrixForm
area[x][u,v]
InputForm[%]
```

```
a:= 1; b:= 2;
NIntegrate[area[x][u,v], {u,-1,1}, {v,-1,1}]
```

■ N12.4.1

N12.5

The default option explained in N12.0 automatically provides two-sided coloring of the following quadrics.

```
z[x_]:= ParametricPlot3D[col[x][r,θ], {r,0,1}, {θ,0,2π},
  PlotPoints->{10,36}, ViewPoint->{2,-4,-1}];
pair:= z/@polarize/@{paraboloid[-1,-1], paraboloid[1,-1]}
S @ GraphicsArray[pair]
```

Figure 12.4

The next definitions are used in N12.5.1 to plot Figures 12.5 and 12.6.

```
cylinder[d_, γ_][u_,v_]:= γ[u] + v d
helicoid[a_,b_,c_][u_,v_]:= {a v Cos[u], b v Sin[u], c u}
```

N12.5.1

N12.6

Here is a famous minimal surface that will be studied again in Chapter 16:

```
enneper[u_, v_]:=
  {u - u^3/3 + u v^2, -v + v^3/3 - v u^2, u^2 - v^2}
```

```
Module[{xx,z,p,zz},
  r= 2;
  xx= col[enneper];
  z[p_]= ParametricPlot3D[xx[u,v], {u,-r,r}, {v,-r,r}, ViewPoint->p];
  zz:= {z[{1,-4,2}], z[{2,0,2}]};
  S @ GraphicsArray[zz]
]
```

Figure 12.7

N12.6.1

Chapter 13

Shape and Curvature

In this chapter, we study the relationship between the geometry of a regular surface $\mathcal{M}$ in 3-dimensional space $\mathbb{R}^3$ and the geometry of $\mathbb{R}^3$ itself. The basic tool is the *shape operator* defined in Section 13.1. The shape operator at a point $\mathbf{p}$ of $\mathcal{M}$ is a self-adjoint linear transformation S of the tangent space $\mathcal{M}_\mathbf{p}$ that measures how $\mathcal{M}$ bends in different directions. The shape operator can also be considered to be (minus) the differential of the Gauss map of $\mathcal{M}$ (Proposition 13.5), and its effect is illustrated using tangent vectors to coordinate curves on the surface and the image of these vectors on the sphere (Figure 13.1).

In Section 13.2, we define a variant of the shape operator called *normal curvature*. Given a tangent vector $\mathbf{v_p}$ to a surface $\mathcal{M}$, the normal curvature $k(\mathbf{v_p})$ is a real number that measures how $\mathcal{M}$ bends in the direction $\mathbf{v_p}$. This number is easy to understand geometrically because it is the curvature of the *plane curve* formed by the intersection of $\mathcal{M}$ with the plane passing through $\mathbf{v_p}$ meeting $\mathcal{M}$ perpendicularly.

Techniques for computing the shape operator and normal curvature are given in Section 13.3. The eigenvalues of the shape operator at $\mathbf{p} \in \mathcal{M}$ are studied at the end of that section. They turn out to be the maximum and minimum of the normal curvature at $\mathbf{p}$, the so-called *principal curvatures*, and the corresponding eigenvectors are orthogonal. The principal curvatures are graphed for the monkey saddle in Figure 13.6.

The most important curvature functions of a surface in $\mathbb{R}^3$ are the *Gaussian curvature* and the *mean curvature*, both defined in Section 13.4. This leads to a unified discussion of the *first*, *second* and *third fundamental forms*. Two separate techniques for computing curvatures from the parametric representation of a surface are described in Section 13.5, which includes calculations from Notebook 13 with reference to Monge patches and other examples.

Section 13.6 is devoted to a global curvature theorem, while we discuss how to compute the curvatures of nonparametric surfaces in Section 13.7.

13.1 The Shape Operator

We want to measure how a regular surface $\mathcal{M}$ bends in $\mathbb{R}^3$. A good way to do this is to estimate how the surface normal $\mathbf{U}$ changes from point to point. We use a linear operator called the *shape operator* to calculate the bending of $\mathcal{M}$. The shape operator came into wide use after its introduction in O'Neill's book [ON1]; however, it occurs much earlier, for example, implicitly in Blaschke's classical book [Blas2],[1] and explicitly in [BuBu].[2]

The shape operator applied to a tangent vector $\mathbf{v_p}$ is the negative of the derivative of $\mathbf{U}$ in the direction $\mathbf{v_p}$:

Definition 13.1. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface, and let $\mathbf{U}$ be a surface normal to $\mathcal{M}$ defined in a neighborhood of a point $\mathbf{p} \in \mathcal{M}$. For a tangent vector $\mathbf{v_p}$ to $\mathcal{M}$ at $\mathbf{p}$ we put*

$$S(\mathbf{v_p}) = -\boldsymbol{D}_{\mathbf{v}}\mathbf{U}.$$

Then S is called the ***shape operator****.*

Derivatives of vector fields were discussed in Section 9.5, and regular surfaces in Section 10.3. The precise definition of $\boldsymbol{D}_{\mathbf{v}}\mathbf{U}$ relies on Lemma 9.40, page 281, since $\mathbf{U}$ is not defined away from the surface $\mathcal{M}$.

It is easy to see that the shape operator of a plane is identically zero at all points of the plane. For a nonplanar surface the surface normal $\mathbf{U}$ will twist and turn from point to point, and S will be nonzero. At any point of an orientable regular surface there are two choices for the unit normal: $\mathbf{U}$ and $-\mathbf{U}$. The shape operator corresponding to $-\mathbf{U}$ is the negative of the shape operator corresponding to $\mathbf{U}$. If $\mathcal{M}$ is nonorientable, we have seen that a surface normal $\mathbf{U}$ cannot be defined continuously on all of $\mathcal{M}$. This does not matter in the present chapter, because all calculations involving $\mathbf{U}$ are local, so it suffices to perform them on an open subset of $\mathcal{M}$, such as the image of a regular patch, where $\mathbf{U}$ *is* defined continuously.

---

[1] 

Wilhelm Johann Eugen Blaschke (1885–1962). Austrian-German mathematician. In 1919 he was appointed to a chair in Hamburg, where he built an important school of differential geometry.

[2] Around 1900 the Gibbs–Heaviside vector analysis notation (which one can read about in the interesting book [Crowe]) spread to differential geometry, although its use was controversial for another 30 years. Blaschke's [Blas2] was one of the first differential geometry books to use vector analysis. The book coauthored by Burali-Forti and Burgatti [BuBu] contains an amusing attack on those resisting the new vector notation. In the multivolume works of Darboux [Darb1] and Bianchi [Bian] most formulas are written component by component. Compact vector notation, of course, is indispensable nowadays both for humans and for computers.

Recall (Lemma 10.26, page 299) that any point $\mathbf{q}$ in the domain of definition of a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ has a neighborhood $\mathcal{U}_\mathbf{q}$ of $\mathbf{q}$ such that $\mathbf{x}(\mathcal{U}_\mathbf{q})$ is a regular surface. Therefore, the shape operator of a regular patch is also defined. Conversely, we can use patches on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ to calculate the shape operator of $\mathcal{M}$.

The next lemmas establish some elementary properties.

Lemma 13.2. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then*

$$S(\mathbf{x}_u) = -\mathbf{U}_u \qquad \text{and} \qquad S(\mathbf{x}_v) = -\mathbf{U}_v.$$

Proof. Fix v and define a curve $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}(u) = \mathbf{x}(u, v)$. Then by Lemma 9.40, page 281, we have

$$S\big(\mathbf{x}_u(u, v)\big) = S\big(\boldsymbol{\alpha}'(u)\big) = -\boldsymbol{D}_{\alpha'(u)}\mathbf{U} = -(\mathbf{U} \circ \boldsymbol{\alpha})'(u).$$

But $(\mathbf{U} \circ \boldsymbol{\alpha})'$ is just $\mathbf{U}_u$. Similarly, $S(\mathbf{x}_v) = -\mathbf{U}_v$. ∎

Lemma 13.3. *At each point $\mathbf{p}$ of a regular surface $\mathcal{M} \subset \mathbb{R}^3$, the shape operator is a linear map*

$$S\colon \mathcal{M}_\mathbf{p} \longrightarrow \mathcal{M}_\mathbf{p}.$$

Proof. That S is linear follows from the fact that $\boldsymbol{D}_{a\mathbf{v}+b\mathbf{w}} = a\boldsymbol{D}_\mathbf{v} + b\boldsymbol{D}_\mathbf{w}$. To prove that S maps $\mathcal{M}_\mathbf{p}$ into $\mathcal{M}_\mathbf{p}$ (instead of merely into $\mathbb{R}^3_\mathbf{p}$), we differentiate the equation $\mathbf{U} \cdot \mathbf{U} = 1$ and use (9.12) on page 276:

$$0 = \mathbf{v}_\mathbf{p}[\mathbf{U} \cdot \mathbf{U}] = 2(\boldsymbol{D}_\mathbf{v}\mathbf{U}) \cdot \mathbf{U}(\mathbf{p}) = -2S(\mathbf{v}_\mathbf{p}) \cdot \mathbf{U}(\mathbf{p}),$$

for any tangent vector $\mathbf{v}_\mathbf{p}$. Since $S(\mathbf{v}_\mathbf{p})$ is perpendicular to $\mathbf{U}(\mathbf{p})$, it must be tangent to $\mathcal{M}$; that is, $S(\mathbf{v}_\mathbf{p}) \in \mathcal{M}_\mathbf{p}$. ∎

Next, we find an important relation between the shape operator of a surface and the acceleration of a curve on the surface.

Lemma 13.4. *If $\boldsymbol{\alpha}$ is a curve on a regular surface $\mathcal{M} \subset \mathbb{R}^3$, then*

$$\boldsymbol{\alpha}'' \cdot \mathbf{U} = S(\boldsymbol{\alpha}') \cdot \boldsymbol{\alpha}'.$$

Proof. We restrict the vector field $\mathbf{U}$ to the curve $\boldsymbol{\alpha}$ and use Lemma 9.40. Since $\boldsymbol{\alpha}(t) \in \mathcal{M}$ for all t, the velocity $\boldsymbol{\alpha}'$ is always tangent to $\mathcal{M}$, so

$$\boldsymbol{\alpha}' \cdot \mathbf{U} = 0.$$

When we differentiate this equation and use Lemmas 13.2 and 13.3, we obtain

$$\boldsymbol{\alpha}'' \cdot \mathbf{U} = -\mathbf{U}' \cdot \boldsymbol{\alpha}' = S(\boldsymbol{\alpha}') \cdot \boldsymbol{\alpha}'. \quad ∎$$

Observe that $\boldsymbol{\alpha}'' \cdot \mathbf{U}$ can be interpreted geometrically as the *component of the acceleration of $\boldsymbol{\alpha}$ that is perpendicular to $\mathcal{M}$.*

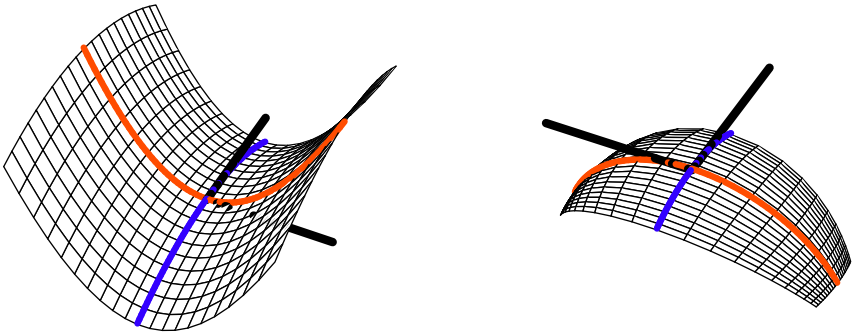


Figure 13.1: Coordinate curves and tangent vectors, together with (on the right) their Gauss images

Lemma 13.2 allows us to illustrate $-S$ by its effect on tangent vectors to coordinate curves to $\mathcal{M}$. These are mapped to the tangent vectors to the corresponding curves on the unit sphere defined by $\mathbf{U}$. This is expressed more invariantly by the next result that asserts that $-S$ is none other than the tangent map of the Gauss map.

Proposition 13.5. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$ oriented by a unit normal vector field $\mathbf{U}$. View $\mathbf{U}$ as the Gauss map $\mathcal{M} \to S^2(1)$, where $S^2(1)$ denotes the unit sphere in $\mathbb{R}^3$. If $\mathbf{v_p}$ is a tangent vector to $\mathcal{M}$ at $\mathbf{p} \in \mathcal{M}$, then $\mathbf{U}_*(\mathbf{v_p})$ is parallel to $-S(\mathbf{v_p}) \in \mathcal{M}_\mathbf{p}$.*

Proof. By Lemma 9.10, page 269, we have

$$\mathbf{U}_*(\mathbf{v_p}) = \big(\mathbf{v_p}[u_1],\ \mathbf{v_p}[u_2],\ \mathbf{v_p}[u_3]\big)_{\mathbf{U(p)}}.$$

On the other hand, Lemma 9.28, page 275, implies that

$$-S(\mathbf{v_p}) = \boldsymbol{D}_\mathbf{v}\mathbf{U} = \big(\mathbf{v_p}[u_1],\ \mathbf{v_p}[u_2],\ \mathbf{v_p}[u_3]\big)_\mathbf{p}.$$

Since the vectors $(\mathbf{v_p}[u_1], \mathbf{v_p}[u_2], \mathbf{v_p}[u_3])_{\mathbf{U(p)}}$ and $(\mathbf{v_p}[u_1], \mathbf{v_p}[u_2], \mathbf{v_p}[u_3])_\mathbf{p}$ are parallel in $\mathbb{R}^3$, the lemma follows. ∎

We conclude this introductory section by noting a fundamental relationship between shape operators of surfaces and Euclidean motions of $\mathbb{R}^3$.

Theorem 13.6. *Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be an orientation-preserving Euclidean motion, and let $\mathcal{M}_1$ and $\mathcal{M}_2$ be oriented regular surfaces such that $F(\mathcal{M}_1) = \mathcal{M}_2$. Then*

(i) *the unit normals* $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *of* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *can be chosen so that* $F_*(\mathbf{U}_1) = \mathbf{U}_2$;

(ii) *the shape operators* S_1 *and* S_2 *of the two surfaces* (*with the choice of unit normals given by* (i)) *are related by* $S_2 \circ F_* = F_* \circ S_1$.

Proof. Since F_* preserves lengths and inner products, it follows that $F_*(\mathbf{U}_1)$ is perpendicular to $\mathcal{M}_2$ and has unit length. Hence $F_*(\mathbf{U}_1) = \pm\mathbf{U}_2$; we choose the plus sign at all points of $\mathcal{M}_2$. Let $\mathbf{p} \in \mathcal{M}_1$ and $\mathbf{v_p} \in \mathcal{M}_{1\mathbf{p}}$, and let $\mathbf{w}_{F(\mathbf{p})} = F_*(\mathbf{v_p})$. A Euclidean motion is an affine transformation, so by Lemma 9.35, page 278, we have

$$\begin{aligned}(S_2 \circ F_*)(\mathbf{v_p}) = S_2(\mathbf{w}_{F(\mathbf{p})}) = -\boldsymbol{D}_{\mathbf{w}}\mathbf{U}_2 &= -\boldsymbol{D}_{\mathbf{w}}F_*(\mathbf{U}_1) \\ &= -F_*(\boldsymbol{D}_{\mathbf{v}}\mathbf{U}_1) = (F_* \circ S_1)(\mathbf{v_p}).\end{aligned}$$

Because $\mathbf{v_p}$ is arbitrary, we have $S_2 \circ F_* = F_* \circ S_1$. ∎

13.2 Normal Curvature

Although the shape operator does the job of measuring the bending of a surface in different directions, frequently it is useful to have a real-valued function, called the ***normal curvature***, which does the same thing. We shall define it in terms of the shape operator, though it is worth bearing in mind that normal curvature is explicitly a much older concept (see [Euler3], [Meu] and Corollary 13.20).

First, we need to make precise the notion of direction on a surface.

Definition 13.7. *A* ***direction*** ℓ *on a regular surface* $\mathcal{M}$ *is a* 1*-dimensional subspace of* (*that is, a line through the origin in*) *a tangent space to* $\mathcal{M}$.

A nonzero vector $\mathbf{v_p}$ in a tangent space $\mathcal{M}_\mathbf{p}$ determines a unique 1-dimensional subspace ℓ, so we can use the terminology "the direction $\mathbf{v_p}$" to mean ℓ, provided the sign of $\mathbf{v_p}$ is irrelevant.

Definition 13.8. *Let* $\mathbf{u_p}$ *be a tangent vector to a regular surface* $\mathcal{M} \subset \mathbb{R}^3$ *with* $\|\mathbf{u_p}\| = 1$. *Then the* ***normal curvature of*** $\mathcal{M}$ ***in the direction*** $\mathbf{u_p}$ *is*

$$k(\mathbf{u_p}) = S(\mathbf{u_p}) \cdot \mathbf{u_p}.$$

More generally, if $\mathbf{v_p}$ *is any nonzero tangent vector to* $\mathcal{M}$ *at* $\mathbf{p}$, *we put*

$$k(\mathbf{v_p}) = \frac{S(\mathbf{v_p}) \cdot \mathbf{v_p}}{\|\mathbf{v_p}\|^2}. \tag{13.1}$$

If ℓ is a direction in a tangent space $\mathcal{M}_\mathbf{p}$ to a regular surface $\mathcal{M} \subset \mathbb{R}^3$, then $k(\mathbf{v_p})$ is easily seen to be the same for all nonzero tangent vectors $\mathbf{v_p}$ in ℓ (this is Exercise 9). Therefore, we call the common value of the normal curvature the ***normal curvature of the direction*** ℓ.

Let us single out two kinds of directions.

Definition 13.9. *Let ℓ be a direction in a tangent space $\mathcal{M}_\mathbf{p}$, where $\mathcal{M} \subset \mathbb{R}^3$ is a regular surface. If the normal curvature of ℓ is zero, we say that ℓ is an* ***asymptotic direction****. Similarly, if the normal curvature vanishes on a tangent vector $\mathbf{v_p}$ to $\mathcal{M}$, we say that $\mathbf{v_p}$ is an* ***asymptotic vector****. An* ***asymptotic curve*** *on $\mathcal{M}$ is a curve whose trace lies on $\mathcal{M}$ and whose tangent vector is everywhere asymptotic.*

Asymptotic curves will be studied in detail in Chapter 18.

Definition 13.10. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$ and let $\mathbf{p} \in \mathcal{M}$. The maximum and minimum values of the normal curvature k of $\mathcal{M}$ at $\mathbf{p}$ are called the* ***principal curvatures*** *of $\mathcal{M}$ at $\mathbf{p}$ and are denoted (in either order) by k_1, k_2. Unit vectors $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{M}_\mathbf{p}$ at which these extreme values occur are called* ***principal vectors****. The corresponding directions are called* ***principal directions****. A* ***principal curve*** *on $\mathcal{M}$ is a curve whose trace lies on $\mathcal{M}$ and whose tangent vector is everywhere principal.*

The principal curvatures measure the maximum and minimum bending of a regular surface $\mathcal{M}$ at each point $\mathbf{p} \in \mathcal{M}$. Principal curves will be studied again in Section 15.2, and in more detail in Chapter 19.

There is an important interpretation of normal curvature of a regular surface as the curvature of a space curve.

Lemma 13.11. (Meusnier) *Let $\mathbf{u_p}$ be a unit tangent vector to $\mathcal{M}$ at $\mathbf{p}$, and let $\boldsymbol{\beta}$ be a unit-speed curve in $\mathcal{M}$ with $\boldsymbol{\beta}(0) = \mathbf{p}$ and $\boldsymbol{\beta}'(0) = \mathbf{u_p}$. Then*

$$k(\mathbf{u_p}) = \boldsymbol{\kappa}[\boldsymbol{\beta}](0)\cos\theta, \tag{13.2}$$

where $\boldsymbol{\kappa}[\boldsymbol{\beta}](0)$ is the curvature of $\boldsymbol{\beta}$ at 0, and θ is the angle between the normal $\mathbf{N}(0)$ of $\boldsymbol{\beta}$ and the surface normal $\mathbf{U}(\mathbf{p})$. Thus all curves lying on a surface $\mathcal{M}$ and having the same tangent line at a given point $\mathbf{p} \in \mathcal{M}$ have the same normal curvature at $\mathbf{p}$.

Proof. Suppose that $\boldsymbol{\kappa}[\boldsymbol{\beta}](0) \neq 0$. By Lemma 13.4 and Theorem 7.10, page 197, we have

$$\begin{aligned} k(\mathbf{u_p}) = S(\mathbf{u_p}) \cdot \mathbf{u_p} &= \boldsymbol{\beta}''(0) \cdot \mathbf{U}(\mathbf{p}) \\ &= \boldsymbol{\kappa}[\boldsymbol{\beta}](0)\mathbf{N}(0) \cdot \mathbf{U}(\mathbf{p}) = \boldsymbol{\kappa}[\boldsymbol{\beta}](0)\cos\theta. \end{aligned} \tag{13.3}$$

In the exceptional case that $\boldsymbol{\kappa}[\boldsymbol{\beta}](0) = 0$, the normal $\mathbf{N}(0)$ is not defined, but we still have $k(\mathbf{u_p}) = 0$. ∎

To understand the meaning of normal curvature geometrically, we need to find curves on a surface to which we can apply Lemma 13.11.

Definition 13.12. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface and $\mathbf{u_p}$ a unit tangent vector to $\mathcal{M}$. Denote by $\Pi\big(\mathbf{u_p}, \mathbf{U}(\mathbf{p})\big)$ the plane determined by $\mathbf{u_p}$ and the surface normal $\mathbf{U}(\mathbf{p})$. The* ***normal section*** *of $\mathcal{M}$ in the $\mathbf{u_p}$ direction is the intersection of $\Pi\big(\mathbf{u_p}, \mathbf{U}(\mathbf{p})\big)$ and $\mathcal{M}$.*

Corollary 13.13. *Let $\boldsymbol{\beta}$ be a unit-speed curve which lies in the intersection of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ and one of its normal sections Π through $\mathbf{p} \in \mathcal{M}$. Assume that $\boldsymbol{\beta}(0) = \mathbf{p}$ and put $\mathbf{u_p} = \boldsymbol{\beta}'(0)$. Then the normal curvature $k(\mathbf{u_p})$ of $\mathcal{M}$ and the curvature of $\boldsymbol{\beta}$ are related by*

$$k(\mathbf{u_p}) = \pm\boldsymbol{\kappa}[\boldsymbol{\beta}](0). \tag{13.4}$$

Proof. We may assume that $\boldsymbol{\kappa}[\boldsymbol{\beta}](0) \neq 0$, for otherwise (13.4) is an obvious consequence of (13.2). Since $\boldsymbol{\beta}$ has unit speed, $\boldsymbol{\kappa}[\boldsymbol{\beta}](0)\mathbf{N}(0) = \boldsymbol{\beta}''(0)$ is perpendicular to $\boldsymbol{\beta}'(0)$. On the other hand, both $\mathbf{U}(\mathbf{p})$ and $\mathbf{N}(0)$ lie in Π, so the only possibility is $\mathbf{N}(0) = \pm\mathbf{U}(\mathbf{p})$. Hence $\cos\theta = \pm 1$ in (13.2), and so we obtain (13.4). ∎

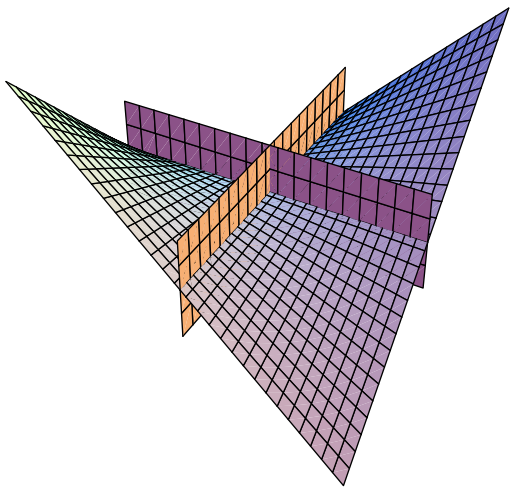


Figure 13.2: Normal sections to a paraboloid through asymptotic curves

Corollary 13.13 gives an excellent method for estimating normal curvature visually. For a regular surface $\mathcal{M} \subset \mathbb{R}^3$, suppose we want to know the normal curvature in various directions at $\mathbf{p} \in \mathcal{M}$. Each unit vector $\mathbf{u_p} \in \mathcal{M}_\mathbf{p}$, together with the surface normal $\mathbf{U}(\mathbf{p})$, determines a plane $\Pi\big(\mathbf{u_p}, \mathbf{U}(\mathbf{p})\big)$. Then the normal section in the direction $\mathbf{u_p}$ is the intersection of $\Pi\big(\mathbf{u_p}, \mathbf{U}(\mathbf{p})\big)$ and $\mathcal{M}$. This is a plane curve in $\Pi\big(\mathbf{u_p}, \mathbf{U}(\mathbf{p})\big)$ whose curvature is given by (13.4). There are three cases:

- If $k(\mathbf{u_p}) > 0$, then the normal section is bending in the same direction as $\mathbf{U}(\mathbf{p})$. Hence in the $\mathbf{u_p}$ direction $\mathcal{M}$ is bending toward $\mathbf{U}(\mathbf{p})$.
- If $k(\mathbf{u_p}) < 0$, then the normal section is bending in the opposite direction from $\mathbf{U}(\mathbf{p})$. Hence in the $\mathbf{u_p}$ direction $\mathcal{M}$ is bending away from $\mathbf{U}(\mathbf{p})$.
- If $k(\mathbf{u_p}) = 0$, then the curvature of the normal section vanishes at $\mathbf{p}$ so the normal to a curve in the normal section is undefined. It is impossible to conclude that there is no bending of $\mathcal{M}$ in the $\mathbf{u_p}$ direction since $\boldsymbol{\kappa}[\boldsymbol{\beta}]$ might vanish only at $\mathbf{p}$. But in some sense the bending is small.

As the unit tangent vector $\mathbf{u_p}$ turns, the surface may bend in different ways. A good example of this occurs at the center point of a hyperbolic paraboloid.

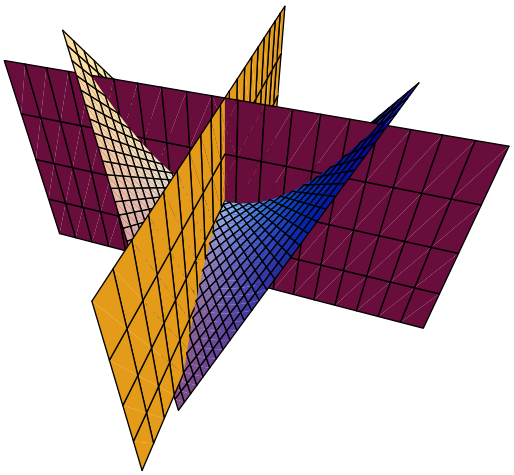


Figure 13.3: Normal sections to a paraboloid through principal curves

In Figure 13.2, both normal sections intersect the hyperbolic paraboloid in straight lines, and the normal curvature determined by each of these sections vanishes. These straight lines are in fact asymptotic curves, whereas the sections shown in Figure 13.3 intersect the surface in curves tangent to principal directions (recall Definitions 13.9 and 13.10). In the second case, the normal curvature determined by one section is positive, and that determined by the other is negative.

The normal sections at the center of a monkey saddle are similar to those of the hyperbolic paraboloid, but more complicated. In this case, there are three asymptotic directions passing through the center point $\mathbf{o}$ of the monkey saddle. It is this fact that forces S to vanish as a linear transformation of the tangent space $T_{\mathbf{o}}\mathcal{M}$ at the point $\mathbf{o}$ itself.

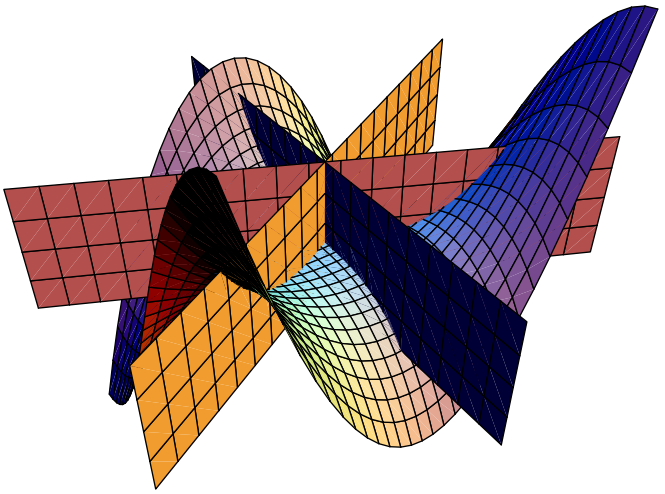


Figure 13.4: Normal sections to a monkey saddle

13.3 Calculation of the Shape Operator

Symmetric linear transformations are much easier to work with than general linear transformations. We shall exploit this in developing the theory of the shape operator, which fortunately falls into this category.

*Lemma 13.14. The shape operator of a regular surface $\mathcal{M}$ is **symmetric** or **self-adjoint**, meaning that*

$$S(\mathbf{v_p}) \cdot \mathbf{w_p} = \mathbf{v_p} \cdot S(\mathbf{w_p})$$

for all tangent vectors $\mathbf{v_p}, \mathbf{w_p}$ to $\mathcal{M}$.

Proof. Let $\mathbf{x}$ be an injective regular patch on $\mathcal{M}$. We differentiate the formula $\mathbf{U} \cdot \mathbf{x}_u = 0$ with respect to v and obtain

$$0 = \frac{\partial}{\partial v}(\mathbf{U} \cdot \mathbf{x}_u) = \mathbf{U}_v \cdot \mathbf{x}_u + \mathbf{U} \cdot \mathbf{x}_{uv}, \tag{13.5}$$

where $\mathbf{U}_v$ is the derivative of the vector field $v \mapsto \mathbf{U}(u, v)$ along any v-parameter curve. Since $\mathbf{U}_v = -S(\mathbf{x}_v)$, equation (13.5) becomes

$$S(\mathbf{x}_v) \cdot \mathbf{x}_u = \mathbf{U} \cdot \mathbf{x}_{uv}. \tag{13.6}$$

Similarly,

$$S(\mathbf{x}_u) \cdot \mathbf{x}_v = \mathbf{U} \cdot \mathbf{x}_{vu}. \tag{13.7}$$

From (13.6), (13.7) and the fact that $\mathbf{x}_{uv} = \mathbf{x}_{vu}$ we get

$$S(\mathbf{x}_u) \cdot \mathbf{x}_v = \mathbf{U} \cdot \mathbf{x}_{vu} = \mathbf{U} \cdot \mathbf{x}_{uv} = S(\mathbf{x}_v) \cdot \mathbf{x}_u. \tag{13.8}$$

The proof is completed by expressing $\mathbf{v_p}, \mathbf{w_p}$ in terms of $\mathbf{x}_u, \mathbf{x}_v$ and using linearity. ▌

Definition 13.15. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then*

$$\begin{cases} e \;=\; -\mathbf{U}_u \cdot \mathbf{x}_u \;=\; \mathbf{U} \cdot \mathbf{x}_{uu}, \\ f \;=\; -\mathbf{U}_v \cdot \mathbf{x}_u \;=\; \mathbf{U} \cdot \mathbf{x}_{uv} \;=\; \mathbf{U} \cdot \mathbf{x}_{vu} \;=\; -\mathbf{U}_u \cdot \mathbf{x}_v, \\ g \;=\; -\mathbf{U}_v \cdot \mathbf{x}_v \;=\; \mathbf{U} \cdot \mathbf{x}_{vv}. \end{cases} \tag{13.9}$$

Classically, e, f, g are called the ***coefficients of the second fundamental form*** of $\mathbf{x}$.

In Section 12.1 we wrote the metric as

$$ds^2 = E\,du^2 + 2F\,du dv + G\,dv^2,$$

and E, F, G are called the ***coefficients of the first fundamental form*** of $\mathbf{x}$. The quantity

$$e\,du^2 + 2f\,du dv + g\,dv^2$$

has a more indirect interpretation, given on page 402. The notation e, f, g is that used in most classical differential geometry books, though many authors use L, M, N in their place. Incidentally, Gauss used the notation D, D', D'' for the respective quantities $e\sqrt{EG - F^2}$, $f\sqrt{EG - F^2}$, $g\sqrt{EG - F^2}$.

Theorem 13.16. (The Weingarten[3] equations) *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then the shape operator S of $\mathbf{x}$ is given in terms of the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$ by*

$$\begin{cases} -S(\mathbf{x}_u) \;=\; \mathbf{U}_u \;=\; \dfrac{fF - eG}{EG - F^2}\mathbf{x}_u + \dfrac{eF - fE}{EG - F^2}\mathbf{x}_v, \\ -S(\mathbf{x}_v) \;=\; \mathbf{U}_v \;=\; \dfrac{gF - fG}{EG - F^2}\mathbf{x}_u + \dfrac{fF - gE}{EG - F^2}\mathbf{x}_v. \end{cases} \tag{13.10}$$

Proof. Since $\mathbf{x}$ is regular, and $\mathbf{x}_u$ and $\mathbf{x}_v$ are linearly independent, we can write

$$\begin{cases} -S(\mathbf{x}_u) = \mathbf{U}_u = a_{11}\mathbf{x}_u + a_{21}\mathbf{x}_v, \\ -S(\mathbf{x}_v) = \mathbf{U}_v = a_{12}\mathbf{x}_u + a_{22}\mathbf{x}_v, \end{cases} \tag{13.11}$$



[3] Julius Weingarten (1836–1910). Professor at the Technische Universität in Berlin. A surface for which there is a definite functional relation between the principal curvatures is called a *Weingarten surface*.

for some functions $a_{11}, a_{21}, a_{12}, a_{22}$, which we need to compute. We take the scalar product of each of the equations in (13.11) with $\mathbf{x}_u$ and $\mathbf{x}_v$, and obtain

$$\text{(13.12)} \qquad \begin{cases} -e = Ea_{11} + Fa_{21}, \\ -f = Fa_{11} + Ga_{21}, \\ -f = Ea_{12} + Fa_{22}, \\ -g = Fa_{12} + Ga_{22}. \end{cases}$$

Equations (13.12) can be written more concisely in terms of matrices:

$$-\begin{pmatrix} e & f \\ f & g \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix};$$

hence

$$\begin{aligned} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= -\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} e & f \\ f & g \end{pmatrix} \\ &= \frac{-1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} e & f \\ f & g \end{pmatrix} \\ &= \frac{-1}{EG - F^2} \begin{pmatrix} Ge - Ff & Gf - Fg \\ -Fe + Ef & -Ff + Eg \end{pmatrix}. \end{aligned}$$

The result follows from (13.11). ▌

Although S is a symmetric linear operator, its matrix (a_{ij}) relative to $\{\mathbf{x}_u, \mathbf{x}_v\}$ need not be symmetric, because $\mathbf{x}_u$ and $\mathbf{x}_v$ are not in general perpendicular to one another.

There is also a way to express the normal curvature in terms of E, F, G and e, f, g:

Lemma 13.17. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface and let $\mathbf{p} \in \mathcal{M}$. Let $\mathbf{x}$ be an injective regular patch on $\mathcal{M}$ with $\mathbf{p} = \mathbf{x}(u_0, v_0)$. Let $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$ and write*

$$\mathbf{v_p} = a\,\mathbf{x}_u(u_0, v_0) + b\,\mathbf{x}_v(u_0, v_0).$$

Then the normal curvature of $\mathcal{M}$ in the direction $\mathbf{v_p}$ is

$$k(\mathbf{v_p}) = \frac{e\,a^2 + 2f\,ab + g\,b^2}{E\,a^2 + 2F\,ab + G\,b^2}.$$

Proof. We have $\|\mathbf{v_p}\|^2 = \|a\,\mathbf{x}_u + b\,\mathbf{x}_v\|^2 = a^2E + 2abF + b^2G$, and

$$S(\mathbf{v_p}) \cdot \mathbf{v_p} = \big(a\,S(\mathbf{x}_u) + b\,S(\mathbf{x}_v)\big) \cdot (a\,\mathbf{x}_u + b\,\mathbf{x}_v) = a^2e + 2abf + b^2g.$$

The result follows from (13.1). ▌

Eigenvalues of the Shape Operator

We first recall an elementary version of the spectral theorem in linear algebra.

Lemma 13.18. *Let V be a real n-dimensional vector space with an inner product and let $A\colon V \to V$ be a linear transformation that is self-adjoint with respect to the inner product. Then the eigenvalues of A are real and A is diagonalizable: there is an orthonormal basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of V such that*

$$A\mathbf{e}_j = \lambda_j \mathbf{e}_j \qquad \text{for } j = 1, \ldots, n.$$

Lemma 13.14 tells us that the shape operator S is a self-adjoint linear operator on each tangent space to a regular surface in $\mathbb{R}^3$, so the eigenvalues of S must be real. These eigenvalues are important geometric quantities associated with each regular surface in $\mathbb{R}^3$. Instead of proving Lemma 13.18 in its full generality, we prove it for the special case of the shape operator.

Lemma 13.19. *The eigenvalues of the shape operator S of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ at $\mathbf{p} \in \mathcal{M}$ are precisely the principal curvatures k_1 and k_2 of $\mathcal{M}$ at $\mathbf{p}$. The corresponding unit eigenvectors are unit principal vectors, and vice versa. If $k_1 = k_2$, then S is scalar multiplication by their common value. Otherwise, the eigenvectors $\mathbf{e}_1, \mathbf{e}_2$ of S are perpendicular, and S is given by*

$$S\mathbf{e}_1 = k_1\mathbf{e}_1, \qquad S\mathbf{e}_2 = k_2\mathbf{e}_2. \tag{13.13}$$

Proof. Consider the normal curvature as a function $k\colon S^1_{\mathbf{p}} \to \mathbb{R}$, where $S^1_{\mathbf{p}}$ is the set of unit tangent vectors in the tangent space $\mathcal{M}_{\mathbf{p}}$. Since $S^1_{\mathbf{p}}$ is a circle, it is compact, and so k achieves its maximum at some unit vector, call it $\mathbf{e}_1 \in S^1_{\mathbf{p}}$. Choose $\mathbf{e}_2$ to be any vector in $S^1_{\mathbf{p}}$ perpendicular to $\mathbf{e}_1$, so $\{\mathbf{e}_1, \mathbf{e}_2\}$ is an orthonormal basis of $\mathcal{M}_{\mathbf{p}}$ and

$$\begin{cases} S\mathbf{e}_1 = (S\mathbf{e}_1 \cdot \mathbf{e}_1)\mathbf{e}_1 + (S\mathbf{e}_1 \cdot \mathbf{e}_2)\mathbf{e}_2, \\ S\mathbf{e}_2 = (S\mathbf{e}_2 \cdot \mathbf{e}_1)\mathbf{e}_1 + (S\mathbf{e}_2 \cdot \mathbf{e}_2)\mathbf{e}_2. \end{cases} \tag{13.14}$$

Define a function $\mathbf{u} = \mathbf{u}(\theta)$ by setting $\mathbf{u}(\theta) = \mathbf{e}_1 \cos\theta + \mathbf{e}_2 \sin\theta$, and write $k(\theta) = k\big(\mathbf{u}(\theta)\big)$. Then

$$k(\theta) = (S\mathbf{e}_1 \cdot \mathbf{e}_1)\cos^2\theta + 2(S\mathbf{e}_1 \cdot \mathbf{e}_2)\sin\theta\cos\theta + (S\mathbf{e}_2 \cdot \mathbf{e}_2)\sin^2\theta, \tag{13.15}$$

so that

$$\frac{d}{d\theta}k(\theta) = 2(S\mathbf{e}_2 \cdot \mathbf{e}_2 - S\mathbf{e}_1 \cdot \mathbf{e}_1)\sin\theta\cos\theta + 2(S\mathbf{e}_1 \cdot \mathbf{e}_2)(\cos^2\theta - \sin^2\theta).$$

In particular,

$$0 = \frac{dk}{d\theta}(0) = 2S\mathbf{e}_1 \cdot \mathbf{e}_2, \tag{13.16}$$

because $k(\theta)$ has a maximum at $\theta = 0$. Then (13.14) and (13.16) imply (13.13). From (13.13) it follows that both $\mathbf{e}_1$ and $\mathbf{e}_2$ are eigenvectors of S, and from (13.15) it follows that the principal curvatures of $\mathcal{M}$ at $\mathbf{p}$ are the eigenvalues of S. Hence the lemma follows. ∎

The principal curvatures determine the normal curvature completely:

Corollary 13.20. (Euler) *Let $k_1(\mathbf{p}), k_2(\mathbf{p})$ be the principal curvatures of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ at $\mathbf{p} \in \mathcal{M}$, and let $\mathbf{e}_1, \mathbf{e}_2$ be the corresponding unit principal vectors. Let θ denote the oriented angle from $\mathbf{e}_1$ to $\mathbf{u}_\mathbf{p}$, so that $\mathbf{u}_\mathbf{p} = \mathbf{e}_1 \cos\theta + \mathbf{e}_2 \sin\theta$. Then the normal curvature $k(\mathbf{u}_\mathbf{p})$ is given by*

$$k(\mathbf{u}_\mathbf{p}) = k_1(\mathbf{p}) \cos^2\theta + k_2(\mathbf{p}) \sin^2\theta. \tag{13.17}$$

Proof. Since $S(\mathbf{e}_1) \cdot \mathbf{e}_2 = 0$, (13.15) reduces to (13.17). ∎

13.4 Gaussian and Mean Curvature

The notion of the curvature of a surface is a great deal more complicated than the notion of curvature of a curve. Let $\boldsymbol{\alpha}$ be a curve in $\mathbb{R}^3$, and let $\mathbf{p}$ be a point on the trace of $\boldsymbol{\alpha}$. The curvature of $\boldsymbol{\alpha}$ at $\mathbf{p}$ measures the rate at which $\boldsymbol{\alpha}$ leaves the tangent *line* to $\boldsymbol{\alpha}$ at $\mathbf{p}$. By analogy, the curvature of a surface $\mathcal{M} \subset \mathbb{R}^3$ at $\mathbf{p} \in \mathcal{M}$ should measure the rate at which $\mathcal{M}$ leaves the tangent *plane* to $\mathcal{M}$ at $\mathbf{p}$. But for surfaces a difficulty arises that was not present for curves: although a curve can separate from one of its tangent lines in only two directions, a surface separates from one of its tangent planes in infinitely many directions. In general, the rate of departure of a surface from one of its tangent planes depends on the direction.

There are several competing notions for the curvature of a surface in $\mathbb{R}^3$:

- the normal curvature k;
- the principal curvatures k_1, k_2;
- the mean curvature H;
- the Gaussian curvature K.

We defined normal curvature and the principal curvatures of a surface $\mathcal{M} \subset \mathbb{R}^3$ in Section 13.2. In the present section, we give the definitions of the Gaussian and mean curvatures; these are the most important functions in surface theory.

First, we recall some useful facts from linear algebra. If $S\colon V \to V$ is a linear transformation on a vector space V, we may define the ***determinant*** and ***trace*** of S, written $\det S$ and $\operatorname{tr} S$, merely as the determinant and trace of the matrix

A representing S with respect to any chosen basis. If P is an invertible matrix describing a change in basis then S is represented by $P^{-1}AP$ with respect to the new basis, but standard properties of the determinant and trace functions ensure that

$$\begin{aligned} \det(P^{-1}AP) &= \det A, \\ \operatorname{tr}(P^{-1}AP) &= \operatorname{tr} A, \end{aligned} \tag{13.18}$$

so that our definitions are independent of the choice of basis.

Definition 13.21. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$. The* ***Gaussian curvature*** *K and* ***mean curvature*** *H of $\mathcal{M}$ are the functions $K, H\colon \mathcal{M} \to \mathbb{R}$ defined by*

$$K(\mathbf{p}) = \det\bigl(S(\mathbf{p})\bigr) \qquad \text{and} \qquad H(\mathbf{p}) = \tfrac{1}{2}\operatorname{tr}\bigl(S(\mathbf{p})\bigr). \tag{13.19}$$

Note that although the shape operator S and the mean curvature H depend on the choice of unit normal $\mathbf{U}$, the Gaussian curvature K is independent of that choice. The name "mean curvature" is due to Germain.[4]

Definition 13.22. *A* ***minimal surface*** *in $\mathbb{R}^3$ is a regular surface for which the mean curvature vanishes identically. A regular surface is* ***flat*** *if and only if its Gaussian curvature vanishes identically.*

We shall see in Chapter 16 that surfaces of minimal area are indeed minimal in the sense of Definition 13.22.

The Gaussian curvature permits us to distinguish four kinds of points on a surface.

Definition 13.23. *Let $\mathbf{p}$ be a point on a regular surface $\mathcal{M} \subset \mathbb{R}^3$. We say that*

- $\mathbf{p}$ *is* ***elliptic*** *if $K(\mathbf{p}) > 0$ (equivalently, k_1 and k_2 have the same sign);*
- $\mathbf{p}$ *is* ***hyperbolic*** *if $K(\mathbf{p}) < 0$ (equivalently, k_1 and k_2 have opposite signs);*
- $\mathbf{p}$ *is* ***parabolic*** *if $K(\mathbf{p}) = 0$, but $S(\mathbf{p}) \neq \mathbf{0}$ (equivalently, exactly one of k_1 and k_2 is zero);*
- $\mathbf{p}$ *is* ***planar*** *if $K(\mathbf{p}) = 0$ and $S(\mathbf{p}) = \mathbf{0}$ (equivalently, $k_1 = k_2 = 0$).*

It is usually possible to glance at almost any surface and recognize which points are elliptic, hyperbolic, parabolic or planar. Consider, for example, the paraboloids shown in Figure 12.10 on page 375. Not surprisingly, all the points

---

4 

Sophie Germain (1776–1831). French mathematician, best known for her work on elasticity and Fermat's last theorem. Germain (under the pseudonym "M. Blanc") corresponded with Gauss regarding her results in geometry and number theory.

on the left-hand elliptical paraboloid are elliptic, and all those on the right-hand hyperbolic paraboloid are hyperbolic. Calculations from the next section show that the monkey surface (13.25) has all its points hyperbolic except for its central point, $\mathbf{o} = (0, 0, 0)$, which is planar. This corresponds to the fact, illustrated in Figure 13.5, that the Gaussian curvature achieves its maximum value of 0 at $\mathbf{o}$.

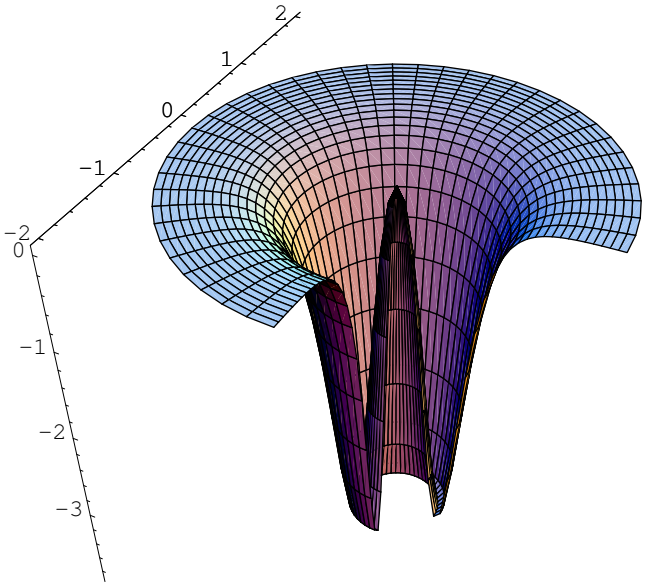


Figure 13.5: Gaussian curvature of the monkey saddle

There are two especially useful ways of choosing a basis of a tangent space to a surface in $\mathbb{R}^3$. Each gives rise to important formulas for the Gaussian and mean curvatures, which are presented in turn by the following proposition and subsequent theorem.

Proposition 13.24. *The Gaussian curvature and mean curvature of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ are related to the principal curvatures by*

$$K = k_1 k_2 \qquad \textit{and} \qquad H = \tfrac{1}{2}(k_1 + k_2).$$

Proof. If we choose an orthonormal basis of eigenvectors of S for $\mathcal{M}_{\mathbf{p}}$, the matrix of S with respect to this basis is diagonal so that

$$K = \det \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} = k_1 k_2$$

$$H = \tfrac{1}{2} \operatorname{tr} \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} = \tfrac{1}{2}(k_1 + k_2). \quad \blacksquare$$

Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface. The Gaussian and mean curvatures are functions $K, H\colon \mathcal{M} \to \mathbb{R}$; we have written $K(\mathbf{p})$ and $H(\mathbf{p})$ for their values at $\mathbf{p} \in \mathcal{M}$. We need a slightly different notation for a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$. Strictly speaking, K and H are functions defined on $\mathbf{x}(\mathcal{U})$. However, we follow conventional notation and abbreviate $K \circ \mathbf{x}$ to K and $H \circ \mathbf{x}$ to H. Thus $K(u, v)$ and $H(u, v)$ will denote the values of the Gaussian and mean curvatures at $\mathbf{x}(u, v)$.

Theorem 13.25. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. The Gaussian curvature and mean curvature of $\mathbf{x}$ are given by the formulas*

$$K = \frac{eg - f^2}{EG - F^2}, \tag{13.20}$$

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}, \tag{13.21}$$

where e, f, g are the coefficients of the second fundamental form relative to $\mathbf{x}$, and E, F, G are the coefficients of the first fundamental form.

Proof. This time we compute K and H using the basis $\{\mathbf{x}_u, \mathbf{x}_v\}$, and the matrix

$$\frac{-1}{EG - F^2} \begin{pmatrix} Ge - Ff & Gf - Fg \\ -Fe + Ef & -Ff + Eg \end{pmatrix}$$

of Theorem 13.16. Taking the determinant and half the trace yields

$$K = \frac{(fF - eG)(fF - gE) - (eF - fE)(gF - fG)}{(EG - F^2)^2} = \frac{eg - f^2}{EG - F^2},$$

and

$$H = -\frac{(fF - eG) + (fF - gE)}{2(EG - F^2)} = \frac{eG - 2fF + gE}{2(EG - F^2)}. \qquad \blacksquare$$

The importance of Proposition 13.24 is theoretical, that of Theorem 13.25 more practical. Usually, one uses Theorem 13.25 to compute K and H, and afterwards Proposition 13.24 to find the principal curvatures. More explicitly,

Corollary 13.26. *The principal curvatures k_1, k_2 are the roots of the quadratic equation*

$$x^2 - 2Hx + K = 0.$$

Thus we can choose k_1, k_2 so that

$$k_1 = H + \sqrt{H^2 - K} \qquad \text{and} \qquad k_2 = H - \sqrt{H^2 - K}. \tag{13.22}$$

Corollary 13.27. *Suppose that $\mathcal{M} \subset \mathbb{R}^3$ has negative Gaussian curvature K at $\mathbf{p}$. Then*:

(i) *there are exactly two asymptotic directions at $\mathbf{p}$, and they are bisected by the principal directions*;

(ii) *the two asymptotic directions at $\mathbf{p}$ are perpendicular if and only if the mean curvature H of $\mathcal{M}$ vanishes at $\mathbf{p}$.*

Proof. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be unit principal vectors corresponding to $k_1(\mathbf{p})$ and $k_2(\mathbf{p})$. Then $K(\mathbf{p}) = k_1(\mathbf{p})k_2(\mathbf{p}) < 0$ implies that $k_1(\mathbf{p})$ and $k_2(\mathbf{p})$ have opposite signs. Thus, there exists θ with $0 < \theta < \pi/2$ such that

$$\tan^2\theta = -\frac{k_1(\mathbf{p})}{k_2(\mathbf{p})}.$$

Put $\mathbf{u}_\mathbf{p}(\theta) = \mathbf{e}_1 \cos\theta + \mathbf{e}_2 \sin\theta$. Then (13.17) implies that $\mathbf{u}_\mathbf{p}(\theta)$ and $\mathbf{u}_\mathbf{p}(-\theta)$ are linearly independent asymptotic vectors at $\mathbf{p}$. The angle between $\mathbf{u}_\mathbf{p}(\theta)$ and $\mathbf{u}_\mathbf{p}(-\theta)$ is 2θ, and it is clear that $\mathbf{e}_1$ bisects the angle between $\mathbf{u}_\mathbf{p}(\theta)$ and $\mathbf{u}_\mathbf{p}(-\theta)$. This proves (i). For (ii) we observe that $H(\mathbf{p}) = 0$ if and only if θ equals $\pm\pi/4$, up to integral multiples of π. ∎

The Three Fundamental Forms

In classical differential geometry, there are frequent references to the "second fundamental form" of a surface in $\mathbb{R}^3$, a notion that is essentially equivalent to the shape operator S. Such references can for example be found in the influential textbook [Eisen1] of Eisenhart.[5]

Definition 13.28. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$. The* ***second fundamental form*** *is the symmetric bilinear form $\mathbf{II}$ on a tangent space $\mathcal{M}_\mathbf{p}$ given by*

$$\mathbf{II}(\mathbf{v}_\mathbf{p}, \mathbf{w}_\mathbf{p}) = S(\mathbf{v}_\mathbf{p}) \cdot \mathbf{w}_\mathbf{p}$$

for $\mathbf{v}_\mathbf{p}, \mathbf{w}_\mathbf{p} \in \mathcal{M}_\mathbf{p}$.

Since there is a second fundamental form, there must be a ***first fundamental form***. It is nothing but the inner product between tangent vectors:

$$\mathbf{I}(\mathbf{v}_\mathbf{p}, \mathbf{w}_\mathbf{p}) = \mathbf{v}_\mathbf{p} \cdot \mathbf{w}_\mathbf{p}.$$



[5] Luther Pfahler Eisenhart (1876–1965). American differential geometer and dean at Princeton University.

Note that the first fundamental form $\mathbf{I}$ can in a sense be defined whether or not the surface is in $\mathbb{R}^3$; this is the basis of theory to be discussed in Chapter 26.

The following lemma is an immediate consequence of the definitions.

Lemma 13.29. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then*

$$\mathbf{I}(a\,\mathbf{x}_u + b\,\mathbf{x}_v,\ a\,\mathbf{x}_u + b\,\mathbf{x}_v) = E\,a^2 + 2F\,ab + G\,b^2,$$
$$\mathbf{II}(a\,\mathbf{x}_u + b\,\mathbf{x}_v,\ a\,\mathbf{x}_u + b\,\mathbf{x}_v) = e\,a^2 + 2f\,ab + g\,b^2.$$

The normal curvature is therefore given by

$$k(\mathbf{v_p}) = \frac{\mathbf{II}(\mathbf{v_p}, \mathbf{v_p})}{\mathbf{I}(\mathbf{v_p}, \mathbf{v_p})}$$

for any nonzero tangent vector $\mathbf{v_p}$.

This lemma explains why we call E, F, G the coefficients of the first fundamental form, and e, f, g the coefficients of the second fundamental form.

Finally, there is also a ***third fundamental form*** $\mathbf{III}$ for a surface in $\mathbb{R}^3$ given by

$$\mathbf{III}(\mathbf{v_p}, \mathbf{w_p}) = S(\mathbf{v_p}) \cdot S(\mathbf{w_p})$$

for $\mathbf{v_p}, \mathbf{w_p} \in \mathcal{M_p}$. Note that $\mathbf{III}$, in contrast to $\mathbf{II}$, does not depend on the choice of surface normal $\mathbf{U}$. The third fundamental form $\mathbf{III}$ contains no new information, since it is expressible in terms of $\mathbf{I}$ and $\mathbf{II}$.

Lemma 13.30. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface. Then the following relation holds between the first, second and third fundamental forms of $\mathcal{M}$:*

$$\mathbf{III} - 2H\,\mathbf{II} + K\,\mathbf{I} = 0, \tag{13.23}$$

where H and K denote the mean curvature and Gaussian curvature of $\mathcal{M}$.

Proof. Although (13.23) follows from Corollary 13.26 and the *Cayley–Hamilton*[6] *Theorem* (which states that a matrix satisfies its own characteristic polynomial), we prefer to give a direct proof.

First, note that the product $H\,\mathbf{II}$ is independent of the choice of surface normal $\mathbf{U}$. Hence (13.23) makes sense whether or not $\mathcal{M}$ is orientable. To



6 Sir William Rowan Hamilton (1805–1865). Irish mathematician, best known for having been struck with the concept of quaternions as he crossed Brougham Bridge in Dublin (see Chapter 23), and for his work in dynamics.

prove it, we observe that since its left-hand side is a symmetric bilinear form, it suffices to show that for each $\mathbf{p} \in \mathcal{M}$ and some basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ of $\mathcal{M}_\mathbf{p}$ we have

$$(\mathbf{III} - 2H\,\mathbf{II} + K\,\mathbf{I})(\mathbf{e}_i, \mathbf{e}_j) = 0, \tag{13.24}$$

for $i, j = 1, 2$. We choose $\mathbf{e}_1$ and $\mathbf{e}_2$ to be linearly independent principal vectors at $\mathbf{p}$. Then

$$(\mathbf{III} - 2H\,\mathbf{II} + K\,\mathbf{I})(\mathbf{e}_1, \mathbf{e}_2) = 0$$

because each term vanishes separately. Furthermore,

$$(\mathbf{III} - 2H\,\mathbf{II} + K\,\mathbf{I})(\mathbf{e}_i, \mathbf{e}_i) = k_i^2 - (k_1 + k_2)k_i + k_1 k_2 = 0$$

for $i = 1$ and 2, as required. ∎

13.5 More Curvature Calculations

In this section, we show how to compute by hand the Gaussian curvature K and the mean curvature H for a monkey saddle and a torus. Along the way we compute the coefficients of their first and second fundamental forms.

The Monkey Saddle

For the surface parametrized by

$$\mathbf{x}(u, v) = \mathsf{monkeysaddle}(u, v) = (u,\ v,\ u^3 - 3uv^2), \tag{13.25}$$

and described on page 304, we easily compute

$$\mathbf{x}_u(u, v) = (1, 0, 3u^2 - 3v^2), \qquad \mathbf{x}_v(u, v) = (0, 1, -6uv),$$

$$\mathbf{x}_{uu}(u, v) = (0, 0, 6u), \quad \mathbf{x}_{uv}(u, v) = (0, 0, -6v), \quad \mathbf{x}_{vv}(u, v) = (0, 0, -6u).$$

Therefore,

$$E = 1 + 9(u^2 - v^2)^2, \quad F = -18uv(u^2 - v^2), \quad G = 1 + 36u^2v^2,$$

and by inspection a unit surface normal is

$$\mathbf{U} = \frac{(-3u^2 + 3v^2,\ 6uv,\ 1)}{\sqrt{1 + 9u^4 + 18u^2v^2 + 9v^4}},$$

so that

$$\begin{aligned} e &= \mathbf{U} \cdot \mathbf{x}_{uu} = \frac{6u}{\sqrt{1 + 9u^4 + 18u^2v^2 + 9v^4}}, \\ f &= \mathbf{U} \cdot \mathbf{x}_{uv} = \frac{-6v}{\sqrt{1 + 9u^4 + 18u^2v^2 + 9v^4}}, \\ g &= \mathbf{U} \cdot \mathbf{x}_{vv} = \frac{-6u}{\sqrt{1 + 9u^4 + 18u^2v^2 + 9v^4}}. \end{aligned}$$

Theorem 13.25 yields

$$K = \frac{-36(u^2+v^2)}{(1+9u^4+18u^2v^2+9v^4)^2}, \quad H = \frac{-27u^5+54u^3v^2+81uv^4}{(1+9u^4+18u^2v^2+9v^4)^{3/2}}.$$

A glance at these expressions shows that $\mathbf{o} = (0,0,0)$ is a planar point of the monkey saddle and that every other point is hyperbolic. Furthermore, the Gaussian curvature of the monkey saddle is invariant under all rotations about the z-axis (this is Exercise 9 on page 547), even though the monkey saddle itself does *not* have this property.

The principal curvatures are determined by Corollary 13.26, and are easy to plot. Figure 13.6 shows graphically their singular nature at the point $\mathbf{o}$, which contrasts with the surface itself sandwiched in the middle.

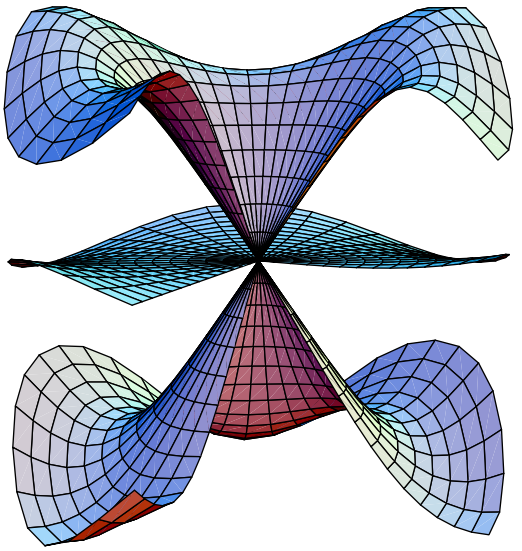


Figure 13.6: Principal curvatures of the monkey saddle

Alternative Formulas

Classically, the standard formulas for computing K and H for a patch $\mathbf{x}$ are (13.20) and (13.21). It is usually too tedious to compute K and H by hand in one step. Therefore, the functions E, F, G and e, f, g need to be computed before any of the curvature functions are calculated. The computation of E, F, G is straightforward: first one computes the first derivatives $\mathbf{x}_u$ and $\mathbf{x}_v$ and then the dot products $E = \mathbf{x}_u \cdot \mathbf{x}_u$, $F = \mathbf{x}_u \cdot \mathbf{x}_v$ and $G = \mathbf{x}_v \cdot \mathbf{x}_v$.

There are two methods for computing e, f, g. The direct approach using the definitions necessitates computing the surface normal via equation (10.11) on

page 295, and then using the definition (13.9). The other method, explained in the next lemma, avoids computation of the surface normal; it uses instead the vector triple product which can be computed as a determinant.

Lemma 13.31. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then*

$$e = \frac{[\mathbf{x}_{uu}\ \mathbf{x}_u\ \mathbf{x}_v]}{\sqrt{EG - F^2}}, \quad f = \frac{[\mathbf{x}_{uv}\ \mathbf{x}_u\ \mathbf{x}_v]}{\sqrt{EG - F^2}}, \quad g = \frac{[\mathbf{x}_{vv}\ \mathbf{x}_u\ \mathbf{x}_v]}{\sqrt{EG - F^2}}.$$

Proof. From (7.4) and (13.9) it follows that

$$e = \mathbf{x}_{uu} \cdot \mathbf{U} = \mathbf{x}_{uu} \cdot \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = \frac{[\mathbf{x}_{uu}\ \mathbf{x}_u\ \mathbf{x}_v]}{\sqrt{EG - F^2}}.$$

The other formulas are proved similarly. ∎

On the other hand, at least theoretically, we can compute K and H for a regular patch $\mathbf{x}$ directly in terms of the first and second derivatives of $\mathbf{x}$. Here are the relevant formulas.

Corollary 13.32. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. The Gaussian and mean curvatures of $\mathcal{M}$ are given by the formulas*

$$K = \frac{[\mathbf{x}_{uu}\ \mathbf{x}_u\ \mathbf{x}_v][\mathbf{x}_{vv}\ \mathbf{x}_u\ \mathbf{x}_v] - [\mathbf{x}_{uv}\ \mathbf{x}_u\ \mathbf{x}_v]^2}{\big(\|\mathbf{x}_u\|^2\|\mathbf{x}_v\|^2 - (\mathbf{x}_u \cdot \mathbf{x}_v)^2\big)^2}, \tag{13.26}$$

$$H = \frac{[\mathbf{x}_{uu}\ \mathbf{x}_u\ \mathbf{x}_v]\|\mathbf{x}_v\|^2 - 2[\mathbf{x}_{uv}\ \mathbf{x}_u\ \mathbf{x}_v](\mathbf{x}_u \cdot \mathbf{x}_v) + [\mathbf{x}_{vv}\ \mathbf{x}_u\ \mathbf{x}_v]\|\mathbf{x}_u\|^2}{2\big(\|\mathbf{x}_u\|^2\|\mathbf{x}_v\|^2 - (\mathbf{x}_u \cdot \mathbf{x}_v)^2\big)^{3/2}}. \tag{13.27}$$

Proof. The equations follow from (13.20), (13.21) and Lemma 13.31 when we write out e, f, g and E, F, G explicitly in terms of dot products. ∎

Equations 13.26 and 13.27 are used effectively in Notebook 13. They are usually too complicated for hand calculation, but we did use them below in the case of the torus.

It is neither enlightening nor useful to write out the formulas for the principal curvatures in terms of E, F, G, e, f, g in general. However, there is one special case when such formulas for k_1, k_2 are worth noting.

Corollary 13.33. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch for which $f = F = 0$. With respect to this patch, the principal curvatures are e/E and g/G.*

Proof. When $F = f = 0$, the Weingarten equations (13.10) reduce to

$$S(\mathbf{x}_u) = \frac{e}{E}\mathbf{x}_u \qquad \text{and} \qquad S(\mathbf{x}_v) = \frac{g}{G}\mathbf{x}_v.$$

By definition, e/E and g/G are the eigenvalues of the shape operator S. ∎

The Sphere

We compute the principal curvatures of the patch

$$\mathbf{x}(u,v) = \mathsf{sphere}[a](u,v) = \big(a\cos v\cos u,\ a\cos v\sin u,\ a\sin v\big) \tag{13.28}$$

of the sphere with center $\mathbf{o} = (0,0,0)$ and radius a. We find that

$$\begin{aligned}\mathbf{x}_u(u,v) &= \big(-a\cos v\sin u,\ a\cos v\cos u,\ 0\big),\\ \mathbf{x}_v(u,v) &= \big(-a\sin v\cos u,\ -a\sin v\sin u,\ a\cos v\big),\end{aligned}$$

and so

$$E = a^2\cos^2 v, \qquad F = 0, \qquad G = a^2.$$

Furthermore,

$$\begin{aligned}\mathbf{x}_{uu}(u,v) &= \big(-a\cos v\cos u,\ -a\cos v\sin u,\ 0\big),\\ \mathbf{x}_{uv}(u,v) &= \big(\ a\sin v\sin u,\ -a\sin v\cos u,\ 0\big),\\ \mathbf{x}_{vv}(u,v) &= \big(-a\cos v\cos u,\ -a\cos v\sin u,\ -a\sin v\big) = -\mathbf{x}(u,v),\end{aligned}$$

and Lemma 13.31 yields

$$e = \frac{\det\begin{pmatrix} -a\cos v\cos u & -a\cos v\sin u & 0\\ -a\cos v\sin u & a\cos v\cos u & 0\\ -a\sin v\cos u & -a\sin v\sin u & a\cos v\end{pmatrix}}{a^2\cos v} = -a\cos^2 v.$$

Changing just the first row of the determinant gives $f = 0$ and $g = -a$. Therefore,

$$K = a^{-2}, \qquad H = -a^{-1},$$

$$k_1 = -a^{-1} = k_2,$$

and the corresponding Weingarten matrix of the shape operator is

$$\begin{pmatrix} a^{-1} & 0\\ 0 & a^{-1}\end{pmatrix}.$$

The Circular Torus

We compute the Gaussian and mean curvatures of the patch

$$\mathbf{x}(u,v) = \mathsf{torus}[a,b,b](u,v) = \big((a+b\cos v)\cos u,\ (a+b\cos v)\sin u,\ b\sin v\big),$$

representing a torus with circular sections rather than the more general case on pages 210 and 305.

Setting $a = 0$ reduces the parametrization of the torus to that of the sphere (13.28). For this reason, the calculations are only slightly more involved than those above, though we now assume that $a > b > 0$ and jump to the results

$$e = -\cos v(a + b\cos v), \qquad f = 0, \qquad g = -b,$$

and

$$K = \frac{\cos v}{b(a + b\cos v)}, \qquad H = -\frac{a + 2b\cos v}{2b(a + b\cos v)},$$

$$k_1 = -\frac{\cos v}{a + b\cos v}, \qquad k_2 = -\frac{1}{b}.$$

It follows that the Gaussian curvature K of the torus vanishes along the curves given by $v = \pm\pi/2$. These are the two circles of contact, when the torus is held between two planes of glass. These circles consist exclusively of parabolic points, since the angle featured in (13.2) on page 390 is $\pi/2$, and the normal curvature cannot change sign.

The set of hyperbolic points is $\{\mathbf{x}(u, v) \mid \frac{1}{2}\pi < v < \frac{3}{2}\pi\}$, and the set of elliptic points is $\{\mathbf{x}(u, v) \mid -\frac{1}{2}\pi < v < \frac{1}{2}\pi\}$. This situation can be illustrated using commands in Notebook 13 that produce different colors according to the sign of K.

The Astroidal Ellipsoid

If we modify the standard parametrization of the ellipsoid given on page 313 by replacing each coordinate by its cube, we obtain the ***astroidal ellipsoid***

$$\mathsf{astell}[a, b, c](u, v) = \big((a\cos u\cos v)^3,\ (b\sin u\cos v)^3,\ (c\sin v)^3\big).$$

Therefore, $\mathsf{astell}[a, a, a]$ has the nonparametric equation

$$x^{2/3} + y^{2/3} + z^{2/3} = a^2,$$

and is called the ***astroidal sphere***. Figure 13.7 depicts it touching an ordinary sphere.

Notebook 13 computes the Gaussian curvature of the astroidal sphere, which is given by

$$K = \frac{1024\sec^4 v}{9a^6\left(-18 + 2\cos 4u + \cos(4u - 2v) + 14\cos 2v + \cos(4u + 2v)\right)^2}.$$

Surprisingly, this function is continuous on the edges of the astroid, and is singular only at the vertices. This is confirmed by Figure 13.8.

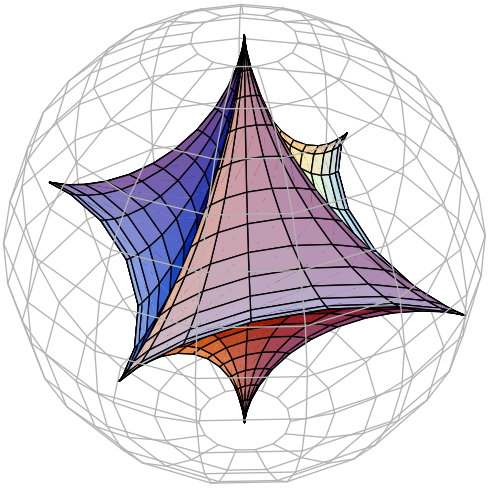


Figure 13.7: The astroidal sphere

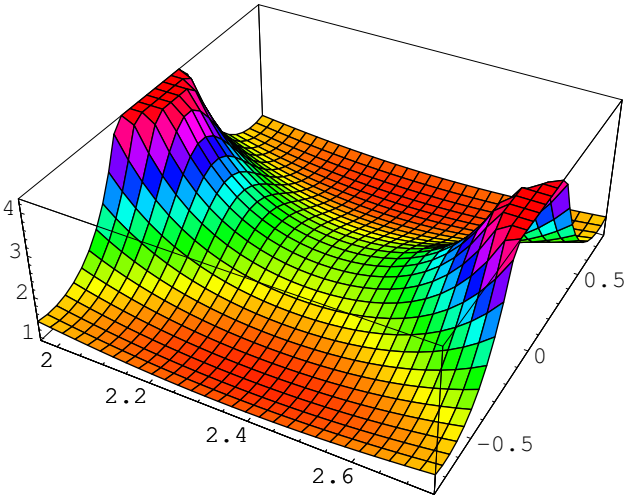


Figure 13.8: Gaussian curvature of a portion of the astroidal sphere

Monge Patches

It is not hard to compute directly the Gaussian and mean curvatures of graphs of functions of two variables; see, for example, [dC1, pages 162–163]. However, we can use computations from Notebook 13 to verify the following results:

Lemma 13.34. *For a Monge patch* $(u,v) \mapsto \big(u,\ v,\ h(u,v)\big)$ *we have*

$$E = 1 + h_u^2, \qquad F = h_u h_v, \qquad G = 1 + h_v^2,$$

$$e = \frac{h_{uu}}{(1 + h_u^2 + h_v^2)^{1/2}}, \qquad f = \frac{h_{uv}}{(1 + h_u^2 + h_v^2)^{1/2}}, \qquad g = \frac{h_{vv}}{(1 + h_u^2 + h_v^2)^{1/2}},$$

$$K = \frac{h_{uu}h_{vv} - h_{uv}^2}{(1 + h_u^2 + h_v^2)^2}, \qquad H = \frac{(1 + h_v^2)h_{uu} - 2h_u h_v h_{uv} + (1 + h_u^2)h_{vv}}{2(1 + h_u^2 + h_v^2)^{3/2}}.$$

See Exercise 18 for the detailed general computations.

Once the formulas of Lemma 13.34 have been stored in Notebook 13, they can be applied to specific functions. For example, suppose we wish to determine the curvature of the graph of the function

$$p_{m,n}(u,v) = u^m v^n, \tag{13.29}$$

for fixed positive integers m, n. One quickly discovers that

$$K = \frac{m(1-m-n)nu^{2m-2}v^{2n-2}}{\big(1 + m^2u^{2m-2}v^{2n-2} + n^2u^{2m}v^{2n-2}\big)^2}.$$

We see from this formula that the graph has nonpositive Gaussian curvature at all points, provided $m + n > 1$. For definiteness, let us illustrate what happens when $m = 2$ and $n = 4$.

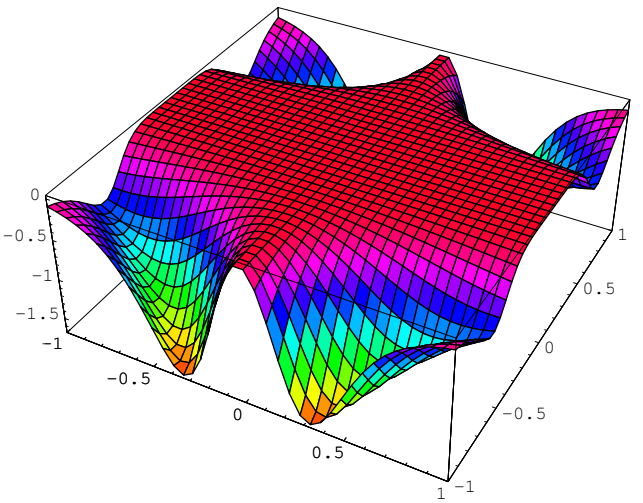


Figure 13.9: Gaussian curvature of the graph of $(u,v) \mapsto u^2v^4$

Figure 13.9 shows that the Gaussian curvature of $p_{2,4}$ is everywhere nonpositive. It is easier to see that all points are either hyperbolic or planar by plotting p in polar coordinates; this we do in Figure 13.10.

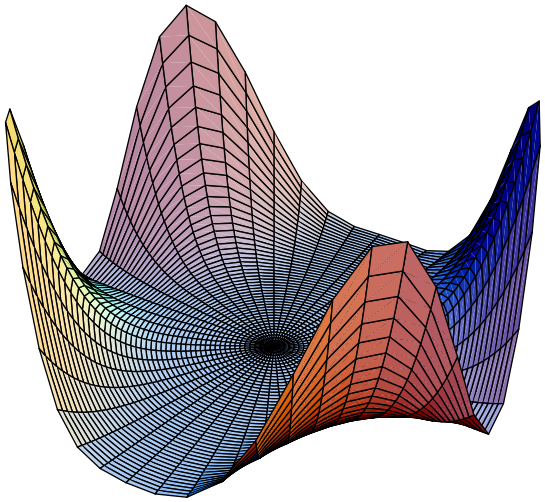


Figure 13.10: The graph of $p_{2,4}$ in polar coordinates

13.6 A Global Curvature Theorem

We recall the following fact about compact subsets of $\mathbb{R}^n$ (see page 373):

Lemma 13.35. *Let $\mathcal{R}$ be a compact subset of $\mathbb{R}^n$, and let $f\colon \mathcal{R} \to \mathbb{R}$ be a continuous function. Then f assumes its maximum value at some point $\mathbf{p} \in \mathcal{R}$.*

This basic result was in fact already used in the proof of Lemma 13.19. For its proof, see [Buck, page 74].

Intuitively, it is reasonable to assume that for each compact surface $\mathcal{M} \subset \mathbb{R}^3$, there is a point $\mathbf{p}_0 \in \mathcal{M}$ that is furthest from the origin, and that at $\mathbf{p}_0$ the surface bends towards the origin. Thus it appears that the Gaussian curvature K of $\mathcal{M}$ is positive at $\mathbf{p}_0$. We now prove that this is indeed the case, using standard facts concerning a maximum of a differentiable function of one variable.

Theorem 13.36. *If $\mathcal{M}$ is a compact regular surface in $\mathbb{R}^3$, there is a point $\mathbf{p} \in \mathbb{R}^3$ at which the Gaussian curvature K is strictly positive.*

Proof. Define $f\colon \mathbb{R}^3 \to \mathbb{R}$ by $f(\mathbf{p}) = \|\mathbf{p}\|^2$. Then f is continuous (in fact, differentiable), since it can be expressed in terms of the natural coordinate functions of $\mathbb{R}^3$ as $f = u_1^2 + u_2^2 + u_3^2$. By Lemma 13.35, f assumes its maximum value at some point $\mathbf{p}_0 \in \mathcal{M}$. Let $\mathbf{v} \in \mathcal{M}_{\mathbf{p}_0}$ be a unit vector, and choose a unit-speed curve $\boldsymbol{\alpha}\colon (a, b) \to \mathcal{M}$ such that $a < 0 < b$, $\boldsymbol{\alpha}(0) = \mathbf{p}_0$, and $\boldsymbol{\alpha}'(0) = \mathbf{v}$.

Since the function $g\colon (a,b) \to \mathbb{R}$ defined by $g = f \circ \boldsymbol{\alpha}$ has a maximum at 0, it follows that

$$g'(0) = 0 \qquad \text{and} \qquad g''(0) \leqslant 0.$$

But $g(t) = \boldsymbol{\alpha}(t) \cdot \boldsymbol{\alpha}(t)$, so that

$$0 = g'(0) = 2\boldsymbol{\alpha}'(0) \cdot \boldsymbol{\alpha}(0) = 2\mathbf{v} \cdot \mathbf{p}_0. \tag{13.30}$$

In (13.30), $\mathbf{v}$ can be an arbitrary unit tangent vector, and so $\mathbf{p}_0$ must be normal to $\mathcal{M}$ at $\mathbf{p}_0$. Clearly, $\mathbf{p}_0 \neq 0$, so that (13.30) implies that $\mathbf{p}_0/\|\mathbf{p}_0\|$ is a unit normal vector to $\mathcal{M}$ at $\mathbf{p}_0$. Furthermore,

$$0 \geqslant g''(0) = 2\boldsymbol{\alpha}''(0) \cdot \boldsymbol{\alpha}(0) + 2\boldsymbol{\alpha}'(0) \cdot \boldsymbol{\alpha}'(0) = 2\big(\boldsymbol{\alpha}''(0) \cdot \mathbf{p}_0 + 1\big),$$

so that $\boldsymbol{\alpha}''(0) \cdot \mathbf{p}_0 \leqslant -1$, or

$$k(\mathbf{v}) = \boldsymbol{\alpha}''(0) \cdot \frac{\mathbf{p}_0}{\|\mathbf{p}_0\|} \leqslant -\frac{1}{\|\mathbf{p}_0\|}, \tag{13.31}$$

where $k(\mathbf{v})$ is the normal curvature determined by the tangent vector $\mathbf{v}$ and the unit normal vector $\mathbf{p}_0/\|\mathbf{p}_0\|$. In particular, the principal curvatures of $\mathcal{M}$ at $\mathbf{p}_0$ (with respect to $\mathbf{p}_0/\|\mathbf{p}_0\|$) satisfy

$$k_1(\mathbf{p}_0), k_2(\mathbf{p}_0) \leqslant -\frac{1}{\|\mathbf{p}_0\|}.$$

This implies that the Gaussian curvature of $\mathcal{M}$ at $\mathbf{p}_0$ satisfies

$$K(\mathbf{p}_0) = k_1(\mathbf{p}_0)k_2(\mathbf{p}_0) \geqslant \frac{1}{\|\mathbf{p}_0\|^2} > 0. \blacksquare$$

Noncompact surfaces of positive Gaussian curvature exist (see Exercise 15). On the other hand, for surfaces of nonpositive curvature, we can deduce the following result:

Corollary 13.37. *Any surface in $\mathbb{R}^3$ whose Gaussian curvature is everywhere nonpositive must be noncompact.*

13.7 Nonparametrically Defined Surfaces

So far we have discussed computing the curvature of a surface from its parametric representation. In this section we show how in some cases the curvature can be computed from the nonparametric form of a surface.

Lemma 13.38. *Let $\mathbf{p}$ be a point on a regular surface $\mathcal{M} \subset \mathbb{R}^3$, and let $\mathbf{v_p}$ and $\mathbf{w_p}$ be tangent vectors to $\mathcal{M}$ at $\mathbf{p}$. Then the Gaussian and mean curvatures of $\mathcal{M}$ at $\mathbf{p}$ are related to the shape operator by the formulas*

$$S(\mathbf{v_p}) \times S(\mathbf{w_p}) \;=\; K(\mathbf{p})\,\mathbf{v_p} \times \mathbf{w_p}, \tag{13.32}$$

$$S(\mathbf{v_p}) \times \mathbf{w_p} + \mathbf{v_p} \times S(\mathbf{w_p}) \;=\; 2H(\mathbf{p})\,\mathbf{v_p} \times \mathbf{w_p}. \tag{13.33}$$

Proof. First, assume that $\mathbf{v_p}$ and $\mathbf{w_p}$ are linearly independent. Then we can write

$$S(\mathbf{v_p}) = a\mathbf{v_p} + b\mathbf{w_p} \qquad \text{and} \qquad S(\mathbf{w_p}) = c\mathbf{v_p} + d\mathbf{w_p},$$

so that

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

is the matrix of S with respect to $\mathbf{v_p}$ and $\mathbf{w_p}$. It follows from (13.19), page 398, that

$$\begin{aligned} S(\mathbf{v_p}) \times \mathbf{w_p} + \mathbf{v_p} \times S(\mathbf{w_p}) &= (a\mathbf{v_p} + b\mathbf{w_p}) \times \mathbf{w_p} + \mathbf{v_p} \times (c\mathbf{v_p} + d\mathbf{w_p}) \\ &= (a+d)\,\mathbf{v_p} \times \mathbf{w_p} \\ &= (\operatorname{tr} S(\mathbf{p}))\,\mathbf{v_p} \times \mathbf{w_p} \;=\; 2H(\mathbf{p})\,\mathbf{v_p} \times \mathbf{w_p}, \end{aligned}$$

proving (13.33) in the case that $\mathbf{v_p}$ and $\mathbf{w_p}$ are linearly independent.

If $\mathbf{v_p}$ and $\mathbf{w_p}$ are linearly dependent, they are still the limits of linearly independent tangent vectors. Since both sides of (13.33) are continuous in $\mathbf{v_p}$ and $\mathbf{w_p}$, we get (13.33) in the general case. Equation (13.32) is proved by the same method (see Exercise 16). ∎

Theorem 13.39. *Let $\mathbf{Z}$ be a nonvanishing vector field on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ which is everywhere perpendicular to $\mathcal{M}$. Let $\mathbf{V}$ and $\mathbf{W}$ be vector fields tangent to $\mathcal{M}$ such that $\mathbf{V} \times \mathbf{W} = \mathbf{Z}$. Then*

$$K \;=\; \frac{[\mathbf{Z}\ \boldsymbol{D}_{\mathbf{V}}\mathbf{Z}\ \boldsymbol{D}_{\mathbf{W}}\mathbf{Z}]}{\|\mathbf{Z}\|^4}, \tag{13.34}$$

$$H \;=\; \frac{[\mathbf{Z}\ \mathbf{W}\ \boldsymbol{D}_{\mathbf{V}}\mathbf{Z}] - [\mathbf{Z}\ \mathbf{V}\ \boldsymbol{D}_{\mathbf{W}}\mathbf{Z}]}{2\|\mathbf{Z}\|^3}. \tag{13.35}$$

Proof. Let $\mathbf{U} = \mathbf{Z}/\|\mathbf{Z}\|$; then (9.5), page 267, implies that

$$\boldsymbol{D}_{\mathbf{V}}\mathbf{U} = \frac{\boldsymbol{D}_{\mathbf{V}}\mathbf{Z}}{\|\mathbf{Z}\|} + \mathbf{V}\left[\frac{1}{\|\mathbf{Z}\|}\right]\mathbf{Z}.$$

Therefore,

$$S(\mathbf{V}) = -\boldsymbol{D}_{\mathbf{V}}\mathbf{U} = \frac{-\boldsymbol{D}_{\mathbf{V}}\mathbf{Z}}{\|\mathbf{Z}\|} + \mathbf{N}_{\mathbf{V}},$$

where $\mathbf{N}_{\mathbf{V}}$ is a vector field normal to $\mathcal{M}$. By Lemma 13.38 we have

$$\begin{aligned} K\,\mathbf{V} \times \mathbf{W} &= S(\mathbf{V}) \times S(\mathbf{W}) \\ &= \left(\frac{-\boldsymbol{D}_{\mathbf{V}}\mathbf{Z}}{\|\mathbf{Z}\|} + \mathbf{N}_{\mathbf{V}}\right) \times \left(\frac{-\boldsymbol{D}_{\mathbf{W}}\mathbf{Z}}{\|\mathbf{Z}\|} + \mathbf{N}_{\mathbf{W}}\right). \end{aligned} \tag{13.36}$$

Since $\mathbf{N_V}$ and $\mathbf{N_W}$ are linearly dependent, it follows from (13.36) that

$$K\mathbf{Z} = \frac{\boldsymbol{D}_{\mathbf{V}}\mathbf{Z} \times \boldsymbol{D}_{\mathbf{W}}\mathbf{Z}}{\|\mathbf{Z}\|^2} + \text{some vector field tangent to } \mathcal{M}.$$

Taking the scalar product of both sides with $\mathbf{Z}$ yields (13.34). Equation (13.35) is proved in a similar fashion. ▌

In order to make use of Theorem 13.39, we need an important function that measures the distance from the origin of each tangent plane to a surface.

Definition 13.40. *Let $\mathcal{M}$ be an oriented regular surface in $\mathbb{R}^3$ with surface normal $\mathbf{U}$. Then the* ***support function*** *of $\mathcal{M}$ is the function $h\colon \mathcal{M} \to \mathbb{R}$ given by*

$$h(\mathbf{p}) = \mathbf{p} \cdot \mathbf{U}(\mathbf{p}).$$

Geometrically, $h(\mathbf{p})$ is the distance from the origin to the tangent space $\mathcal{M}_{\mathbf{p}}$.

Corollary 13.41. *Let $\mathcal{M}$ be the surface*

$$\{\, (u_1, u_2, u_3) \in \mathbb{R}^3 \mid f_1 u_1^k + f_2 u_2^k + f_3 u_3^k = 1 \,\},$$

where f_1, f_2, f_3 are constants, not all zero, and k is a nonzero real number. Then the support function, Gaussian curvature and mean curvature of $\mathcal{M}$ are given by

$$h = \frac{1}{\sqrt{f_1^2 u_1^{2k-2} + f_2^2 u_2^{2k-2} + f_3^2 u_3^{2k-2}}}, \tag{13.37}$$

$$K = \frac{(k-1)^2 f_1 f_2 f_3 (u_1 u_2 u_3)^{k-2}}{\left(\sum_{i=1}^{3} f_i^2 u_i^{2k-2} \right)^2} = h^4 (k-1)^2 f_1 f_2 f_3 (u_1 u_2 u_3)^{k-2}, \tag{13.38}$$

$$\begin{aligned} H = {} & \frac{-k+1}{2\left(\sum_{i=1}^{3} f_i^2 u_i^{2k-2} \right)^{\frac{3}{2}}} \Big(f_1 f_2 (u_1 u_2)^{k-2} (f_1 u_1^k + f_2 u_2^k) \\ & + f_2 f_3 (u_2 u_3)^{k-2} (f_2 u_2^k + f_3 u_3^k) + f_3 f_1 (u_3 u_1)^{k-2} (f_3 u_3^k + f_1 u_1^k) \Big). \end{aligned} \tag{13.39}$$

Proof. Let $g(u_1, u_2, u_3) = f_1 u_1^k + f_2 u_2^k + f_3 u_3^k - 1$, so that

$$\mathcal{M} = \{\, \mathbf{p} \in \mathbb{R}^3 \mid g(\mathbf{p}) = 0 \,\}.$$

Then $\mathbf{Z} = \mathsf{grad}\, g$ is a nonvanishing vector field that is everywhere perpendicular to $\mathcal{M}$. Explicitly,

$$\mathbf{Z} = k \sum_{i=1}^{3} f_i u_i^{k-1} \mathbf{U}_i,$$

where u_1, u_2, u_3 are the natural coordinate functions of $\mathbb{R}^3$, and $\mathbf{U}_i = \partial/\partial u_i$ (see Definition 9.20). The vector field $\mathbf{X} = \sum u_i \mathbf{U}_i$ satisfies

$$\mathbf{X} \cdot \mathbf{Z} = k \sum f_i u_i^k, \tag{13.40}$$

in which the summations continue to be over $i = 1, 2, 3$. The support function of $\mathcal{M}$ is given by

$$h = \mathbf{X} \cdot \frac{\mathbf{Z}}{\|\mathbf{Z}\|} = \frac{\sum f_i u_i^k}{\sqrt{\sum f_i^2 u_i^{2k-2}}}.$$

Since $\sum f_i u_i^k$ equals 1 on $\mathcal{M}$, we get (13.37).

Next, let

$$\mathbf{V} = \sum v_i \mathbf{U}_i \quad \text{and} \quad \mathbf{W} = \sum w_i \mathbf{U}_i$$

be vector fields on $\mathbb{R}^3$. Since f_1, f_2, f_3 are constants, we have

$$\boldsymbol{D}_{\mathbf{V}}\mathbf{Z} = k \sum \mathbf{V}[f_i u_i^{k-1}]\mathbf{U}_i = k(k-1) \sum f_i v_i u_i^{k-2} \mathbf{U}_i,$$

and similarly for $\mathbf{W}$. Therefore, the triple product $[\mathbf{Z}\ \boldsymbol{D}_{\mathbf{V}}\mathbf{Z}\ \boldsymbol{D}_{\mathbf{W}}\mathbf{Z}]$ equals

$$\det \begin{pmatrix} k f_1 u_1^{k-1} & k f_2 u_2^{k-1} & k f_3 u_3^{k-1} \\ k(k-1) f_1 v_1 u_1^{k-2} & k(k-1) f_2 v_2 u_2^{k-2} & k(k-1) f_3 v_3 u_3^{k-2} \\ k(k-1) f_1 w_1 u_1^{k-2} & k(k-1) f_2 w_2 u_2^{k-2} & k(k-1) f_3 w_3 u_3^{k-2} \end{pmatrix}$$

$$= k^3 (k-1)^2 f_1 f_2 f_3 u_1^{k-2} u_2^{k-2} u_3^{k-2} [\mathbf{X}\ \mathbf{V}\ \mathbf{W}].$$

Now we choose $\mathbf{V}$ and $\mathbf{W}$ so that they are tangent to $\mathcal{M}$ and $\mathbf{V} \times \mathbf{W} = \mathbf{Z}$. Using (13.34), we obtain

$$K = \frac{k^3 (k-1)^2 f_1 f_2 f_3 u_1^{k-2} u_2^{k-2} u_3^{k-2} \mathbf{X} \cdot \mathbf{Z}}{\|\mathbf{Z}\|^4} = \frac{(k-1)^2 f_1 f_2 f_3 u_1^{k-2} u_2^{k-2} u_3^{k-2}}{\left(\sum_{i=1}^{3} f_i^2 u_i^{2k-2} \right)^2}.$$

This proves (13.38). The proof of (13.39) is similar. ∎

Computations in Notebook 13 yield three special cases of Corollary 13.41.

Corollary 13.42. The support function and Gaussian curvature of

(i) *the ellipsoid* $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$,

(ii) *the hyperboloid of one sheet* $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$,

(iii) *the hyperboloid of two sheets* $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = -1$

are given in each case by

$$h = \left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-1/2} \qquad \textit{and} \qquad K = \pm\frac{h^4}{a^2b^2c^2},$$

where the minus sign only applies in (ii).

We shall return to these quadrics in Section 19.6, where we describe geometrically useful parametrizations of them. We conclude the present section by considering a ***superquadric***, which is a surface of the form

$$f_1x^k + f_2y^k + f_3z^k = 1,$$

where k is different from 2. We do the special case $k = 2/3$, which is an astroidal ellipsoid (see page 407).

Corollary 13.43. The Gaussian curvature of the superquadric

$$f_1x^{2/3} + f_2y^{2/3} + f_3z^{2/3} = 1$$

is given by

$$K = \frac{f_1f_2f_3}{9\Big(f_3^2(xy)^{2/3} + f_2^2(xz)^{2/3} + f_1^2(yz)^{2/3}\Big)^2}.$$

13.8 Exercises

M **1.** Plot the graph of $p_{2,3}$, defined by (13.29), and its Gaussian curvature.

M **2.** Plot the following surfaces and describe in each case (without additional calculation) the set of elliptic, hyperbolic, parabolic, and planar points:

(a) a sphere,

(b) an ellipsoid,

(c) an elliptic paraboloid,

(d) a hyperbolic paraboloid,

(e) a hyperboloid of one sheet,

(f) a hyperboloid of two sheets.

This exercise continues on page 450. Notebook 13 includes a program to color regions of a surface according to the sign of the Gaussian curvature.

3. By studying the plots of the surfaces listed in Exercise 2, describe the general shape of the image of their Gauss maps.

4. Compute the coefficients of the first fundamental form, those of the second fundamental form, the unit normal, the mean curvature, and the principal curvatures of the following surfaces:

(a) the elliptical torus,

(b) the helicoid,

(c) Enneper's minimal surface.

Refer to the exercises on page 377.

5. Compute the first fundamental form, the second fundamental form, the unit normal, the Gaussian curvature, the mean curvature, and the principal curvatures of the patch

$$\mathbf{x}(u, v) = \big(u^2 + v,\ v^2 + u,\ u\,v\big).$$

6. The ***translation surface*** determined by curves $\boldsymbol{\alpha}, \boldsymbol{\gamma}\colon (a, b) \to \mathbb{R}^3$ is the patch

$$(u, v) \mapsto \boldsymbol{\alpha}(u) + \boldsymbol{\gamma}(v).$$

It is the surface formed by moving $\boldsymbol{\alpha}$ parallel to itself in such a way that a point of the curve moves along $\boldsymbol{\gamma}$. Show that the second fundamental form of this patch satisfies $f = 0$.

7. Explain the difference between the translation surface formed by a circle and a lemniscate lying in perpendicular planes, and the twisted surface formed from a lemniscate according to Section 11.6.

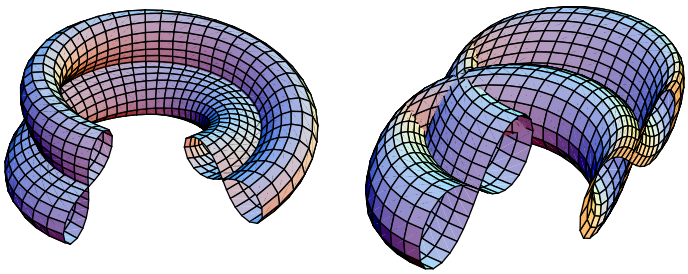


Figure 13.11: Translation and twisted surface formed by a lemniscate

8. Compute the first fundamental form, the second fundamental form, the unit normal, the Gaussian curvature, the mean curvature, and the principal curvatures of the patch

$$\mathbf{x}_n(u, v) = \big(u^n,\ v^n,\ u\,v\big).$$

For $n = 3$, see the picture on page 291.

9. Prove the statement immediately following Definition 13.8.

10. Show that the first fundamental form of the Gauss map of a patch $\mathbf{x}$ coincides with the third fundamental form of $\mathbf{x}$.

11. Show that an orientation-preserving Euclidean motion $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ leaves unchanged both principal curvatures and principal vectors.

12. Show that the ***Bohemian dome*** defined by

$$\mathsf{bohdom}[a,b,c,d](u,v) = \big(a\cos u,\ b\sin u + c\cos v,\ d\sin v\big)$$

is the translation surface of two ellipses (see Exercise 6). Compute the Gaussian curvature, the mean curvature, and the principal curvatures of $\mathsf{bohdom}[a,b,b,a]$.

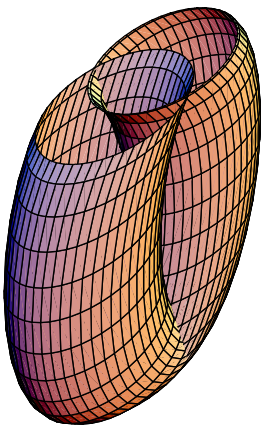


Figure 13.12: The Bohemian dome $\mathsf{bohdom}[1,2,2,3]$

13. Show that the mean curvature $H(\mathbf{p})$ of a surface $\mathcal{M} \subset \mathbb{R}^3$ at $\mathbf{p} \in \mathcal{M}$ is given by

$$H(\mathbf{p}) = \frac{1}{\pi}\int_0^\pi k(\theta)\,d\theta,$$

where $k(\theta)$ denotes the normal curvature, as in the proof of Lemma 13.19.

14. (Continuation) Let n be an integer larger than 2 and for $0 \leqslant i \leqslant n-1$, put $\theta_i = \psi + 2\pi i/n$, where ψ is some angle. Show that

$$H(\mathbf{p}) = \frac{1}{n}\sum_{i=0}^{n-1} k(\theta_i).$$

15. Give examples of a noncompact surface

(a) whose Gaussian curvature is negative,

(b) whose Gaussian curvature is identically zero,

(c) whose Gaussian curvature is positive,

(d) containing elliptic, hyperbolic, parabolic, and planar points.

16. Prove equation (13.32).

17. Show that there are no compact minimal surfaces in $\mathbb{R}^3$.

M **18.** Find the coefficients of the first, second and third fundamental forms to the following surfaces:

(a) the elliptical torus defined on page 305,

(b) the helicoid defined on page 376,

(c) Enneper's minimal surface defined on page 378,

(d) a Monge patch defined on page 302.

19. Determine the principal curvatures of the Whitney umbrella, defined on page 307. Its mean curvature is shown in Figure 13.13.

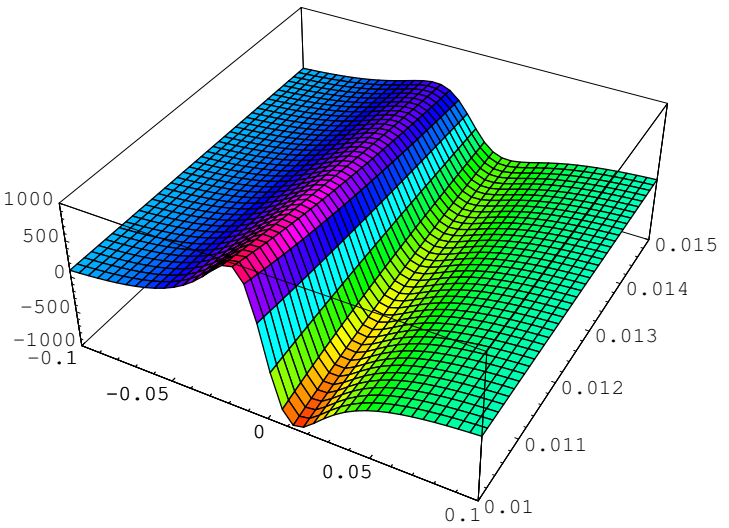


Figure 13.13: Mean curvature of the Whitney umbrella

M **20.** Find the formulas for the coefficients e, f, g of the second fundamental forms of the following surfaces parametrized in Chapter 11: the Möbius strip, the Klein bottle, Steiner's Roman surface, the cross cap.

M **21.** Find the formulas for the Gaussian and mean curvatures of the surfaces of the preceding exercise.

22. Prove

Lemma 13.44. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface, and suppose that $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ are coherently-oriented patches on $\mathcal{M}$ with $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ nonempty. Let $\mathbf{x}^{-1} \circ \mathbf{y} = (\bar{u}, \bar{v})\colon \mathcal{U} \cap \mathcal{V} \to \mathcal{U} \cap \mathcal{V}$ be the associated change of coordinates, so that*

$$\mathbf{y}(u, v) = \mathbf{x}\big(\bar{u}(u, v), \bar{v}(u, v)\big).$$

Let $e_{\mathbf{x}}, f_{\mathbf{x}}, g_{\mathbf{x}}$ denote the coefficients of the second fundamental form of $\mathbf{x}$, and let $e_{\mathbf{y}}, f_{\mathbf{y}}, g_{\mathbf{y}}$ denote the coefficients of the second fundamental form of $\mathbf{y}$. Then

$$\left\{\begin{aligned} e_{\mathbf{y}} &= e_{\mathbf{x}}\left(\frac{\partial \bar{u}}{\partial u}\right)^2 + 2f_{\mathbf{x}}\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial u} + g_{\mathbf{x}}\left(\frac{\partial \bar{v}}{\partial u}\right)^2, \\ f_{\mathbf{y}} &= e_{\mathbf{x}}\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{u}}{\partial v} + f_{\mathbf{x}}\left(\frac{\partial \bar{u}}{\partial u}\frac{\partial \bar{v}}{\partial v} + \frac{\partial \bar{u}}{\partial v}\frac{\partial \bar{v}}{\partial u}\right) + g_{\mathbf{x}}\frac{\partial \bar{v}}{\partial u}\frac{\partial \bar{v}}{\partial v}, \\ g_{\mathbf{y}} &= e_{\mathbf{x}}\left(\frac{\partial \bar{u}}{\partial v}\right)^2 + 2f_{\mathbf{x}}\frac{\partial \bar{u}}{\partial v}\frac{\partial \bar{v}}{\partial v} + g_{\mathbf{x}}\left(\frac{\partial \bar{v}}{\partial v}\right)^2. \end{aligned}\right. \tag{13.41}$$

23. Prove equation (13.39) of Corollary 13.41.

Notebook 13

■ N13.0

This notebook contains programs for computing the shape operator and various curvatures. For most surfaces in 3-space, calculation of shape and curvature is more efficiently done by computer, with the added advantage that the resulting functions can be immediately plotted and interpreted. Usually, their visualization provides much more insight than the formulas themselves. A good example is the plot of the principal curvatures of the monkey saddle in Figure 13.6.

N13.1 contains a miniprogram to plot tangent vectors to coordinate curves, and this is used in Figure 13.1 to illustrate the differential of the Gauss map. The remainder of the notebook is devoted to a comparison of various definitions, in parallel to the text of Chapter 13. The most effective definitions for computing the Gauss and mean curvature are given in N13.5, and correspond to Corollary 13.32.

The following surfaces been have introduced previously.

```
torus[a_,b_,c_][u_, v_]:=
  {(a + b Cos[v])Cos[u], (a + b Cos[v])Sin[u], c Sin[v]}
sphere[a_]:= torus[0,a,a]
paraboloid[a_,b_][u_,v_]:= {u, v, a u^2+b v^2}
monkeysaddle[u_,v_]:= {u, v, u^3-3u v^2}
```

We shall consider another hyperbolic paraboloid, parametrized by

```
hyperbolicparaboloid[u_,v_]:= {u, v, u v}
```

Indeed, it is convenient to use the abbreviations

```
hp:= hyperbolicparaboloid
ms:= monkeysaddle
pa:= paraboloid[1,-1]
```

We shall make early use of the following command, introduced in N7.3, to visibly place space curves on surfaces. It includes new default settings for Thickness and Hue.

```
highlight[α_, h_:1/20, k_:3/200][t_]:=
  Append[α[t], {Thickness[k],Hue[h]}]
```

For Figure 13.7, we need the WireFrame construction from the package

```
<<Graphics`Shapes`
```

Subsection N13.0.1 includes the program unitnormal, originally introduced in N10.2, and the program metric introduced in N12.1. Both will be used in N13.3.

- N13.0.1

N13.1

The following program displays coordinate curves and their tangent vectors. Applying this to a patch and its unit normal illustrates the process that defines the shape operator.

```
coortang[x_,r_]:=
Module[{k,α,β,αα,ββ,o,p,q,y,z},
  k= {0,0,.01};
  α[t_]= x[t,0] + k;
  β[t_]= x[0,t] + k;
  αα= highlight[α];
  ββ= highlight[β, 7/10];
  z[1]= ParametricPlot3D[x[u,v], {u,-r,r}, {v,-r,r},
    PlotPoints->21];
  z[2]= ParametricPlot3D[αα[t], {t,-r,r}];
  z[3]= ParametricPlot3D[ββ[t], {t,-r,r}];
  o= x[0,0];
  p= o + α'[0]/L[α'[0]];
  q= o + β'[0]/L[β'[0]];
  y[4]= {Thickness[.02], Line[{o,p}], Line[{o,q}]};
  y[5]= {PointSize[.05], Point[o]};
  z[i_]= y[i]//Graphics3D;
  Show[Array[z,4], Boxed->False, Shading->False]
]
```

Examples of its use can be displayed by adding S@ to the above command.

```
coortang[hp,2]
```

```
S @ Module[{x,y},
  x[u_,v_]={u, v, u^2 - (v-.3)^2/4};
  y[u_,v_]= unitnormal[x][u,v];
  GraphicsArray[{coortang[x,1], coortang[y,1]}]
]
```

Figure 13.1

A different perspective is obtained by changing the plot range as follows:

```
coortang[hp,.5]
```

■ N13.2

In the next figure, p1 and p2 represent planes. Reducing the number of PlotPoints is a useful device for simplifying the mesh that is constructed by *Mathematica*.

```
p1:= ParametricPlot3D[{u,0,v}, {u,-3,3}, {v,-3,3},
  PlotPoints->{Automatic,5}]
p2:= ParametricPlot3D[{0,u,v}, {u,-3,3}, {v,-3,3},
  PlotPoints->{Automatic,5}]
zz:= ParametricPlot3D[hp[u,v], {u,-3,3}, {v,-3,3}]
S @ Show[p1, p2, zz, BoxRatios->{1,1,1}]
```

Figure 13.2

Figures 13.3 and 13.4 are produced in much the same way, and instructions are hidden in the next subsection.

■ N13.2.1

■ N13.3

The next cell contains a program for the shape operator that functions in the same way as metric, loaded in N13.0.1.

```
shape[x_][u_,v_]:=
Module[{xuu,xuv,xvv,U,V},
  Nx= unitnormal[x][u,v];
  si= Simplify[#]/.{U->u, V->v} &;
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  eee= xuu.Nx//si;
  fff=  xuv.Nx//si;
  ggg= xvv.Nx//si;
  {{eee,fff}, {fff,ggg}}
]
```

It computes the coefficients eee,fff,ggg (notation for e,f,g), and displays the resulting second fundamental form as a matrix. It is based on the definitions involving second derivatives of a patch, and accordingly makes essential use of unitnormal. Once the program has been run, the coefficients of the shape operator can be called one by one:

```
shape[pa][u,v]
eee
fff
ggg
```

Simplifying the matrix for the sphere requires a bit more effort:

```
shape[sphere[1]][u,v]
%//FullSimplify
%//PowerExpand
%//MatrixForm
```

The shape operator is represented by *minus* the Weingarten matrix relative to the tangent vectors defined by the chart in question. This observation is the basis of the following command for the Weingarten matrix.

```
weingarten[x_][u_,v_]:=
Module[{M1,M2},
  M1= metric[x][u,v];
  M2= shape[x][u,v];
  -M2.Inverse[M1]//Simplify
]
```

The principal curvatures are *minus* the eigenvalues of the Weingarten matrix. The next output provides and simplifies one of the principal curvatures of the monkey saddle.

```
Wms:= weingarten[ms][u,v]
Eigenvalues[-Wms][[1]]
%//Simplify
```

■ N13.4

The next two programs compute the Gauss and mean curvature using the usual formulas (Theorem 13.25) in terms of the coefficients ee,ff,gg (standing for E,F,G) and eee,fff,ggg (standing for e,f,g), and their associated symmetric matrices.

```
gausscurv[x_][u_,v_]:=
Module[{M1,M2},
  M1:= metric[x][u,v];
  M2:= shape[x][u,v];
  Simplify[Det[M2]/Det[M1]]
]
```

```
meancurv[x_][u_,v_]:=
Module[{M1},
  M1:= metric[x][u,v];
  shape[x][u,v];
  Simplify[(ee ggg - 2 ff fff + gg eee)/(2Det[M1])]
]
```

The Gauss and mean curvatures could have been computed from the Weingarten matrix found in N13.3, and we can easily check that the two approaches are consistent:

```
Whp:= weingarten[hp][u,v]
Whp//MatrixForm
Det[Whp] - gausscurv[hp][u,v] //Simplify
```

```
Sum[Whp[[i]][[i]],{i,2}] + 2 meancurv[hp][u,v]
```

Here is a simpler numerical example:

```
gausscurv[pa][1,2]
```

We can easily compute and plot the Gaussian and mean curvatures of the monkey saddle. As usual, add S@ to display the graphics.

```
K= gausscurv[ms]
H= meancurv[ms]
Plot3D[K[u,v], {u,-1,1}, {v,-1,1}, PlotPoints->40]
Plot3D[H[u,v], {u,-1,1}, {v,-1,1}, PlotPoints->40]
```

The next cell uses the polarize command, from N10.4 and N13.0.1, to display the Gaussian curvature in a more symmetrical way. Labels on the y- and z-axes confirm that the plot is to scale, a fact dictated by the option BoxRatios->Automatic set in N13.0.1.

```
KK[r_,θ_]:= {r Cos[θ], r Sin[θ], polarize[K][r,θ]}
S @ ParametricPlot3D[KK[r,θ], {r,0,2}, {θ,0,3π/2},
  PlotPoints->{30,40}, ViewPoint->{1,-2,2},
  Axes->{False,True,True}, ImageSize->400]
```

Figure 13.5

■ N13.5

We first compute the first and second fundamental form of the monkey saddle.

```
metric[ms][u,v]//MatrixForm
ee
ff
gg
```

```
shape[ms][u,v]//MatrixForm
eee
fff
ggg
```

The same computations are repeated (though not displayed) when running the next cell.

```
gausscurv[ms][u,v]
meancurv[ms][u,v]
```

Corollary 13.32 yields the following self-contained and more computationally effective curvature programs. They are based upon the realization of normal components of second derivatives of the chart as dot products, computed as determinants. We immediately check their consistency for the monkey saddle.

```
gcurv[x_][u_,v_]:=
Module[{xu,xv,xuu,xuv,xvv,U,V,A,B},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  B= Det[{xuu,xu,xv}]Det[{xvv,xu,xv}] - Det[{xuv,xu,xv}]^2;
  A= (xu.xu)(xv.xv) - (xu.xv)^2;
  Simplify[B/A^2]/.{U->u,V->v}
]
```

```
gausscurv[ms][u,v] - gcurv[ms][u,v]//Simplify
```

```
mcurv[x_][u_,v_]:=
Module[{xu,xv,xuu,xuv,xvv,U,V,A,B},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  B= Det[{xuu,xu,xv}](xv.xv) - 2Det[{xuv,xu,xv}](xu.xv)
    + Det[{xvv,xu,xv}](xu.xu);
  A= (xu.xu)(xv.xv) - (xu.xv)^2;
  Simplify[B/(2A^(3/2))]/.{U->u,V->v}
]
```

```
meancurv[ms][u,v] - mcurv[ms][u,v]//Simplify
```

In practice, we prefer to compute principal curvatures using these programs rather than the method in N13.3:

```
princurv[x_][u_,v_]:=
Module[{K,H},
  K= gcurv[x][u,v];
  H= mcurv[x][u,v];
  k1= H + Sqrt[Simplify[H^2-K]];
  k2= H - Sqrt[Simplify[H^2-K]];
  {k1,k2}
]
```

```
princurv[ms][1,0]
k1
```

We next define the astell surface, and check consistency of the two Gaussian curvature programs for the special case in which a=b=c.

```
astell[a_,b_,c_][u_,v_]:=
  {a Cos[u] Cos[v], b Sin[u]Cos[v], c Sin[v]}^3
```

```
as:= astell[1,1,1]
as[u,v]
K1= gausscurv[as][u,v]//Timing
K2= gcurv[as][u,v]//Timing
K1 - K2//FullSimplify
```

```
Module[{ms,xy},
  ms:= monkeysaddle;
  princurv[ms][u,v];
  ss= {u->r Cos[θ], v->r Sin[θ]};
  xy= {r Cos[θ], r Sin[θ]};
  tri:= {ms@@xy, Append[xy, k1/.ss], Append[xy, k2/.ss]};
  S @ ParametricPlot3D[tri//Evaluate, {r,0,.6}, {θ,0,3π/2},
    PlotPoints->{15,30}, BoxRatios->{1,1,1},
    Axes->None, Boxed->False, ViewPoint->{1.5,-3,.6}]
]
```

Figure 13.6

Here is a way to plot a torus of revolution indicating regions of positive, negative and zero Gaussian curvature:

```
Module[{x,f},
  x= torus[8,3,3];
  f[v_]= FaceForm[Hue[N[If[v<Pi/2,.2,.8]]]];
  xx[u_,v_]= Append[x[u,v], f[v]];
  ParametricPlot3D[xx[u,v], {u,0,2π}, {v,-π/2,3π/2},
    PlotPoints->{40,40}, Lighting->False]
]
```

The astroidal sphere is plotted next, and its curvature in N13.5.1.

```
Module[{sp,z},
  sp:= sphere[1];
  z[1]= ParametricPlot3D[as[u,v], {u,0,2π}, {v,-π/2,π/2}];
  z[2]= WireFrame[{GrayLevel[.7], Sphere[1,15,15]}];
  S[Array[z,2]]
]
```

Figure 13.7

N13.5.1

N13.7

Here are ready-made formulas for the Gauss and mean curvatures of the so-called *superquadrics*:

```
gcurvsq[k_][f1_,f2_,f3_][u1_,u2_,u3_]:=
Module[{F,U,n,B},
  F= f1 f2 f3;
  U= u1 u2 u3;
  n=2k-2;
  B= f1^2 u1^n+f2^2 u2^n+f3^2 u3^n;
  (k-1)^2 F U^(k-2) / B^2
]
```

```
gcurvsq[k][1,1,1][x,y,z]
```

```
mcurvsq[k_][f1_,f2_,f3_][u1_,u2_,u3_]:=
Module[{m,n,A,B},
  m= k-2;
  n= 2k-2;
  A= f1 f2 (u1 u2)^m(f1 u1^k+f2 u2^k)+
    f2 f3(u2 u3)^m(f2 u2^k+f3 u3^k)+
    f3 f1(u3 u1)^m(f3 u3^k+f1 u1^k);
  B= f1^2 u1^n+f2^2 u2^n+f3^2 u3^n;
  -(k-1)A / (2 B^(3/2))
]
```

```
mcurvsq[k][1,1,1][x,y,z]
```

We can now list the functions for ordinary quadrics:

```
mcurvsq[2][1/a^2, 1/b^2, -1/c^2][u1,u2,u3]
%//Simplify//PowerExpand
```

```
gcurvsq[2][1/a^2, 1/b^2, -1/c^2][u1,u2,u3]
%//Simplify//PowerExpand
```

```
gcurvsq[2/3][f1,f2,f3][u1,u2,u3]
%//Together//PowerExpand
```

■ N13.8

The following plane curve was first defined in N2.2.

```
lemniscate[t_]:= {Cos[t], Sin[t]Cos[t]}/(1+Sin[t]^2)
```

It gives rise to the following surface using the command twist defined in N11.6 and N13.0.1.

```
twist[lemniscate, 2, 1/2][u,v]//Simplify
```

This is used for Figure 13.11, whose program contains a plot z[3], consisting of a set of individual ellipses, that has not been displayed in the book. It can be seen by replacing the "2" in the penultimate line by "3" so that the Array becomes one of three objects.

```
Module[{α,γ,ta,x,z},
  α[t_]= Prepend[lemniscate[t],0];
  γ[t_]= {Cos[t], Sin[t], 0};
  ta= Table[γ[2π k/12]+α[t], {k,12}];
  x= twist[lemniscate, 2, 1/2];
  z[1]:= ParametricPlot3D[x[u,v], {u,-π/8, 3π/2}, {v,0,2π},
    PlotPoints->{40,30}];
  z[2]:= ParametricPlot3D[α[u]+γ[v], {u,0, 2π}, {v,0,3π/2},
    PlotPoints->{40,30}];
  z[3]:= ParametricPlot3D[ta//Evaluate, {t,0,2π},
    PlotPoints->40];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->-.2]
]
```

Figure 13.11

The definition of the Bohemian dome below is followed by pictures that illustrate its construction:

```
bohdom[a_,b_,c_,d_][u_,v_]:=
  {a Cos[u], b Sin[u] + c Cos[v], d Sin[v]}
```

```
x:= bohdom[1,2,2,3]
ParametricPlot3D[x[u, v], {u,0,π}, {v,-π,0}];
```

```
x:= bohdom[1,2,2,3]
r:= 5
S @ ParametricPlot3D[x[u, v], {u,0,2π}, {v,-2π,0},
   PlotPoints->{50, 30}, PlotRange->{{-r,r},{-r,r},{-r,2.7}}]
```

Figure 13.12

A final subsection computes and plots curvature functions for the Whitney umbrella.

■ N13.8.1

Chapter 14

Ruled Surfaces

We describe in this chapter the important class of surfaces, consisting of those which contain infinitely many straight lines. The most obvious examples of ruled surfaces are cones and cylinders (see pages 314 and 375).

Ruled surfaces are arguably the easiest of all surfaces to parametrize. A chart can be defined by choosing a curve and a vector field along that curve, and the resulting parametrization is linear in one coordinate. This is the content of Definition 14.1, that provides a model for the definition of both the helicoid and the Möbius strip (described on pages 376 and 339).

Several quadric surfaces, including the hyperbolic paraboloid and the hyperboloid of one sheet, are also ruled, though this fact does not follow so readily from their definition. These particular quadric surfaces are in fact "doubly ruled" in the sense that they admit two one-parameter families of lines. In Section 14.1, we explain how to parametrize them in such a way as to visualize the straight line rulings. We also define the Plücker conoid and its generalizations.

It is a straightforward matter to compute the Gaussian and mean curvature of a ruled surface. This we do in Section 14.2, quickly turning attention to *flat* ruled surfaces, meaning ruled surfaces with zero Gaussian curvature. There are three classes of such surfaces, the least obvious but most interesting being the class of *tangent developables*.

Tangent developables are the surfaces swept out by the tangent lines to a space curve, and the two halves of each tangent line effectively divide the surface into two sheets which meet along the curve in a singular fashion described by Theorem 14.7. This behavior is examined in detail for Viviani's curve, and the linearity inherent in the definition of a ruled surface enables us to highlight the useful role played by certain plane curves in the description of tangent developables.

In Section 14.4, we shall study a class of surfaces that will help us to better understand how the Gaussian curvature varies on a ruled surface.

14.1 Definitions and Examples

A ruled surface is a surface generated by a straight line moving along a curve.

Definition 14.1. *A* ***ruled surface*** $\mathcal{M}$ *in* $\mathbb{R}^3$ *is a surface which contains at least one* 1*-parameter family of straight lines. Thus a ruled surface has a parametrization* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *of the form*

$$\mathbf{x}(u, v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\gamma}(u), \tag{14.1}$$

where $\boldsymbol{\alpha}$ *and* $\boldsymbol{\gamma}$ *are curves in* $\mathbb{R}^3$*. We call* $\mathbf{x}$ *a* ***ruled patch****. The curve* $\boldsymbol{\alpha}$ *is called the* ***directrix*** *or* ***base curve*** *of the ruled surface, and* $\boldsymbol{\gamma}$ *is called the* ***director curve****. The* ***rulings*** *are the straight lines* $v \mapsto \boldsymbol{\alpha}(u) + v\boldsymbol{\gamma}(u)$*.*

In using (14.1), we shall assume that $\boldsymbol{\alpha}'$ is never zero, and that $\boldsymbol{\gamma}$ is not identically zero.

We shall see later that any straight line in a surface is necessarily an asymptotic curve, pointing as it does in a direction for which the normal curvature vanishes (see Definition 13.9 on page 390 and Corollary 18.6 on page 560). It follows that the rulings are asymptotic curves. Sometimes a ruled surface $\mathcal{M}$ has two distinct ruled patches on it, so that a ruling of one patch does not belong to the other patch. In this case, we say that $\mathbf{x}$ is ***doubly ruled***.

We now investigate a number of examples.

The Helicoid and Möbius Strip Revisited

We can rewrite the definition (12.27) as

$$\mathsf{helicoid}[a, b](u, v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\gamma}(u),$$

where

$$\begin{cases} \boldsymbol{\alpha}(u) = (0,\ 0,\ bu), \\ \boldsymbol{\gamma}(u) = a(\cos u,\ \sin u,\ 0). \end{cases} \tag{14.2}$$

In this way, $\mathsf{helicoid}[a, b]$ is a ruled surface whose base curve has the z-axis as its trace, and director curve $\boldsymbol{\gamma}$ that describes a circle.

In a similar fashion, our definition (11.3), page 339, of the Möbius strip becomes

$$\mathsf{moebiusstrip}(u, v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\gamma}(u),$$

where

$$\begin{cases} \boldsymbol{\alpha}(u) = (\cos u,\ \sin u,\ 0), \\ \boldsymbol{\gamma}(u) = \Big(\cos\dfrac{u}{2}\cos u,\ \cos\dfrac{u}{2}\sin u,\ \sin\dfrac{u}{2}\Big). \end{cases} \tag{14.3}$$

This time, it is the circle that is the *base* curve of the ruled surface. The Möbius strip has a director curve $\boldsymbol{\gamma}$ that lies on a unit sphere, shown in two equivalent ways in Figure 14.1.

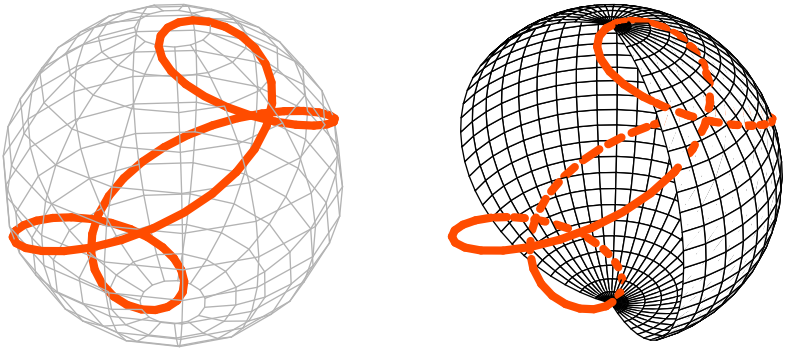


Figure 14.1: Director curve for the Möbius strip

The curve $\boldsymbol{\gamma}$ is not one that we encountered in the study of curves on the sphere in Section 8.4. It does however have the interesting property that whenever $\mathbf{p}$ belongs to its trace, so does the antipodal point $-\mathbf{p}$.

The Hyperboloid of One Sheet

The *elliptical hyperboloid of one sheet* is defined nonparametrically by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{14.4}$$

Planes perpendicular to the z-axis intersect the surface in ellipses, while planes parallel to the z-axis intersect it in hyperbolas. We gave the standard parametrization of the hyperboloid of one sheet on page 313, but this has the disadvantage of not showing the rulings.

Let us show that the hyperboloid of one sheet is a doubly-ruled surface by finding two ruled patches on it. This can be done by fixing $a, b, c > 0$ and defining

$$\mathbf{x}^{\pm}(u, v) = \boldsymbol{\alpha}(u) \pm v\,\boldsymbol{\alpha}'(u) + v\,(0, 0, c),$$

where

$$\boldsymbol{\alpha}(u) = \mathsf{ellipse}[a, b](u) = (a\cos u,\ b\sin u,\ 0) \tag{14.5}$$

is the standard parametrization of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

in the xy-plane. It is readily checked that

$$\mathbf{x}^{\pm}(u, v) = \Big(a(\cos u \mp v \sin u),\ b(\sin u \pm v \cos u),\ cv\Big), \tag{14.6}$$

and that both $\mathbf{x}^+$ and $\mathbf{x}^-$ are indeed parametrizations of the hyperboloid (14.4). Hence the elliptic hyperboloid is doubly ruled; in both cases, the base curve can be taken to be the ellipse (14.5). We also remark that $\mathbf{x}^+$ can be obtained from $\mathbf{x}^-$ by simultaneously changing the signs of the parameter c and the variable v.

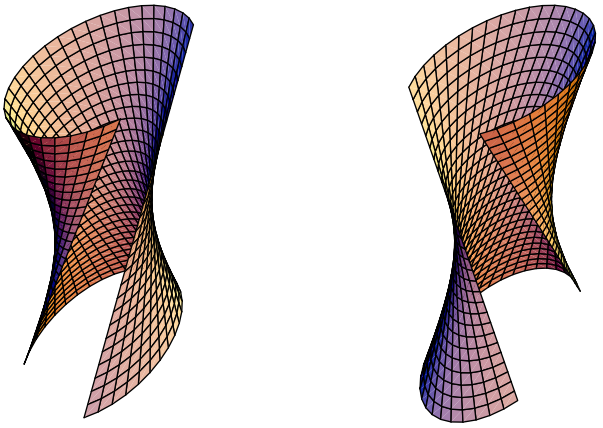


Figure 14.2: Rulings on a hyperboloid of one sheet

The Hyperbolic Paraboloid

The hyperbolic paraboloid is defined nonparametrically by

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2} \tag{14.7}$$

(though the constants a, b here are different from those in (10.16)). It is doubly ruled, since it can be parametrized in the two ways

$$\begin{aligned} \mathbf{x}^{\pm}(u,v) &= (au,\ 0,\ u^2) + v(a,\ \pm b,\ 2u) \\ &= (a(u+v),\ \pm bv,\ u^2 + 2uv). \end{aligned}$$

Although we have tacitly assumed that $b > 0$, both parametrizations can obviously be obtained from the same formula by changing the sign of b, a fact exploited in Notebook 14.

A special case corresponds to taking $a = b$ and carrying out a rotation by $\pi/2$ about the z-axis, so as to define new coordinates

$$\begin{cases} \bar{x} &= \dfrac{1}{\sqrt{2}}(x-y) \\ \bar{y} &= \dfrac{1}{\sqrt{2}}(x+y), \\ \bar{z} &= z. \end{cases}$$

This transforms (14.7) into the equation $a^2\bar{z} = 2\bar{x}\bar{y}$.

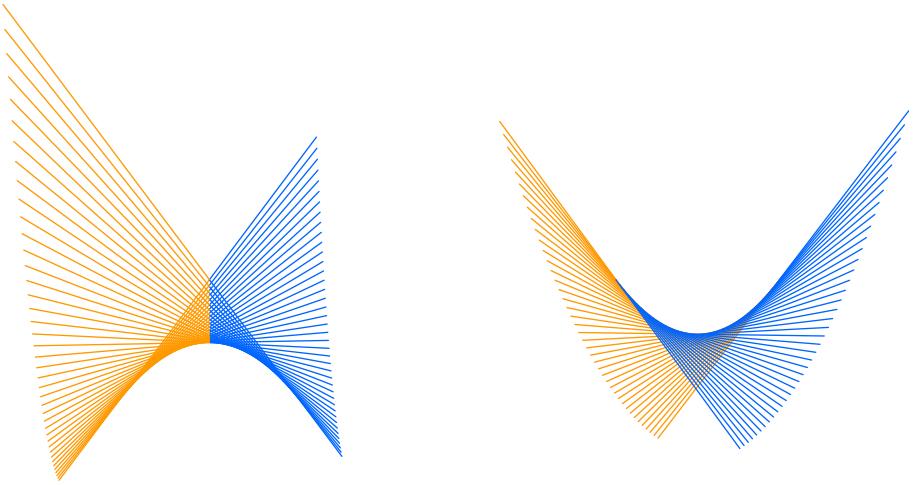


Figure 14.3: Rulings on a hyperbolic paraboloid

Plücker's Conoid

The surface defined nonparametrically by

$$z = \frac{2xy}{x^2 + y^2}$$

is called ***Plücker's conoid***[1] [BeGo, pages 352,363]. Its Monge parametrization is obviously

$$\mathsf{pluecker}(u, v) = \left(u,\ v,\ \frac{2uv}{u^2 + v^2} \right). \tag{14.8}$$

A computer plot using (14.8) does not reveal any rulings (Figure 14.4, left). To see that this conoid is in fact ruled, one needs to convert (u, v) to polar coordinates, as explained in Section 10.4. Let us write

$$\begin{aligned} \mathsf{pluecker}(r\cos\theta,\ r\sin\theta) &= (r\cos\theta,\ r\sin\theta,\ 2\cos\theta\sin\theta) \\ &= (0,\ 0,\ \sin 2\theta) + r(\cos\theta,\ \sin\theta, 0). \end{aligned}$$

Thus, the z-axis acts as base curve and the circle $\theta \mapsto (\cos\theta, \sin\theta)$ as director curve for the parametrization in terms of (r, θ). Using this parametrization, the rulings are clearly visible passing through the z-axis (Figure 14.4, right).

---

[1] 

Julius Plücker (1801–1868). German mathematician. Until 1846 Plücker's original research was in analytic geometry, but starting in 1846 as professor of physics in Bonn, he devoted his energies to experimental physics for nearly twenty years. At the end of his life he returned to mathematics, inventing line geometry.

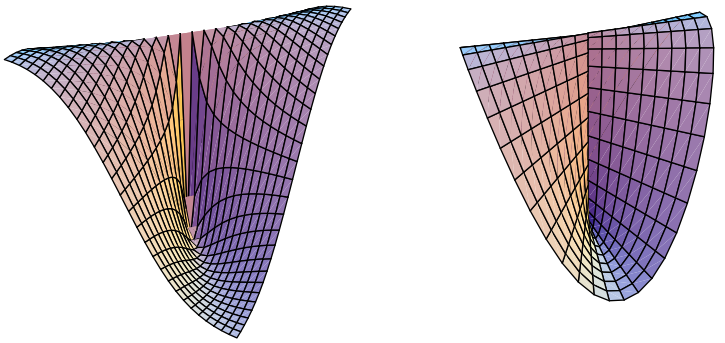


Figure 14.4: Parametrizations of the Plücker conoid

It is now an easy matter to define a generalization of Plücker's conoid that has n folds instead of 2:

$$\text{plueckerpolar}[n](r,\theta) = (r\cos\theta,\ r\sin\theta,\ \sin n\theta). \tag{14.9}$$

Each surface plueckerpolar$[n]$ is a ruled surface with the rulings passing through the z-axis. For a generalization of (14.9) that includes a variant of the monkey saddle as a special case, see Exercise 5.

Even more general than (14.9) is the ***right conoid***, which is a ruled surface with rulings parallel to a plane and passing through a line that is perpendicular to the plane. For example, if we take the plane to be the xy-plane and the line to be the z-axis, a right conoid will have the form

$$\text{rightconoid}[\vartheta,h](u,v) = \big(v\cos\vartheta(u),\ v\sin\vartheta(u),\ h(u)\big). \tag{14.10}$$

This is investigated in Notebook 14 and illustrated in Figure 14.12 on page 449.

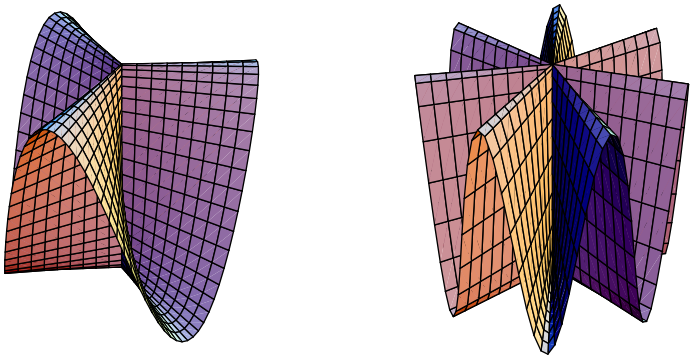


Figure 14.5: The conoids plueckerpolar$[n]$ with $n = 3$ and 7

14.2 Curvature of a Ruled Surface

Lemma 14.2. *The Gaussian curvature of a ruled surface $\mathcal{M} \subset \mathbb{R}^3$ is everywhere nonpositive.*

Proof. If $\mathbf{x}$ is a ruled patch on $\mathcal{M}$, then $\mathbf{x}_{vv} = 0$; consequently $g = 0$. Hence it follows from Theorem 13.25, page 400, that

$$K = \frac{-f^2}{EG - F^2} \leqslant 0. \quad \blacksquare \tag{14.11}$$

As an example, we consider the parametrization (14.3) of the Möbius strip. The coefficients of its first and second fundamental forms, its Gaussian and mean curvatures are all computed in Notebook 14. For example, with $a = 2$ as in Figure 14.6, we obtain

$$K(u,v) = -\frac{16}{\left(16 + 3v^2 + 16v\cos\frac{u}{2} + 2v^2\cos u\right)^2}$$

$$H(u,v) = -2\frac{\left(8 + 2v^2 + 8v\cos\frac{u}{2} + v^2\cos u\right)\sin\frac{u}{2}}{\left(16 + 3v^2 + 16v\cos\frac{u}{2} + 2v^2\cos u\right)^{3/2}}.$$

It is clear that, for this particular parametrization, K is never zero. The right-hand plot of K in Figure 14.6 exhibits its minima, which correspond to definite regions of the strip which are especially distorted.

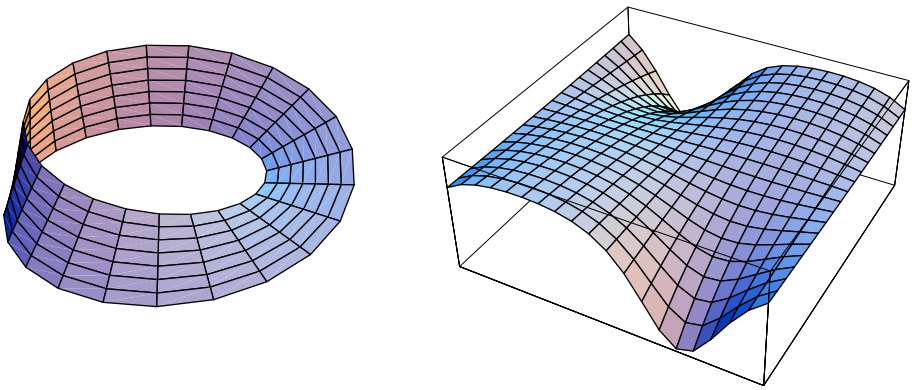


Figure 14.6: Curvature of a Möbius strip

Recall that a ***flat surface*** is a surface whose Gaussian curvature is everywhere zero. Such a surface is classically called a ***developable surface***. An immediate consequence of (14.11) is the following criterion for flatness.

Corollary 14.3. *$\mathcal{M}$ is flat if and only if $f = 0$.*

The most obvious examples of flat surfaces, other than the plane, are circular cylinders and cones (see the discussion after Figure 14.7). Paper models of both can easily be constructed by bending a sheet of paper and, as we shall see in Section 17.2, this operation leaves the Gaussian curvature unchanged and identically zero. Circular cylinders and cones are subsumed into parts (ii) and (iii) of the following list of flat ruled surfaces.

Definition 14.4. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface. Then*:

(i) *$\mathcal{M}$ is said to be the **tangent developable** of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ if $\mathcal{M}$ can be parametrized as*

$$\mathbf{x}(u,v) = \boldsymbol{\alpha}(u) + v\,\boldsymbol{\alpha}'(u); \tag{14.12}$$

(ii) *$\mathcal{M}$ is a **generalized cylinder** over a curve $\boldsymbol{\alpha}\colon (a,b) \mapsto \mathbb{R}^3$ if $\mathcal{M}$ can be parametrized as*

$$\mathbf{y}(u,v) = \boldsymbol{\alpha}(u) + v\,\mathbf{q},$$

where $\mathbf{q} \in \mathbb{R}^3$ is a fixed vector;

(iii) *$\mathcal{M}$ is a **generalized cone** over a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$, provided $\mathcal{M}$ can be parametrized as*

$$\mathbf{z}(u,v) = \mathbf{p} + v\,\boldsymbol{\alpha}(u),$$

*where $\mathbf{p} \in \mathbb{R}^3$ is fixed (it can be interpreted as the **vertex** of the cone).*

Our next result gives criteria for the regularity of these three classes of surfaces. We base it on Lemma 10.18.

Lemma 14.5. (i) *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a regular curve whose curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ is everywhere nonzero. The tangent developable $\mathbf{x}$ of $\boldsymbol{\alpha}$ is regular everywhere except along $\boldsymbol{\alpha}$.*

(ii) *A generalized cylinder $\mathbf{y}(u,v) = \boldsymbol{\alpha}(u) + v\,\mathbf{q}$ is regular wherever $\boldsymbol{\alpha}' \times \mathbf{q}$ does not vanish.*

(iii) *A generalized cone $\mathbf{z}(u,v) = \mathbf{p} + v\,\boldsymbol{\alpha}(u)$ is regular wherever $v\,\boldsymbol{\alpha} \times \boldsymbol{\alpha}'$ is nonzero, and is never regular at its vertex.*

Proof. For a tangent developable $\mathbf{x}$, we have

$$(\mathbf{x}_u \times \mathbf{x}_v)(u,v) = (\boldsymbol{\alpha}' + v\,\boldsymbol{\alpha}'') \times \boldsymbol{\alpha}' = v\,\boldsymbol{\alpha}'' \times \boldsymbol{\alpha}'. \tag{14.13}$$

If $\boldsymbol{\kappa}[\boldsymbol{\alpha}] \neq 0$, then $\boldsymbol{\alpha}'' \times \boldsymbol{\alpha}'$ is everywhere nonzero by (7.26). Thus (14.13) implies that $\mathbf{x}$ is regular whenever $v \neq 0$. The other statements have similar proofs. ∎

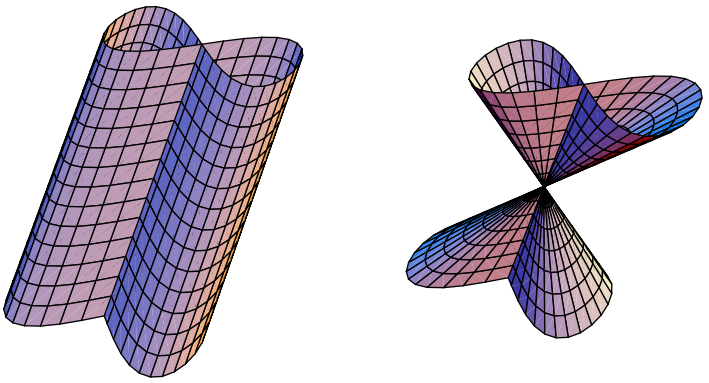


Figure 14.7: Cylinder and cone over a figure eight

Figure 14.7 illustrates cases (ii) and (iii) of Definition 14.4.

Proposition 14.6. *If $\mathcal{M}$ is a tangent developable, a generalized cylinder or a generalized cone, then $\mathcal{M}$ is flat.*

Proof. By Corollary 14.3, it suffices to show that $f = 0$ in each of the three cases. For a tangent developable $\mathbf{x}$, we have $\mathbf{x}_u = \boldsymbol{\alpha}' + v\boldsymbol{\alpha}''$, $\mathbf{x}_v = \boldsymbol{\alpha}'$ and $\mathbf{x}_{uv} = \boldsymbol{\alpha}''$; hence the triple product numerator of

$$f = \frac{[\boldsymbol{\alpha}''\ (\boldsymbol{\alpha}' + v\boldsymbol{\alpha}'')\ \boldsymbol{\alpha}']}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$$

is zero. It is obvious that $f = 0$ for a generalized cylinder $\mathbf{y}$, since $\mathbf{y}_{uv} = 0$. Finally, for a generalized cone, we compute

$$f = \frac{[\boldsymbol{\alpha}'\ v\boldsymbol{\alpha}'\ \boldsymbol{\alpha}]}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = 0. \ \blacksquare$$

The general developable surface is in some sense the union of tangent developables, generalized cylinders and generalized cones. This remark is explained in the paragraph directly after the proof of Theorem 14.15.

The ***normal surface*** and the ***binormal surface*** to a space curve $\boldsymbol{\alpha}$ can be defined by mimicking the construction of the tangent developable. It suffices to replace the tangent vector $\boldsymbol{\alpha}'(u)$ in (14.12) by the unit normal or binormal vector $\mathbf{N}(u)$ or $\mathbf{B}(u)$. However, unlike a tangent developable, the normal and binormal surfaces are not in general flat.

Consider the normal and binormal surfaces to Viviani's curve, defined in Section 7.5. Output from Notebook 14 gives the following expression for the binormal surface of viviani[1]:

$$\left(1+\cos u+\frac{v\left(3\sin\frac{u}{2}+\sin\frac{3u}{2}\right)}{\sqrt{26+6\cos u}},\ \frac{-2\sqrt{2}\,v\cos^3\frac{u}{2}}{\sqrt{13+3\cos u}}+\sin u,\ \frac{2\sqrt{2}\,v}{\sqrt{13+3\cos u}}+2\sin\frac{u}{2}\right).$$

Figure 14.8 shows parts of the normal and binormal surfaces together, with the small gap representing Viviani's curve itself. The curvature of these surfaces is computed in Notebook 14, and shown to be nonzero.

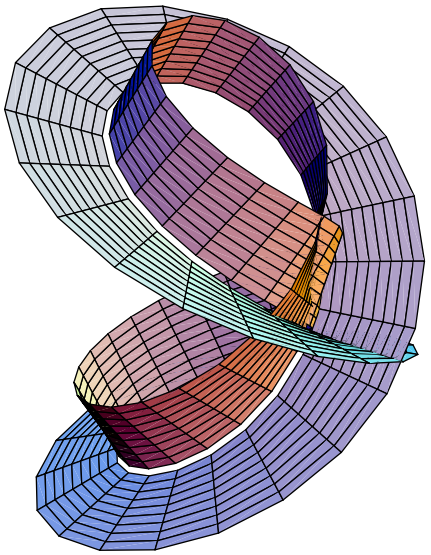


Figure 14.8: Normal and binormal surfaces to Viviani's curve

14.3 Tangent Developables

Consider the surface

$$\mathsf{tandev}[\boldsymbol{\alpha}] = \boldsymbol{\alpha}(u) + v\,\boldsymbol{\alpha}'(u);$$

this was first defined in (14.12), though we now use notation from Notebook 14. We know from Lemma 14.5 that $\mathsf{tandev}[\boldsymbol{\alpha}]$ is singular along the curve $\boldsymbol{\alpha}$. We prove next that it is made up of two sheets which meet along the trace of $\boldsymbol{\alpha}$ in a sharp edge, called the ***edge of regression***.

Theorem 14.7. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a unit-speed curve with $a < 0 < b$, and let $\mathbf{x}$ be the tangent developable of $\boldsymbol{\alpha}$. Suppose that $\boldsymbol{\alpha}$ is differentiable at 0 and that the curvature and torsion of $\boldsymbol{\alpha}$ are nonzero at 0. Then the intersection of the trace of $\mathbf{x}$ with the plane perpendicular to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(0)$ is approximated by a semicubical parabola with a cusp at $\boldsymbol{\alpha}(0)$.*

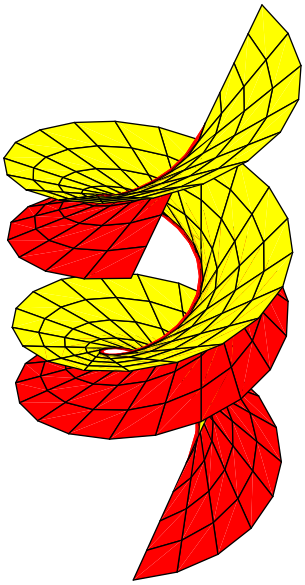


Figure 14.9: Tangent developable to a circular helix

Proof. Since $\boldsymbol{\alpha}$ has unit speed, we have $\boldsymbol{\alpha}'(s) = \mathbf{T}(s)$. The Frenet formulas (Theorem 7.10 on page 197) tell us that $\boldsymbol{\alpha}''(s) = \boldsymbol{\kappa}(s)\mathbf{N}(s)$ and moreover

$$\begin{aligned}\boldsymbol{\alpha}'''(s) &= -\boldsymbol{\kappa}(s)^2\mathbf{T}(s) + \boldsymbol{\kappa}'(s)\mathbf{N}(s) + \boldsymbol{\kappa}(s)\boldsymbol{\tau}(s)\mathbf{B}(s),\\ \boldsymbol{\alpha}''''(s) &= -3\boldsymbol{\kappa}(s)\boldsymbol{\kappa}'(s)\mathbf{T}(s) + \big(-\boldsymbol{\kappa}(s)^3 + \boldsymbol{\kappa}''(s) - \boldsymbol{\kappa}(s)\boldsymbol{\tau}(s)^2\big)\mathbf{N}(s)\\ &\qquad + \big(2\boldsymbol{\kappa}'(s)\boldsymbol{\tau}(s) + \boldsymbol{\kappa}(s)\boldsymbol{\tau}'(s)\big)\mathbf{B}(s).\end{aligned}$$

We next substitute these formulas into the power series expansion

$$\boldsymbol{\alpha}(s) = \boldsymbol{\alpha}_0 + s\boldsymbol{\alpha}_0' + \frac{s^2}{2}\boldsymbol{\alpha}_0'' + \frac{s^3}{6}\boldsymbol{\alpha}_0''' + \frac{s^4}{24}\boldsymbol{\alpha}_0'''' + O\big(s^5\big)\,,$$

where the subscript $_0$ denotes evaluation at $s = 0$. We get

$$\begin{aligned}\boldsymbol{\alpha}(s) = \boldsymbol{\alpha}_0 \;+\; &\mathbf{T}_0\left(s - \frac{s^3}{6}{\boldsymbol{\kappa}_0}^2 - \frac{s^4}{8}\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0'\right)\\ +\; &\mathbf{N}_0\left(\frac{s^2}{2}\boldsymbol{\kappa}_0 + \frac{s^3}{6}\boldsymbol{\kappa}_0' + \frac{s^4}{24}\big(-{\boldsymbol{\kappa}_0}^3 + \boldsymbol{\kappa}_0'' - \boldsymbol{\kappa}_0{\boldsymbol{\tau}_0}^2\big)\right)\\ +\; &\mathbf{B}_0\left(\frac{s^3}{6}\boldsymbol{\kappa}_0\boldsymbol{\tau}_0 + \frac{s^4}{24}\big(2\boldsymbol{\kappa}_0'\boldsymbol{\tau}_0 + \boldsymbol{\kappa}_0\boldsymbol{\tau}_0'\big)\right) + O\big(s^5\big)\,.\end{aligned} \tag{14.14}$$

This is a significant formula that gives a local expression for a space curve relative to a fixed Frenet frame; see [dC1, §1–6]. Taking the derivative of (14.14)

gives

$$\begin{aligned}\boldsymbol{\alpha}'(s) \;=\; \mathbf{T}_0\left(1-\frac{s^2}{2}{\boldsymbol{\kappa}_0}^2-\frac{s^3}{2}\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0'\right)\\ +\,\mathbf{N}_0\left(s\,\boldsymbol{\kappa}_0+\frac{s^2}{2}\boldsymbol{\kappa}_0'+\frac{s^3}{6}\big(-{\boldsymbol{\kappa}_0}^3+\boldsymbol{\kappa}_0''-\boldsymbol{\kappa}_0{\boldsymbol{\tau}_0}^2\big)\right)\\ +\,\mathbf{B}_0\left(\frac{s^2}{2}\boldsymbol{\kappa}_0\boldsymbol{\tau}_0+\frac{s^3}{6}\big(2\boldsymbol{\kappa}_0'\boldsymbol{\tau}_0+\boldsymbol{\kappa}_0\boldsymbol{\tau}_0'\big)\right)+O\big(s^4\big)\,.\end{aligned}$$

Combining this with (14.14), we conclude that the tangent developable of $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}\mathbf{x}(u,v) \;&=\; \boldsymbol{\alpha}(u)+v\,\boldsymbol{\alpha}'(u)\\ &=\; \boldsymbol{\alpha}_0+\mathbf{T}_0\left(u-\frac{u^3}{6}{\boldsymbol{\kappa}_0}^2+v-\frac{u^2v}{2}{\boldsymbol{\kappa}_0}^2-\frac{u^3v}{2}\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0'\right)\\ &+\mathbf{N}_0\left(\frac{u^2}{2}\boldsymbol{\kappa}_0+\frac{u^3}{6}\boldsymbol{\kappa}_0'+u\,v\,\boldsymbol{\kappa}_0+\frac{u^2v}{2}\boldsymbol{\kappa}_0'+\frac{u^3v}{6}\big(-{\boldsymbol{\kappa}_0}^3+\boldsymbol{\kappa}_0''-\boldsymbol{\kappa}_0{\boldsymbol{\tau}_0}^2\big)\right)\\ &+\mathbf{B}_0\left(\Big(\frac{u^3}{6}+\frac{u^2v}{2}\Big)\boldsymbol{\kappa}_0\boldsymbol{\tau}_0+\frac{u^3v}{6}\big(2\boldsymbol{\kappa}_0'\boldsymbol{\tau}_0+\boldsymbol{\kappa}_0\boldsymbol{\tau}_0'\big)\right)+O\big(u^4\big)\,.\end{aligned}$$

We want to determine the intersection of $\mathbf{x}$ with the plane perpendicular to $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}_0$. Therefore, we set the above coefficient of $\mathbf{T}_0$ equal to zero and solve for v to give

$$v \;=\; -\frac{u-\frac{1}{6}u^3{\boldsymbol{\kappa}_0}^2+O\big(u^4\big)}{1-\frac{1}{2}u^2{\boldsymbol{\kappa}_0}^2-\frac{1}{2}u^3\boldsymbol{\kappa}_0\boldsymbol{\kappa}_0'+O\big(u^4\big)} \;=\; -u-{\boldsymbol{\kappa}_0}^2\frac{u^3}{3}+O\big(u^4\big)\,.$$

Substituting back this value yields the following power series expansions:

$$\left\{\begin{array}{l} x \;=\; \text{coefficient of } \mathbf{N}_0 \;=\; -\dfrac{u^2}{2}\boldsymbol{\kappa}_0-\cdots\\[2ex] y \;=\; \text{coefficient of } \mathbf{B}_0 \;=\; -\dfrac{u^3}{3}\boldsymbol{\kappa}_0\boldsymbol{\tau}_0+\cdots\end{array}\right. \tag{14.15}$$

By hypothesis, $\boldsymbol{\kappa}_0\neq 0\neq\boldsymbol{\tau}_0$, so that we can legitimately ignore higher order terms. Doing this, the plane curve described by (14.15) is approximated by the implicit equation

$$8{\boldsymbol{\tau}_0}^2\,x^3+9\boldsymbol{\kappa}_0\,y^2=0,$$

which is a semicubical parabola (see page 25). ▮

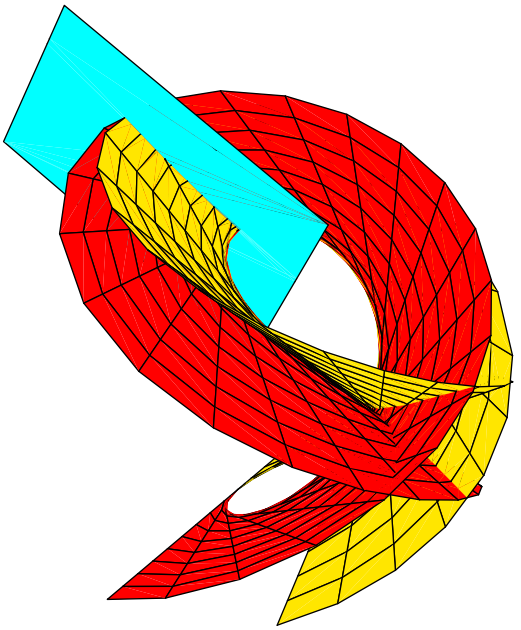


Figure 14.10: Tangent developable to Viviani's curve

Figures 14.10 illustrates a case where the hypotheses of Theorem 14.7 are not valid. It shows the plane generated by the normal $\mathbf{N}$ and binormal $\mathbf{B}$ unit vectors to Viviani's curve

$$\boldsymbol{\alpha}(t) = \Big(1 - \cos t,\ -\sin t,\ 2\cos\frac{t}{2}\Big), \qquad -3\pi \leqslant t \leqslant \pi \tag{14.16}$$

at $\boldsymbol{\alpha}(0) = (0, 0, 2)$. We have reparametrized the curve (7.35) on page 207 so that the point at which the torsion vanishes has parameter value 0 rather than π (see Figure 7.7). We can find the intersection of this plane with the tangent developable explicitly, using the following general considerations.

We seek points $\mathbf{x}(u, v) = \boldsymbol{\alpha}(u) + v\boldsymbol{\alpha}'(u)$ on the tangent developable surface satisfying the orthogonality condition

$$(\mathbf{x}(u, v) - \boldsymbol{\alpha}(0)) \cdot \boldsymbol{\alpha}'(0) = 0.$$

The resulting equation is linear in v, and has solution

$$v = v(u) = \frac{(\boldsymbol{\alpha}(0) - \boldsymbol{\alpha}(u)) \cdot \boldsymbol{\alpha}'(0)}{\boldsymbol{\alpha}'(u) \cdot \boldsymbol{\alpha}'(0)}.$$

The cuspidal section is now represented by the plane curve (x, y) where

$$\begin{cases} x(u) = (\mathbf{x}(u, v(u)) - \boldsymbol{\alpha}(0)) \cdot \mathbf{N}, \\ y(u) = (\mathbf{x}(u, v(u)) - \boldsymbol{\alpha}(0)) \cdot \mathbf{B}. \end{cases}$$

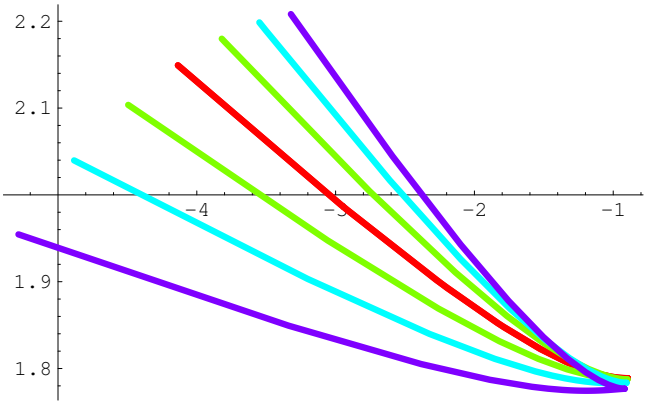


Figure 14.11: Viviani cusps for $t = 0.3$, 0.2, 0.1 and (in center) $t = 0$

Carrying out the calculations for (14.16) in Notebook 14 at $t = 0$ shows that (x, y) equals

$$\frac{1}{2\sqrt{5}\cos u}\left(-4 + 4\cos u - 3\cos\frac{u}{2} - \cos\frac{3u}{2},\ -2 + 2\cos u + 6\cos\frac{u}{2} + 2\cos\frac{3u}{2}\right).$$

It follows that $(x(-u),\, y(-u)) = (x(u),\, y(u))$, and so the two branches of the cusp coincide. The situation is illustrated by Figure 14.11 for neighboring values of t. For a fuller discussion of the two sheets that form the tangent developable, see [Spivak, vol 3, pages 207–213] and [Stru2, pages 66–73].

The following lemma is easy to prove.

Lemma 14.8. *Let $\boldsymbol{\beta}\colon (c, d) \to \mathbb{R}^3$ be a unit-speed curve. The metric of the tangent developable* tandev$[\boldsymbol{\beta}]$ *depends only on the curvature of $\boldsymbol{\beta}$. Explicitly:*

$$E = 1 + v^2\boldsymbol{\kappa}[\boldsymbol{\beta}]^2, \qquad F = 1, \qquad G = 1.$$

The helix on page 200 has the same constant curvature as a circle of radius $(a^2 + b^2)/a$. Lemmas 12.7 and 14.8 can then be used to provide a local isometry between a portion of the plane (the tangent developable to the circle) and the tangent developable to the helix.

14.4 Noncylindrical Ruled Surfaces

We shall now impose the following hypothesis on a class of ruled surfaces to be studied further.

Definition 14.9. *A ruled surface parametrized by $\mathbf{x}(u, v) = \boldsymbol{\beta}(u) + v\boldsymbol{\gamma}(u)$ is said to be* ***noncylindrical*** *provided $\boldsymbol{\gamma} \times \boldsymbol{\gamma}'$ never vanishes.*

The rulings are always changing directions on a noncylindrical ruled surface. We show how to find a useful reference curve on a noncylindrical ruled surface. This curve, called a striction curve, is a generalization of the edge of regression of a tangent developable.

Lemma 14.10. *Let $\widetilde{\mathbf{x}}$ be a parametrization of a noncylindrical ruled surface of the form $\widetilde{\mathbf{x}}(u,v) = \boldsymbol{\beta}(u) + v\boldsymbol{\gamma}(u)$. Then $\widetilde{\mathbf{x}}$ has a reparametrization of the form*

$$\mathbf{x}(u,v) = \boldsymbol{\sigma}(u) + v\,\boldsymbol{\delta}(u), \tag{14.17}$$

where $\|\boldsymbol{\delta}\| = 1$ and $\boldsymbol{\sigma}' \cdot \boldsymbol{\delta}' = 0$. The curve $\boldsymbol{\sigma}$ is called the ***striction curve*** *of $\widetilde{\mathbf{x}}$.*

Proof. Since $\boldsymbol{\gamma}\times\boldsymbol{\gamma}'$ is never zero, $\boldsymbol{\gamma}$ is never zero. We define a reparametrization $\widetilde{\widetilde{\mathbf{x}}}$ of $\widetilde{\mathbf{x}}$ by

$$\widetilde{\widetilde{\mathbf{x}}}(u,v) = \widetilde{\mathbf{x}}\Big(u,\ \frac{v}{\|\boldsymbol{\gamma}(u)\|}\Big) = \boldsymbol{\beta}(u) + \frac{v\boldsymbol{\gamma}(u)}{\|\boldsymbol{\gamma}(u)\|}.$$

Clearly, $\widetilde{\widetilde{\mathbf{x}}}$ has the same trace as $\widetilde{\mathbf{x}}$. If we put $\boldsymbol{\delta}(u) = \boldsymbol{\gamma}(u)/\|\boldsymbol{\gamma}(u)\|$, then

$$\widetilde{\widetilde{\mathbf{x}}}(u,v) = \boldsymbol{\beta}(u) + v\,\boldsymbol{\delta}(u).$$

Furthermore, $\|\boldsymbol{\delta}(u)\| = 1$ and so $\boldsymbol{\delta}(u)\cdot\boldsymbol{\delta}'(u) = 0$.

Next, we need to find a curve $\boldsymbol{\sigma}$ such that $\boldsymbol{\sigma}'(u)\cdot\boldsymbol{\delta}'(u) = 0$. To this end, we write

$$\boldsymbol{\sigma}(u) = \boldsymbol{\beta}(u) + t(u)\boldsymbol{\delta}(u) \tag{14.18}$$

for some function $t = t(u)$ to be determined. We differentiate (14.18), obtaining

$$\boldsymbol{\sigma}'(u) = \boldsymbol{\beta}'(u) + t'(u)\boldsymbol{\delta}(u) + t(u)\boldsymbol{\delta}'(u).$$

Since $\boldsymbol{\delta}(u)\cdot\boldsymbol{\delta}'(u) = 0$, it follows that

$$\boldsymbol{\sigma}'(u)\cdot\boldsymbol{\delta}'(u) = \boldsymbol{\beta}'(u)\cdot\boldsymbol{\delta}'(u) + t(u)\boldsymbol{\delta}'(u)\cdot\boldsymbol{\delta}'(u).$$

Since $\boldsymbol{\gamma}\times\boldsymbol{\gamma}'$ never vanishes, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ are always linearly independent, and consequently $\boldsymbol{\delta}'$ never vanishes. Thus if we define t by

$$t(u) = -\frac{\boldsymbol{\beta}'(u)\cdot\boldsymbol{\delta}'(u)}{\|\boldsymbol{\delta}'(u)\|^2}, \tag{14.19}$$

we get $\boldsymbol{\sigma}'(u)\cdot\boldsymbol{\delta}'(u) = 0$. Now define

$$\mathbf{x}(u,v) = \widetilde{\widetilde{\mathbf{x}}}(u,\ t(u)+v).$$

Then $\mathbf{x}(u,v) = \boldsymbol{\beta}(u) + (t(u)+v)\boldsymbol{\delta}(u) = \boldsymbol{\sigma}(u) + v\,\boldsymbol{\delta}(u)$, so that $\mathbf{x}$, $\widetilde{\mathbf{x}}$ and $\widetilde{\widetilde{\mathbf{x}}}$ all have the same trace, and $\mathbf{x}$ satisfies (14.17). ∎

Lemma 14.11. *The striction curve of a noncylindrical ruled surface* $\mathbf{x}$ *does not depend on the choice of base curve.*

Proof. Let $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\beta}}$ be two base curves for $\mathbf{x}$. In the notation of the previous proof, we may write

$$\boldsymbol{\beta}(u) + v\,\boldsymbol{\delta}(u) = \widetilde{\boldsymbol{\beta}}(u) + w(v)\boldsymbol{\delta}(u) \tag{14.20}$$

for some function $w = w(v)$. Let $\boldsymbol{\sigma}$ and $\widetilde{\boldsymbol{\sigma}}$ be the corresponding striction curves. Then

$$\boldsymbol{\sigma}(u) = \boldsymbol{\beta}(u) - \frac{\boldsymbol{\beta}'(u)\cdot\boldsymbol{\delta}'(u)}{\|\boldsymbol{\delta}'(u)\|^2}\boldsymbol{\delta}(u)$$

and

$$\widetilde{\boldsymbol{\sigma}}(u) = \widetilde{\boldsymbol{\beta}}(u) - \frac{\widetilde{\boldsymbol{\beta}}'(u)\cdot\boldsymbol{\delta}'(u)}{\|\boldsymbol{\delta}'(u)\|^2}\boldsymbol{\delta}(u),$$

so that

$$\boldsymbol{\sigma} - \widetilde{\boldsymbol{\sigma}} = \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} - \frac{(\boldsymbol{\beta}' - \widetilde{\boldsymbol{\beta}}')\cdot\boldsymbol{\delta}'}{\|\boldsymbol{\delta}'\|^2}\boldsymbol{\delta}. \tag{14.21}$$

On the other hand, it follows from (14.20) that

$$\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} = (w(v) - v)\boldsymbol{\delta}. \tag{14.22}$$

The result follows by substituting (14.22) and its derivative into (14.21).

There is a nice geometric interpretation of the striction curve $\boldsymbol{\sigma}$ of a ruled surface $\mathbf{x}$, which we mention without proof. Let $\varepsilon > 0$ be small. Since nearby rulings are not parallel to each other, there is a unique point $P(\varepsilon)$ on the straight line $v \mapsto \mathbf{x}(u, v)$ that is closest to the line $v \mapsto \mathbf{x}(u+\varepsilon,\, v)$. Then $P(\varepsilon) \to \boldsymbol{\sigma}(u)$ as $\varepsilon \to 0$. This follows because (14.19) is the equation that results from minimizing $\|\boldsymbol{\sigma}'\|$ to first order.

Definition 14.12. *Let* $\mathbf{x}$ *be a noncylindrical ruled surface given by* (14.17). *Then the* ***distribution parameter*** *of* $\mathbf{x}$ *is the function* $p = p(u)$ *defined by*

$$p = \frac{[\boldsymbol{\sigma}'\ \boldsymbol{\delta}\ \boldsymbol{\delta}']}{\boldsymbol{\delta}'\cdot\boldsymbol{\delta}'}. \tag{14.23}$$

Whilst the definition of $\boldsymbol{\sigma}$ requires $\boldsymbol{\sigma}'$ to be perpendicular to $\boldsymbol{\delta}'$, the function p measures the component of $\boldsymbol{\sigma}'$ perpendicular to $\boldsymbol{\delta} \times \boldsymbol{\delta}'$.

Lemma 14.13. *Let* $\mathcal{M}$ *be a noncylindrical ruled surface, parametrized by a patch* $\mathbf{x}$ *of the form* (14.17). *Then* $\mathbf{x}$ *is regular whenever* $v \neq 0$, *or when* $v = 0$ *and* $p(u) \neq 0$. *Furthermore, the Gaussian curvature of* $\mathbf{x}$ *is given in terms of its distribution parameter by*

$$K = \frac{-p(u)^2}{\big(p(u)^2 + v^2\big)^2}. \tag{14.24}$$

Also,

$$E = \|\boldsymbol{\sigma}'\|^2 + v^2\|\boldsymbol{\delta}'\|^2, \qquad F = \boldsymbol{\sigma}' \cdot \boldsymbol{\delta}, \qquad G = 1,$$
$$EG - F^2 = (p^2 + v^2)\|\boldsymbol{\delta}'\|^2,$$

and

$$g = 0, \qquad f = \frac{p\|\boldsymbol{\delta}'\|}{\sqrt{p^2 + v^2}}.$$

Proof. First, we observe that both $\boldsymbol{\sigma}' \times \boldsymbol{\delta}$ and $\boldsymbol{\delta}'$ are perpendicular to both $\boldsymbol{\delta}$ and $\boldsymbol{\sigma}'$. Therefore, $\boldsymbol{\sigma}' \times \boldsymbol{\delta}$ must be a multiple of $\boldsymbol{\delta}'$, and

$$\boldsymbol{\sigma}' \times \boldsymbol{\delta} = p\boldsymbol{\delta}' \qquad \text{where} \qquad p = \frac{[\boldsymbol{\sigma}'\ \boldsymbol{\delta}\ \boldsymbol{\delta}']}{\boldsymbol{\delta}' \cdot \boldsymbol{\delta}'}.$$

Since $\mathbf{x}_u = \boldsymbol{\sigma}' + v\boldsymbol{\delta}'$ and $\mathbf{x}_v = \boldsymbol{\delta}$, we have

$$\mathbf{x}_u \times \mathbf{x}_v = p\boldsymbol{\delta}' + v\boldsymbol{\delta}' \times \boldsymbol{\delta},$$

so that

$$\|\mathbf{x}_u \times \mathbf{x}_v\|^2 = \|p\boldsymbol{\delta}'\|^2 + \|v\boldsymbol{\delta}' \times \boldsymbol{\delta}\|^2 = (p^2 + v^2)\|\boldsymbol{\delta}'\|^2.$$

It is now clear that the regularity of $\mathbf{x}$ is as stated.

Next, $\mathbf{x}_{uv} = \boldsymbol{\delta}'$ and $\mathbf{x}_{vv} = 0$, so that $g = 0$ and

$$f = \frac{[\mathbf{x}_{uv}\ \mathbf{x}_u\ \mathbf{x}_v]}{\|\mathbf{x}_u \times \mathbf{x}_v\|} = \frac{\boldsymbol{\delta}' \cdot (p\boldsymbol{\delta}' + v\boldsymbol{\delta}' \times \boldsymbol{\delta})}{\sqrt{p^2 + v^2}\|\boldsymbol{\delta}'\|} = \frac{p\|\boldsymbol{\delta}'\|}{\sqrt{p^2 + v^2}}.$$

Therefore,

$$K = \frac{-f^2}{\|\mathbf{x}_u \times \mathbf{x}_v\|^2} = \frac{-\left(\dfrac{p\|\boldsymbol{\delta}'\|}{\sqrt{p^2 + v^2}}\right)^2}{(p^2 + v^2)\|\boldsymbol{\delta}'\|^2},$$

which simplifies into (14.24). ∎

Equation (14.24) tells us that the Gaussian curvature of a noncylindrical ruled surface is generally negative. However, more can be said.

Corollary 14.14. *Let $\mathcal{M}$ be a noncylindrical ruled surface given by* (14.17) *with distribution parameter p, and Gaussian curvature $K(u, v)$.*

(i) *Along a ruling (so u is fixed), $K(u, v) \to 0$ as $v \to \infty$.*

(ii) *$K(u, v) = 0$ if and only if $p(u) = 0$.*

(iii) *If p never vanishes, then $K(u, v)$ is continuous and $|K(u, v)|$ assumes its maximum value $1/p^2$ at $v = 0$.*

Proof. All of these statements follow from (14.24). ∎

Next, we prove a partial converse of Proposition 14.6.

Theorem 14.15. *Let* $\mathbf{x}(u,v) = \boldsymbol{\beta}(u) + v\,\boldsymbol{\delta}(u)$ *with* $\|\boldsymbol{\delta}(u)\| = 1$ *parametrize a flat ruled surface* $\mathcal{M}$.

(i) *If* $\boldsymbol{\beta}'(u) \equiv 0$, *then* $\mathcal{M}$ *is a cone.*

(ii) *If* $\boldsymbol{\delta}'(u) \equiv 0$, *then* $\mathcal{M}$ *is a cylinder.*

(iii) *If both* $\boldsymbol{\beta}'$ *and* $\boldsymbol{\delta}'$ *never vanish, then* $\mathcal{M}$ *is the tangent developable of its striction curve.*

Proof. Parts (i) and (ii) are immediate from the definitions, so it suffices to prove (iii). We can assume that $\boldsymbol{\beta}$ is a unit-speed striction curve, so that

$$\boldsymbol{\beta}' \cdot \boldsymbol{\delta}' \equiv 0. \tag{14.25}$$

Since $K \equiv 0$, it follows from (14.24) and (14.23) that

$$[\boldsymbol{\beta}'\ \boldsymbol{\delta}\ \boldsymbol{\delta}'] \equiv 0. \tag{14.26}$$

Then (14.25) and (14.26) imply that $\boldsymbol{\beta}'$ and $\boldsymbol{\delta}$ are collinear. ∎

Of course, cases (i), (ii) and (iii) of Theorem 14.15 do not exhaust all of the possibilities. If there is a clustering of the zeros $\boldsymbol{\beta}$ or $\boldsymbol{\delta}$, the surface can be complicated. In any case, away from the cluster points a developable surface is the union of pieces of cylinders, cones and tangent developables. Indeed, the following result is proved in [Krey1, page 185]. Every flat ruled patch $(u,v) \mapsto \mathbf{x}(u,v)$ can be subdivided into sufficiently small u-intervals so that the portion of the surface corresponding to each interval is a subset of one of the following: a plane, a cylinder, a cone, a tangent developable.

Examples of Striction Curves

The parametrization (12.27) of the circular helicoid can be rewritten as

$$\mathbf{x}(u,v) = (0,0,b\,u) + a\,v(\cos u, \sin u, 0),$$

which shows that it is a ruled surface. The striction curve $\boldsymbol{\sigma}$ and director curve $\boldsymbol{\delta}$ are given by

$$\boldsymbol{\sigma}(u) = (0,\ 0,\ b\,u) \qquad \text{and} \qquad \boldsymbol{\delta}(u) = (\cos u,\ \sin u,\ 0).$$

Then

$$\mathbf{x}\Big(u,\ \frac{v}{a}\Big) = \boldsymbol{\sigma}(u) + v\,\boldsymbol{\delta}(u) = (v\cos u,\ v\sin u,\ b\,u),$$

and the distribution parameter assumes the constant value b.

The hyperbolic paraboloid (10.12), page 296, when parametrized as

$$\mathbf{x}(u, v) = (u, 0, 0) + v(0, 1, u),$$

has $\boldsymbol{\sigma}(u) = (u, 0, 0)$ as its striction curve and

$$\boldsymbol{\delta}(u) = \frac{(0, 1, u)}{\sqrt{1 + u^2}}$$

as its director curve. Thus

$$\mathbf{x}(u,\ v\sqrt{1 + u^2}) = \boldsymbol{\sigma}(u) + v\,\boldsymbol{\delta}(u),$$

and the distribution parameter is given by $p(u) = 1 + u^2$.

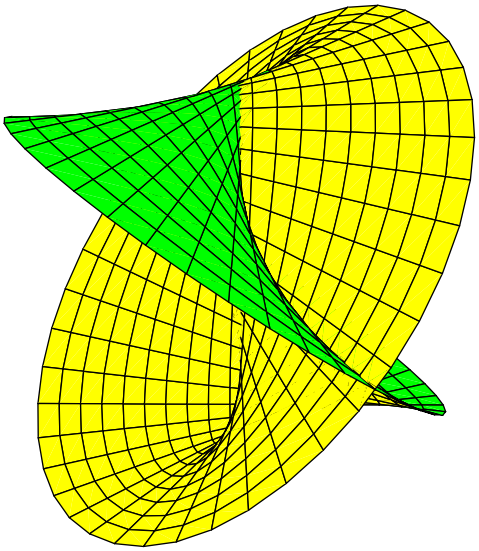


Figure 14.12: The right conoid

14.5 Exercises

1. Compute the Gaussian and mean curvatures of the generalized hyperbolic paraboloid defined and plotted on page 434.

M **2.** Compute the Gaussian and mean curvatures of the right conoid defined by equation (14.10) and illustrated in Figure 14.12.

3. Explain why Lemma 14.2 is also a consequence of the fact (mentioned on page 432) that a ruled surface has asymptotic curves, namely, the rulings.

M **4.** Further to Exercise 2 of the previous chapter, describe the sets of elliptic, hyperbolic, parabolic, and planar points for the following surfaces:

(a) a cylinder over an ellipse,
(b) a cylinder over a parabola,
(c) a cylinder over a hyperbola,
(d) a cylinder over $y = x^3$,
(e) a cone over a circle.

M **5.** Compute the Gaussian and mean curvature of the patch

$$\mathsf{plueckerpolar}[m, n, a](r, \theta) = a\big(r\cos\theta,\ r\sin\theta,\ r^m \sin n\theta\big)$$

that generalizes plueckerpolar on page 436. Relate the case $m = n = 3$ to the monkey saddle on page 304.

6. Complete the proof of Lemma 14.5 and prove Lemma 14.8.

M **7.** ***Wallis's conical edge***[2] is defined by

$$\mathsf{wallis}[a, b, c](u, v) = \Big(v\cos u,\ v\sin u,\ c\sqrt{a^2 - b^2\cos^2 u}\Big).$$

Show that wallis$[a, b, c]$ is a right conoid, as in (14.10). Figure 14.13 illustrates the case $a = 1 - c$ and $b = \sqrt{3}$. Compute and plot its Gaussian and mean curvatures.

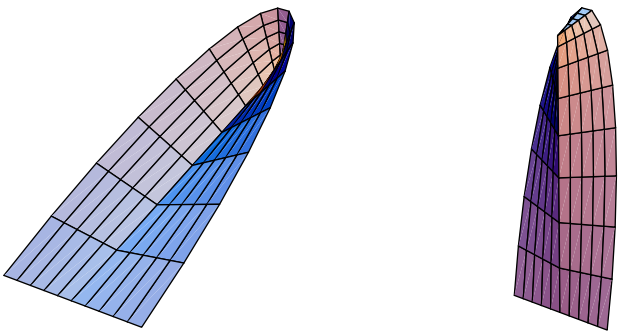


Figure 14.13: Back and front of Wallis's conical edge



2 John Wallis (1616–1703). English mathematician. Although ordained as a minister, Wallis was appointed Savilian professor of geometry at Oxford in 1649. He was one of the first to use Cartesian methods to study conic sections instead of employing the traditional synthetic approach. The sign ∞ for infinity (probably adapted from the late Roman symbol for 1000) was first introduced by Wallis.

8. Show that the parametrization $\mathsf{exptwist}[a, c]$ of the exponentially twisted helicoid described on page 563 of Chapter 18 is a ruled patch. Find the rulings.

M **9.** Plot the tangent developable to the twisted cubic defined on page 202.

M **10.** Plot the tangent developable to the Viviani curve defined on page 207 for the complete range $-2\pi \leqslant t \leqslant 2\pi$ (see Figure 14.10).

M **11.** Plot the tangent developable to the bicylinder defined on page 214.

M **12.** Carry out the calculations of Theorem 14.7 by computer.

M **13.** Show that the normal surface to a circular helix is a helicoid. Draw the binormal surface to a circular helix and compute its Gaussian curvature.

M **14.** For any space curve $\boldsymbol{\alpha}$, there are other surfaces which lie between the normal surface and the binormal surface. Consider

$$\mathsf{perpsurf}[\phi, \boldsymbol{\alpha}](u, v) = \boldsymbol{\alpha}(u) + v \cos\phi \mathbf{N}(u) + \sin\phi \mathbf{B}(u),$$

where $\mathbf{N}, \mathbf{B}$ are the normal and binormal vector fields to $\boldsymbol{\alpha}$. Clearly, $\mathsf{perpsurf}[0, \boldsymbol{\alpha}]$ is the normal surface of $\boldsymbol{\alpha}$, whilst $\mathsf{perpsurf}[\frac{\pi}{2}, \boldsymbol{\alpha}]$ is its binormal surface. Use perpsurf to construct several of these intermediate surfaces for a helix.

15. Determine the striction curve $\boldsymbol{\sigma}$ of an elliptical hyperboloid of one sheet with rulings defined by the patch $\mathbf{x}^+$ in (14.6). Is $\boldsymbol{\sigma}$ a coordinate curve of the patch $\mathbf{x}^+$?

Notebook 14

N14.0

This notebook is devoted to the modeling of surfaces formed from straight lines defined in one way or another by a space curve.

We single out two particular space curves, defined in Notebook 7, which will be useful for constructing ruled surfaces.

```
helix[a_,b_][t_]:= {a Cos[t], a Sin[t], b t}
viviani[a_][t_]:= a{1+Cos[t], Sin[t], 2Sin[t/2]}
```

The Möbius strip, studied in Chapter 11, is a ready-made example of a ruled surface that fits into the theory of Chapter 14. Here is its parametrization from N11.3:

```
moebiusstrip[a_][u_,v_]:= {a Cos[u]+v Cos[u/2]Cos[u],
  a Sin[u]+v Cos[u/2]Sin[u], v Sin[u/2]}
```

The following packages will be useful in enhancing certain figures.

```
<<Graphics`Shapes`
<<Graphics`Colors`
```

This notebook will also use the programs gcurv and mcurv of N13.5 to compute the Gaussian and mean curvature. They are loaded in N14.0.1, together with some now familiar settings and commands.

N14.0.1

N14.1

The following parametrization duplicates that of Definition 14.1, and needs little extra explanation. Suffice it to say that α is the base curve or *directrix,* and γ the so-called *director curve.*

```
ruled[α_,γ_][u_,v_]:= α[u] + v γ[u]
```

For the Möbius strip parametrized in N14.0, the director curve is given by

```
moebiusstrip[0][u,1]
```

The next program then constructs two separate representations of this director curve. The second uses the WireFrame construction from the Shapes package.

```
Module[{γ,γγ},
  γ[u_]:= moebiusstrip[0][u,1];
  γγ= highlight[γ];
  z1= ParametricPlot3D[sphere[1][u,v],
    {u,3π/8,13π/8}, {v,-π/2,π/2}, Shading->False];
  z2= ParametricPlot3D[γγ[t], {t,0,4π}];
  z3= WireFrame[{GrayLevel[.7], Sphere[1,15,15]}];
  z4= Show[{z1,z2}, ViewPoint->{2,2,2}];
  z5= Show[{z2,z3}, ViewPoint->{2,2,2}];
  S @ GraphicsArray[{z5,z4}, GraphicsSpacing->-.2]
]
```

Figure 14.1

Parametrizations of hyperboloids based on rulings are given by

```
ellipticalhyperboloidplus[a_,b_,c_][u_,v_]:=
  {a Cos[u] - a v Sin[u],  b v Cos[u] + b Sin[u], c v}
ellipticalhyperboloidminus[a_,b_,c_][u_,v_]:=
  {a Cos[u]+a v Sin[u], -b v Cos[u] + b Sin[u], c v}
```

The next computation shows that the "minus" chart could easily have been defined in terms of the "plus" one, which incidentally was already defined in N10.6.

```
Clear[a,c]
ellipticalhyperboloidplus[a,b,c][u,v]-
  ellipticalhyperboloidminus[a,b,-c][u,-v]
```

We provide abbreviations for two examples, in preparation for an examination of their curvature.

```
ehp:= ellipticalhyperboloidplus[1,2,3]
ehm:= ellipticalhyperboloidminus[1,2,3]
```

```
gcurv[ehp][u,v]
mcurv[ehp][u,v]
gcurv[ehp][u,v] - gcurv[ehm][u,-v]
mcurv[ehp][u,v] - mcurv[ehm][u,-v]
```

```
z1:= ParametricPlot3D[ehp[u,v], {u,0,3π/2}, {v,-1.5,1.5}]
z2:= ParametricPlot3D[ehm[u,v], {u,0,3π/2}, {v,-1.5,1.5}]
S @ GraphicsArray[{z1,z2}, GraphicsSpacing->.4]
```

Figure 14.2

The following chart of a hyperbolic paraboloid is more complicated than previous ones, but it will allow us to display various rulings.

```
genhypparab[a_,b_][u_,v_]:= {a(u+v), b v, u^2+2u v}
```

```
r:=1
gp:= genhypparab[1,1]
gm:= genhypparab[1,-1]
z1:= ParametricPlot3D[gp[u,v], {u,-r,r}, {v,-r,r},
  ViewPoint->{-1,3,1}]
z2:= ParametricPlot3D[gm[u,v], {u,-r,r}, {v,-r,r},
  ViewPoint->{1,3,1}]
GraphicsArray[{z1,z2}]
```

Add S@ to the last line above. Here is a special case, the subject of Figure 14.3.

```
gh:= genhypparab[3,-3]
```

```
zm[k_]:= ParametricPlot3D[gh[u,v], {v,-2,2},
  {u, k/15, (3k+1)/45}, ViewPoint->{1,-5,0}]
Show[Table[zm[k],{k,-15,15}]]
```

Straight line strips of the surface can be seen by replacing Show by S in the above cell. The result is not however very satisfactory, and it is much more effective to plot the straight lines directly:

```
Module[{y,tm,z},
  y[k_]= {Hue[.6], Line[{gh[k/10,-3], gh[k/10,0]}], Hue[.1],
    Line[{gh[k/10,0], gh[k/10,3]}]};
  tm= Graphics3D@Table[y[k], {k,-20,20}];
  z[1]= Show[tm, ViewPoint->{1.5,0,0}, Boxed->False];
  z[2]= Show[tm, ViewPoint->{0,9,0}, Boxed->False];
  S @ GraphicsArray[Array[z,2]]
]
```

Figure 14.3

Plücker's conoid is defined by the Monge patch

```
pluecker[u_,v_]:= {u, v, 2 u v/(u^2+v^2)}
```

In polar coordinates, it becomes

```
Clear[r,θ]
pl[u_,v_]:= Simplify[polarize[pluecker][u,v]]
pl[r,θ]
```

The relatively simple instructions for plotting Figures 14.4 and 14.5 are hidden below.

■ N14.1.1

■ N14.2

To study the Möbius strip, we first abbreviate its curvature functions, to be computed using the programs loaded in N14.0.1.

```
mss:= moebiusstrip[2]
K:= gcurv[mss]
H:= mcurv[mss]
```

Output from the next cell was extracted for insertion into the text of Section 14.2:

```
K[u,v]
```

```
H[u,v]
```

```
a:= .6
z1:= ParametricPlot3D[mss[u,v], {u,0,2π}, {v,-a,a},
  PlotPoints->{24,8}]
z2:= Plot3D[K[u,v], {u,π,5π}, {v,-a,a}, Ticks->None,
  PlotPoints->20]
S @ GraphicsArray[{z1,z2}, GraphicsSpacing->0]
```

Figure 14.6

The following command provides the image of a function α with one more coordinate by inserting the quantity a as a new nth argument:

```
inscurve[n_,a_,α_][t_]:= Insert[α[t], a, n]
```

As an example, we convert the plane figure eight curve (from N4.3) into a space curve by merely setting its z-coordinate equal to zero:

```
eight[t_]:= {Sin[t], Sin[t]Cos[t]}
inscurve[3,0,eight][t]
```

Generalized cylinders and cones are defined by

```
cylinder[d_, γ_][u_,v_]:= γ[u] + v d
cone[p_, γ_][u_,v_]:= (1-v)p + v γ[u]
```

The above commands are combined to construct surfaces based on a figure eight curve.

```
z1:= ParametricPlot3D[cylinder[{0,1,1}, inscurve[3,0,eight]][u,v],
  {u,0,2π}, {v,0,2}, PlotPoints->{40,15}]
z2:= ParametricPlot3D[cone[{0,0,0},
  inscurve[3, 2, (2eight[#]+{1,1}/2)&]][u,v], {u,0,2π}, {v,-3,3},
  PlotPoints->{40,15}]
S @ GraphicsArray[{z1,z2}];
```

Figure 14.7

The next three definitions are taken from N7.2, and are then used immediately afterwards in the construction of normal and binormal surfaces. We did not include them in N14.0.1, as they are only used in the next few cells.

```
tangent[α_][t_]:=
Module[{A,tt}, A= α'[tt]; A/L[A] /. tt->t];
```

```
binormal[α_][t_]:= binormal[α][t]=
Module[{A,tt}, A= Cross[α'[tt], α''[tt]]; A/L[A] /. tt->t];
```

```
normal[α_][t_]:= normal[α][t]= Cross[binormal[α][t], tangent[α][t]]
```

```
normalsurf[α_][u_,v_]:= α[u] + v normal[α][u]
binormalsurf[α_][u_,v_]:= α[u] + v binormal[α][u]
```

Examples follow, including a TEX version of a parametrization of the binormal surface to Viviani's curve that is printed in Section 14.2.

```
ns:= normalsurf[viviani[1]]
ns[u,v]//Simplify
```

```
bs:= binormalsurf[viviani[1]]
bs[u,v]//Simplify
TeXForm[%]
```

The next computation confirms that the normal surface is not flat!

```
gcurv[ns][0,v]
```

```
mcurv[ns][1,v]//N
```

Figure 14.8 plots normal and binormal surface together, purposely leaving a small gap where one can imagine Viviani's curve to lie.

```
α:= viviani[1]
z1:= ParametricPlot3D[normalsurf[α][u,v], {u,-2π,2π}, {v,-1,-.05},
  PlotPoints->{40,10}]
z2:= ParametricPlot3D[binormalsurf[α][u,v], {u,-2π,2π}, {v,.05,1},
  PlotPoints->{40,10}]
S @ {z1,z2}
```

Figure 14.8

■ N14.3

The tangent developable of a space curve α is the ruled surface, whose director curve is the unit tangent vector field to α:

```
tandev[α_][u_,v_]:=
Module[{uu},
  α[uu]+v α'[uu]/L[α'[uu]]/.uu->u
]
```

```
ruled[α,α'][u,v]
tandev[α][u,v]
```

The FaceForm command enables one to specify different colors for separate sides of a surface, and the colors themselves can be conveniently chosen using the Colors package loaded in N14.0. In order for this to have an effect, it is necessary to insert the option Lighting->False. The same technique was used to define the command col in N12.0, but we do not bother to use this abbreviation in Notebook 14.

```
Module[{x,xx},
  x= tandev[helix[1,.6]];
  xx[u_,v_]= Append[x[u,v], FaceForm[Red,Yellow]];
  S @ ParametricPlot3D[xx[u,v], {u,0,4π}, {v,-3,3},
    PlotPoints->{40,10}, Lighting->False]
]
```

Figure 14.9

```
ParametricPlot3D[tandev[helix[1,0.4]][u,v], {u,0,4π}, {v,-2,2},
  PlotPoints->{40,10}]
```

The next hidden subsection contains the analysis of the tangent developable to Viviani's curve at the end of Section 14.3.

■ N14.3.1

■ N14.4

The following cells perform calculations to back up the work in Section 14.4, and are self-explanatory.

```
δ[u_]:= {0,1,u}/Sqrt[1+u^2]
σ[u_]:= {u,0,0}
Simplify[Det[{σ'[u], δ[u], δ'[u]}]/δ'[u].δ'[u]]
```

```
x:= Module[{α,γ},
  α[t_]= {t,t^2,t^3};
  γ[t_]= {1,0,0};
  x= ruled[α, γ]
]
x[u,v]
```

```
gcurv[x][u,v]
mcurv[x][u,v]
```

The next definitions are used in the construction of the right conoid, illustrated in Figure 14.12. But we first keep the construction more general; the curve β is merely a parametrization of the z-axis by means of a function h, while γ corresponds to a unit vector rotating around the axis.

```
Clear[β,γ,h]
β[u_]:= {0,0,h[u]}
γ[θ_][u_]:= {Cos[θ[u]], Sin[θ[u]], 0}
x:= ruled[β, γ[θ]]
```

It is easy to find the curvature of the resulting ruled surface:

```
gcurv[x][u,v]
mcurv[x][u,v]
```

Choosing appropriate functions h,θ gives rise to the right conoid:

```
h[u_]:= 2Sin[u]
θ[u_]:= u
xx[u_,v_]:= Append[x[u,v], FaceForm[Green,Yellow]]
S @ ParametricPlot3D[xx[u,v], {u,-π,π}, {v,-2,2},
  PlotPoints->{40,20}, PlotRange->All, Lighting->False,
  ViewPoint->{2.2,-1.5,2}];
```

Figure 14.12

■ N14.5

Here is a direct definition of Wallis's conical edge:

```
wallis[a_,b_,c_][u_,v_]:=
  {v Cos[u], v Sin[u], c Sqrt[a^2-b^2 Cos[u]^2]}
```

We can permute coordinates as follows:

```
wa:= wallis[2,2,3]
wa[u,v]
wb[u_,v_]= Reverse[RotateLeft[wa[u,v]]]
```

```
Module[{z},
  z[x_]:= ParametricPlot3D[x[u,v], {u,0,π}, {v,0,1},
    PlotPoints->{18,6}];
  S @ GraphicsArray[{z[wb], z[wa]}]
]
```

Figure 14.13

Chapter 15

Surfaces of Revolution and Constant Curvature

Surfaces of revolution form the most easily recognized class of surfaces. We know that ellipsoids and hyperboloids are surfaces of revolution provided that two of their axes are equal; this is evident from the figures on pages 312 and 314. Similarly for tori and paraboloids. Indeed, many objects from everyday life such as cans, table glasses, and furniture legs are surfaces of revolution. The process of lathing wood produces surfaces of revolution by its very nature.

We begin an analytic study of surfaces of revolution by defining *standard parametrizations* whose coordinate curves are parallels and meridians. Then, in order to provide a fresh example, we define the catenoid, an important minimal surface. We compare it with a hyperboloid of revolution, which (only) at first sight it resembles. We also display a surface obtained by rotating a plane curve with assigned curvature (Figure 15.3).

We point out in Section 15.2 that parallels and meridians are principal curves, being tangent to the directions determined by the principal curvatures. We pursue this topic analytically in Section 15.3, in which formulas are given for the Gaussian and mean curvatures of a surface of revolution. All the accompanying computations are carried out in Notebook 15. More curvature formulas are given in Section 15.4 for the generalized helicoids, which constitute a class of surfaces encompassing both helicoids and surfaces of revolution.

While the concept of a surface of revolution, like that of a ruled surface, may be understood in terms of elementary geometry, some of the most interesting surfaces of revolution are those of *constant Gaussian curvature*. We develop this theme in the second half of the chapter by exhibiting surfaces of revolution with constant positive and negative Gaussian curvature in Sections 15.5 and 15.6 respectively. Understanding their equations requires at least cursory knowledge

of the *elliptic integrals* of Legendre,[1] which we mention in passing.

We shall see that, other than the sphere, there are two types of surfaces of revolution with constant positive curvature, one shaped like an American football, and the other like a barrel. There are also three types in the negative case, the most famous of which is the pseudosphere, the surface of revolution obtained by rotating the tractrix on page 51. Their equations will be studied further in Chapter 21.

In Section 15.7, we take a glimpse at more exotic surfaces with constant Gaussian curvature. We define a flat generalized helicoid, thereby extending a discussion of flat ruled surfaces in Section 14.2. Then we illustrate the surfaces of Kuen and Dini, whose definitions will be justified mathematically in Sections 21.7 and 21.8.

15.1 Surfaces of Revolution

A surface of revolution is formed by revolving a plane curve about a line in $\mathbb{R}^3$. More precisely:

Definition 15.1. *Let Π be a plane in $\mathbb{R}^3$, let ℓ be a line in Π, and let $\mathscr{C}$ be a point set in Π. When $\mathscr{C}$ is rotated in $\mathbb{R}^3$ about ℓ, the resulting point set $\mathcal{M}$ is called the* ***surface of revolution*** *generated by $\mathscr{C}$, which is called the* ***profile curve****. The line ℓ is called the* ***axis of revolution*** *of $\mathcal{M}$.*

For convenience, we shall choose Π to be the xz-plane and ℓ to be the z-axis. We shall assume that the point set $\mathscr{C}$ has a parametrization $\boldsymbol{\alpha}\colon (a,b) \to \mathscr{C}$ that is differentiable. Write $\boldsymbol{\alpha} = (\varphi, \psi)$.

Definition 15.2. *The patch* $\mathsf{surfrev}[\boldsymbol{\alpha}]\colon (0,2\pi)\times(a,b) \to \mathbb{R}^3$ *defined by*

$$\mathsf{surfrev}[\boldsymbol{\alpha}](u,v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big) \tag{15.1}$$

is called the ***standard parametrization*** *of the surface of revolution $\mathcal{M}$.*

The patch is of necessity defined with the angle u lying in the open interval $(0,2\pi)$, which means that the standard parametrization must be combined with an analogous patch in order to cover the surface. One usually assumes $\varphi(v) > 0$ in (15.1) to ensure that the profile curve does not cross the axis of revolution,

---



[1] Adrien Marie Legendre (1752–1833). French mathematician, who made numerous contributions to number theory and to the theory of elliptic functions. In 1782, Legendre determined the attractive force for certain solids of revolution by introducing an infinite series of polynomials that are now called Legendre polynomials. In his 3-volume work *Traité des fonctions elliptiques* (1825, 1826, 1830), Legendre founded the theory of elliptic integrals.

to avoid singular points; however, we do not make this assumption initially. In particular, the curvature formulas that we derive hold in the general case.

A meridian on the earth is a great circle that passes through the north and south poles. A parallel is a circle on the earth parallel to the equator. These notions extend to an arbitrary surface of revolution.

Definition 15.3. *Let $\mathscr{C}$ be a point set in a plane $\Pi \subset \mathbb{R}^3$, and let $\mathcal{M}[\mathscr{C}]$ be the surface of revolution in $\mathbb{R}^3$ generated by revolving $\mathscr{C}$ about a line $\ell \subset \Pi$. A* ***meridian*** *on $\mathcal{M}[\mathscr{C}]$ is the intersection of $\mathcal{M}[\mathscr{C}]$ with a plane containing the axis ℓ of the surface of revolution. A* ***parallel*** *on $\mathcal{M}[\mathscr{C}]$ is the intersection of $\mathcal{M}[\mathscr{C}]$ with a plane orthogonal to ℓ.*

For a surface parametrized by (15.1), the parallels

$$u \mapsto \mathsf{surfrev}[\boldsymbol{\alpha}](u, v_0) = \big(\varphi(v_0)\cos u,\ \varphi(v_0)\sin u,\ \psi(v_0)\big) \tag{15.2}$$

and the meridians

$$v \mapsto \mathsf{surfrev}[\boldsymbol{\alpha}](u_0, v) = \big(\varphi(v)\cos u_0,\ \varphi(v)\sin u_0,\ \psi(v)\big) \tag{15.3}$$

are coordinate curves.

Many of the computer-generated plots of surfaces of revolution display (polygonal approximations of) the parallels and meridians by default. This phenomenon is evident in Figure 15.1, which shows a surface of revolution flanked by its parallels (on the left) and meridians (on the right).

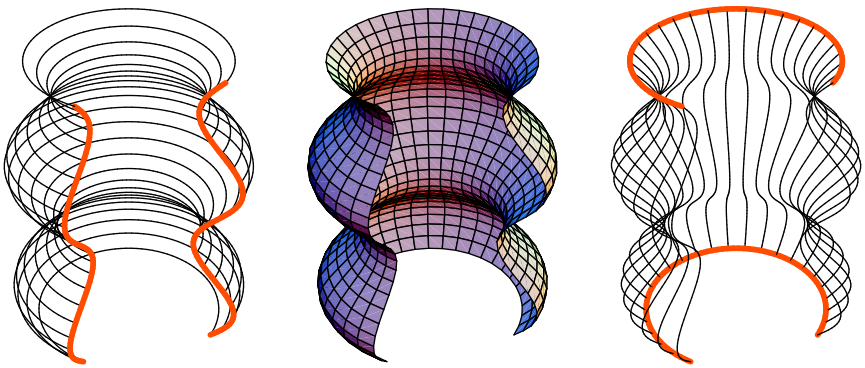


Figure 15.1: Parallels and meridians on a surface of revolution generated by the curve $t \mapsto \big(2 + \frac{1}{2}\sin 2t,\ t\big)$

The quantities $\varphi(v_0)$ and $\psi(v_0)$ in (15.2) have geometric interpretations: $|\varphi(v_0)|$ represents the radius of the parallel, whereas $\psi(v_0)$ can be interpreted as the distance (measured positively or negatively) of the center of the same parallel from the origin.

The Catenoid

We have already discussed various basic surfaces of revolution. Next, we consider the ***catenoid***, the surface of revolution generated by a catenary (see Section 2.4). The catenoid has a standard parametrization

$$\mathsf{catenoid}[c](u,v) = \Big(c\cos u\cosh\frac{v}{c},\ c\sin u\cosh\frac{v}{c},\ v\Big).$$

It is easy to compute the principal curvatures of the catenoid, with the result that

$$k_1 = -k_2 = \frac{1}{c\Big(\cosh\frac{v}{c}\Big)^2}.$$

The Gaussian curvature is therefore

$$K = \frac{-1}{c^2\Big(\cosh\frac{v}{c}\Big)^4},$$

whereas the mean curvature H vanishes. It follows that the catenoid is a minimal surface, a concept defined on page 398, and we shall prove in Section 16.3 that any surface of revolution which is also a minimal surface is contained in a catenoid or a plane.

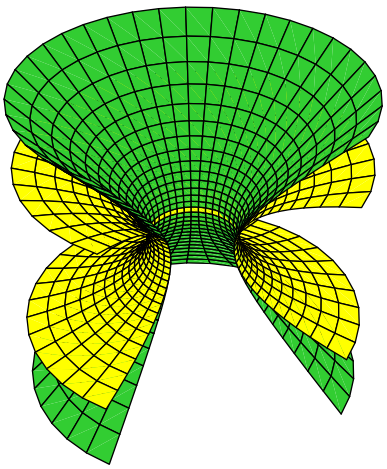


Figure 15.2: Hyperboloid of revolution and (lighter) the catenoid

The shape of a small portion of the catenoid around the equator is similar to that of the one-sheeted hyperboloid of revolution around the equator, but large portions are quite different, as are their Gauss maps. We know that the image of the Gauss map of the whole hyperboloid omits disks around the north and south

poles (see Figure 11.2 on page 336 and the discussion there). By contrast, the image of the Gauss map of the whole catenoid omits only the north and south poles. This is demonstrated in Notebook 15. There is a well-known relationship between the catenoid and the *helicoid*, discussed in Section 16.4.

Revolving Curves with Prescribed Curvature

One can also consider the surfaces formed by revolving curves with specified curvature. This provides a means of rendering "solid" many of the curves studied in Chapter 6. An example is illustrated in Figure 15.3; others are given in Notebook 15.

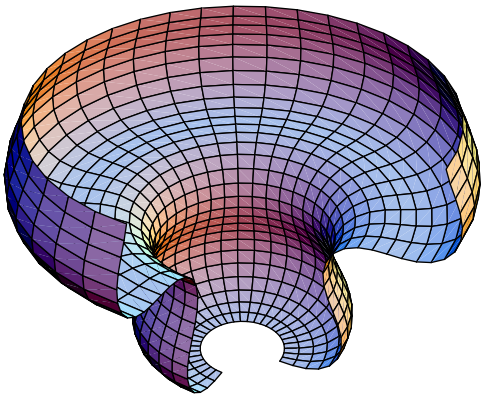


Figure 15.3: Surface of revolution whose meridian has $\boldsymbol{\kappa 2}(s) = \sin s$

15.2 Principal Curves

In this section, we take the first step in the study of curves with particular properties that lie on surfaces. Later, in Chapter 18, we shall study other classes, namely asymptotic curves and geodesics.

For simplicity, we shall deal only with orientable surfaces. For such a surface $\mathcal{M}$, we choose a globally-defined surface unit normal $\mathbf{U}$. Recall

Definition 15.4. *A curve $\boldsymbol{\alpha}$ on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is called a* ***principal curve*** *if and only if the velocity $\boldsymbol{\alpha}'$ always points in a principal direction. Thus,*

$$S(\boldsymbol{\alpha}') = k_i\, \boldsymbol{\alpha}',$$

where S denotes the shape operator of $\mathcal{M}$ with respect to $\mathbf{U}$, and k_i ($i = 1$ or 2) is a principal curvature of $\mathcal{M}$.

A useful characterization of a principal vector is provided by

Lemma 15.5. *A nonzero tangent vector* $\mathbf{v_p}$ *to a regular surface* $\mathcal{M} \subset \mathbb{R}^3$ *is principal if and only if*

$$S(\mathbf{v_p}) \times \mathbf{v_p} = \mathbf{0}.$$

Hence a curve $\boldsymbol{\alpha}$ *on* $\mathcal{M}$ *is a principal curve if and only if* $S(\boldsymbol{\alpha}') \times \boldsymbol{\alpha}' = \mathbf{0}$.

Proof. If $S(\mathbf{v_p}) = k_i \mathbf{v_p}$, then $S(\mathbf{v_p}) \times \mathbf{v_p} = k_i \mathbf{v_p} \times \mathbf{v_p} = \mathbf{0}$. Conversely, if $S(\mathbf{v_p}) \times \mathbf{v_p} = \mathbf{0}$, then $S(\mathbf{v_p})$ and $\mathbf{v_p}$ are linearly dependent. ∎

More often than not, principal curves can be found by geometrical considerations. For example, we shall construct them in Chapter 19 using triply orthogonal families of surfaces. Another example is the next theorem, due to Terquem[2] and Joachimsthal[3] [Joach1], which provides a simple but useful criterion for the intersection of two surfaces to be a principal curve on both.

Theorem 15.6. *Let* $\boldsymbol{\alpha}$ *be a curve whose trace lies in the intersection of regular surfaces* $\mathcal{M}_1, \mathcal{M}_2 \subset \mathbb{R}^3$. *Denote by* $\mathbf{U}_i$ *the unit surface normal to* $\mathcal{M}_i$, $i = 1, 2$. *Suppose that along* $\boldsymbol{\alpha}$ *the surfaces* $\mathcal{M}_1$, $\mathcal{M}_2$ *meet at a constant angle*; *that is,* $\mathbf{U}_1 \cdot \mathbf{U}_2$ *is constant along* $\boldsymbol{\alpha}$. *Then* $\boldsymbol{\alpha}$ *is a principal curve in* $\mathcal{M}_1$ *if and only if it is a principal curve in* $\mathcal{M}_2$.

The converse to this theorem is the subject of Exercise 4.

Proof. Along the curve of intersection we have

$$0 = \frac{d}{dt}(\mathbf{U}_1 \cdot \mathbf{U}_2) = \left(\frac{d}{dt}\mathbf{U}_1\right) \cdot \mathbf{U}_2 + \mathbf{U}_1 \cdot \left(\frac{d}{dt}\mathbf{U}_2\right). \tag{15.4}$$

Suppose that the curve of intersection $\boldsymbol{\alpha}$ is a principal curve in $\mathcal{M}_1$. Then

$$\frac{d}{dt}\mathbf{U}_1 = -k_1 \boldsymbol{\alpha}', \tag{15.5}$$

where k_1 is a principal curvature on $\mathcal{M}_1$. But $\boldsymbol{\alpha}'$ is also orthogonal to $\mathbf{U}_2$, so we conclude from (15.4) and (15.5) that

$$\mathbf{U}_1 \cdot \left(\frac{d}{dt}\mathbf{U}_2\right) = 0. \tag{15.6}$$

Since $d\mathbf{U}_2/dt$ is also perpendicular to $\mathbf{U}_2$, it follows from (15.6) that

$$\frac{d}{dt}\mathbf{U}_2 = -k_2 \boldsymbol{\alpha}'$$

for some k_2. In other words, $\boldsymbol{\alpha}$ is a principal curve in $\mathcal{M}_2$. ∎

---

[2]Olry Terquem (1782–1862). French mathematician and religious correspondent. Known for his study of the nine point circle of a given triangle, and also for suggesting radical reforms aimed at impoving the standing of the Jewish community in France.

[3]Ferdinand Joachimsthal (1818–1861). German mathematician, student of Kummer and professor at Halle and Breslau. Joachimsthal was a great teacher; his book [Joach2] was one of the first to explain the results of the Monge school and Gauss.

As an important application of Theorem 15.6, we find principal curves on a surface of revolution.

Theorem 15.7. *Suppose the surface of revolution $\mathcal{M}[\boldsymbol{\alpha}]$ generated by a plane curve $\boldsymbol{\alpha}$ is a regular surface. Then the meridians and parallels on $\mathcal{M}[\boldsymbol{\alpha}]$ are principal curves.*

Proof. Each meridian is sliced from $\mathcal{M}[\boldsymbol{\alpha}]$ by a plane Π_m containing the axis of rotation of $\mathcal{M}[\boldsymbol{\alpha}]$. For $\mathbf{p} \in \mathcal{M}[\boldsymbol{\alpha}] \cap \Pi_m$ it is clear that the surface normal $\mathbf{U}(\mathbf{p})$ of $\mathcal{M}[\boldsymbol{\alpha}]$ lies in Π_m. Hence $\mathbf{U}(\mathbf{p})$ and the unit surface normal of Π_m are orthogonal. Therefore, Theorem 15.6 implies that the meridians are principal curves of $\mathcal{M}[\boldsymbol{\alpha}]$.

Next, let Π_p be a plane orthogonal to the axis of $\mathcal{M}[\boldsymbol{\alpha}]$. By rotational symmetry, the unit surface normal $\mathbf{U}$ of $\mathcal{M}[\boldsymbol{\alpha}]$ makes a constant angle with the unit surface normal of Π_p. Again, Theorem 15.6 implies that the parallels are principal curves. ∎

Principal curves give rise to an important class of patches for which curvature computations are especially simple.

Definition 15.8. *A **principal patch** is a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ for which the curves*

$$u \mapsto \mathbf{x}(u, v) \qquad \text{and} \qquad v \mapsto \mathbf{x}(u, v)$$

are principal curves.

It now follows that the standard parametrization (15.1) of a surface of revolution is a principal patch. This fact will be derived independently in the next section, using

Lemma 15.9. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a patch.*

(i) *If $F = 0 = f$ at all points of $\mathcal{U}$, then $\mathbf{x}$ is a principal patch.*

(ii) *If $\mathbf{x}$ is a principal patch with distinct principal curvatures, then $F = 0 = f$ on $\mathcal{U}$.*

Proof. If $F = 0 = f$, the Weingarten equations (Theorem 13.16, page 394) imply that $\mathbf{x}_u$ and $\mathbf{x}_v$ are eigenvectors of the shape operator. Conversely, if $\mathbf{x}$ is principal, then both $\mathbf{x}_u$ and $\mathbf{x}_v$ are eigenvectors of the shape operator. If in addition the principal curvatures are distinct, Lemma 13.19 (page 396) implies that $F = 0 = f$. ∎

In the case of the plane or sphere, the principal curvatures are equal and every point of the surface is said to be ***umbilic***. Moreover, there are infinitely many principal curves through each point of the plane or sphere. This situation will be discussed further in Chapter 19.

15.3 Curvature of a Surface of Revolution

First, we compute the coefficients of the first and second fundamental forms, and also the unit surface normal for a general surface of revolution.

Lemma 15.10. *Let $\mathcal{M}$ be a surface of revolution with profile curve $\boldsymbol{\alpha} = (\varphi, \psi)$. Let $\mathbf{x} = \mathsf{surfrev}[\boldsymbol{\alpha}]$ be the standard parametrization* (15.1) *of $\mathcal{M}$. Then*

$$E = \varphi^2, \qquad F = 0, \qquad G = \varphi'^2 + \psi'^2. \tag{15.7}$$

Thus $\mathbf{x}$ is regular wherever φ and $\varphi'^2 + \psi'^2$ are nonzero. When this is the case,

$$e = \frac{-|\varphi|\psi'}{\sqrt{\varphi'^2 + \psi'^2}}, \qquad f = 0, \qquad g = \frac{(\operatorname{sign}\varphi)(\varphi''\psi' - \varphi'\psi'')}{\sqrt{\varphi'^2 + \psi'^2}}, \tag{15.8}$$

and the unit surface normal is

$$\mathbf{U}(u, v) = \frac{\operatorname{sign}\varphi}{\sqrt{\varphi'^2 + \psi'^2}}\big(\psi' \cos u,\ \psi' \sin u,\ -\varphi'\big). \tag{15.9}$$

Proof. From (15.1) it follows that the first partial derivatives of $\mathbf{x}$ are given by

$$\begin{cases} \mathbf{x}_u = \big(-\varphi(v)\sin u,\ \varphi(v)\cos u,\ 0\big), \\ \mathbf{x}_v = \big(\varphi'(v)\cos u,\ \varphi'(v)\sin u,\ \psi'(v)\big). \end{cases} \tag{15.10}$$

Then (15.7) is immediate from (15.10) and the definitions of E, F, G. The unit normal

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$$

can now be expressed in the form (15.9).

Next, it is necessary to write down the second partial derivatives of $\mathbf{x}$, and use Lemma 13.31 on page 405 to find e, f, g. We omit the calculations, which are easily checked. ∎

Recalling Lemma 15.9, we obtain another proof of the following result.

Corollary 15.11. *The standard parametrization* (15.1) *of a surface of revolution is a principal patch.*

At this point, one could compute the Gaussian curvature K by means of Theorem 13.25. However, we prefer to compute the principal curvatures first. We know that the coordinate curves (15.2) and (15.3) are principal curves. Hence the principal curvatures for a surface of revolution have special meaning, so we denote them by k_p, k_m instead of the usual k_1, k_2. To be specific, k_p is the curvature of the parallel (15.2) and k_m is the curvature of the meridian (15.3) (the reverse of alphabetical order!).

Theorem 15.12. *The principal curvatures of a surface of revolution parametrized by* (15.1) *are given by*

$$\text{(15.11)} \qquad \begin{cases} k_p = \dfrac{e}{E} = \dfrac{-\psi'}{|\varphi|\sqrt{\varphi'^2+\psi'^2}}, \\ k_m = \dfrac{g}{G} = \dfrac{(\operatorname{sign}\varphi)(\varphi''\psi'-\varphi'\psi'')}{(\varphi'^2+\psi'^2)^{3/2}}. \end{cases}$$

The Gaussian curvature is given by

$$\text{(15.12)} \qquad K = \frac{-\psi'^2\varphi''+\varphi'\psi'\psi''}{\varphi(\varphi'^2+\psi'^2)^2},$$

and the mean curvature by

$$\text{(15.13)} \qquad H = \frac{\varphi(\varphi''\psi'-\varphi'\psi'')-\psi'(\varphi'^2+\psi'^2)}{2|\varphi|(\varphi'^2+\psi'^2)^{3/2}}.$$

Proof. Since $F = f = 0$, it follows that

$$\left\{ \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|}, \frac{\mathbf{x}_v}{\|\mathbf{x}_v\|} \right\}$$

forms an orthonormal basis which diagonalizes the shape operator S wherever $\mathbf{x}$ is regular. Hence by Corollary 13.33 on page 405,

$$\text{(15.14)} \qquad S(\mathbf{x}_u) = \frac{e}{E}\mathbf{x}_u \qquad \text{and} \qquad S(\mathbf{x}_v) = \frac{g}{G}\mathbf{x}_v.$$

Thus, $k_p = e/E$, $k_m = g/G$, and (15.11) follows from (15.7) and (15.8). Finally, (15.12) and (15.13) follow from the relations $K = k_p k_m$, $H = \frac{1}{2}(k_p + k_m)$. ∎

We know from the geometric description of the normal curvature given in Section 13.2 that k_m at $\mathbf{p} \in \mathcal{M}$ coincides, up to sign, with the curvature $\boldsymbol{\kappa 2}$ of the meridian through $\mathbf{p}$. This is confirmed overleaf.

Corollary 15.13. *For a surface of revolution, the functions*

$$K,\ H,\ k_p,\ k_m,\ E, F, G,\ e, f, g$$

are all constant along parallels.

Proof. All these functions are expressible in terms of φ and ψ and their derivatives. But φ and ψ do not depend on the angle u. ∎

We may in theory choose a profile curve to have unit speed. In this case, the formulas we have derived so far for a surface of revolution simplify considerably.

Corollary 15.14. *Let* $\mathbf{x}$ *be the standard parametrization* (15.1) *of a surface of revolution in* $\mathbb{R}^3$ *whose profile curve* $\boldsymbol{\alpha} = (\varphi, \psi)$ *has unit speed. Then*

$$E = \varphi^2, \qquad F = 0, \qquad G = 1,$$
$$e = -|\varphi|\psi', \quad f = 0, \quad g = (\operatorname{sign}\varphi)(\varphi''\psi' - \varphi'\psi''),$$
$$k_p = \frac{-\psi'}{|\varphi|}, \qquad k_m = (\operatorname{sign}\varphi)(\varphi''\psi' - \varphi'\psi''),$$
$$2H = (\operatorname{sign}\varphi)(\varphi''\psi' - \varphi'\psi'') - \frac{\psi'}{|\varphi|}, \qquad K = \frac{-\varphi''}{\varphi}.$$

In particular, the signed curvature $\boldsymbol{\kappa 2}[\boldsymbol{\alpha}]$ *equals* $\pm k_m$.

Proof. Since $1 = \|\boldsymbol{\alpha}'\| = \sqrt{\varphi'^2 + \psi'^2}$, all formulas, except for the last, are immediate from (15.7), (15.8), (15.11). That for K follows from (15.12) and the equality $\varphi'\varphi'' + \psi'\psi'' = 0$. The last statement follows from Newton's equation

$$\boldsymbol{\kappa 2}[\boldsymbol{\alpha}] = \frac{-\psi'\varphi'' + \varphi'\psi''}{(\varphi'^2 + \psi'^2)^{3/2}}$$

on page 15. ∎

15.4 Generalized Helicoids

Both helicoids and surfaces of revolution are examples of the class of surfaces known as generalized helicoids, which were first studied by Minding[4] in 1839 (see [Mind]).

Definition 15.15. *Let* Π *be a plane in* $\mathbb{R}^3$, ℓ *be a line in* Π, *and* $\mathscr{C}$ *be a point set in* Π. *Suppose that* $\mathscr{C}$ *is rotated in* $\mathbb{R}^3$ *about* ℓ *and simultaneously displaced parallel to* ℓ *so that the speed of displacement is proportional to the speed of rotation. Then the resulting point set* $\mathcal{M}$ *is called the* ***generalized helicoid*** *generated by* $\mathscr{C}$, *which is called the* ***profile curve*** *of* $\mathcal{M}$. *The line* ℓ *is called the* ***axis*** *of* $\mathcal{M}$. *The ratio of the speed of displacement to the speed of rotation is called the* ***slant*** *of the generalized helicoid* $\mathcal{M}$.

The Euclidean motion (consisting of a simultaneous translation and rotation) used in this definition is called a ***screw motion***. Clearly, a surface of revolution



4 Ernst Ferdinand Adolf Minding (1806–1885). German professor, later dean of the faculty, at the University of Dorpat (now Tartu) in Estonia. Minding was a self-taught mathematician. While a school teacher, he studied for his doctorate, which was awarded by Halle for a thesis on approximating the values of double integrals. In 1864, Minding became a Russian citizen and in the same year was elected to the St. Petersburg Academy. In the 1830s Minding was one of the first mathematicians to use Gauss's approach to the differential geometry of surfaces.

is a generalized helicoid of slant 0, and a generalized helicoid reduces to an ordinary helicoid when the profile curve is $t \mapsto (bt,\, 0)$.

Darboux ([Darb2, Volume 1, page 128]) put it very well when he observed that a surface of revolution has an important kinematic property: when it rotates about its axis, it glides over itself. In more modern language, we would say that a surface of revolution is the orbit of a curve in a plane Π by a 1-parameter group of rotations of $\mathbb{R}^3$ about a line in Π. Generalized helicoids enjoy a similar property: when a generalized helicoid is subjected to a screw motion, it glides over itself. Thus, a generalized helicoid is the orbit of a curve in a plane Π by a 1-parameter group of screw motions of $\mathbb{R}^3$ about a line in Π.

Just as we did for surfaces of revolution on page 462, we choose Π to be the xz-plane and ℓ to be the z-axis. We shall assume that the point set $\mathscr{C}$ has a parametrization $\boldsymbol{\alpha}\colon (a, b) \to \mathscr{C}$ that is differentiable. This allows us to give the alternative

Definition 15.16. *Let $\boldsymbol{\alpha}$ be a plane curve and write $\boldsymbol{\alpha}(t) = (\varphi(t), \psi(t))$. Then the generalized helicoid with profile curve $\boldsymbol{\alpha}$ and slant c is the surface in $\mathbb{R}^3$ parametrized by*

$$\mathsf{genhel}[c, \boldsymbol{\alpha}](u, v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ cu + \psi(v)\big).$$

To summarize, a generalized helicoid is generated by a plane curve, which is rotated about a fixed axis and at the same time translated in the direction of the axis with speed a constant multiple of the speed of rotation.

It is also possible to define the notion of meridian for a generalized helicoid.

Definition 15.17. *Let $\mathscr{C}$ be a point set in a plane $\Pi \subset \mathbb{R}^3$, and let $\mathcal{M}[\mathscr{C}, c]$ be the generalized helicoid in $\mathbb{R}^3$ of slant c generated by revolving $\mathscr{C}$ about a line $\ell \subset \Pi$ and simultaneously displacing it parallel to ℓ. A **meridian** on $\mathcal{M}[\mathscr{C}, c]$ is the intersection of $\mathcal{M}[\mathscr{C}, c]$ with a plane containing the axis ℓ.*

The generalization to generalized helicoids of the notion of parallel given on page 463 is more complicated. For a surface of revolution a parallel is a circle; for a generalized helicoid the corresponding curve is a helical curve, which we call a ***parallel helical curve***. The v-parameter curves are the meridians and the u-parameter curves are the parallel helical curves. In contrast to the situation with a surface of revolution, the meridians and parallel helical curves will not be perpendicular to one another for a generalized helicoid of nonzero slope. In Theorem 15.25, page 483, we shall show that a generalized helicoid with the property that its meridians *are* principal curves must either be a surface of revolution or Dini's surface illustrated in Figure 15.13.

Generalized helicoids can be considered to be "twisted" surfaces of revolution, though this is unrelated to the construction on page 348. For example, Figure 15.4 is evidently a twisted version of Figure 15.1; the slant is $1/2$.

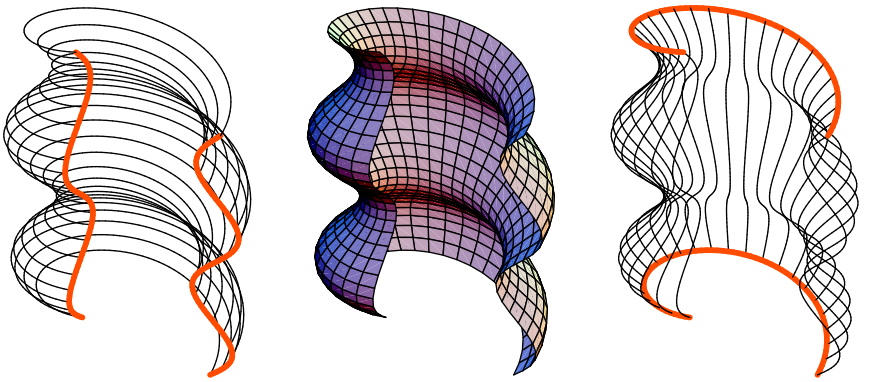


Figure 15.4: A generalized helicoid together with its parallel helical curves and meridians

The proof of the following lemma is an easy extension of the proof of Lemma 15.10, and the calculations are carried out in Notebook 15.

Lemma 15.18. *Let* $\mathbf{x}$ *be a generalized helicoid in* $\mathbb{R}^3$ *whose profile curve is* $\boldsymbol{\alpha} = (\varphi, \psi)$. *Then*

$$E = \varphi^2 + c^2, \qquad F = c\,\psi', \qquad G = \varphi'^2 + \psi'^2.$$

Therefore, $\mathbf{x}$ *is regular wherever*

$$EG - F^2 = \varphi^2(\varphi'^2 + \psi'^2) + c^2\varphi'^2$$

is nonzero. In this case, setting $D = \sqrt{EG - F^2}$, *we have*

$$e = \frac{-\varphi^2\psi'}{D}, \qquad f = \frac{c\,\varphi'^2}{D}, \qquad g = \frac{\varphi(\varphi''\psi' - \varphi'\psi'')}{D}.$$

Furthermore

$$K = \frac{-c^2\varphi'^4 + \varphi^3\psi'(-\psi'\varphi'' + \varphi'\psi'')}{D^4}, \tag{15.15}$$

and

$$H = \frac{-2c^2\varphi'^2\psi' - \varphi^2\psi'(\varphi'^2 + \psi'^2) + \varphi(c^2 + \varphi^2)(\psi'\varphi'' - \varphi'\psi'')}{2\,D^3}.$$

We know that the sphere $S^2(a)$ is a surface of revolution parametrized by

$$\mathsf{sphere}[a](u, v) = \big(a\cos v\cos u,\ a\cos v\sin u,\ a\sin v\big); \tag{15.16}$$

it has constant positive Gaussian curvature $K = 1/a^2$. We call the twisted version of the sphere the ***corkscrew surface*** or ***twisted sphere***. Explicitly, it is given by

$$\mathsf{twisphere}[a,b](u,v) = \big(a\cos u\cos v,\ a\sin u\cos v,\ a\sin v + bu\big).$$

In Notebook 15, the Gaussian curvature of the twisted sphere is found to be

$$\frac{4a^2\cos^4 v - 4b^2\sin^4 v}{\big(a^2+b^2+(a^2-b^2)\cos 2v\big)^2},$$

and simplifies to $(\cos 2v)/a^2$ if $b = a$. This formula is used to color the left-hand side of Figure 15.5.

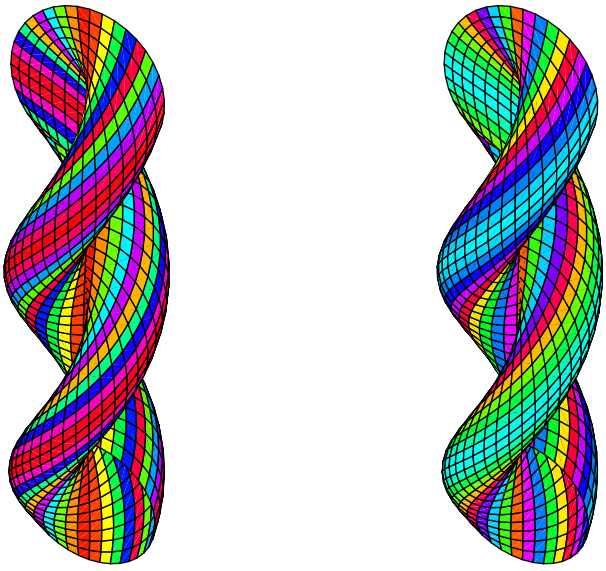


Figure 15.5: Corkscrew surfaces colored by Gaussian curvature (left) and mean curvature (right)

15.5 Surfaces of Constant Positive Curvature

The sphere is not the only surface in $\mathbb{R}^3$ that has constant curvature. We can take a spherical cap made up of some thin inelastic material, for example, half of a ping-pong ball. Then the spherical cap can be bent without stretching into many different shapes. Since there is no stretching involved and the Gaussian curvature is an isometric invariant (see Theorem 17.5 on page 536), a bent cap also has constant positive curvature. On the other hand, it is clear intuitively that a whole ping-pong ball cannot be bent. Indeed, Liebmann's Theorem (Theorem 19.13, page 603) states that the sphere $S^2(a)$ is rigid in this sense.

To find other surfaces of revolution in $\mathbb{R}^3$ that have constant positive curvature, we proceed backwards: we assume that we are given a surface of revolution $\mathcal{M}$ with constant positive curvature and seek restrictions on a parametrization $\mathbf{x}$ of $\mathcal{M}$. First, we determine the profile curves of a surface of constant positive Gaussian curvature.

Theorem 15.19. *Let $\mathcal{M}$ be a surface of revolution whose Gaussian curvature is a positive constant $1/a^2$, where $a > 0$. Then $\mathcal{M}$ is part of a surface parametrized by a patch $\mathbf{x}$ of the form*

$$\mathbf{x}(u, v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big),$$

where

$$\begin{cases} \varphi(v) = b\cos\dfrac{v}{a}, \\ \psi(v) = \displaystyle\int_0^{v/a} \sqrt{a^2 - b^2\sin^2 t}\, dt, \end{cases} \tag{15.17}$$

for some constant $b > 0$. The parameter v has one of the following ranges:

$$\begin{aligned} &\text{if} \quad b = a, \quad \text{then} \quad -\frac{\pi a}{2} \leqslant v \leqslant \frac{\pi a}{2}; \\ &\text{if} \quad b < a, \quad \text{then} \quad -\infty \leqslant v \leqslant \infty; \\ &\text{if} \quad b > a, \quad \text{then} \quad -a\arcsin\frac{a}{b} \leqslant v \leqslant a\arcsin\frac{a}{b}. \end{aligned}$$

The patch $\mathbf{x}$ is regular at (u, v) if and only if $\varphi(v) \neq 0$, that is, $v \neq (n + \frac{1}{2})\pi a$.

Proof. Assume without loss of generality that $\mathbf{x} = \mathsf{surfrev}[\boldsymbol{\alpha}]$ is given by (15.1), and that the profile curve $\boldsymbol{\alpha} = (\varphi, \psi)$ has unit speed, so that $\varphi'^2 + \psi'^2 = 1$. If $\mathcal{M}$ has constant positive curvature $1/a^2$, then Corollary 15.14, page 470, implies that φ satisfies the differential equation

$$\varphi'' + \frac{1}{a^2}\varphi = 0, \tag{15.18}$$

whose general solution is $\varphi(v) = b\cos((v/a) + c)$. Without loss of generality, we can assume that $c = 0$; this amounts to translating the profile curve along the axis of revolution so that the profile curve is farthest from the axis of revolution when $v = 0$. Also, by taking a mirror image if necessary, we may assume that $b > 0$. Thus we get the first equation of (15.17).

We can assume that $\psi'(v) \geqslant 0$ for all v; otherwise, replace v by $-v$. Then $\varphi'^2 + \psi'^2 = 1$ implies that

$$a\,\psi'(v) = \sqrt{a^2 - b^2\sin^2\frac{v}{a}}, \tag{15.19}$$

and when we integrate this equation from 0 to v, and then rescale the variable, we get the second equation of (15.17).

In order that $\psi(v)$ be well defined, it is necessary that the quantity under the square root in (15.19) be nonnegative. For $b < a$ this quantity is always positive, so $\psi(v)$ is defined for all v. When $b = a$ the profile curve is a part of a circle; in that case the requirement that $-\pi/2 \leqslant v/a \leqslant \pi/2$ ensures that the profile curve does not overlap itself and is thus a semicircle. If $b > a$ then (15.19) is defined if and only if $-a/b \leqslant \sin(v/a) \leqslant a/b$. ∎

When $b = a$, the profile curve is

$$v \mapsto \left(a\cos\frac{v}{a},\ a\sin\frac{v}{a}\right), \qquad -\frac{\pi}{2} \leqslant \frac{v}{a} \leqslant \frac{\pi}{2},$$

which when revolved about the z-axis yields a sphere $S^2(a)$ of radius a. If $b < a$, the profile curve (φ, ψ) makes a shallower arc and first crosses the z-axis when

$$z = \pm \int_0^{\pi/2} \sqrt{a^2 - b^2 \sin^2 t}\, dt.$$

Its length between these two points is πa, which is the same as the length of a semicircle of radius a. But the profile curve can be continued indefinitely, so as to weave its way up and down the z-axis, instead of closing up (a hint of this behavior can be seen in Figure 15.6).

When $b > a$ the profile curve is only defined on the interval

$$-a\arcsin\frac{a}{b} \leqslant v \leqslant a\arcsin\frac{a}{b},$$

and the resulting surface of revolution resembles a barrel when b is moderately larger than a. The closer the ratio a/b is to zero, the larger the "hole" in the middle. The outside surface in Figure 15.7 illustrates the case $a/b = 3/4$.

In conclusion,

Corollary 15.20. *Let $\mathcal{S}(a, b)$ be the surface of revolution whose profile curve is $\boldsymbol{\alpha} = (\varphi, \psi)$ with φ, ψ given by* (15.17).

(i) *$\mathcal{S}(a, a)$ is an ordinary sphere of radius a.*

(ii) (*Spindle type*) *If $0 < b < a$, then $\mathcal{S}(a, b)$ is a surface of revolution that resembles an infinite string of beads, each of which is shaped like a football with its vertices on the axis of revolution.*

(iii) (*Bulge type*) *If $0 < a < b$, then $\mathcal{S}(a, b)$ is barrel-shaped and does not meet the axis of revolution.*

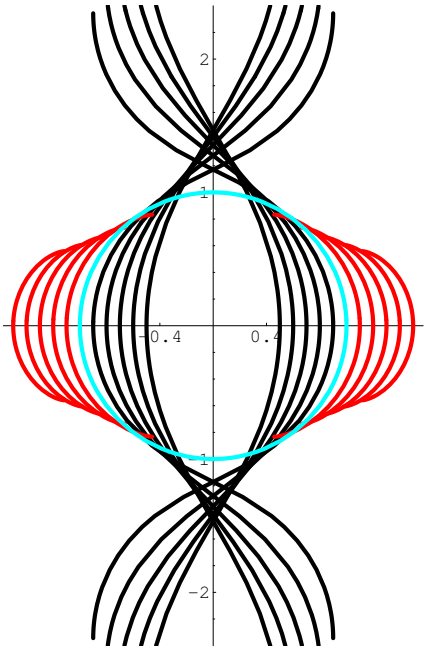


Figure 15.6: Profile curves with $a \leqslant b$ and $b \leqslant a$

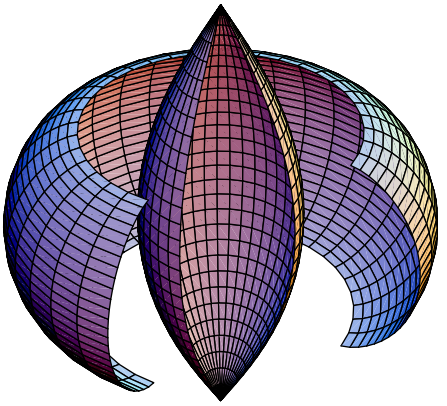


Figure 15.7: The surfaces $\mathcal{S}(a, b)$ with $a/b = 3/4$ and 2

Elliptic Integrals

We shall make a slight detour into the complicated subject of elliptic functions and integrals, which will be useful when we study surfaces of revolution of constant negative curvature.

Definition 15.21. *The* ***elliptic integral of the second kind*** *is defined by*

$$\mathrm{E}\big(\phi \,\big|\, m\big) = \int_0^{\phi} \sqrt{1 - m\sin^2\theta}\, d\theta$$

with $m < 1$*, whereas the* ***complete elliptic integral of the second kind*** *is*

$$\mathrm{E}\Big(\frac{\pi}{2} \,\Big|\, m\Big) = \int_0^{\pi/2} \sqrt{1 - m\sin^2\theta}\, d\theta.$$

The motivation for this definition comes from the formula for the arc length of an ellipse, given in Exercise 9. But we have already made implicit use of it, for in (15.17) we may write

$$\psi(v) = a\,\mathrm{E}\left(\frac{v}{a} \,\middle|\, \frac{b^2}{a^2}\right),$$

after having brought a factor of a outside the square root.

Notice that $\mathrm{E}\big(\phi \,\big|\, 1\big) = \sin\phi$; thus $\mathrm{E}\big(\phi \,\big|\, m\big)$ can be considered to be a generalization of the sine function. The corresponding generalization of the hyperbolic sine function turns out to be $-i\,\mathrm{E}\big(i\phi \,\big|\, -m\big)$ because

$$-i\,\mathrm{E}\big(i\phi \,\big|\, -m\big) = \int_0^{\phi} \sqrt{1 - m\sinh^2\theta}\, d\theta, \tag{15.20}$$

as can be checked by changing variables in the integral and using the identity $\sinh ix = i\sin x$. We shall need (15.20) in the next section.

15.6 Surfaces of Constant Negative Curvature

The determination of the surfaces of revolution of constant negative curvature proceeds along the same lines as that in Section 15.5. However, the resulting surfaces are quite different in appearance.

Theorem 15.22. *Let* $\mathcal{M}$ *be a surface of revolution whose Gaussian curvature is a negative constant* $-1/a^2$*. Then* $\mathcal{M}$ *is part of a surface parametrized by a patch* $\mathbf{x}$ *such that*

$$\mathbf{x}(u, v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big),$$

where the profile curve $\boldsymbol{\alpha} = (\varphi, \psi)$ *is one of the following types*:

(i) (*Pseudosphere*)

$$\boldsymbol{\alpha}(v) = \begin{cases} \left(a\,e^{-v/a},\ \displaystyle\int_0^v \sqrt{1 - e^{-2t/a}}\, dt\right) & \text{for } 0 \leqslant v < \infty, \\ \left(a\,e^{v/a},\ \displaystyle\int_0^v \sqrt{1 - e^{2t/a}}\, dt\right) & \text{for } -\infty < v < 0. \end{cases} \tag{15.21}$$

(ii) (*Hyperboloid type*)

$$\text{(15.22)}\qquad \begin{cases} \varphi(v) \;=\; b\cosh\dfrac{v}{a}, \\[2ex] \psi(v) \;=\; \displaystyle\int_0^{v/a} \sqrt{a^2 - b^2\sinh^2 t}\, dt = -ia\,\mathrm{E}\left(\frac{iv}{a}\,\middle|\,\frac{-b^2}{a^2}\right), \end{cases}$$

for some constant $b > 0$*, and* v *satisfies* $-a\,\mathrm{arcsinh}\,\dfrac{a}{b} \leqslant v \leqslant a\,\mathrm{arcsinh}\,\dfrac{a}{b}$.

(iii) (*Conic type*)

$$\text{(15.23)}\qquad \begin{cases} \varphi(v) \;=\; b\sinh\dfrac{v}{a}, \\[2ex] \psi(v) \;=\; \displaystyle\int_0^{v/a} \sqrt{a^2 - b^2\cosh^2 t}\, dt \\[2ex] \qquad\qquad = -i\sqrt{a^2-b^2}\,\mathrm{E}\left(\dfrac{iv}{a}\,\middle|\,\dfrac{-b^2}{a^2-b^2}\right), \end{cases}$$

for some constant b *with* $0 < b \leqslant a$*, and* v *must satisfy*

$$-a\,\mathrm{arcsinh}\,\frac{\sqrt{a^2-b^2}}{b} \leqslant v \leqslant a\,\mathrm{arcsinh}\,\frac{\sqrt{a^2-b^2}}{b}.$$

Proof. Without loss of generality, $a > 0$. The general solution of the negative analogue $\varphi'' - \varphi/a^2 = 0$ of (15.18) is

$$\text{(15.24)}\qquad \varphi(v) = Ae^{v/a} + Be^{-v/a}.$$

Case 1. First, suppose that A is zero in (15.24). We can assume that $B > 0$ by reflecting the profile curve about the y-axis if necessary, so that (φ, ψ) becomes $(-\varphi, \psi)$. Moreover, substituting $v \rightsquigarrow v + a\log B - a\log a$, we can assume that $B = a$ and $\varphi(v) = ae^{-v/a}$. Since

$$0 \leqslant \psi'(v)^2 = 1 - \varphi'(v)^2 = 1 - e^{-2v/a},$$

we must have $v \geqslant 0$. Thus we get the first alternative in (15.21). Similarly, $B = 0$ leads to the second alternative.

Next, suppose that A and B are both different from zero in (15.24). Using the change of variables

$$v \rightsquigarrow v + \frac{a}{2}\log\left|\frac{B}{A}\right|$$

if necessary, we may assume that $|A| = |B|$.

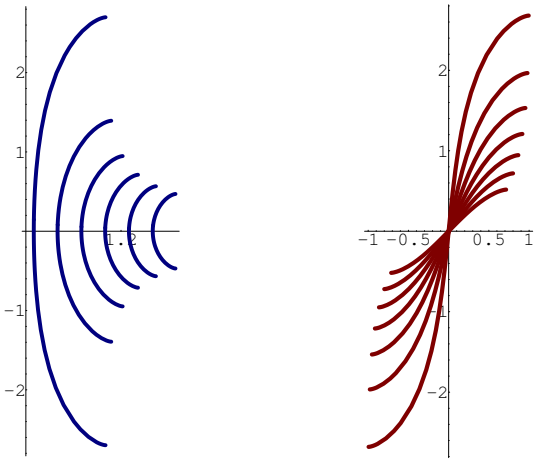


Figure 15.8: Profile curves of hyperbolic type (left) and conic type (right)

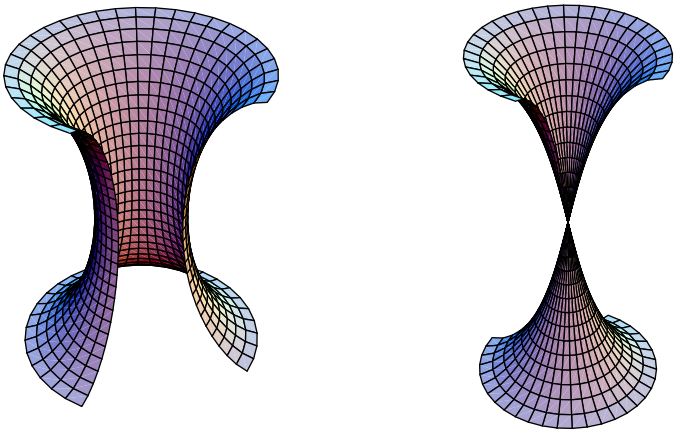


Figure 15.9: The respective surfaces of constant negative curvature

Case 2. When $A = B$, we can (using a mirror image of the profile curve if necessary) assume that $A > 0$. Then

$$\varphi(v) = A\big(e^{v/a} + e^{-v/a}\big) = 2A\cosh\frac{v}{a}.$$

We put $b = 2A$ and obtain (15.22).

Case 3. If $A = -B$, we can (changing v to $-v$ if necessary) assume that $A > 0$. Thus (15.24) becomes

$$\varphi(v) = A\big(e^{v/a} - e^{-v/a}\big) = 2A\sinh\frac{v}{a}.$$

We put $b = 2A$ and obtain (15.23). ▮

Corollary 15.23. *The surface of revolution whose profile curve is given by* (15.21) *is a* ***pseudosphere*** *or* ***tractoid***; *that is, the surface of revolution of a tractrix.*

Proof. This is an immediate consequence of Lemma 2.1, page 52, and part (i) of Theorem 15.22. ▮

In the light of this corollary, equation (2.16) gives rise to the parametrization

$$\mathsf{pseudosphere}[a](u,v) = a\left(\cos u \sin v,\ \sin u \sin v,\ \cos v + \log\left(\tan\frac{v}{2}\right)\right).$$

In Notebook 15, we verify that this surface does indeed have constant curvature equal to $-1/a^2$.

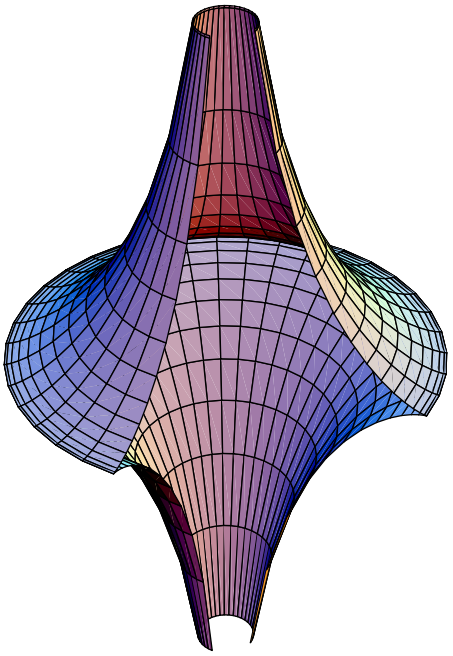


Figure 15.10: The pseudosphere

15.7 More Examples of Constant Curvature

In this final section, we determine the *flat* generalized helicoids and introduce the surfaces of Dini and Kuen, which will be studied in greater detail by transform methods in Chapter 21.

Flat Generalized Helicoids

Recall the parametrization of a generalized helicoid given in Definition 15.16.

Theorem 15.24. *A generalized helicoid is flat if and only if its profile curve can be parametrized as* $\boldsymbol{\alpha}(t) = (t,\, \psi(t))$, *where*

$$\pm\psi(t) = t\sqrt{a^2 - \frac{c^2}{t^2}} + c\,\arcsin\frac{c}{at}. \tag{15.25}$$

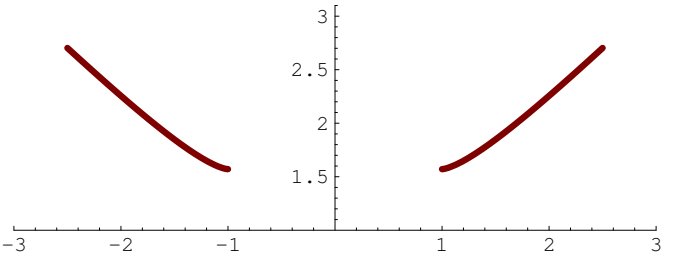


Figure 15.11: Zero-curvature profile curve

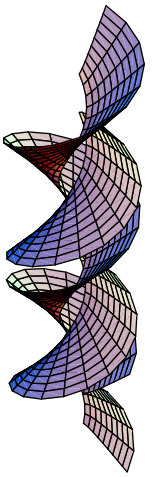


Figure 15.12: Zero-curvature generalized helicoid

Proof. Equation (15.15), page 472, implies that a generalized helicoid is flat if and only if its profile curve $\boldsymbol{\alpha} = (\varphi, \psi)$ satisfies the differential equation

$$\varphi^3(\varphi'\psi'\psi'' - \psi'^2\varphi'') - c^2\varphi'^4 = 0. \tag{15.26}$$

The profile curve can be parametrized so that $\varphi(t) = t$; then (15.26) reduces to

$$t^3\psi'\psi'' - c^2 = 0.$$

Thus,

$$\frac{d}{dt}(\psi')^2 = \frac{2c^2}{t^3},$$

and

$$\psi'(t) = \pm\sqrt{a^2 - \frac{c^2}{t^2}}, \tag{15.27}$$

where a^2 is a constant and $a > 0$. Then (15.27) can integrated to give (15.25), apart from an irrelevant constant of integration (see Notebook 15). ▮

In Notebook 15, we verify that the generalized helicoid generated by

$$t \mapsto \big(t,\ \psi(t)\big),$$

with ψ defined in (15.25), does indeed have zero Gaussian curvature.

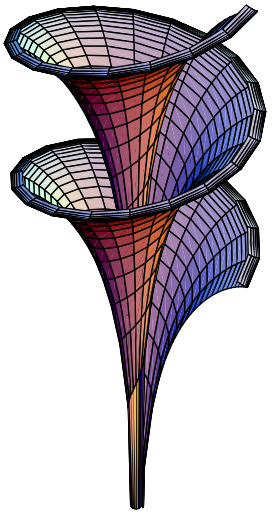


Figure 15.13: Dini's surface

Dini's Surface

The twisted version of (15.21) for $v > 0$ is associated with the name of Dini.[5] The explicit parametrization of Dini's surface is obtained by applying the genhel construction on page 471 to a tractrix (see page 51), and is therefore

$$\mathsf{dini}[a,b](u,v) = \Big(a\cos u\sin v,\ a\sin u\sin v,\ a\Big(\cos v + \log\big(\tan\frac{v}{2}\big)\Big) + cu\Big).$$



5 Ulisse Dini (1845–1918). Italian mathematician who worked mainly in Pisa. He made fundamental contributions to surface theory and real analysis. His statue can be found off the Piazza dei Cavalieri, not far from the leaning tower.

In contrast to the corkscrew surface, Dini's has *constant curvature*, a fact that will be explained in Chapter 21. Figure 15.13 shows clearly that it is a "twisted deformation" of the pseudosphere.

The following theorem is taken from volume 1 page 353 of Luigi Bianchi's classical text,[6] *Lezioni di Geometria Differenziale* [Bian].

Theorem 15.25. *Let $\mathcal{M}$ be a generalized helicoid with the property that the meridians are principal curves. Then $\mathcal{M}$ is part of Dini's surface.*

Proof. Without loss of generality, we can suppose that the profile curve of the generalized helicoid is of the form $\boldsymbol{\alpha}(t) = (t,\ \psi(t))$, so that

$$\mathsf{genhel}[c, \boldsymbol{\alpha}](u, v) = \big(v\cos u,\ v\sin u,\ cu + \psi(v)\big).$$

The unit normal $\mathbf{U}$ to $\mathsf{genhel}[c, \boldsymbol{\alpha}]$ is given by

$$\mathbf{U}(u, v) = \frac{\big(-c\sin u + v\psi'(v)\cos u,\ c\cos u + v\psi'(v)\sin u,\ -v\big)}{\sqrt{c^2 + v^2 + v^2\psi'(v)^2}},$$

a formula extracted from Notebook 15. The meridian $v \mapsto \mathsf{genhel}[c, \boldsymbol{\alpha}](u, v)$ is the intersection of the generalized helicoid with a plane Π through its axis. The unit normal to this plane is given by

$$\mathbf{V}(u, v) = (-\sin u,\ \cos u,\ 0).$$

Now suppose that $\boldsymbol{\alpha}$ is a principal curve on the generalized helicoid. Since $\boldsymbol{\alpha}$ is automatically a principal curve on Π, Theorem 15.26 (which is Exercise 4) implies that the vector fields $\mathbf{U}$ and $\mathbf{V}$ meet at a constant angle σ along $\boldsymbol{\alpha}$. Thus

$$\cos\sigma = \mathbf{U}\cdot\mathbf{V} = \frac{c}{\sqrt{c^2 + v^2 + v^2\psi'(v)^2}}. \tag{15.28}$$

Then (15.28) implies that ψ satisfies the differential equation

$$v^2\big(1 + \psi'(v)^2\big) = c^2\tan^2\sigma.$$

But this is equivalent to the differential equation (2.17) on page 51, so $\boldsymbol{\alpha}$ is a tractrix. Hence the generalized helicoid is a surface of Dini. ∎



6 Luigi Bianchi (1856–1928). Italian mathematician who worked mainly in Pisa. Although he is most remembered for the "Bianchi identities" (see page 864), he also made fundamental contributions to surface theory.

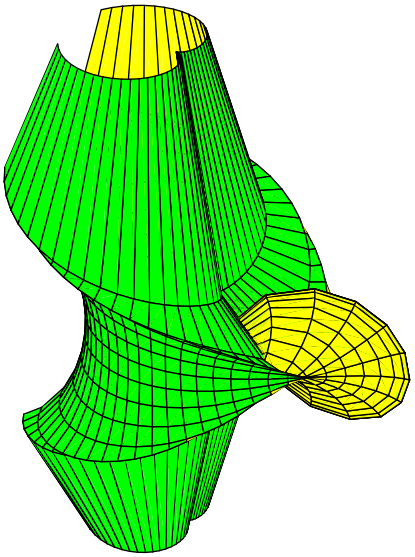


Figure 15.14: Kuen's surface

Kuen's Surface

A more complicated surface of constant negative curvature is that of Kuen[7] [Kuen]. It can be parametrized by

$$\left(\frac{(\cos u + u\sin u)\sin v}{1+u^2\sin^2 v},\ \frac{(\sin u - u\cos u)\sin v}{1+u^2\sin^2 v},\ \frac{1}{2}\log\bigl(\tan\tfrac{v}{2}\bigr)+\frac{\cos v}{1+u^2\sin^2 v}\right).$$

A computation in Notebook 15 shows that this surface has Gaussian curvature

$$K = -4.$$

It also turns out that, for this parametrization, the "mixed" coefficients F and f of both the first and second fundamental forms vanish identically. As a consequence, the u and v parameter curves are principal; these are the curves visible in Figure 15.14. We shall explain this fact and see exactly how the equation for Kuen's surface arises in Section 21.7.

An extraordinary plaster model was made of this surface; plate 86 of [Fischer] is a photo of this plaster model. Reckziegel has given an excellent description of Kuen's surface (see pages 30–41 of the Commentaries to [Fischer]).

---

[7] Th. Kuen used Bianchi's parametrization to make a plaster model of his surface (see [Kuen]).

15.8 Exercises

M **1.** Show that the surface of revolution generated by the graph of a function $h\colon \mathbb{R} \to \mathbb{R}$ can be parametrized by

$$\mathbf{x}(u,v) = \big(v\cos u,\ v\sin u,\ h(v)\big), \qquad 0 < u < 2\pi.$$

Determine the associated functions $E, F, G,\ e, f, g,\ k_p, k_m, K$, in terms of the function h.

2. Find the formulas for the Gaussian, mean and principal curvatures of a surface of revolution with profile curve $\boldsymbol{\alpha}(t) = (t,\ \psi(t))$.

M **3.** Plot the surfaces of revolution corresponding to each of the following curves: cycloid, cissoid, logarithmic spiral, lemniscate, cardioid, astroid, deltoid, nephroid, triangle.

4. Prove the following converse to Theorem 15.6 that was used to prove Theorem 15.25.

Theorem 15.26. Let $\boldsymbol{\alpha}$ be a curve which lies on the intersection of regular surfaces $\mathcal{M}_1, \mathcal{M}_2 \subset \mathbb{R}^3$. Suppose that $\boldsymbol{\alpha}$ is a principal curve in $\mathcal{M}_1$ and also in $\mathcal{M}_2$. Then the normals to $\mathcal{M}_1$ and $\mathcal{M}_2$ meet at a constant angle along $\boldsymbol{\alpha}$.

5. With reference to Lemma 15.9, show that the mapping $\mathbf{x}\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$\mathbf{x}(u,v) = (u+v,\ u,\ 0)$$

is a principal patch for which $F \neq 0$.

M **6.** Compute the Gaussian curvature and mean curvature of the generalized helicoid of a catenary. Plot the generalized helicoid of catenary[1] with slant $1/2$.

7. Show that a surface of revolution whose Gaussian curvature is zero is a part of a plane, circular cone, or circular cylinder. These are therefore the only *flat* surfaces of revolution.

M **8.** Plot the elliptic functions

$$\phi \mapsto \mathrm{E}\Big(\phi \,\Big|\, \frac{n}{5}\Big) \qquad \text{and} \qquad \phi \mapsto -i\,\mathrm{E}\Big(i\phi \,\Big|\, -\frac{n}{5}\Big)$$

for $n = 0, 1, 2, 3, 4, 5$.

9. Find a formula for the length of the ellipse $(x/a)^2 + (y/b)^2 = 1$ in terms of the elliptic integral E.

10. A surface of revolution is formed by moving a curve $\boldsymbol{\gamma}$ in a plane Π_1 about a circle in a plane Π_2, where Π_1 and Π_2 are perpendicular. There is a more general construction in which the circle is replaced by an arbitrary curve $\boldsymbol{\alpha}$ in Π_2:

$$\mathsf{gensurfrev}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](u, v) = \big(\boldsymbol{\alpha}_1(u)\boldsymbol{\gamma}_1(v),\ \boldsymbol{\alpha}_2(u)\boldsymbol{\gamma}_2(v),\ \boldsymbol{\alpha}_3(u)\boldsymbol{\gamma}_3(v)\big).$$

Determine when this parametrization is a principal patch.

M **11.** Compute the Gaussian curvature K and the mean curvature H of the generalized surface of revolution formed by moving an eight curve along an eight curve. This is represented in Figure 15.15 (left), while the function K is plotted on the right.

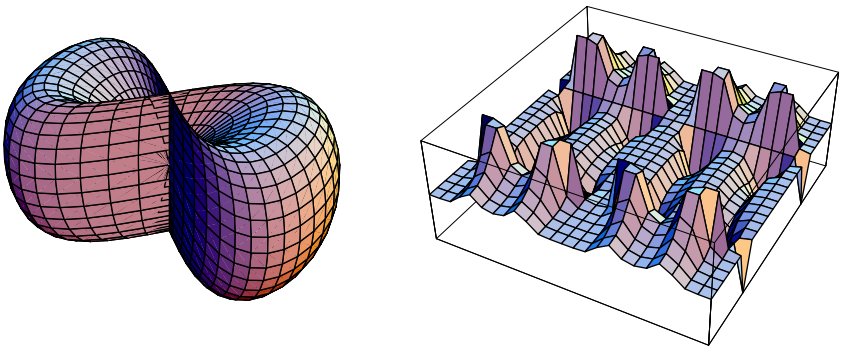


Figure 15.15: A generalized surface of revolution and its Gaussian curvature

M **12.** Find the principal curvatures and the Gauss map of a pseudosphere.

M **13.** ***Sievert's surface*** is defined by $\mathbf{x} = (x_1, x_2, x_3)$, where

$$\begin{cases} x_1 = \dfrac{2\sqrt{2 + 2\sin^2 u}\,\sin v}{2 - \sin^2 v\,\cos^2 u}\cos\Big(-\dfrac{u}{\sqrt{2}} + \arctan(\sqrt{2}\tan u)\Big) \\ x_2 = \dfrac{2\sqrt{2 + 2\sin^2 u}\,\sin v}{2 - \sin^2 v\,\cos^2 u}\sin\Big(-\dfrac{u}{\sqrt{2}} + \arctan(\sqrt{2}\tan u)\Big), \\ x_3 = \log\Big(\tan\dfrac{v}{2}\Big) + \dfrac{4\cos v}{2 - \sin^2 v\cos^2 u}. \end{cases}$$

Show that it has constant positive Gaussian curvature (see [Siev] and [Fischer, Commentary, page 38]). Investigate the more general patch,

sievert[a], depending upon a parameter a (equal to 1 above) defined in Notebook 15. Figure 15.16 illustrates the case $a = 4/5$.

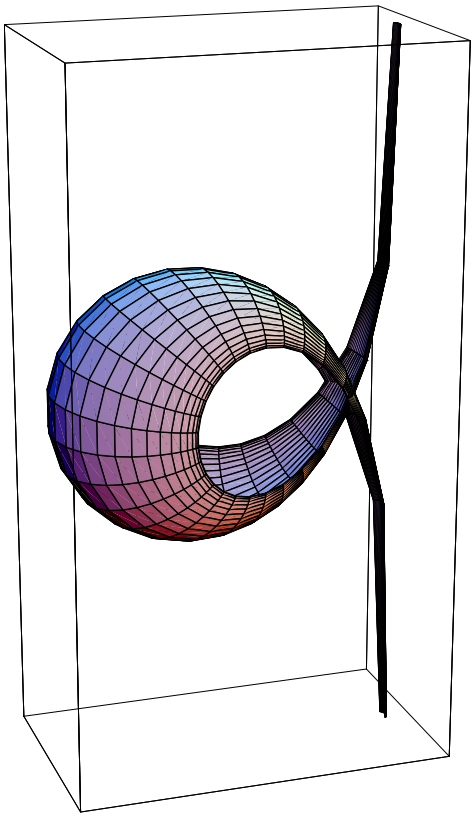


Figure 15.16: The surface sievert[0.8]

14. Suppose that $\mathbf{x}$ is a generalized helicoid in $\mathbb{R}^3$ with slant c whose profile curve $\boldsymbol{\alpha} = (\varphi, \psi)$ has unit speed. Simplify the equations of Lemma 15.18 that apply in this case.

15. Fill in the details of the proof of Theorem 15.25.

Notebook 15

N15.0

The first part of this notebook is concerned with the construction of surfaces of revolution, obtained by starting from a plane curve. As well as illustrating specific examples such as the hyperboloid and catenoid in N15.1, we carry out the construction in general so as to compute the first and second fundamental forms of the resulting surface in terms of the curve in N15.3.

A similar analysis is carried out for the generalized helicoids in N15.4, at the end of which we plot a twisted version of the sphere colored by both its Gaussian and its mean curvature (Figure 15.5).

We then channel attention to the description of surfaces of revolution with constant curvature, dealing in turn with the positive case in N15.5 and the negative case in N15.6. The surfaces of Dini and Kuen are plotted in N15.7.

In this notebook, we shall need the plane curves

```
circ[a_][t_]:= a{Cos[t], Sin[t]}
eight[t_]:= {Sin[t], Sin[t]Cos[t]}
tractrix[a_][t_]:= a{Sin[t], Cos[t]+Log[Tan[t/2]]}
```

and the hyperboloid of one sheet, originally defined in N10.7:

```
hyperboloid1[a_,b_,c_][u_,v_]:=
  {a Cosh[v]Cos[u], b Cosh[v]Sin[u], c Sinh[v]}
```

We load the packages

```
<<Graphics`Shapes`
<<Graphics`Colors`
```

We make frequent use of the previously-defined programs unitnormal, metric, shape, gcurv, mcurv, all of which are loaded next.

N15.0.1

■ N15.1

The standard parametrization of a surface of revolution, given in the text, is easily translated into *Mathematica*. Observe first that the components of a plane curve α at a given point, with parameter value v, are α[v][[1]] and α[v][[2]].

```
surfrev[α_][u_,v_]:= {Cos[u]α[v][[1]], Sin[u]α[v][[1]], α[v][[2]]}
```

We next provide a general definition for α that will help us to recover various formulas in Chapter 15. This definition will remain in force throughout the notebook.

```
α[t_]:= {φ[t], ψ[t]}
```

This is consistent with the notation of Definition 15.2, which corresponds to the output of the following cell:

```
surfrev[α][u,v]
```

The next example returns a parametrization of the torus of revolution. The first argument of surfrev indicates a function that maps a variable # to the expression preceding the symbol &.

```
surfrev[circ[b][#]+{a,0} &][u, v]
```

To build up the objects in Figure 15.1, we first define and plot the surface of revolution formed from the curve β involving the sine function:

```
β[t_]:= {2 + Sin[2t]/2, t}
x:= surfrev[β]
z[2]= ParametricPlot3D[x[u,v], {u,0,3π/2}, {v,-π,π}]
```

Its meridians are defined next, with the top and bottom ones highlighted in red. This process is repeated for the parallels, and gives us the complete Figure 15.1.

```
f[u_]:= x[u,π]
g[u_]:= x[u,-π]
ff:= highlight[f]
gg:= highlight[g]
mer[v_]:= Table[x[u,v], {u,0,3π/2,π/16}]
y1:= ParametricPlot3D[mer[v]//Evaluate, {v,-π,π}]
y2:= ParametricPlot3D[{ff[u],gg[u]}//Evaluate, {u,0,3π/2}]
z[3] = Show[{y1,y2}]
```

```
h[v_]:= x[0,v]
k[v_]:= x[3π/2,v]
hh:= highlight[h]
kk:= highlight[k]
par[u_]:= Table[x[u, v], {v,-π,π,π/12}]
y3:= ParametricPlot3D[par[u]//Evaluate, {u,0,3π/2}]
y4:= ParametricPlot3D[{hh[v],kk[v]}//Evaluate, {v,-π,π}]
z[1]:= Show[{y3,y4}]
S[GraphicsArray[Array[z, 3]], GraphicsSpacing->-.2]
```

Figure 15.1

The catenoid is obtained by rotating the catenary, parametrized in N4.0.

```
catenary[a_][t_]:= {a Cosh[t/a], t}
```

```
ParametricPlot[catenary[1][t], {t,-2,2}];
```

Its standard patch is

```
surfrev[catenary[c]][u,v]
```

We use output from the last cell to define the catenoid directly, making explicit its dependence upon the parameter c:

```
catenoid[c_][u_,v_]:= {c Cos[u]Cosh[v/c], c Cosh[v/c]Sin[u], v}
```

The next program illustrates the WireFrame command, part of the Shapes package, loaded in N15.0.

```
z1:= ParametricPlot3D[catenoid[1][u,v], {u,0,2π}, {v,-2,2}]
z2:= Show[WireFrame[z1]]
GraphicsArray[{z1,z2}]
```

The catenoid can be compared with the hyperboloid of one sheet. In this program, we use Append to color the respective surfaces, a procedure explained in N12.0.

```
Module[{x,xx,y,yy},
  x= hyperboloid1[1,1,1];
  xx[u_,v_]= Append[x[u,v], FaceForm[LimeGreen]];
  y= catenoid[1];
  yy[u_,v_]= Append[y[u,v], FaceForm[Yellow]];
  S @ ParametricPlot3D[{xx[u,v],yy[u,v]}//Evaluate,
       {u,0,3π/2}, {v,-2,2}, Lighting->False]
]
```

Figure 15.2

N15.1.1

N15.3

Lemma 15.10 can be verified by using the programs metric and unitnormal from Notebook 12, included in N15.0.1.

```
x:= surfrev[α]
x[u,v]
metric[x][u,v]//MatrixForm
```

```
unitnormal[x][u,v]
%//PowerExpand
```

```
shape[x][u,v]
PowerExpand[eee]
PowerExpand[ggg]
```

For a surface of revolution, the principal curvatures are given by

```
kp= eee/ee
km= ggg/gg
```

It is tempting to simplify these expressions using PowerExpand, but that would yield incorrect results when φ is negative, as the correct formulas for kp and km involve Abs[φ] and Sign[φ] respectively. In any case, we may verify the above ouput by computing Gaussian and mean curvature in two independent ways:

```
kp km//Simplify
gcurv[x][u,v]
```

```
(kp+km)/2
mcurv[x][u,v]
% - %%//Simplify
```

We next compare the curvature of the hyperboloid of one sheet and the catenoid.

```
x:= hyperboloid1[1,1,1]
gcurv[x][u,v]
mcurv[x][u,v]
%//PowerExpand
```

```
y:= catenoid[1]
gcurv[y][u,v]
mcurv[y][u,v]
```

■ N15.4

The definition of the generalized helicoid is

```
genhel[α_,c_][u_,v_]:=
  {Cos[u]α[v][[1]], Sin[u]α[v][[1]], c u + α[v][[2]]}
```

The next cells lead up to the construction of Figure 15.4 and follow almost exactly the same method that was adopted for the analogous Figure 15.1.

```
β[t_]:= {2 + Sin[2t]/2, t}
y:= genhel[β,.5]
z[2]= ParametricPlot3D[y[u,v], {u,0,3π/2}, {v,-π,π}]
```

```
f[u_]:= y[u,π]
g[u_]:= y[u,-π]
ff:= highlight[f]
gg:= highlight[g]
mer[v_]:= Table[y[u,v], {u,0,3π/2,π/16}]
y1:= ParametricPlot3D[Evaluate[mer[v]], {v,-π,π}]
y2:= ParametricPlot3D[Evaluate[{ff[u],gg[u]}], {u,0,3π/2}]
z[3]= Show[{y1,y2}]
```

The order in which z[1],z[2],z[3] are defined enables the parallels (defined below) to be displayed on the left, and meridians (defined above) on the right.

```
h[v_]:= y[0,v]
k[v_]:= y[3π/2,v]
hh:= highlight[h]
kk:= highlight[k]
par[u_]:= Table[y[u,v], {v,-π,π,π/12}]
y3:= ParametricPlot3D[Evaluate[par[u]], {u,0,3π/2}]
y4:= ParametricPlot3D[Evaluate[{hh[v],kk[v]}], {v,-π,π}]
z[1]= Show[{y3,y4}]
S @ GraphicsArray[Array[z, 3], GraphicsSpacing->-.2]
Clear[z]
```

Figure 15.4

We now work with the general twisted surface

```
y:= genhel[α, c]
y[u,v]
```

Lemma 15.18 is verified by output from the next two cells. A semicolon after shape prevents the matrix being displayed, as it is more helpful to view its components eee,fff,ggg line by line.

```
metric[y][u,v]//MatrixForm
dd= Sqrt[ee gg - ff^2]
```

```
shape[y][u,v];
eee
fff
ggg
```

```
dd^4 gcurv[y][u,v]//Simplify
2dd^3 mcurv[y][u,v]//Simplify//Expand
```

The "twisted" sphere has a parametrization

```
twisphere[a_,b_][u_,v_]:=
  {a Cos[u]Cos[v], a Sin[u]Cos[v], a Sin[v] + b u}
```

We shall compute its Gaussian curvature, and then extract the numerator for simplification.

```
gcurv[twisphere[a,b]][u,v]//Simplify
num= Part[%, 3]
```

```
num//TrigExpand
%/.{Cos[v]->c, Sin[v]->Sqrt[1-c^2]}//Simplify
```

Much less work is needed when `a=b`:

```
gcurv[twisphere[a,a]][u,v]
```

We define a matrix `A` to rotate the twisted sphere by $\pi/2$ in the xz-plane.

```
ts:= twisphere[1,1]
A:= {{0,0,-1},{0,1,0},{1,0,0}}
A.ts[u,v]
```

```
ParametricPlot3D[A.ts[u,v], {u,0,2π}, {v,-π,π}];
```

Here the twisted sphere is colored by Gaussian and mean curvature:

```
Module[{ts,f,y,z},
  ts:= twisphere[1,1];
  f[1]:= gcurv[ts][u,v];
  f[2]:= mcurv[ts][u,v];
  y[i_]:= Append[ts[u,v], Hue[f[i]]];
  z[i_]:= ParametricPlot3D[y[i], {u,0,2π}, {v,-π,π},
    Lighting->False, PlotPoints->{40,40}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->.7]
]
```

Figure 15.5

■ N15.5

Here is a selection of elliptic function graphs, which can be displayed by adding S@ at the start of each line.

```
Plot[Evaluate[Table[EllipticE[φ, m], {m,0,1,.2}]], {φ,-π/2,π/2}]
Plot[Evaluate[Table[-(I EllipticE[I φ,-m]), {m,0,1,.2}]], {φ,-.8,.8}]
Plot[Evaluate[Table[EllipticE[φ, m], {m,0,1,.2}]] , {φ,-π/2,4π}]
```

The curve in Theorem 15.19 is defined in terms of elliptic functions as

```
cpcprofile[pm_,a_,b_][t_]:=
  {pm b Cos[t/a], a EllipticE[t/a, b^2/a^2]}
```

The corresponding surface of revolution is therefore given by

```
cpcsurfrev[a_,b_][u_,v_]:=
  {b Cos[u]Cos[v/a], b Sin[u]Cos[v/a], a EllipticE[v/a, b^2/a^2]}
```

In order to display the extended profile curves for b<a in Figure 15.6, we set parameter range {t,-π,π}, and (subsequently) a plot range of {-2.4,2.4}.

```
ft[t_]:= Flatten[Table[{cpcprofile[1,1,s][t], cpcprofile[-1,1,s][t]},
  {s, .5, .9, .1}], 1]
zz1:= ParametricPlot[ft[t]//Evaluate, {t,-π,π},
  Ticks->{{-.4, .4}, Automatic}, PlotStyle->{Thickness[0.01]}]
```

```
ParametricPlot[ft[t]//Evaluate, {t,-π,π}]
```

The unique closed circular profile is shown in a different color.

```
pr[c_]:= {cpcprofile[1,1,c][t], cpcprofile[-1,1,c][t]}
z[c_,h_]:= ParametricPlot[pr[c]//Evaluate,
  {t, -ArcSin[1/c], ArcSin[1/c]},
   PlotStyle->{{Hue[h],Thickness[1/100]}}]
zz2:=  Show[Table[z[c,0], {c, 1.1, 1.5, .1}]]
S[{zz1, zz2, z[1,.5]}, ImageSize->300, PlotRange->{-2.4,2.4}]
```

Figure 15.6

```
yy1:= ParametricPlot3D[cpcsurfrev[1,1/2][u,v],
  {u,0,3π/2}, {v,-π/2,π/2}]
yy2:= ParametricPlot3D[cpcsurfrev[1,1.3][u,v],
  {u,0,3π/2}, {v,-ArcSin[1/1.3], ArcSin[1/1.3]}]
S @ {yy1, yy2}
```

Figure 15.7

■ N15.6

Here are the definitions of hyperbolic and conic type profile curves:

```
cnchyprofile[pm_,a_,b_][t_]:=
  {pm b Cosh[t/a], -I a EllipticE[I t/a,-b^2/a^2]}
cnccoprofile[pm_,a_,b_][t_]:=
  {pm b Sinh[t/a], -I Sqrt[a^2-b^2]EllipticE[I t/a,-b^2/(a^2-b^2)]}
```

For computational purposes, it is best to express the second coordinate in terms of *Mathematica*'s built-in elliptic functions. But the following example relates one of these to the appropriate integral.

```
Clear[a,b]
f[v_]:= Integrate[Sqrt[a^2 - b^2 Sinh[tt]^2], {tt,0,v/a},
  GenerateConditions->False]
f[v]//PowerExpand
f[0]
```

The GenerateConditions option is needed to stop the system worrying about the nature of a,b, though it turns out to be simpler to use an indefinite integral that satisfies the same initial condition:

```
g[v_]:= Integrate[Sqrt[a^2 - b^2 Sinh[tt]^2], tt]/.tt->v/a
g[v]//PowerExpand
g[0]
```

```
a:= 1;
b:= 1;
r:= .99 ArcSinh[1]
Plot[f[t], {t,-r,r}]
```

We can easily plot corresponding surfaces of revolution using the command surfrev from the start of N15.1:

```
Module[{x},
  x= surfrev[cnchyprofile[1,1,.1]];
  ParametricPlot3D[x[u,v], {u,0,3π/2},
    {v, -ArcSinh[1/0.1], ArcSinh[1/0.1]}]
]
```

The pseudosphere is obtained by revolving a tractrix:

```
pseudosphere[a_][u_,v_]:=
  a{Cos[u]Sin[v], Sin[u]Sin[v], Cos[v]+Log[Tan[v/2]]}
```

```
pseudosphere[a][u,v] - surfrev[tractrix[a]][u,v]
```

```
gcurv[pseudosphere[a]][u,v]
```

The hidden subsection below contains instructions for Figures 15.8 and 15.9.

■ N15.6.1

■ N15.7

Zero-curvature profiles curves for generalized helicoids can be generated by solving the relevant differential equation. For simplicity, we first take a=c=1:

```
DSolve[ψ'[t]==Sqrt[1-1/t^2], ψ[t], t]
```

We claim that the general solution is given by

```
zcprofile[pm_,a_,c_][t_]:=
  {t, pm(t Sqrt[a^2 - c^2/t^2] + c ArcSin[c/(a t)])}
```

Reconciling this expression with the differential equation above provides a model exercise in computer simplification. The latter is achieved in a 3-step process that begins with the Together command to express the derivative of the purported solution over a common denominator:

```
f:= zcprofile[1,a,c]
f'[t]//Together
%//PowerExpand
%//Simplify
```

A way to plot a solution that consists of separate branches for t negative and t positive is given by

```
z1:= ParametricPlot[Evaluate[zcprofile[-1,1,1][t]], {t,-2.5,-1}]
z2:= ParametricPlot[Evaluate[zcprofile[1,1,1][t]], {t, 1, 2.5}]
S[{z1,z2}, PlotRange->{{-3,3},{1,3.1}}, ImageSize->450]
```

Figure 15.11

Here is a flat generalized helicoid:

```
x:= genhel[zcprofile[1,a,c], c]
x[u,v]
```

It requires some effort to manipulate its Gaussian curvature to zero:

```
gcurv[x][u,v]
%//Together
%//PowerExpand
%//Simplify
```

The parametrization of Dini's surface is used to compute its curvature:

```
dini[a_,b_][u_,v_]:=
  {a Cos[u]Sin[v], a Sin[u]Sin[v], a(Cos[v]+Log[Tan[v/2]]) + b u}
```

```
gcurv[dini[a,b]][u,v]
mcurv[dini[a,b]][u,v]//PowerExpand
```

Here is Kuen's surface and its first fundamental form:

```
kuen[a_][u_,v_]:=
Module[{A,x,y,z},
  A= 1+u^2 Sin[v]^2;
  x= (Cos[u]+u Sin[u])Sin[v]/A;
  y= (Sin[u] - u Cos[u])Sin[v]/A;
  z= Log[Tan[v/2]]/2 + Cos[v]/A;
  a{x,y,z}
]
```

```
x:= kuen[1]
metric[x][u,v]
```

The next cell verifies that fff, like ff, is identically zero. Computing the matrix of the second fundamental form would take longer.

```
D[x[u,v],u,v].unitnormal[x][u,v]
%//Simplify
```

```
gcurv[kuen[a]][u,v]
mcurv[kuen[a]][u,v]
```

N15.7.1

N15.8

The definition in Exercise 10 translates into

```
gensurfrev[α_,γ_][u_,v_]:=
  {α[u][[1]]γ[v][[1]], α[u][[2]]γ[v][[1]], γ[v][[2]]}
```

This command is applied to determine a patch on the surface that results when both the curves α and γ are figure eights. We can then compute and plot the Gaussian curvature of this generalized surface of revolution.

```
x:= gensurfrev[eight,eight]
x[u,v]
```

```
gcurv[x][u,v]
K[u_,v_]:= gcurv[x][u,v]
```

```
ϵ:= .001
z1:= ParametricPlot3D[x[u,v], {u,0,2π}, {v,-π/2,π/2},
  PlotPoints->{48, 30}]
z2:= Plot3D[K[u,v]//Evaluate, {u,ϵ,2π-ϵ}, {v,-π/2,π/2},
  Axes->False]
S @ GraphicsArray[{z1,z2}]//Timing
```

Figure 15.15

A final subsection includes a definition of Sievert's surface and a verification that its Gaussian curvature is constant.

N15.8.1

Chapter 16

A Selection of Minimal Surfaces

The origins of minimal surface theory can be traced back to 1744 with Euler's paper [Euler2], and to 1760 with Lagrange's paper [Lag]. Euler showed that the catenoid is a minimal surface, and Lagrange wrote down the partial differential equation that must be satisfied for a surface of the form $z = f(x, y)$ to be minimal. In 1776, Meusnier[1] rediscovered the catenoid and also showed that the helicoid is a minimal surface [Meu]. The mathematical world had to wait over 50 years until other examples were found by Scherk. These include the surfaces now called "Scherk's minimal surface," "Scherk's fifth minimal surface," and a family of surfaces that includes both the catenoid and the helicoid.

The importance of minimal surfaces as those of least potential surface energy was illustrated by the experiments of Plateau,[2] who dipped wires in the form of space curves into a solution of soapy water and glycerin, thus realizing minimal surfaces experimentally. Plateau's problem is that of determining the minimal surfaces through a given curve. It has been formulated and studied in great generality by many mathematicians.

In this chapter, we begin a study of minimal surfaces by showing that a minimal surface is a critical point of the area function in an appropriate sense (Theorem 16.4). In Section 16.2, we provide the first of a series of new examples

[1] Jean Babtiste Meusnier de la Place (1754–1793). French mathematician, a student of Monge. He was a general in the revolutionary army and died of battle wounds.

[2] 

Joseph Antoine Ferdinand Plateau (1801–1883). Belgian physicist. His thesis concerned the impressions that light can have on the eye. Unfortunately, this led him to stare into the bright sun, which had an adverse effect on his sight. In 1840, he began a series of experiments with surfaces for which he became famous.

of minimal surfaces by investigating the celebrated deformation from the helicoid to the catenoid. This construction also provides examples of local isometries of surfaces, a concept defined in Chapter 12. In Section 16.3, we prove that any surface of revolution which is also a minimal surface is contained in a catenoid or a plane.

The minimal surfaces of Enneper, Catalan, and Henneberg are described in Section 16.4. The second is intimately related to a cycloid curve that it contains, and this example anticipates the definition of geodesic torsion in Section 17.4 and the theory of geodesics in Chapter 18. A discussion of Scherk's surface in Section 16.5 begins with the minimal surface equation for a Monge patch.

In Section 16.6 we show that the Gauss map of a minimal surface is conformal, an important result that lies at the heart of much modern research into minimal surfaces [Os1]. *Isothermal coordinates* for surfaces are introduced in Section 16.7, and provide the foundation for our further study of minimal surfaces using a complex variable in Chapter 22.

16.1 Normal Variation

On page 398, we defined a minimal surface in $\mathbb{R}^3$ as a surface whose mean curvature vanishes. This is Lagrange's 1760 definition. A more intuitive meaning of minimal surface is the surface of least area among a family of surfaces having the same boundary. In order to show how these two definitions coincide, we define the normal variation of a surface $\mathcal{M}$ in $\mathbb{R}^3$ to be a family of surfaces $t \mapsto \mathcal{M}(t)$ representing how $\mathcal{M}$ changes when pulled in a normal direction. Let $A(t)$ denote the area of $\mathcal{M}(t)$. We show that the mean curvature of $\mathcal{M}$ vanishes if and only if the first derivative of $t \mapsto A(t)$ vanishes at $\mathcal{M}$.

First, we make precise the notion of normal variation. The notion of a bounded subset was discussed on page 373, and the general formula for the area of a bounded region of a surface in $\mathbb{R}^n$ was given by (12.24). In the case $n = 3$, (12.22) on the foot of page 372 yields

Lemma 16.1. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$, and $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ a regular patch. For a bounded region $\mathcal{Q}$ in $\mathcal{U}$, the area of $\mathbf{x}(\mathcal{Q})$ is given by*

$$\mathsf{area}(\mathbf{x}(\mathcal{Q})) = \iint_{\mathcal{Q}} \|\mathbf{x}_u \times \mathbf{x}_v\| \, du dv.$$

We wish to study how area changes with a small perturbation of a surface. The simplest perturbation is one that is normal to the surface.

Definition 16.2. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch, and choose a bounded region $\mathcal{Q} \subset \mathcal{U}$. Suppose that $h\colon \mathcal{Q} \to \mathbb{R}$ is differentiable and $\varepsilon > 0$. Let $\mathbf{U}$ denote a unit*

vector field such that $\mathbf{U}(u,v)$ *is normal to* $\mathbf{x}(\mathcal{Q})$ *at* $\mathbf{x}(u,v)$ *for all* $(u,v) \in \mathcal{U}$. *Then the* ***normal variation*** *of* $\mathbf{x}$ *and* $\mathcal{Q}$, *determined by* h, *is the map*

$$\mathbf{X}\colon (-\varepsilon,\varepsilon) \times \mathcal{Q} \longrightarrow \mathbb{R}^3$$

given by

$$\mathbf{X}_t(u,v) = \mathbf{x}(u,v) + t\,h(u,v)\mathbf{U}(u,v) \tag{16.1}$$

for $(u,v) \in \mathcal{Q}$ *and* $-\varepsilon < t < \varepsilon$.

It follows from this definition that $\mathbf{X}_t$ is a patch for each t with $-\varepsilon < t < \varepsilon$ for sufficiently small ε. Let

$$\begin{cases} E(t) \;=\; (\mathbf{X}_t)_u \cdot (\mathbf{X}_t)_u, \\ F(t) \;=\; (\mathbf{X}_t)_u \cdot (\mathbf{X}_t)_v, \\ G(t) \;=\; (\mathbf{X}_t)_v \cdot (\mathbf{X}_t)_v. \end{cases} \tag{16.2}$$

Then $E = E(0)$, $F = F(0)$, $G = G(0)$, and by Lemma 16.1 the area of $\mathbf{X}_t(\mathcal{Q})$ is given by

$$A(t) = \iint_{\mathcal{Q}} \sqrt{E(t)G(t) - F(t)^2}\,du dv. \tag{16.3}$$

We compute the derivative at zero of the function $\mathcal{A}(t)$.

Lemma 16.3. We have

$$A'(0) = -2\iint_{\mathcal{Q}} h\,H\sqrt{EG - F^2}\,du dv, \tag{16.4}$$

where H *denotes the mean curvature of* $\mathcal{M}$.

Proof. Differentiating (16.1) with respect to u and v gives

$$\begin{cases} (\mathbf{X}_t)_u = \mathbf{x}_u + t\,h_u\mathbf{U} + t\,h\,\mathbf{U}_u, \\ (\mathbf{X}_t)_v = \mathbf{x}_v + t\,h_v\mathbf{U} + t\,h\,\mathbf{U}_v. \end{cases} \tag{16.5}$$

From (13.9), page 394, (16.5) and the definitions (16.2), it follows that

$$\begin{aligned} E(t) \;&=\; (\mathbf{X}_t)_u \cdot (\mathbf{X}_t)_u \\ &=\; (\mathbf{x}_u + t\,h_u\mathbf{U} + t\,h\,\mathbf{U}_u) \cdot (\mathbf{x}_u + t\,h_u\mathbf{U} + t\,h\,\mathbf{U}_u) \\ &=\; E + 2t\,h\,\mathbf{x}_u \cdot \mathbf{U}_u + O\big(t^2\big) \\ &=\; E - 2t\,h\,e + O\big(t^2\big)\,. \end{aligned} \tag{16.6}$$

Similarly,

$$\begin{aligned} F(t) \;&=\; F - 2t\,h\,f + O\big(t^2\big)\,, \\ G(t) \;&=\; G - 2t\,h\,g + O\big(t^2\big)\,. \end{aligned} \tag{16.7}$$

From (16.6), (16.7) and Theorem 13.25, page 400, we get

$$\begin{aligned} E(t)G(t) - F(t)^2 &= \big(E - 2the + O(t^2)\big)\big(G - 2thg + O(t^2)\big) \\ &\qquad - \big(F - 2thf + O(t^2)\big)^2 \\ &= EG - F^2 - 2th(Eg - 2Ff + Ge) + O(t^2) \\ &= \big(EG - F^2\big)(1 - 4thH) + O(t^2)\,, \end{aligned}$$

and hence

$$\begin{aligned} \sqrt{E(t)G(t) - F(t)^2} &= \sqrt{(EG - F^2)(1 - 4thH) + O(t^2)} \\ &= \sqrt{EG - F^2}\sqrt{1 - 4thH + O(t^2)} \\ &= \sqrt{EG - F^2}\big(1 - 2thH\big) + O(t^2)\,. \end{aligned}$$

Combined with (16.3), we obtain

$$\begin{aligned} A(t) &= \iint_{\mathcal{Q}} \Big(\sqrt{EG - F^2}\big(1 - 2thH\big) + O(t^2)\Big)\,du dv \\ &= \iint_{\mathcal{Q}} \sqrt{EG - F^2}\,du dv - 2t\iint_{\mathcal{Q}} hH\sqrt{EG - F^2}\,du dv + O(t^2)\,. \end{aligned}$$

When we differentiate with respect to t and evaluate the resulting expression at $t = 0$, we obtain (16.4). ∎

Theorem 16.4. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch, and choose a bounded region $\mathcal{Q} \subset \mathcal{U}$. Then $\mathbf{x}$ is minimal on $\mathcal{Q}$ if and only if $A'(0) = 0$ for a normal variation of $\mathbf{x}$ and $\mathcal{Q}$ with respect to any $h\colon \mathcal{Q} \to \mathbb{R}$.*

Proof. If H is identically zero for $\mathbf{x}$, then (16.4) implies that $A'(0) = 0$ for any h. Conversely, suppose that $A'(0) = 0$ for any differentiable function $h\colon \mathcal{Q} \to \mathbb{R}$ but that there is $\mathbf{q} \in \mathcal{Q}$ for which $H(\mathbf{q}) \neq 0$. Choose h such that $h(\mathbf{q}) = H(\mathbf{q})$, with h identically zero outside of a small neighborhood of $\mathbf{q}$ on which $hH \geqslant 0$. But now (16.4) implies that $A'(0) < 0$. This contradiction shows that $H(\mathbf{q}) = 0$. Since $\mathbf{q}$ is arbitrary, $\mathbf{x}$ is minimal. ∎

We have said nothing about the second derivative of A at 0, so that a minimal surface, although a critical point of A, may not actually be a *minimum*. More details of this approach, as well as extensions of the results of this section, can be found in [Nits, pages 90–116] and [Oprea2, chapter 4].

16.2 Deformation from the Helicoid to the Catenoid

We have already considered the helicoid on page 376 and the catenoid on page 464. In fact, these two surfaces constitute the initial and final points of a

deformation through isometric minimal surfaces. For each t with $0 \leqslant t \leqslant \pi/2$, define

$$\begin{aligned}\mathbf{z}[t](u,v) = \cos t\,\big(\sinh v \sin u,\ -\sinh v \cos u,\ u\big) \\ +\sin t\,\big(\cosh v \cos u,\ \cosh v \sin u,\ v\big).\end{aligned} \tag{16.8}$$

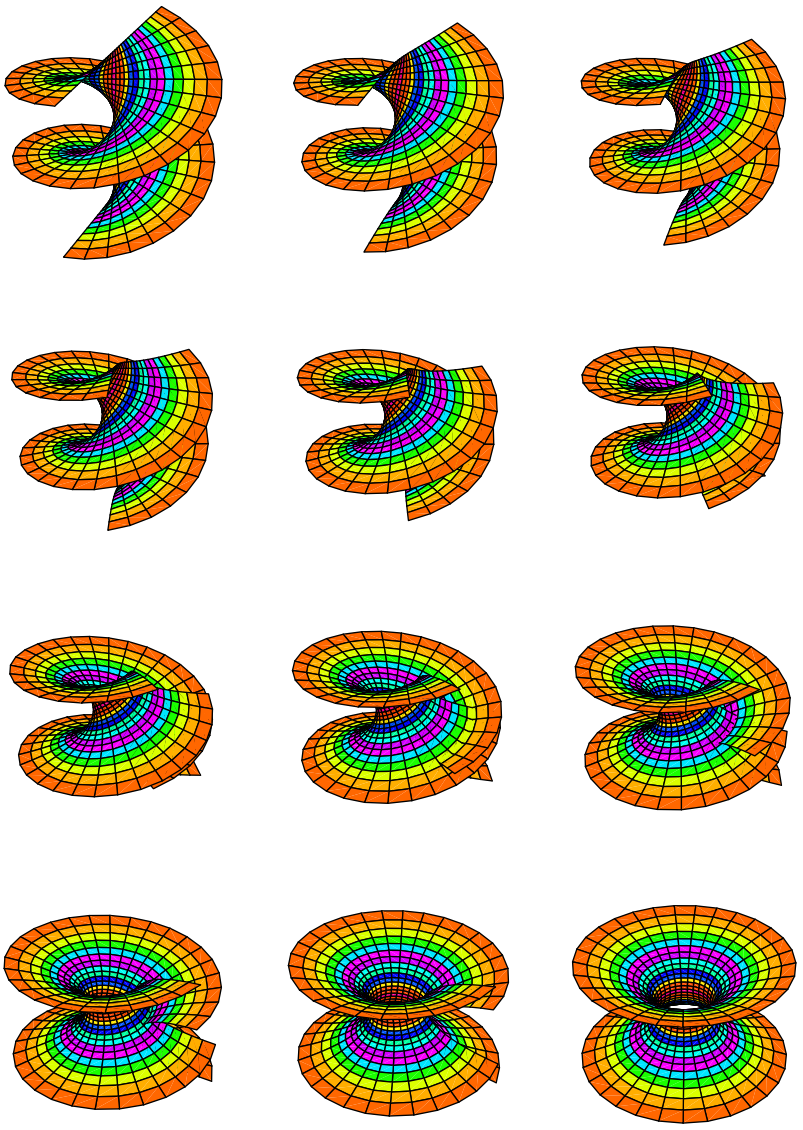


Figure 16.1: $\mathbf{z}\left[\dfrac{n\pi}{22}\right]$ with $0 \leqslant n \leqslant 11$

The exact relation between this family of surfaces and the helicoid and catenoid is provided by the equations

$$\begin{aligned} \mathbf{z}[0](u,v) &= \mathsf{helicoid}[1,1]\left(u-\frac{\pi}{2},\,\sinh v\right)+\left(0,0,\frac{\pi}{2}\right),\\ \mathbf{z}\big[\frac{\pi}{2}\big](u,v) &= \mathsf{catenoid}[1](u,v), \end{aligned}$$

verified in Notebook 16.

Theorem 16.5. *The 1-parameter family of surfaces* (16.8) *is a deformation from the helicoid to the catenoid such that* $\mathbf{z}[0]$ *is* (*a reparametrization of*) *a helicoid and* $\mathbf{z}[\pi/2]$ *is a catenoid. Furthermore, each* $\mathbf{z}[t]$ *is a minimal surface which is locally isometric to* $\mathbf{z}[0]$. *In particular, the helicoid is locally isometric to the catenoid.*

Proof. Let $E(t), F(t), G(t)$ denote the coefficients of the first fundamental form of $\mathbf{z}[t]$. An easy calculation (by hand or with the help of Notebook 16) shows that

$$E(t) = \cosh^2 v = G(t), \qquad F(t) = 0.$$

In particular, $E(t), F(t), G(t)$ are constant functions of t. The result follows from Lemma 12.7 on page 366. ▌

It is now an easy matter to plot accurate pictures of the deformation $t \mapsto \mathbf{z}[t]$ like those of Figure 16.1, that are colored by a function of Gaussian curvature. The immersion $\mathbf{z}[t]$ obviously does depend on t; mathematically, this is actually a special case of a subsequent result on page 724. Neither the catenoid nor the helicoid intersects itself; however, every intermediate surface $\mathbf{z}[t]$ has self-intersections. It is also true that asymptotic curves on the helicoid are gradually transformed into principal curves on the catenoid (see page 390 for the definitions, and page 725 for the result).

Figure 16.2 illustrates two particular regions of the helicoid and catenoid that correspond under the deformation. The boundaries of these regions include a helix and a circle, with respective parametrizations

$$\begin{aligned} \boldsymbol{\alpha}(u) &= \big(\sinh 1 \sin u,\ -\sinh 1 \cos u,\ u\,\big), \qquad 0 \leqslant u < 2\pi,\\ \boldsymbol{\gamma}(u) &= \big(\cosh 1 \cos u,\ \cosh 1 \sin u,\ 1\,\big), \qquad 0 \leqslant u < 2\pi. \end{aligned}$$

The length of the helix is

$$\int_0^{2\pi} |\boldsymbol{\alpha}'(u)|du = 2\pi\sqrt{\sinh^2 1 + 1},$$

and that of the circle is

$$\int_0^{2\pi} |\boldsymbol{\gamma}'(u)|\,du = 2\pi \cosh 1.$$

The fact that these two lengths are equal is consistent with the statement that the helix maps isometrically onto the circle.

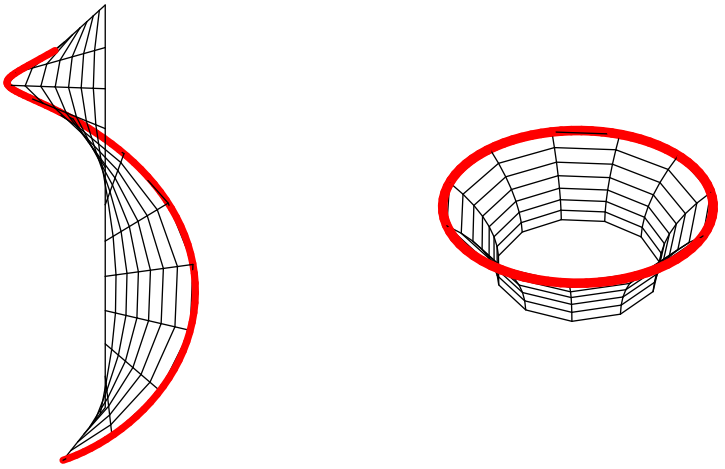


Figure 16.2: Isometric regions of the helicoid and catenoid

16.3 Minimal Surfaces of Revolution

The catenoid is the only member of the family $\mathbf{z}[t]$ that is a surface of revolution. We next prove that the catenoid and the plane are in fact essentially the only surfaces of revolution that are simultaneously minimal surfaces. The following result is therefore an analogue of Exercise 7 of Chapter 15, that dealt with the case of zero *Gaussian* curvature.

Theorem 16.6. *A surface of revolution $\mathcal{M}$ which is a minimal surface is contained in either a plane or a catenoid.*

Proof. Let $\mathbf{x}$ be a patch whose trace is contained in $\mathcal{M}$, and let $\boldsymbol{\alpha} = (\varphi, \psi)$ be the profile curve. Then $\mathbf{x}$ is given by (15.1). There are three cases:

Case 1. ψ' is identically 0. Then ψ is constant, so that $\boldsymbol{\alpha}$ is a horizontal line and $\mathcal{M}$ is part of a plane perpendicular to the axis of revolution.

Case 2. ψ' is never 0. Then by the Inverse Function Theorem, ψ has an inverse ψ^{-1}. Define

$$\widetilde{\boldsymbol{\alpha}}(t) = \boldsymbol{\alpha}\big(\psi^{-1}(t)\big) = \big(h(t),\ t\big),$$

where $h = \varphi \circ \psi^{-1}$, and a new patch $\mathbf{y}$ by

$$\mathbf{y}(u, v) = \big(h(v)\cos u,\ h(v)\sin u,\ v\big).$$

Since $\widetilde{\boldsymbol{\alpha}}$ is a reparametrization of $\boldsymbol{\alpha}$, it follows that $\mathbf{x}$ and $\mathbf{y}$ have the same trace. Thus it suffices to show that the surface of revolution $\mathbf{y}$ is part of a catenoid.

In the notation of page 469, equations (15.11) reduce to

$$\text{(16.9)} \qquad \begin{cases} k_p = \dfrac{e}{E} = \dfrac{-1}{h\sqrt{h'^2+1}}, \\ k_m = \dfrac{g}{G} = \dfrac{h''}{(h'^2+1)^{3/2}}. \end{cases}$$

It follows from the assumption that $H = 0$ and (16.9) that h must satisfy the differential equation

$$\text{(16.10)} \qquad h''h = 1 + h'^2.$$

To solve (16.10), we first rewrite it as

$$\frac{2h'h''}{1+h'^2} = \frac{2h'}{h}.$$

Integrating both sides yields

$$\log(1+h'^2) = \log(h^2) - \log(c^2)$$

for some constant $c \neq 0$, and exponentiating we obtain

$$\text{(16.11)} \qquad 1 + h'^2 = \Big(\frac{h}{c}\Big)^2.$$

The first-order differential equation (16.11) can be written as

$$\text{(16.12)} \qquad \frac{h'/c}{\sqrt{(h/c)^2 - 1}} = \frac{1}{c}.$$

Both sides of (16.12) can be integrated to yield

$$\operatorname{arccosh}\frac{h}{c} = \frac{v}{c} + b.$$

Thus the solution of (16.10) is

$$h(v) = c\cosh\Big(\frac{v}{c} + b\Big),$$

and so $\mathcal{M}$ is part of a catenoid.

Case 3. ψ' is zero at some points, but nonzero at others. In fact, this case cannot occur. Suppose, for example, that $\psi'(v_0) = 0$, but $\psi'(v) > 0$ for $v < v_0$. By Case 2, the profile curve is a catenary for $v < v_0$, whose slope is given by φ'/ψ'. Then $\psi'(v_0) = 0$ implies that the slope becomes infinite at v_0. But this is impossible, since the profile curve is the graph of the function $\cosh$. ▮

Nicely complementing Theorem 16.6 is a theorem due to Catalan, which asserts that the *helicoid* is the only minimal surface other than the plane which is also ruled (see [dC1]).

16.4 More Examples of Minimal Surfaces

Enneper's Minimal Surface

One of the simplest minimal surfaces is the one found by Enneper in 1864 (see [Enn1]). It is defined by

$$\mathsf{enneper}(u,v) = \left(u - \frac{u^3}{3} + uv^2,\ -v + \frac{v^3}{3} - vu^2,\ u^2 - v^2\right).$$

Determining its first and second fundamental forms was the object of Exercise 6 of Chapter 12 (see Figure 12.13 on page 378) and Exercise 4 of Chapter 13.

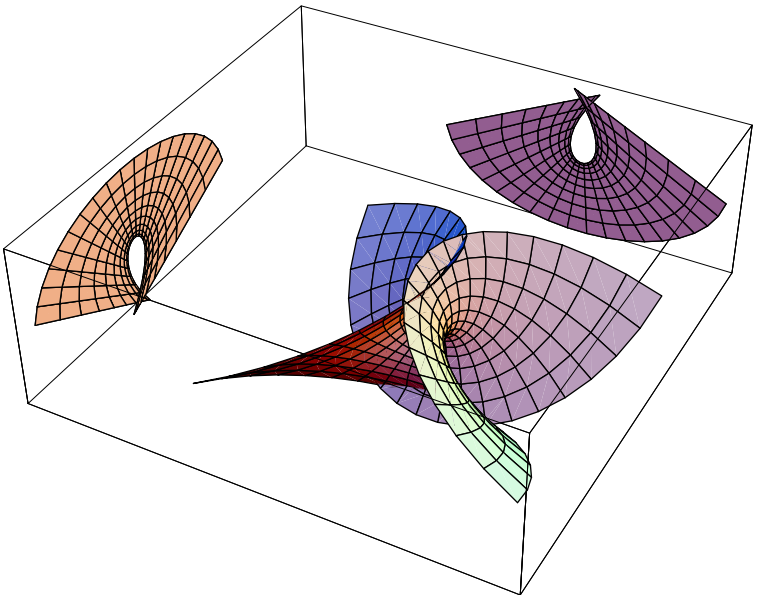


Figure 16.3: Enneper's surface with shadows

It is easy to verify directly that H is identically zero. Firstly,

$$\begin{aligned}\mathbf{x}_u &= (1 - u^2 + v^2,\ -2uv,\ 2u),\\ \mathbf{x}_v &= (2uv,\ -1 - u^2 + v^2,\ -2v),\end{aligned}$$

whence

$$E = \left(1 + u^2 + v^2\right)^2 = G, \qquad F = 0. \tag{16.13}$$

Setting $\rho = 1 + u^2 + v^2$ helps further calculations, such as

$$\mathbf{x}_u \times \mathbf{x}_v = \left(2u\rho,\ 2v\rho,\ \rho^2 - 2\rho\right).$$

However, we do not need this normal vector, since the vanishing of the mean curvature already follows from (16.13) and the equations

$$\mathbf{x}_{uu} = -2(u,\ v,\ -1) = -\mathbf{x}_{vv}.$$

In spite of the simplicity of its definition, Enneper's surface is complicated because of self-intersections. These are best visualized by projecting the surface to the coordinate planes as shown in Figure 16.3.

Catalan's Minimal Surface

The minimal surface of Catalan[3] is parametrized by

$$\mathbf{x}(u,v) = \Big(u - \sin u \cosh v,\ 1 - \cos u \cosh v,\ -4\sin\frac{u}{2}\sinh\frac{v}{2}\Big). \tag{16.14}$$

We defer to Notebook 16 for the verification that the mean curvature H is identically zero.

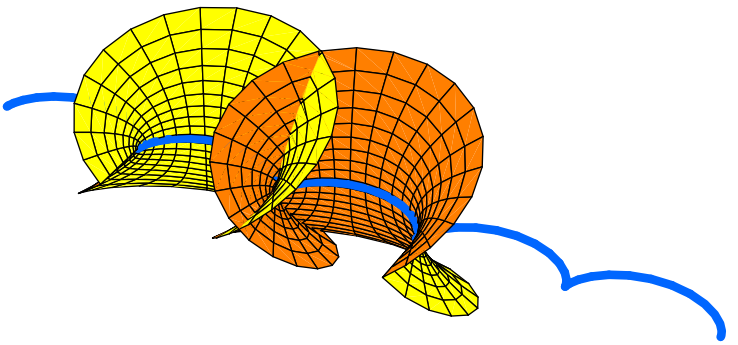


Figure 16.4: Catalan's minimal surface

The intersection of the plane $z = 0$ with the trace of (16.14) contains the coordinate curve

$$\boldsymbol{\alpha}(u) = \mathbf{x}(u,0) = \big(u - \sin u,\ 1 - \cos u,\ 0\big), \tag{16.15}$$

that defines a cycloid in the xy-plane. Much of the interest in Catalan's surface derives from its properties in relation to this curve, which we now investigate. We denote by s an arc length function along the curve, whose speed is therefore

$$\frac{ds}{du} = |\boldsymbol{\alpha}'(u)| = \sqrt{2 - 2\cos u} = 2\Big|\sin\frac{u}{2}\Big|. \tag{16.16}$$



3 Eugène Charles Catalan (1814–1894). Belgian mathematician, who had difficulty obtaining a position in France because of his left-wing views. Catalan's constant is

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} \approx 0.915966.$$

One can use this to write down s as a function of u, though we shall not need an explicit formula.

Consider the vectors $\mathbf{x}_u$, $\mathbf{x}_v$ tangent to the surface at points of the curve. The first is the tangent vector to the curve itself, since

$$\mathbf{x}_u(u,0) = \boldsymbol{\alpha}'(u) = \big(1-\cos u,\ \sin u,\ 0\big).$$

By contrast,

$$\mathbf{x}_v(u,0) = \Big(0,\ 0,\ -2\sin\frac{u}{2}\Big),$$

and so this second vector is orthogonal to the plane containing the trace of $\boldsymbol{\alpha}$. The acceleration vector $\boldsymbol{\alpha}''(s)$ is perpendicular to

$$\boldsymbol{\alpha}'(s) = \boldsymbol{\alpha}'(u)\frac{du}{ds},$$

since the latter has constant norm. But $\boldsymbol{\alpha}''(s)$ must lie in the xy-plane, and is thus orthogonal to $\mathbf{x}_v$. We have therefore established the next result without any real calculation.

Lemma 16.7. *The acceleration $\boldsymbol{\alpha}''(s)$ of the curve (16.15) is everywhere parallel to the normal vector $\mathbf{x}_u \times \mathbf{x}_v$ of the surface (16.14).*

In general, given a curve $\boldsymbol{\alpha}$ whose trace lies on a surface, the tangential component of $\boldsymbol{\alpha}''(s)$ defines the so-called ***geodesic curvature***, which in the next chapter we prove depends only on the first fundamental form. The curve $\boldsymbol{\alpha}$ is called ***pregeodesic*** if its acceleration vector $\boldsymbol{\alpha}''(s)$ satisfies the conclusion of Lemma 16.7. The term ***geodesic*** is reserved for the case in which the curve's parameter is a constant times arc length; for then the mapping $s \mapsto \boldsymbol{\alpha}(s)$ satisfies the so-called geodesic equations that are further explained in Chapter 18.

An interesting problem in minimal surface theory is the determination of a minimal surface that contains a given curve as a geodesic (see Section 22.6) or as an asymptotic or principal curve (see page 390 and [Nits, page 140]).

Henneberg's Minimal Surface

The following elementary parametrization of the minimal surface of Henneberg[4] is given on page 144 of [Nits]:

$$\mathsf{henneberg}(u,v) = \Big(2\sinh u\cos v - \tfrac{2}{3}\sinh 3u\cos 3v,\\ 2\sinh u\sin v + \tfrac{2}{3}\sinh 3u\sin 3v,\ 2\cosh 2u\cos 2v\Big).$$



4 Ernst Lebrecht Henneberg (1850–1922). German mathematician. Professor at the University of Darmstadt.

This patch fails to be regular precisely at the points $(0, n\pi/2)$, where n is an integer. Note that (16.4) is periodic in v with period 2π, and that

$$\mathsf{henneberg}(u, v) = \mathsf{henneberg}(-u, v + \pi). \tag{16.17}$$

Hence for any region $\mathcal{U}$ in the right half-plane $\{\,(p, q) \mid p > 0\,\}$ there will be a region $\widetilde{\mathcal{U}}$ in the left half-plane $\{\,(p, q) \mid p < 0\,\}$ whose image under henneberg is the same subset of $\mathbb{R}^3$. But (16.17) implies that the unit normal $\mathbf{U}$ of henneberg satisfies

$$\mathbf{U}(u, v) = -\mathbf{U}(-u, v + \pi).$$

Hence $\mathsf{henneberg}(\mathcal{U})$ and $\mathsf{henneberg}(\widetilde{\mathcal{U}})$ will have opposite orientations. An instance of this is illustrated by the different shadings on the left and right of Figure 16.5.

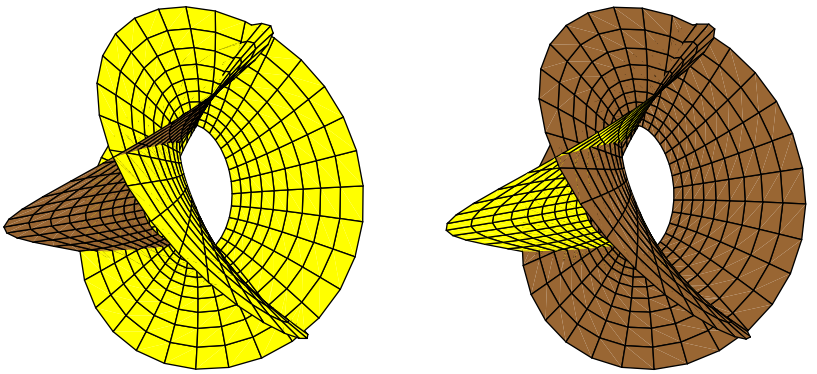


Figure 16.5: The henneberg images of $\{(p, q) \mid \frac{1}{3} \leqslant |p| \leqslant \frac{5}{6},\ -\pi \leqslant q < \pi\,\}$

In Chapter 22, we shall show how to associate to a minimal surface $\mathcal{M}$ a ***conjugate minimal surface*** $\widetilde{\mathcal{M}}$, and an isometric deformation $\mathbf{z}[t]$ from $\mathcal{M}$ to $\widetilde{\mathcal{M}}$ through minimal surfaces that generalizes equation (16.8). It turns out that the conjugate of Henneberg's surface is given by

$$\begin{aligned}\mathsf{hennebergconj}(u, v) = \Big(&2\cosh u \sin v - \tfrac{2}{3}\cosh 3u \sin 3v,\\ &2\cosh u \cos v + \tfrac{2}{3}\cosh 3u \cos 3v,\ 2\sinh 2u \sin 2v\Big),\end{aligned}$$

with a parametrization that is obtained from henneberg by merely interchanging

$$\sinh u \leftrightarrow \cosh u \qquad \text{and} \qquad \sin v \leftrightarrow \cos v$$

(see Exercise 12 on page 754). The resulting surface is apparently more complicated to visualize (Figure 16.6). Notebook 16 contains a verification that its mean curvature vanishes.

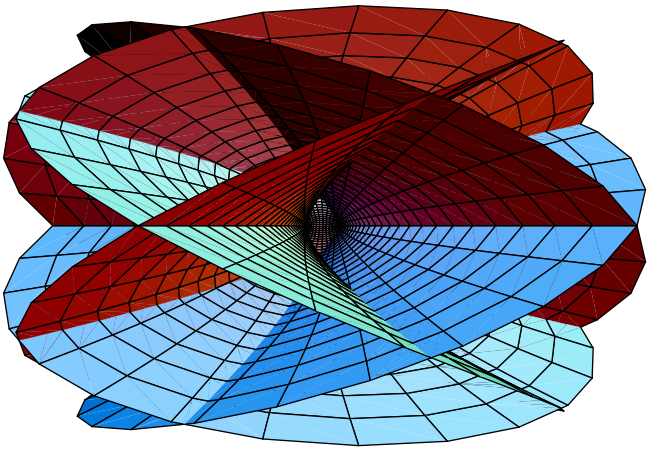


Figure 16.6: A realization of the conjugate Henneberg minimal surface

16.5 Monge Patches and Scherk's Minimal Surface

We have the following immediate consequence of Lemma 13.34, page 409:

Lemma 16.8. *A Monge patch* $(u,v) \mapsto (u,v,h(u,v))$ *is a minimal surface if and only if*

$$(1+h_v^2)h_{uu} - 2h_u h_v h_{uv} + (1+h_u^2)h_{vv} = 0. \tag{16.18}$$

One can hope to find interesting examples of minimal surfaces by assuming that h has a special form. In 1835 Scherk[5] determined the minimal surfaces of the form

$$(u,v) \mapsto \big(u,v,f(u)+g(v)\big)$$

(see [Scherk], [BaCo]).

Theorem 16.9. *If a Monge patch* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *with* $h(u,v) = f(u)+g(v)$ *is a minimal surface, then either* $\mathcal{M}$ *is part of a plane or there are constants* a, c_1, c_2, c_3, c_4*, with* $a \neq 0$*, such that*

$$\begin{aligned} f(u) &= -\frac{1}{a}\log\big(\cos(au+c_1)\big) + c_2, \\ g(v) &= \frac{1}{a}\log\big(\cos(av+c_3)\big) + c_4. \end{aligned} \tag{16.19}$$

[5]Heinrich Ferdinand Scherk (1798–1885). German mathematician, who also worked in number theory. He studied with Bessel in Königsberg, and was elected rector of the University of Kiel three times, but was forced to leave in 1848 by the Danes.

Proof. When $h(u,v) = f(u) + g(v)$, we have

$$h_{uu} = f''(u), \qquad h_{uv} = 0, \qquad h_{vv} = g''(v).$$

Hence equation (16.18) reduces to

$$\frac{f''(u)}{1+f'(u)^2} = \frac{-g''(v)}{1+g'(v)^2}. \tag{16.20}$$

Here u and v are independent variables, so each side of (16.20) must equal a constant, call it a. If $a = 0$, then both f and g are linear, so that $\mathcal{M}$ is part of a plane. Otherwise, the two equations

$$\frac{f''(u)}{1+f'(u)^2} = a = \frac{-g''(v)}{1+g'(v)^2} \tag{16.21}$$

are easily solved by integrating twice. The result is (16.19). ∎

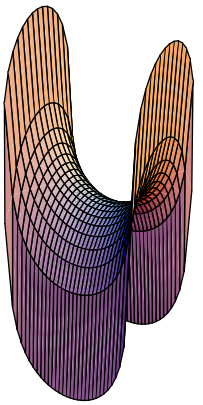


Figure 16.7: A piece of Scherk's minimal surface

As suggested by Theorem 16.9, we define ***Scherk's minimal surface*** by

$$\mathsf{scherk}[a](u,v) = \left(u,\ v,\ \frac{1}{a}\log\left(\frac{\cos a\,v}{\cos a\,u}\right)\right).$$

For simplicity, we also take $a = 1$. Figure 16.7 shows the image by this patch of (almost) the open rectangle $-\pi/2 < u, v < \pi/2$. In fact, $\mathsf{scherk}[1](u,v)$ is well defined on the set

$$\mathcal{R} = \big\{\ (u,v)\ \mid\ \cos u \cos v > 0\big\}$$

that can be imagined as the union of the black squares on an infinite chess board. To see this, let m, n be integers and consider a single square

$$\mathcal{Q}(m,n) = \left\{\ (x,y)\ \middle|\ m\pi - \frac{\pi}{2} < x < m\pi + \frac{\pi}{2},\quad n\pi - \frac{\pi}{2} < y < n\pi + \frac{\pi}{2}\ \right\}$$

that we color black if $m+n$ is even and white if $m+n$ is odd. Then

$$\mathcal{R} = \bigcup \{\, \mathcal{Q}(m,n) \mid m \text{ and } n \text{ are integers with } m+n \text{ even} \,\}.$$

It is easy to see that

$$\mathsf{scherk}[1](u + 2m\pi, v + 2n\pi) = \mathsf{scherk}[1](u,v)$$

for all real u and v and all integers m and n. Hence the piece of Scherk's surface over a black square $\mathcal{Q}(m,n)$, $m+n$ even, is a translate of the portion over $\mathcal{Q}(0,0)$. We can now plot identical pieces of the surface over the black squares, and these fit together to produce the entire surface.

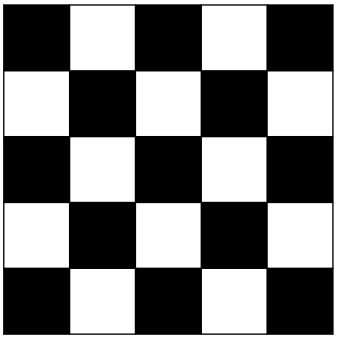


Figure 16.8: Chequered board

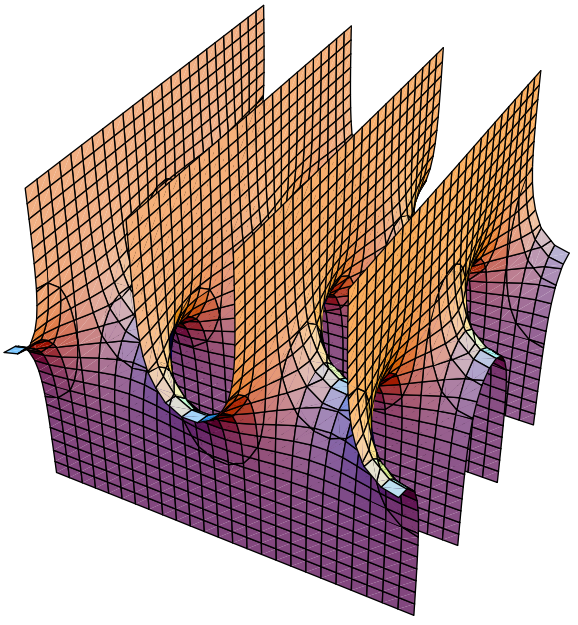


Figure 16.9: Scherk's minimal surface

Actually, Figure 16.9 was obtained in Notebook 16 from an implicit equation of Scherk's surface.

16.6 The Gauss Map of a Minimal Surface

Orientability was defined in Section 11.1 using the notion of a *complex structure* J, and this led to Definition 11.3 of the Gauss map on page 333. We next show that the Gauss map of a *minimal* surface has special properties that leads one to use complex analysis to advantage in the study of minimal surfaces. First, we need the following notion:

Definition 16.10. *Let $\mathcal{M}_1, \mathcal{M}_2$ be oriented regular surfaces in $\mathbb{R}^n$, and let J_1, J_2 be the corresponding complex structures. A map $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ is called*

(i) *a* ***complex map*** *if*

$$\Phi_* \circ J_1 = J_2 \circ \Phi_*. \tag{16.22}$$

(ii) *an* ***anticomplex map*** *if*

$$\Phi_* \circ J_1 = -J_2 \circ \Phi_*. \tag{16.23}$$

An elementary but important property of complex and anticomplex maps (also called ***holomorphic*** and ***antiholomorphic maps***) is

Lemma 16.11. *Let $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a complex or anticomplex map for which Φ_* is nowhere zero. Then Φ is a conformal map.*

Proof. Let $\mathbf{p} \in \mathcal{M}_1$. Choose a nonzero tangent vector $\mathbf{v_p}$ to $\mathcal{M}_1$ at $\mathbf{p}$. We can assume that $\|\Phi_*(\mathbf{v_p})\| = \lambda\|\mathbf{v_p}\|$ for some $\lambda > 0$. Since Φ is complex or anticomplex, abbreviating $(J_1)_\mathbf{p}$ to $J_\mathbf{p}$, we have

$$\|\Phi_*(J_\mathbf{p}\mathbf{v_p})\| = \lambda\|\mathbf{v_p}\| \qquad \text{and} \quad \Phi_*(\mathbf{v_p}) \cdot \Phi_*(J_\mathbf{p}\mathbf{v_p}) = 0.$$

It now follows that

$$\|\Phi_*(a\mathbf{v_p} + bJ_\mathbf{p}\mathbf{v_p})\|^2 = \lambda^2\|a\mathbf{v_p} + bJ_\mathbf{p}\mathbf{v_p}\|^2$$

for all $a, b \in \mathbb{R}$. Since $\{\mathbf{v_p}, J_\mathbf{p}\mathbf{v_p}\}$ is a basis of $\mathcal{M}_\mathbf{p}$, we can set $\lambda = \lambda(\mathbf{p})$ in Definition 12.11 on page 370 to deduce that Φ is conformal. ∎

We have a ready-made example of an anticomplex map.

Theorem 16.12. *The Gauss map of an oriented minimal surface $\mathcal{M} \subset \mathbb{R}^3$ is anticomplex, and*

$$J_\mathbf{p} S_\mathbf{p} = -S_\mathbf{p} J_\mathbf{p} \tag{16.24}$$

for all $\mathbf{p} \in \mathcal{M}$, where $S_\mathbf{p}$ is the shape operator of $\mathcal{M}$ at $\mathbf{p}$.

Proof. Let $\mathbf{p} \mapsto \mathbf{U}(\mathbf{p})$ be the Gauss map of $\mathcal{M}$. Let $\mathbf{p} \in \mathcal{M}$, and let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be an orthonormal basis of $\mathcal{M}_\mathbf{p}$ which diagonalizes $S_\mathbf{p}$, and let k_1, k_2 be the corresponding principal curvatures. The unit vector $\mathbf{U}_\mathbf{p}$ determines a complex structure $J_\mathbf{p}$ on $\mathcal{M}_\mathbf{p}$, via (11.1) on page 332, that satisfies $J_\mathbf{p}\mathbf{e}_1 = \pm\mathbf{e}_2$ and $J_\mathbf{p}\mathbf{e}_2 = \mp\mathbf{e}_1$. We use the fact that $\mathcal{M}$ is a minimal surface, and the resulting equation $k_1 = -k_2$, to compute

$$J_\mathbf{p} S_\mathbf{p}\mathbf{e}_1 = J_\mathbf{p} k_1 \mathbf{e}_1 = \pm k_1 \mathbf{e}_2 = \mp k_2 \mathbf{e}_2 = \mp S_\mathbf{p}\mathbf{e}_2 = -S_\mathbf{p} J_\mathbf{p}\mathbf{e}_1.$$

Similarly, $J_\mathbf{p} S_\mathbf{p}\mathbf{e}_2 = -S_\mathbf{p} J_\mathbf{p}\mathbf{e}_2$. In this way, we have established (16.24). But Lemma 13.5, on page 388, tells us that $S_\mathbf{p}$ is the negative of the tangent map of $\mathbf{U}$ at $\mathbf{p}$. Hence (16.24) implies that the Gauss map is anticomplex. ∎

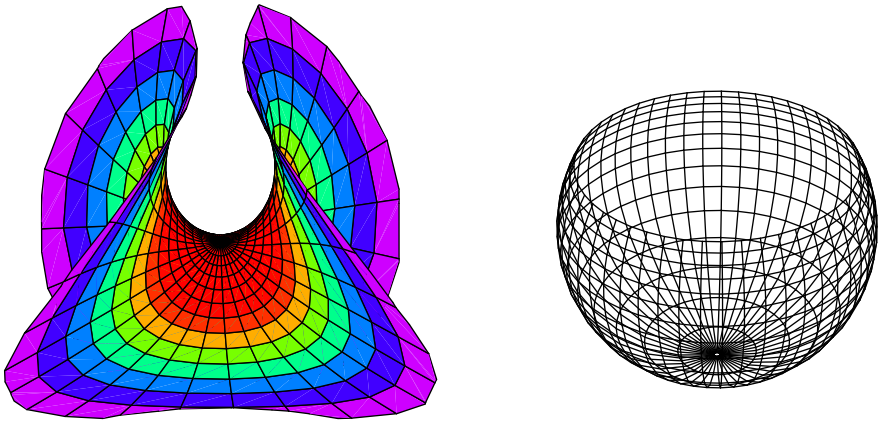


Figure 16.10: Enneper's surface and its Gauss image

The shape operator $S_\mathbf{p}$ at a point of $\mathcal{M}$ vanishes if and only if $k_1 = 0 = k_2$, so that (by definition) $\mathbf{p}$ is a planar point. On a minimal surface this is equivalent to asserting that the Gaussian curvature $K(\mathbf{p})$ is zero. Strictly speaking, we need to exclude such points when applying Lemma 16.11 to the conclusion of Theorem 16.12. We then obtain

Corollary 16.13. *The Gauss map of a minimal surface without planar points is conformal.*

Conformality means that the Gauss map preserves the proportions of infinitesimally small rectangles. When we plot a surface in the usual way with a sufficiently fine mesh, it is divided into lots of small, approximately rectangular shapes. The ratios of the lengths of the sides of these rectangles, and also the angles between the sides must be approximately preserved by the Gauss map. This phenomenon can be examined in Figure 16.10 which plots Enneper's surface along with its image under the Gauss map.

16.7 Isothermal Coordinates

In this final section, we shall define an important type of patch that can be found on an arbitrary surface. It has a special significance for minimal surfaces, and will play an important role in the subsequent study in Chapter 22 in which we shall make more explicit the presence of complex structures.

Definition 16.14. *Let $\mathcal{U} \subseteq \mathbb{R}^2$ be an open subset. A patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is called* ***isothermal*** *if there exists a differentiable function $\lambda\colon \mathcal{U} \to \mathbb{R}$ such that*

$$\mathbf{x}_u \cdot \mathbf{x}_u = \mathbf{x}_v \cdot \mathbf{x}_v = \lambda^2 \qquad \text{and} \qquad \mathbf{x}_u \cdot \mathbf{x}_v = 0. \tag{16.25}$$

We call λ the ***scaling function*** *of the isothermal patch.*

Note that such a patch is regular at all points where λ is nonzero.

The intuitive meaning of an isothermal patch $\mathbf{x}$ can be described as follows. Since $\mathbf{x}_u$ and $\mathbf{x}_v$ have the same length and are orthogonal, an isothermal patch maps an infinitesimal square in $\mathcal{U}$ into an infinitesimal square on its image. A more general patch would transform an infinitesimal square into an infinitesimal quadrilateral. The name "isothermal" is due to Lamé,[6] and dates from 1833.

The following lemma is obvious from the definition of conformal map on page 370.

Lemma 16.15. *A regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is isothermal if and only if it is conformal, considered as a mapping $\mathcal{U} \to \mathbf{x}(\mathcal{U})$.*

Let us first deal with the case of the sphere. Its standard parametrization, given on page 288, is

$$\mathbf{x}(u, v) = a\big(\cos u \cos v,\ \sin u \cos v,\ \sin v\big). \tag{16.26}$$

The metric of this patch was computed to be

$$ds^2 = a^2\big(\cos^2 v\, du^2 + dv^2\big) \tag{16.27}$$

(this is equation (12.26) on page 374), so the patch is certainly *not* isothermal. As a hint of what is in store in Chapter 22, we use complex numbers to factor the right-hand side of (16.27) so as to obtain

$$ds^2 = a^2 \cos^2 v \left(du + \frac{i\,dv}{\cos v}\right)\left(du - \frac{i\,dv}{\cos v}\right). \tag{16.28}$$



6 Gabriel Lamé (1795–1870). French engineer, mathematician and physicist. He worked on a wide variety of topics, such as number theory, differential geometry (we shall mention his work on triply orthogonal systems in Chapter 19), elasticity (where two elastic constants are named after him) and diffusion in crystalline material.

This has the effect of partially separating the variables, and if we set $w = \log(\tan v + \sec v)$, then

$$dw = \frac{dv}{\cos v}.$$

Moreover,

$$e^w + e^{-w} = \frac{\sin v + 1}{\cos v} + \frac{\cos v}{\sin v + 1} = \frac{2}{\cos v},$$

so that $\cos v = \operatorname{sech} w$. Thus, (16.28) becomes

$$ds^2 = \lambda^2(du^2 + dw^2)$$

with $\lambda = a \operatorname{sech} w$. In summary,

Lemma 16.16. *The patch*

$$\mathbf{y}(u, w) = a\big(\cos u \operatorname{sech} w,\ \sin u \operatorname{sech} w,\ \tanh w\big) \tag{16.29}$$

is an isothermal patch on a sphere $S^2(a)$ of radius a.

This patch is a special case of Mercator's parametrization of the ellipsoid given in Exercise 6 of Chapter 10. But a computation in Notebook 16 confirms that it is only isothermal in the spherical case.

It follows from Corollary 16.13 and Lemma 16.16 that the composition

$$\mathbf{p} \mapsto \mathbf{U}(\mathbf{p}) \mapsto \mathbf{y}^{-1}(\mathbf{p}),$$

defined on a suitable open subset of $\mathcal{M}$, is conformal. Its inverse will define an isothermal patch on the minimal surface $\mathcal{M}$. The following well-known result actually guarantees the existence of isothermal patches on an arbitrary surface.

Theorem 16.17. *Let $\mathcal{M}$ be a surface, and suppose ds^2 is a metric on $\mathcal{M}$. Let $\mathbf{p} \in \mathcal{M}$. Then there exists an open set $\mathcal{U} \subset \mathbb{R}^2$ and an isothermal patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ such that $\mathbf{p} \in \mathbf{x}(\mathcal{U})$ and*

$$ds^2 = \lambda^2(du^2 + dv^2).$$

For a proof, see [Bers, page 15], [Os1, page 31] or [FoTu, page 54]. It is also possible to argue in the way that led to Lemma 16.16. For we can factor an arbitrary metric $ds^2 = E\,dp^2 + 2F\,dp\,dq + G\,dq^2$ as

$$ds^2 = \left(\sqrt{E}\,dp + \frac{F + i\sqrt{EG - F^2}}{\sqrt{E}}dq\right)\left(\sqrt{E}\,dp + \frac{F - i\sqrt{EG - F^2}}{\sqrt{E}}dq\right).$$

One can then establish the existence of a complex-valued integrating factor μ for which the differential system

$$\mu(du + i\,dv) = \sqrt{E}\,dp + \frac{F + i\sqrt{EG - F^2}}{\sqrt{E}}dq$$

can be resolved for real-valued functions u, v; see [AhSa, pages 125–126] for details. Setting $\lambda = |\mu|$ gives

$$ds^2 = |\mu|^2(du + i\,dv)(du - i\,dv) = \lambda^2(du^2 + dv^2),$$

and the coordinates u, v provide an isothermal patch.

We finish with a result of special significance for minimal surfaces.

Lemma 16.18. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular isothermal patch with scaling function λ and mean curvature H. Then we have*

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = 2\lambda^2 H\,\mathbf{U}, \tag{16.30}$$

where $\mathbf{U} = (\mathbf{x}_u \times \mathbf{x}_v)/\|\mathbf{x}_u \times \mathbf{x}_v\|$ is the unit normal.

Proof. Since $\mathbf{x}$ is isothermal, we can differentiate equations (16.25), obtaining

$$\mathbf{x}_{uu} \cdot \mathbf{x}_u = \mathbf{x}_{uv} \cdot \mathbf{x}_v \qquad \text{and} \qquad \mathbf{x}_{vv} \cdot \mathbf{x}_u = -\mathbf{x}_{vu} \cdot \mathbf{x}_v.$$

Therefore,

$$(\mathbf{x}_{uu} + \mathbf{x}_{vv}) \cdot \mathbf{x}_u = \mathbf{x}_{uv} \cdot \mathbf{x}_v - \mathbf{x}_{vu} \cdot \mathbf{x}_v = 0.$$

Similarly, $(\mathbf{x}_{uu} + \mathbf{x}_{vv}) \cdot \mathbf{x}_v = 0$. It follows that $\mathbf{x}_{uu} + \mathbf{x}_{vv}$ is normal to $\mathcal{M}$, and so $\mathbf{x}_{uu} + \mathbf{x}_{vv}$ is a multiple of $\mathbf{U}$. To find out which multiple, we use Theorem 13.25, page 400, and the assumption that $\mathbf{x}$ is isothermal to compute

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)} = \frac{e + g}{2\lambda^2} = \frac{(\mathbf{x}_{uu} + \mathbf{x}_{vv}) \cdot \mathbf{U}}{2\lambda^2},$$

so that we get (16.30). ∎

We denote by

$$\boldsymbol{\Delta} = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}$$

the ***Laplacian*** of $\mathbb{R}^2$, as defined in Section 9.5. Extending this to vector-valued functions, we may write

$$\boldsymbol{\Delta}\mathbf{x} = \mathbf{x}_{uu} + \mathbf{x}_{vv},$$

and conclude the chapter with

Corollary 16.19. *A minimal isothermal patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ satisfies the* ***Laplace equation***

$$\boldsymbol{\Delta}\mathbf{x} = 0,$$

and its components are therefore harmonic functions (*see page* 280).

.

16.8 Exercises

1. Prove that the unit normal of the patch $\mathbf{z}[t]$ defined at the top of page 505 does not depend upon the variable t. Show that the coefficients $e(t)$, $f(t)$ and $g(t)$ of the second fundamental form of $\mathbf{z}[t]$ satisfy

$$e(t)(u,v) = -g(t)(u,v) = -\sin t, \qquad f(t)(u,v) = \cos t.$$

More general results can be found in Section 22.1.

M **2.** Check that the surfaces of Henneberg and Catalan are indeed minimal surfaces and compute the Gaussian curvature of each.

3. Prove that the acceleration $\boldsymbol{\alpha}''(s)$ of the curve (16.15) is parallel to the vector $\left(\cos\frac{u}{2}, \sin\frac{u}{2}, 0\right)$, where the parameters u, s are related by (16.16).

4. Verify that the image of the patch (16.29) is contained in the sphere $S^2(a)$. Compute $\mathbf{x}_u$ and $\mathbf{x}_v$ and prove directly that the patch is isothermal.

M **5.** ***Scherk's fifth minimal surface*** is defined implicitly by

(16.31) $$\sin z = \sinh x \sinh y.$$

Show that it indeed is a minimal surface.

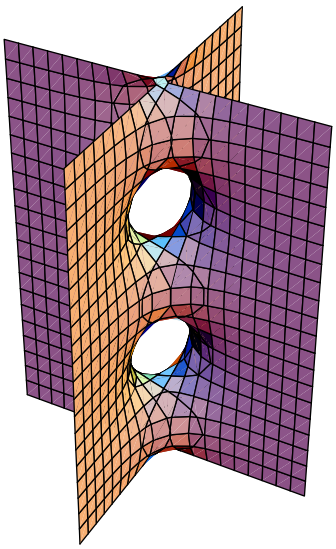


Figure 16.11: Scherk's fifth minimal surface plotted implicitly

6. Show that a surface of revolution is isothermal if and only if $\psi'^2 = \varphi^2 - \varphi'^2$, with notation from page 462. Deduce that catenoid$[a]$ is isothermal if and only if $a = \pm 1$, even though it is a minimal surface for all a.

M **7.** Portions of Scherk's fifth minimal surface are parametrized by

$$\text{scherk5}[a,b,c](u,v) = \big(a \arcsinh u,\ b \arcsinh v,\ c \arcsin(uv)\big),$$

where a, b, c are each equal to ± 1. Show that scherk5$[a,b,c]$ has zero mean curvature, and compute its Gaussian curvature.

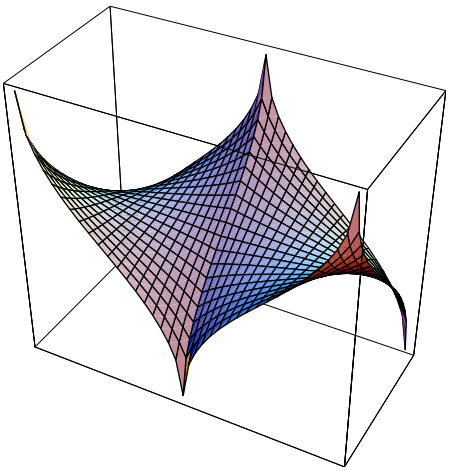


Figure 16.12: Part of Scherk's fifth minimal surface

8. Let $\mathbf{x}$ be an isothermal patch. Show that the Weingarten equations (13.10), page 394, simplify to

$$\text{(16.32)} \qquad \begin{cases} S(\mathbf{x}_u) \;=\; \dfrac{e}{\lambda^2}\mathbf{x}_u + \dfrac{f}{\lambda^2}\mathbf{x}_v, \\[2ex] S(\mathbf{x}_v) \;=\; \dfrac{f}{\lambda^2}\mathbf{x}_u + \dfrac{g}{\lambda^2}\mathbf{x}_v. \end{cases}$$

M **9.** A twisted generalization of Scherk's minimal surface is given by

$$\text{scherk}[a,\theta](u,v) \;=\; \Bigg(\sec 2\theta(u\cos\theta + v\sin\theta),\ \sec 2\theta(u\sin\theta + v\cos\theta),\ \frac{1}{a}\log\Bigg(\frac{\cos\big(\sec 2\theta(u\sin\theta + v\cos\theta)\big)}{\cos\big(\sec 2\theta(u\cos\theta + v\sin\theta)\big)}\Bigg)\Bigg).$$

Show that scherk$[a,\theta]$ is minimal, and compute its Gaussian curvature.

Notebook 16

■ N16.0

This notebook explores the theory and examples of minimal surfaces described in Chapter 16. Much of its motivation derives from the isometric deformation of the helicoid to the catenoid, and in particular the two surfaces

```
catenoid[c_][u_,v_]:= {c Cos[u]Cosh[v/c], c Cosh[v/c]Sin[u], v}
helicoid[a_,c_][u_,v_]:= {a v Cos[u], a v Sin[u], c u}
```

whose relationship is the subject of Section 16.2 and N16.2. The catenoid is a surface of revolution that arises from the solution of the differential equation given in N16.3.

Realizations of the surfaces of Enneper, Catalan, Henneberg are provided in N16.4, and the program mcurv from N13.5 can be used to quickly verify that the mean curvatures do indeed vanish. Scherk's surface is constructed and plotted in N16.5. Another example of the Gauss map is incorporated in N16.6, and N16.7 is based on Mercator's parametrization of the ellipsoid.

This notebook uses Steven Wilkinson's package, which was already exploited for various plots in Notebook 10:

```
<<Graphics`ImplicitPlot3D`
```

We need the standard add-on packages

```
<<Graphics`Graphics3D`
<<Graphics`Colors`
<<Graphics`Shapes`
```

The notebook also makes use of the commands S, L, polarize, highlight, and the programs unitnormal, gcurv, mcurv, all loaded in N16.0.1.

■ N16.0.1

■ N16.2

The deformation from helicoid to catenoid is given by

```
hel2cat[t_][u_,v_]:=
  Cos[t]{Sinh[v]Sin[u], -Sinh[v]Cos[u], u} +
  Sin[t]{Cosh[v]Cos[u], Cosh[v]Sin[u], v}
```

The helicoid and catenoid correspond to particular values of t as each line of the next cell outputs a constant vector:

```
hel2cat[0][u,v] - helicoid[1,1][u-π/2, Sinh[v]]
hel2cat[π/2][u,v] - catenoid[1][u,v]
```

```
ParametricPlot3D[hel2cat[0][u,v], {u,0,2π}, {v,-2,2},
  PlotPoints->{30,30}]
```

The next figure makes use of the program gcurv to compute, and color by, Gaussian curvature. The resulting plots are arranged into a matrix mat whose entry in row i and column j is the plot z[j+3i-4]; it is then displayed with GraphicsArray. If the curvature takes too long to evaluate, one can insert a 0 after the equality sign on the third line to prevent evaluation of the Gaussian curvature.

```
Module[{h,g,y,z},
  h[k_]:= hel2cat[k π/22];
  g[k_]:= gcurv[h[k]];
  fh[X_]:= FaceForm[Hue[X]];
  y[k_][u_,v_]:= Append[h[k][u,v], fh[-10g[k][u,v]]];
  z[k_]:= ParametricPlot3D[y[k][u,v], {u,0,2π}, {v,-2,2},
    Lighting->False];
  mat:= Table[z[j+3i-4], {i,4}, {j,3}];
  S @ GraphicsArray[mat, GraphicsSpacing->-.1]
]
```

Figure 16.1

A similar technique is used to produce the next animation, in which the deformation evolves in a more steady way. Display it by replacing Show by S on the middle line.

```
Module[{z,zz},
  z[t_]= hel2cat[t];
  zz[t_]= Append[z[t][u,v], Hue[(-4gcurv[z[t]][u,v])]];
  Do[Show[ParametricPlot3D[zz[t]//Evaluate, {u,0,2π},
    {v,-2,2}, Lighting->False, BoxRatios->{1,1,1}]],
  {t,0, π/2, π/10}]
]
```

- N16.2.1

N16.3

The differential equation for a minimal surface of revolution can in theory be solved by computer:

```
eq:= h''[t]h[t]==1+h'[t]^2
DSolve[eq, h[t], t]
%[[1,1]]/.{C[1]->1,C[2]->0}
h[t]/.%//Expand
```

N16.4

We investigate the various examples of Section 16.4 in turn. Enneper's surface was first parametrized in N12.6.

```
enneper[u_,v_]:= {u-u^3/3+u v^2, -v+v^3/3-v u^2, u^2-v^2}
```

The next cell outputs the coefficients of the first fundamental form on separate lines.

```
X:= enneper[u,v]
xu:= D[X,u]
xv:= D[X,v]
ff1:= {xu.xu, xu.xv, xv.xv}
ff1//Simplify//ColumnForm
```

A normal vector to Enneper's surface is given by

```
ρ:= 1 + u^2+v^2
Cross[xu,xv]//Simplify
{2u ρ, 2v ρ, ρ^2-2ρ}//Simplify
```

Second derivatives of the chart (and a sum) are

```
xuu= D[X,u,u]
xuv= D[X,u,v]
xvv= D[X,v,v]
xuu + xvv
```

The output already implies that the mean curvature vanishes, though we can compute the mean and Gaussian curvatures without further ado.

```
mcurv[enneper][u,v]
gcurv[enneper][u,v]
```

Charts can be understood using the Shadow command from the Graphics3D package.

```
zz:= ParametricPlot3D[enneper[u,v], {u,-2,2}, {v,-2,2},
  PlotPoints->{20,20}]
S @ Shadow[zz, ZShadow->False]
```

Figure 16.3

Here is another minimal surface:

```
catalan[a_][u_,v_]:=
  {a(u-Sin[u]Cosh[v]), a(1-Cos[u]Cosh[v]), -4a Sin[u/2]Sinh[v/2]}
```

Beware that the *Mathematica* command Catalan gives the sum of an infinite series mentioned in a footnote in Section 16.4.

```
Catalan//N
Sum[(-1)^n/(2n+1)^2, {n,0,10}]//N
```

In the production of Figure 16.4, we use highlight (as defined in N7.3 and N16.0.1) to emphasize the cycloid α formed by intersecting Catalan's surface with the plane z=0.

```
Module[{x,α},
  x= catalan[1];
  α[u_]= x[u,0];
  z1= ParametricPlot3D[Append[x[u,v], FaceForm[Yellow,Orange]],
    {u,2π,6π}, {v,-2,2}, Lighting->False, PlotPoints->{40,20}];
  z2= ParametricPlot3D[highlight[α,.6][u], {u,0,10π}];
  S[z1,z2]
]
```

Figure 16.4

We define Henneberg's minimal surface by

```
henneberg[u_,v_]:={2Sinh[u]Cos[v] - (2/3)Sinh[3u]Cos[3v],
  2Sinh[u]Sin[v]+(2/3)Sinh[3u]Sin[3v], 2Cosh[2u]Cos[2v]}
```

This chart fails to be regular when the following cross product vanishes.

```
X:= henneberg[u,v]
xu:= D[X,u]
xv:= D[X,v]
jac= Cross[xu,xv]//Simplify
```

We check that this can only occur when u=0 and Cos[4v]=0:

```
Solve[jac==0]
```

The unit normal to Henneberg's surface is investigated next.

```
Uhenn:= unitnormal[henneberg]
Uhenn[u,v]//FullSimplify
%//PowerExpand//FullSimplify
```

```
henneberg[u,v]-henneberg[-u,v+π] //Simplify
```

```
Uhenn[u,v]+Uhenn[-u,v+π] //Simplify
```

```
ParametricPlot3D[henneberg[u,v]//Evaluate, {u,0,1},
  {v,-π,π}, PlotPoints->{12,96}, PlotRange->All]
```

N16.4.1

N16.5

The differential equation in the next cell is relevant to the study of Scherk's surface, defined below.

```
DSolve[f'[t]==Tan[c t], f[t] , t]
```

```
scherk[a_][u_,v_]:= {u, v, (1/a)Log[Cos[a v]/Cos[a u]]}
```

```
S @ ParametricPlot3D[scherk[1][u,v], {u,-π/2+.01, π/2-.01},
  {v,-π/2+.01,π/2-0.01}, PlotPoints->{32,16}]
```

Figure 16.7

N16.5.1

N16.6

The Gauss map is determined by a unit normal vector.

```
cat:= catalan[a]
Ucat:= unitnormal[cat]
```

```
Ucat[u,v]
%//PowerExpand//Simplify
```

The intersection of Catalan's surface with the plane z=0 is given by

```
α[u_]= cat[u,0]
```

The vector β[u] below is orthogonal to the tangent vector to the cycloid α in the tangent space to the surface, and we verify that α has no component of acceleration in this direction.

```
β[u_]:= Cross[α'[u], Ucat[u,0]]
%//PowerExpand//Simplify
α''[u].β[u]
```

The Gauss map can be illustrated using the WireFrame construction in the next cell. A similar but more effective example is given for Figure 16.10 inside N16.6.1.

```
en:= ParametricPlot3D[enneper[u,v], {u,-1.5,1.5}, {v,-1.5,1.5},
  ViewPoint->{.17,-.43,3.35}]
un:= ParametricPlot3D[unitnormal[enneper][u,v], {u,-1.7,1.7},
  {v,-1.7,1.7}, ViewPoint->{.17,-.43,3.35},  PlotRange->{-4,4}]
wf:= WireFrame[un]
Show[GraphicsArray[{en, wf}]]
```

- N16.6.1

■ N16.7

Mercator's parametrization of the ellipsoid is

```
mercator[a_,b_,c_][u_,v_]:=
  {a Cos[u]Sech[v], b Sin[u]Sech[v], c Tanh[v]}
```

For a=b=c, it becomes an isothermal chart on the sphere:

```
X:= mercator[a,a,a][u,v]
xu:= D[X,u]
xv:= D[X,v]
{xu.xu, xu.xv, xv.xv}//Simplify
```

But it is *not* in general isothermal:

```
mer:= mercator[1,2,3]
Y:= mer[u,v]
D[Y,u].D[Y,v]//Simplify
Clear[X,Y]
```

```
ParametricPlot3D[mer[u,v], {u,0,2π}, {v,-10,10},
  PlotPoints->{50,80}]
```

■ N16.8

Scherk's fifth surface is given in nonparametric form by

```
scherk5implicit[x_,y_,z_]:= Sin[z] - Sinh[x]Sinh[y]
```

```
S @ ImplicitPlot3D[scherk5implicit[x,y,z]==0,
  {x,-6,6}, {y,-6,6}, {z,-3π,2π},
  PlotPoints->{{7},{3}}, Boxed->False]
```

Figure 16.11

A parametric form of the same surface is

```
scherk5[a_,b_,c_][u_,v_]:=
  {a ArcSinh[u], b ArcSinh[v], c ArcSin[u v]}
```

```
ParametricPlot3D[{scherk5[1,1,1][u,v],
  {0,0,π}+scherk5[1,1,-1][u,v]}//Evaluate, {u,-1,1}, {v,-1,1},
  Boxed->True]
```

```
Simplify[mcurv[scherk5[a,b,c]][u,v]/.{a^2->1,b^2->1,c^2->1}]
Simplify[gcurv[scherk5[a,b,c]][u,v]/.{a^2->1,b^2->1,c^2->1}]
```

■ N16.8.1

Chapter 17

Intrinsic Surface Geometry

The second fundamental form of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ helps to describe precisely how $\mathcal{M}$ sits inside the Euclidean space $\mathbb{R}^3$. The first fundamental form of $\mathcal{M}$, on the other hand, can be used to measure distance on $\mathcal{M}$ without reference to the ambient space. In other words, distance is *intrinsic* to $\mathcal{M}$, and does not require knowledge of the second fundamental form.

It is a remarkable fact that the Gaussian curvature K of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is also intrinsic to $\mathcal{M}$, in spite of the fact that K is the product of two principal curvatures, each of which is nonintrinsic. In order to prove this fact, we establish Brioschi's formula in Section 17.1. It is an explicit (but complicated) formula for K in terms of E, F, G and their first and second partial derivatives. Brioschi's formula is used in Section 17.2 to prove Gauss's *Theorema Egregium*, which states that a surface isometry preserves the Gaussian curvature.

The *Theorema Egregium* has enormous practical value as well as theoretical importance. It implies, for example, that no portion of the earth, which has nonzero Gaussian curvature, can be mapped isometrically onto a plane; indeed, maps of the earth cannot fail to distort distances in one way or another. It also makes it possible to study surfaces in the more abstract setting of Chapter 26.

Christoffel symbols of surfaces are defined and discussed in Section 17.3. In Sections 17.4 and 17.5, we study the *geodesic curvature* $\boldsymbol{\kappa}_g$ and *geodesic torsion* $\boldsymbol{\tau}_g$ of a curve on a surface. The geodesic curvature is the generalization to surfaces of the signed curvature $\boldsymbol{\kappa}\mathbf{2}$ that we used in Chapters 1–5 to study plane curves. An analog of the Frenet formulas (7.12), called the *Darboux formulas*, is established for a curve on a surface.

17.1 Intrinsic Formulas for the Gaussian Curvature

We begin by giving a formula for the Gaussian curvature of a patch $\mathbf{x}$ in terms of the dot products of the first and second partial derivatives of $\mathbf{x}$.

Theorem 17.1. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a patch. Then the Gaussian curvature of $\mathbf{x}$ is given by*

$$K = \frac{1}{(EG-F^2)^2}\left\{ \det\begin{pmatrix} \mathbf{x}_{uu}\cdot\mathbf{x}_{vv} & \mathbf{x}_{uu}\cdot\mathbf{x}_u & \mathbf{x}_{uu}\cdot\mathbf{x}_v \\ \mathbf{x}_u\cdot\mathbf{x}_{vv} & \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_{vv} & \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{pmatrix} - \det\begin{pmatrix} \mathbf{x}_{uv}\cdot\mathbf{x}_{uv} & \mathbf{x}_{uv}\cdot\mathbf{x}_u & \mathbf{x}_{uv}\cdot\mathbf{x}_v \\ \mathbf{x}_u\cdot\mathbf{x}_{uv} & \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_{uv} & \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{pmatrix} \right\}.$$

Proof. It follows from Lemma 13.31, page 405, and Theorem 13.25, page 400, that

$$K = \frac{eg-f^2}{EG-F^2} = \frac{[\mathbf{x}_{uu}\,\mathbf{x}_u\,\mathbf{x}_v][\mathbf{x}_{vv}\,\mathbf{x}_u\,\mathbf{x}_v] - [\mathbf{x}_{uv}\,\mathbf{x}_u\,\mathbf{x}_v]^2}{(EG-F^2)^2}, \tag{17.1}$$

where $[\mathbf{x}_{uu}\,\mathbf{x}_u\,\mathbf{x}_v]$ denotes the vector triple product (page 193). Also, if we write $\mathbf{x} = (x_1, x_2, x_3)$, then

$$\begin{aligned} &[\mathbf{x}_{uu}\,\mathbf{x}_u\,\mathbf{x}_v][\mathbf{x}_{vv}\,\mathbf{x}_u\,\mathbf{x}_v] \qquad (17.2)\\ &= \det\begin{pmatrix} x_{1uu} & x_{2uu} & x_{3uu} \\ x_{1u} & x_{2u} & x_{3u} \\ x_{1v} & x_{2v} & x_{3v} \end{pmatrix} \det\begin{pmatrix} x_{1vv} & x_{2vv} & x_{3vv} \\ x_{1u} & x_{2u} & x_{3u} \\ x_{1v} & x_{2v} & x_{3v} \end{pmatrix} \\ &= \det\left(\begin{pmatrix} x_{1uu} & x_{2uu} & x_{3uu} \\ x_{1u} & x_{2u} & x_{3u} \\ x_{1v} & x_{2v} & x_{3v} \end{pmatrix}\begin{pmatrix} x_{1vv} & x_{1u} & x_{1v} \\ x_{2vv} & x_{2u} & x_{2v} \\ x_{3vv} & x_{3u} & x_{3v} \end{pmatrix}\right) \\ &= \det\begin{pmatrix} \mathbf{x}_{uu}\cdot\mathbf{x}_{vv} & \mathbf{x}_{uu}\cdot\mathbf{x}_u & \mathbf{x}_{uu}\cdot\mathbf{x}_v \\ \mathbf{x}_u\cdot\mathbf{x}_{vv} & \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_{vv} & \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{pmatrix}. \end{aligned}$$

Similarly,

$$[\mathbf{x}_{uv}\,\mathbf{x}_u\,\mathbf{x}_v]^2 = \det\begin{pmatrix} \mathbf{x}_{uv}\cdot\mathbf{x}_{uv} & \mathbf{x}_{uv}\cdot\mathbf{x}_u & \mathbf{x}_{uv}\cdot\mathbf{x}_v \\ \mathbf{x}_u\cdot\mathbf{x}_{uv} & \mathbf{x}_u\cdot\mathbf{x}_u & \mathbf{x}_u\cdot\mathbf{x}_v \\ \mathbf{x}_v\cdot\mathbf{x}_{uv} & \mathbf{x}_v\cdot\mathbf{x}_u & \mathbf{x}_v\cdot\mathbf{x}_v \end{pmatrix}. \tag{17.3}$$

The equation of the theorem follows from (17.1), (17.2) and (17.3). ∎

Except for $\mathbf{x}_{uu} \cdot \mathbf{x}_{vv}$ and $\mathbf{x}_{uv} \cdot \mathbf{x}_{uv}$, we know how to express each of the dot products in the statement of Theorem 17.1 in terms of E, F, G and their derivatives. Next, we show that we can write at least the difference of the remaining two dot products in terms of the second derivatives of E, F, and G.

Lemma 17.2. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a patch. Then*

$$\mathbf{x}_{uu} \cdot \mathbf{x}_{vv} - \mathbf{x}_{uv} \cdot \mathbf{x}_{uv} = -\tfrac{1}{2}E_{vv} + F_{uv} - \tfrac{1}{2}G_{uu}. \tag{17.4}$$

Proof. We have

$$\begin{aligned} \mathbf{x}_{uu} \cdot \mathbf{x}_{vv} - \mathbf{x}_{uv} \cdot \mathbf{x}_{uv} \;=\; & (\mathbf{x}_u \cdot \mathbf{x}_{vv})_u - \mathbf{x}_u \cdot \mathbf{x}_{vvu} \\ & -(\mathbf{x}_u \cdot \mathbf{x}_{uv})_v + \mathbf{x}_u \cdot \mathbf{x}_{uvv}. \end{aligned} \tag{17.5}$$

Since $\mathbf{x}_{vvu} = \mathbf{x}_{uvv}$, equation (17.5) reduces to

$$\begin{aligned} \mathbf{x}_{uu} \cdot \mathbf{x}_{vv} - \mathbf{x}_{uv} \cdot \mathbf{x}_{uv} \;&=\; (\mathbf{x}_u \cdot \mathbf{x}_{vv})_u - (\mathbf{x}_u \cdot \mathbf{x}_{uv})_v \\ &=\; \Big((\mathbf{x}_u \cdot \mathbf{x}_v)_v - \mathbf{x}_{uv} \cdot \mathbf{x}_v\Big)_u - \tfrac{1}{2}(\mathbf{x}_u \cdot \mathbf{x}_u)_{vv} \\ &=\; (\mathbf{x}_u \cdot \mathbf{x}_v)_{vu} - \tfrac{1}{2}(\mathbf{x}_v \cdot \mathbf{x}_v)_{uu} - \tfrac{1}{2}(\mathbf{x}_u \cdot \mathbf{x}_u)_{vv} \\ &=\; -\tfrac{1}{2}E_{vv} + F_{uv} - \tfrac{1}{2}G_{uu}. \end{aligned}$$

∎

Now we are ready to derive the formula of Brioschi[1] (see [Brio]). It is a great deal more complicated than (17.1), but it has the advantage that it expresses the Gaussian curvature of a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ entirely in terms of the *first* fundamental form.

Theorem 17.3. (Brioschi's Formula) *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a patch. Then the Gaussian curvature of $\mathbf{x}$ is given by*

$$K = \frac{1}{(EG - F^2)^2} \left\{ \det \begin{pmatrix} -\frac{1}{2}E_{vv} + F_{uv} - \frac{1}{2}G_{uu} & \frac{1}{2}E_u & F_u - \frac{1}{2}E_v \\ F_v - \frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{pmatrix} \right.$$
$$\left. - \det \begin{pmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{pmatrix} \right\}.$$



[1] Francesco Brioschi (1824–1897). Italian mathematician. Professor first in Padua, then in Milan. Brioschi was one of the founders of modern mathematics in Italy. In addition to his work in differential geometry, Brioschi used elliptical modular functions to solve equations of the fifth and sixth degrees.

Proof. We can rewrite the statement of Theorem 17.1 as

$$K = \frac{1}{(EG-F^2)^2}\left\{ \det\begin{pmatrix} \mathbf{x}_{uu}\cdot\mathbf{x}_{vv} & \frac{1}{2}E_u & F_u-\frac{1}{2}E_v \\ F_v-\frac{1}{2}G_u & E & F \\ \frac{1}{2}G_v & F & G \end{pmatrix} - \det\begin{pmatrix} \mathbf{x}_{uv}\cdot\mathbf{x}_{uv} & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{pmatrix}\right\}.$$

It is possible to expand both of these 3×3 determinants by means of 2×2 determinants. In the resulting expansion each of the expressions $\mathbf{x}_{uu}\cdot\mathbf{x}_{vv}$ and $\mathbf{x}_{uv}\cdot\mathbf{x}_{uv}$ occurs with the same factor $EG-F^2$. Thus we can manipulate the determinants using (17.4) so as to obtain the required formula for K. ∎

Brioschi's formula in its full generality is tedious to use in hand calculations. However, in a frequently occurring case Brioschi's formula simplifies considerably.

Corollary 17.4. *Let $\mathbf{x}\colon \mathcal{U}\to\mathbb{R}^3$ be a patch for which $F=0$. Then the Gaussian curvature of $\mathbf{x}$ is given by*

$$\begin{aligned} K &= \frac{-1}{\sqrt{EG}}\left\{\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right)+\frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right)\right\} \\ &= \frac{-1}{2\sqrt{EG}}\left\{\frac{\partial}{\partial u}\left(\frac{G_u}{\sqrt{EG}}\right)+\frac{\partial}{\partial v}\left(\frac{E_v}{\sqrt{EG}}\right)\right\}. \end{aligned} \tag{17.6}$$

Proof. When $F=0$, Brioschi's Formula reduces to

$$\begin{aligned} K &= \frac{1}{(EG)^2}\left\{ \det\begin{pmatrix} -\frac{1}{2}E_{vv}-\frac{1}{2}G_{uu} & \frac{1}{2}E_u & -\frac{1}{2}E_v \\ -\frac{1}{2}G_u & E & 0 \\ \frac{1}{2}G_v & 0 & G \end{pmatrix} - \det\begin{pmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & 0 \\ \frac{1}{2}G_u & 0 & G \end{pmatrix}\right\} \\ &= \frac{1}{(EG)^2}\Big\{\left(-\tfrac{1}{2}E_{vv}-\tfrac{1}{2}G_{uu}\right)EG \\ &\qquad +\tfrac{1}{4}E_uG_uG+\tfrac{1}{4}E_vG_vE+\tfrac{1}{4}E_v^2G+\tfrac{1}{4}G_u^2E\Big\}, \end{aligned}$$

so that

$$K = -\frac{E_{vv}}{2EG} + \frac{E_v^2}{4E^2G} + \frac{E_vG_v}{4EG^2} - \frac{G_{uu}}{2EG} + \frac{G_u^2}{4EG^2} + \frac{E_uG_u}{4E^2G}.$$

Then (17.6) follows by combining this equation with the expression

$$\begin{aligned}
\frac{1}{\sqrt{EG}}\frac{\partial}{\partial v}\left(\frac{1}{\sqrt{G}}\frac{\partial\sqrt{E}}{\partial v}\right) &= \frac{1}{\sqrt{EG}}\left\{-\frac{G_v}{2G^{\frac{3}{2}}}\frac{\partial\sqrt{E}}{\partial v} + \frac{1}{\sqrt{G}}\frac{\partial^2\sqrt{E}}{\partial v^2}\right\} \\
&= \frac{1}{\sqrt{EG}}\left\{-\frac{G_vE_v}{4G^{\frac{3}{2}}\sqrt{E}} + \frac{1}{2\sqrt{G}}\frac{\partial}{\partial v}\left(E^{-\frac{1}{2}}E_v\right)\right\} \\
&= \frac{1}{\sqrt{EG}}\left\{-\frac{G_vE_v}{4G^{\frac{3}{2}}\sqrt{E}} + \frac{1}{2\sqrt{G}}\left(-\frac{E_v^2}{2E^{\frac{3}{2}}} + \frac{E_{vv}}{\sqrt{E}}\right)\right\} \\
&= -\frac{G_vE_v}{4G^2E} - \frac{E_v^2}{4E^2G} + \frac{E_{vv}}{2EG},
\end{aligned}$$

together with

$$\frac{1}{\sqrt{EG}}\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{E}}\frac{\partial\sqrt{G}}{\partial u}\right) = -\frac{E_uG_u}{4E^2G} - \frac{G_u^2}{4EG^2} + \frac{G_{uu}}{2EG}. \blacksquare$$

Notebook 17 incorporates both the extrinsic formula of Theorem 17.1, and Brioschi's implicit formula in Theorem 17.3. There we verify the consistency of these equations with those of Chapter 13.

17.2 Gauss's Theorema Egregium

We can now prove one of the most celebrated theorems of the 19th century. Mathematicians at the end of the 18th century such as Euler and Monge had used the Gaussian curvature, but only defined as the product of the principal curvatures. Since each principal curvature depends on the particular way the surface is embedded in $\mathbb{R}^3$, there is no obvious reason to suppose that the product of the principal curvatures is intrinsic to $\mathcal{M}$. Gauss's discovery ([Gauss2], published in 1828) that the product of the principal curvatures depends only on the intrinsic geometry of the surface revolutionized differential geometry. This is the ***Theorema Egregium***.

Gauss's proof of the Theorema Egregium is not simple. The difficulty of the proof can be compressed into Brioschi's formula of Theorem 17.3. (A similar formula already occurs in [Gauss1].) As a matter of fact, Brioschi derived his formula in [Brio] in order to provide a new proof of the *Theorema Egregium*. Another proof of the *Theorema Egregium* using geodesic balls was given in the 1848 paper [BDP].

Theorem 17.5. (Gauss's Theorema Egregium[2]) *Let $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a local isometry between regular surfaces $\mathcal{M}_1, \mathcal{M}_2 \subset \mathbb{R}^3$. Denote the Gaussian curvatures of $\mathcal{M}_1$ and $\mathcal{M}_2$ by K_1 and K_2. Then*

$$K_1 = K_2 \circ \Phi.$$

Proof. Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch on $\mathcal{M}_1$, and put $\mathbf{y} = \Phi \circ \mathbf{x}$. Then the restriction $\Phi|\mathbf{x}(\mathcal{U})\colon \mathbf{x}(\mathcal{U}) \to \mathbf{y}(\mathcal{U})$ is a local isometry. Lemma 12.7, page 366, tells us that

$$E_{\mathbf{x}} = E_{\mathbf{y}}, \qquad F_{\mathbf{x}} = F_{\mathbf{y}}, \qquad G_{\mathbf{x}} = G_{\mathbf{y}}.$$

But then $(E_{\mathbf{x}})_u = (E_{\mathbf{y}})_u$, $(F_{\mathbf{x}})_u = (F_{\mathbf{y}})_u$, $(G_{\mathbf{x}})_u = (G_{\mathbf{y}})_u$, and similarly for the other partial derivatives. In particular, Theorem 17.3 implies that $K_{\mathbf{x}} = K_{\mathbf{y}}$, where $K_{\mathbf{x}}$ and $K_{\mathbf{y}}$ denote the Gaussian curvatures of $\mathbf{x}$ and $\mathbf{y}$. Hence

$$K_2 \circ \Phi = K_{\mathbf{y}} \circ \mathbf{y}^{-1} \circ \Phi = K_{\mathbf{x}} \circ \mathbf{x}^{-1} = K_1. \quad \blacksquare$$

For another proof of Theorem 17.5, see page 601.

Examples Illustrating the Theorema Egregium

(i) The patch $\mathbf{x}\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$\mathbf{x}(u, v) = \big(\cos u,\ \sin u,\ v\big)$$

can be considered to be a surface mapping between the plane and a circular cylinder. It is easy to check that $E = G = 1$ and $F = 0$ for both the plane and the parametrization $\mathbf{x}$ of the cylinder; hence $\mathbf{x}$ is a local isometry. On the other hand, $\mathbf{x}$ preserves neither the second fundamental form nor the mean curvature.

(ii) We showed in Section 16.2 that the helicoid is locally isometric to the catenoid. The two surfaces therefore have the same Gaussian curvature, and this fact is graphically illustrated by Figure 16.1 on page 505.

(iii) A sphere and plane are not locally isometric because the Gaussian curvature of a sphere is nonzero, unlike that of the plane. The earth is not exactly a sphere, but still has nonzero Gaussian curvature. As we remarked in the chapter's introduction, this is the precise reason why any flat map of a portion of the earth must distort distances.

---

[2] *Si superficies curva in quamcumque aliam superficiem explicatur, mensura curvaturae in singulis punctis invariata manet.*

(iv) Theorem 17.5 states that a local isometry preserves curvature. There are, however, curvature preserving diffeomorphisms that are not local isometries. We describe a classical example of such a map Φ from the funnel surface funnel$[a, a, b]$ parametrized on page 564 of the next chapter (in the special case that $a=b=1$) to the helicoid parametrized on page 377. It is given by

$$\Phi\big(a\,v\cos u,\ a\,v\sin u,\ b\log v\big) = \big(a\,v\cos u,\ a\,v\sin u,\ b\,u\big).$$

The Gaussian curvatures of the two surfaces are computed in Notebook 17 using Bioschi's formula, and they turn out to be the same, namely

$$\frac{-b^2}{(b^2+a^2v^2)^2}.$$

Hence Φ is a diffeomorphism which preserves curvature. The metric of the funnel surface is given by

$$ds^2 = a^2v^2du^2 + \left(a^2 + \frac{b^2}{v^2}\right)dv^2,$$

while the metric of the helicoid is given by

$$ds^2 = (b^2+a^2v^2)du^2 + a^2dv^2.$$

There is no local isometry between the two surfaces, since the v-parameters would have to correspond to preserve K, but then E would change.

(v) In a similar vein, we noted on page 404 that any rotation about the z-axis preserves the Gaussian curvature of the monkey saddle. It can be checked (see Exercise 9) that only the rotations by an angle which is an integer multiple of $2\pi/3$ are isometries. Like (iv), this example also shows that the converse to Theorem 17.5 is false.

17.3 Christoffel Symbols

Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch, and let

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$$

be its associated unit normal vector. According to the definitions (13.9), page 394, of e, f, g, the normal components of $\mathbf{x}_{uu}, \mathbf{x}_{uv}, \mathbf{x}_{vv}$ are given by

$$\mathbf{x}_{uu}^{\perp} = e\,\mathbf{U}, \qquad \mathbf{x}_{uv}^{\perp} = f\,\mathbf{U}, \qquad \mathbf{x}_{vv}^{\perp} = g\,\mathbf{U}. \tag{17.7}$$

In these formulas, the superscript $^{\perp}$ denotes the component *orthogonal* to the tangent space.

For applications, one also needs the tangential components $\mathbf{x}_{uu}^{\top}, \mathbf{x}_{uv}^{\top}, \mathbf{x}_{vv}^{\top}$ of the patch's second partial derivatives. Since $\mathbf{x}$ is regular, $\mathbf{x}_u$, $\mathbf{x}_v$ and $\mathbf{U}$ form a basis for the tangent space $\mathbb{R}^3_{\mathbf{p}}$ to $\mathbb{R}^3$ at each point $\mathbf{p}$ in the trace of $\mathbf{x}$. It follows that there exist six functions

$$\Gamma_{11}^1, \quad \Gamma_{11}^2, \quad \Gamma_{12}^1, \quad \Gamma_{12}^2, \quad \Gamma_{11}^2, \quad \Gamma_{22}^2$$

such that

$$\left\{ \begin{array}{rcl} \mathbf{x}_{uu} & = & \Gamma_{11}^1 \mathbf{x}_u + \Gamma_{11}^2 \mathbf{x}_v + e\,\mathbf{U}, \\ \mathbf{x}_{uv} & = & \Gamma_{12}^1 \mathbf{x}_u + \Gamma_{12}^2 \mathbf{x}_v + f\,\mathbf{U}, \\ \mathbf{x}_{vv} & = & \Gamma_{22}^1 \mathbf{x}_u + \Gamma_{22}^2 \mathbf{x}_v + g\,\mathbf{U}. \end{array} \right. \tag{17.8}$$

We shall refer to these formulas as the ***Gauss equations***. Given that $\mathbf{U} \cdot \mathbf{U} = 1$, and $\mathbf{U} \cdot \mathbf{x}_u = 0 = \mathbf{U} \cdot \mathbf{x}_v = 0$, the coefficients e, f, g of $\mathbf{U}$ on the far right of (17.8) ensure that (17.7) is satisfied.

Bearing in mind the symmetry of the mixed partial derivatives, one also defines

$$\Gamma_{21}^1 = \Gamma_{12}^1, \qquad \Gamma_{21}^2 = \Gamma_{12}^2,$$

since consistent use of the subscript notation also requires that the equation

$$\mathbf{x}_{vu} = \Gamma_{21}^1 \mathbf{x}_u + \Gamma_{21}^2 \mathbf{x}_v + f\,\mathbf{U}$$

be valid. The functions Γ_{jk}^i are called the ***Christoffel*[3] *symbols***, and were originally defined in [Chris]. The power behind this notation is more evident in the higher-dimensional setting that we shall describe in Chapter 25.

It turns out that the tangential components of $\mathbf{x}_{uu}, \mathbf{x}_{uv}, \mathbf{x}_{vv}$ are ultimately expressible in terms of E, F, G and their first and second partial derivatives. In order to prove this, we first state

Lemma 17.6. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then*

$$\left\{ \begin{array}{ll} \Gamma_{11}^1 E + \Gamma_{11}^2 F = \frac{1}{2} E_u, & \Gamma_{11}^1 F + \Gamma_{11}^2 G = F_u - \frac{1}{2} E_v, \\ \Gamma_{12}^1 E + \Gamma_{12}^2 F = \frac{1}{2} E_v, & \Gamma_{12}^1 F + \Gamma_{12}^2 G = \frac{1}{2} G_u, \\ \Gamma_{22}^1 E + \Gamma_{22}^2 F = F_v - \frac{1}{2} G_u, & \Gamma_{22}^1 F + \Gamma_{22}^2 G = \frac{1}{2} G_v. \end{array} \right. \tag{17.9}$$



[3] Elwin Bruno Christoffel (1829–1900). German mathematician, professor in Strasbourg.

Proof. To determine the tangential coefficients in (17.8), one takes the scalar product of each of the equations in (17.8) with $\mathbf{x}_u$ and $\mathbf{x}_v$, thus giving

$$\begin{aligned}
\Gamma^1_{11}E + \Gamma^2_{11}F &= \mathbf{x}_{uu}\cdot\mathbf{x}_u = \frac{1}{2}\frac{\partial}{\partial u}\|\mathbf{x}_u\|^2 = \tfrac{1}{2}E_u,\\
\Gamma^1_{11}F + \Gamma^2_{11}G &= \mathbf{x}_{uu}\cdot\mathbf{x}_v = \frac{\partial}{\partial u}(\mathbf{x}_u\cdot\mathbf{x}_v) - \mathbf{x}_u\cdot\mathbf{x}_{uv} = F_u - \tfrac{1}{2}E_v,\\
\Gamma^1_{12}E + \Gamma^2_{12}F &= \mathbf{x}_{uv}\cdot\mathbf{x}_u = \frac{1}{2}\frac{\partial}{\partial v}\|\mathbf{x}_u\|^2 = \tfrac{1}{2}E_v,\\
\Gamma^1_{12}F + \Gamma^2_{12}G &= \mathbf{x}_{uv}\cdot\mathbf{x}_v = \frac{1}{2}\frac{\partial}{\partial u}\|\mathbf{x}_v\|^2 = \tfrac{1}{2}G_u,\\
\Gamma^1_{22}E + \Gamma^2_{22}F &= \mathbf{x}_{vv}\cdot\mathbf{x}_u = \frac{\partial}{\partial v}(\mathbf{x}_u\cdot\mathbf{x}_v) - \mathbf{x}_v\cdot\mathbf{x}_{uv} = F_v - \tfrac{1}{2}G_u,\\
\Gamma^1_{22}F + \Gamma^2_{22}G &= \mathbf{x}_{vv}\cdot\mathbf{x}_v = \frac{1}{2}\frac{\partial}{\partial v}\|\mathbf{x}_v\|^2 = \tfrac{1}{2}G_v.
\end{aligned}$$

This establishes (17.9). ∎

We can now state

Theorem 17.7. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a regular patch. Then*

$$\begin{cases}
\Gamma^1_{11} = \dfrac{GE_u - 2FF_u + FE_v}{2(EG-F^2)}, & \Gamma^2_{11} = \dfrac{2EF_u - EE_v - FE_u}{2(EG-F^2)},\\[2ex]
\Gamma^1_{12} = \dfrac{GE_v - FG_u}{2(EG-F^2)}, & \Gamma^2_{12} = \dfrac{EG_u - FE_v}{2(EG-F^2)},\\[2ex]
\Gamma^1_{22} = \dfrac{2GF_v - GG_u - FG_v}{2(EG-F^2)}, & \Gamma^2_{22} = \dfrac{EG_v - 2FF_v + FG_u}{2(EG-F^2)}.
\end{cases}$$

Proof. These equations are easily obtained by solving the system (17.9) in pairs. ∎

Theorem 17.7 will be needed when we study geodesics in Chapter 18 and geodesic curvature in Sections 17.4 and 26.3. In one significant case, the equations simplify greatly:

Corollary 17.8. *Suppose that* $\mathbf{x}$ *is a patch for which* F *is identically zero. Then the Christoffel symbols are given by*

$$\begin{cases}
\Gamma^1_{11} = \dfrac{E_u}{2E}, & \Gamma^2_{11} = \dfrac{-E_v}{2G},\\[2ex]
\Gamma^1_{12} = \dfrac{E_v}{2E}, & \Gamma^2_{12} = \dfrac{G_u}{2G},\\[2ex]
\Gamma^1_{22} = \dfrac{-G_u}{2E}, & \Gamma^2_{22} = \dfrac{G_v}{2G}.
\end{cases}$$

Corollary 17.9. *Let* $\mathbf{x}$ *be a regular patch. Then*

$$\begin{cases} \Gamma^1_{11} + \Gamma^2_{12} = \left(\log\sqrt{EG-F^2}\right)_u, \\ \Gamma^1_{12} + \Gamma^2_{22} = \left(\log\sqrt{EG-F^2}\right)_v. \end{cases} \tag{17.10}$$

Proof. We prove the first equation of (17.10). Using Theorem 17.7, we compute

$$2(EG-F^2)(\Gamma^1_{11}+\Gamma^2_{12}) = GE_u - 2FF_u + EG_u = (EG-F^2)_u.$$

Hence

$$\Gamma^1_{11}+\Gamma^2_{12} = \frac{(EG-F^2)_u}{2(EG-F^2)} = \left(\log\sqrt{EG-F^2}\right)_u. \qquad \blacksquare$$

Here is an application to minimal surfaces. It refers to isothermal patches and the Laplacian Δ, introduced in Section 16.7 at the end of the previous chapter. The relevance here is that isothermal patches satisfy the hypothesis of Corollary 17.8.

Theorem 17.10. *The Gaussian curvature* K *of an isothermal patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *is given in terms of the scaling function* λ *by*

$$K = \frac{-\Delta\log\lambda}{\lambda^2}. \tag{17.11}$$

Furthermore, the Christoffel symbols of $\mathbf{x}$ *are given by*

$$\Gamma^1_{11} = -\Gamma^1_{22} = \Gamma^2_{12} = \frac{\lambda_u}{\lambda} \quad \textit{and} \quad \Gamma^1_{12} = -\Gamma^2_{11} = \Gamma^2_{22} = \frac{\lambda_v}{\lambda}. \tag{17.12}$$

Proof. Formula (17.11) is a special case of (17.6). In detail,

$$\begin{aligned} K &= \frac{-1}{\lambda^2}\left\{\frac{\partial}{\partial u}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial u}\right) + \frac{\partial}{\partial v}\left(\frac{1}{\lambda}\frac{\partial\lambda}{\partial v}\right)\right\} \\ &= \frac{-1}{\lambda^2}\left\{\frac{\partial^2}{\partial u^2}\log\lambda + \frac{\partial^2}{\partial v^2}\log\lambda\right\} = \frac{-\Delta\log\lambda}{\lambda^2}. \end{aligned}$$

Similarly, (17.12) is a special case of Corollary 17.8. $\blacksquare$

17.4 Geodesic Curvature of Curves on Surfaces

Let $\mathcal{M} \subset \mathbb{R}^n$ be a regular surface and $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ a unit-speed curve. We know that the acceleration of $\boldsymbol{\beta}$ in $\mathbb{R}^n$ has a component parallel to $\mathcal{M}$ and a component perpendicular to $\mathcal{M}$. Therefore, we can write

$$\boldsymbol{\beta}''(s) = \boldsymbol{\beta}''(s)^\top + \boldsymbol{\beta}''(s)^\perp \tag{17.13}$$

for $c < s < d$, where $\boldsymbol{\beta}''(s)^\top$ is by definition the component of $\boldsymbol{\beta}''(s)$ tangent to $\mathcal{M}$, and (it follows that) $\boldsymbol{\beta}''(s)^\perp$ is the component parallel to a normal vector to the surface.

Definition 17.11. *The* ***unsigned geodesic curvature*** $\widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}]$ *and the* ***normal curvature*** $\boldsymbol{\kappa}_n[\boldsymbol{\beta}]$ *of a unit-speed curve* $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M} \subset \mathbb{R}^n$ *are given by*

$$\widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}](s) = \|\boldsymbol{\beta}''(s)^\top\| \qquad \text{and} \qquad \boldsymbol{\kappa}_n[\boldsymbol{\beta}](s) = \|\boldsymbol{\beta}''(s)^\perp\|. \tag{17.14}$$

The normal and unsigned geodesic curvatures of an arbitrary curve $\boldsymbol{\alpha}$ *are defined to be the respective curvatures of a unit-speed reparametrization of* $\boldsymbol{\alpha}$.

The expression "normal curvature" was introduced for surfaces in Chapter 13, and we shall see shortly (via Lemma 17.15) that there is a close link between these definitions. We first establish a simple relation between the quantities (17.14) and the curvature $\boldsymbol{\kappa}[\boldsymbol{\beta}]$ of a space curve that was defined in Section 7.2.

Lemma 17.12. *Let* $\mathcal{M} \subset \mathbb{R}^n$ *be a regular surface and* $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ *a unit-speed curve. Then the unsigned geodesic and normal curvatures of* $\boldsymbol{\beta}$ *are related to the curvature* $\boldsymbol{\kappa}[\boldsymbol{\beta}]$ *of* $\boldsymbol{\beta}$ *as a curve in* $\mathbb{R}^n$ *by the formula*

$$\boldsymbol{\kappa}[\boldsymbol{\beta}]^2 = \widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}]^2 + \boldsymbol{\kappa}_n[\boldsymbol{\beta}]^2.$$

Proof. We have

$$\begin{aligned} \boldsymbol{\kappa}[\boldsymbol{\beta}](s)^2 \;&=\; \|\boldsymbol{\beta}''(s)\|^2 \;=\; \|\boldsymbol{\beta}''(s)^\top\|^2 + \|\boldsymbol{\beta}''(s)^\perp\|^2 \\ &= \widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}](s)^2 + \boldsymbol{\kappa}_n[\boldsymbol{\beta}](s)^2. \end{aligned}$$ ∎

Next, we specialize to an oriented regular surface $\mathcal{M} \subset \mathbb{R}^3$ with unit normal $\mathbf{U}$, so that we can define a refined version of $\widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}]$. The associated complex structure of $\mathcal{M}$ is given by $J\mathbf{v}_\mathbf{p} = \mathbf{U} \times \mathbf{v}_\mathbf{p}$, where $\mathbf{v}_\mathbf{p}$ is any tangent vector to $\mathcal{M}$ (page 332). Let $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ be a unit-speed curve. Since $\boldsymbol{\beta}''(s) \cdot \boldsymbol{\beta}'(s) = 0$, it follows that $\boldsymbol{\beta}''(s)^\top$ is a multiple of

$$J\boldsymbol{\beta}'(s) = \mathbf{U} \times \boldsymbol{\beta}'(s)$$

for $c < s < d$.

Definition 17.13. *Let* $\mathcal{M}$ *be an oriented regular surface in* $\mathbb{R}^3$. *The* ***geodesic curvature*** $\boldsymbol{\kappa}_g[\boldsymbol{\beta}]$ *of a unit-speed curve* $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ *is given (for* $c < s < d$*) by*

$$\boldsymbol{\beta}''(s)^\top = \boldsymbol{\kappa}_g[\boldsymbol{\beta}](s) J\boldsymbol{\beta}'(s) = \boldsymbol{\kappa}_g[\boldsymbol{\beta}](s)\, \mathbf{U} \times \boldsymbol{\beta}'(s). \tag{17.15}$$

The geodesic curvature of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}$ *is again defined to be that of the unit-speed reparametrization of* $\boldsymbol{\alpha}$.

To make the general case precise, given $\boldsymbol{\alpha}$ we may compute its arc length $s = s(t)$ measured from any given point $\boldsymbol{\alpha}(c)$, and the inverse function $t = t(s)$. We may then define

$$\begin{aligned} \boldsymbol{\beta}(s) &= \boldsymbol{\alpha}(t(s)), \\ \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t) &= \boldsymbol{\kappa}_g[\boldsymbol{\beta}](s(t)). \end{aligned} \tag{17.16}$$

We return to this situation in the proof of Lemma 17.16.

Here is the refined version of Lemma 17.12:

Corollary 17.14. *Let $\mathcal{M}$ be an oriented regular surface in $\mathbb{R}^3$ and suppose that $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ is a unit-speed curve. Then*

$$|\boldsymbol{\kappa}_g[\boldsymbol{\beta}]| = \widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}] \qquad \text{and} \qquad \boldsymbol{\kappa}[\boldsymbol{\beta}]^2 = \boldsymbol{\kappa}_g[\boldsymbol{\beta}]^2 + \boldsymbol{\kappa}_n[\boldsymbol{\beta}]^2.$$

Proof. We have

$$\begin{aligned} \widetilde{\boldsymbol{\kappa}}_g[\boldsymbol{\beta}](s) \;=\; \|\boldsymbol{\beta}''(s)^\top\| \;&=\; \|\boldsymbol{\kappa}_g[\boldsymbol{\beta}](s) J\boldsymbol{\beta}'(s)\| \\ &=\; |\boldsymbol{\kappa}_g[\boldsymbol{\beta}](s)|\,\|J\boldsymbol{\beta}'(s)\| \;=\; |\boldsymbol{\kappa}_g[\boldsymbol{\beta}](s)|. \end{aligned}$$ ∎

We collect some elementary facts about the curvatures $\boldsymbol{\kappa}_g$ and $\boldsymbol{\kappa}_n$ for the case of a curve in an oriented regular surface $\mathcal{M} \subset \mathbb{R}^3$. First, we show that the normal curvature of a unit-speed curve contains the same geometric information as the normal curvature of $\mathcal{M}$ that we defined in Section 13.2.

Lemma 17.15. *Let $\mathcal{M} \subset \mathbb{R}^3$ be an oriented regular surface, and choose a unit surface normal $\mathbf{U}$ to $\mathcal{M}$. Let $\mathbf{u_p}$ be a unit tangent vector to $\mathcal{M}$ at $\mathbf{p}$. If $\boldsymbol{\beta}$ is any unit-speed curve in $\mathcal{M}$ with $\boldsymbol{\beta}(s) = \mathbf{p}$ and $\boldsymbol{\beta}'(s) = \mathbf{u_p}$, then*

$$\boldsymbol{\kappa}_n[\boldsymbol{\beta}](s) = k(\mathbf{u_p}),$$

where $k(\mathbf{u_p})$ denotes the normal curvature of $\mathcal{M}$ in the direction $\mathbf{u_p}$.

Proof. By definition (see page 389), we have

$$k(\mathbf{u_p}) = S(\mathbf{u_p}) \cdot \mathbf{u_p} = S\big(\boldsymbol{\beta}'(s)\big) \cdot \boldsymbol{\beta}'(s).$$

But according to Lemma 13.4, page 387, we have $S\big(\boldsymbol{\beta}'(s)\big) \cdot \boldsymbol{\beta}'(s) = \boldsymbol{\beta}''(s) \cdot \mathbf{U}$. Because $(\boldsymbol{\beta}''(s) \cdot \mathbf{U})\,\mathbf{U}$ is the component of $\boldsymbol{\beta}''(s)$ perpendicular to $\mathcal{M}$, the lemma follows from the definition of $\boldsymbol{\kappa}_n[\boldsymbol{\beta}]$. ∎

Next, we find simple formulas for the geodesic and normal curvatures when the curve in question does not necessarily have unit speed. Recall that $[\mathbf{u}\,\mathbf{v}\,\mathbf{w}]$ denotes the vector triple product $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$, defined on page 193.

Lemma 17.16. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M} \subset \mathbb{R}^3$ be an arbitrary-speed curve in an oriented regular surface $\mathcal{M}$ in $\mathbb{R}^3$. Let $\mathbf{U}$ be a surface normal to $\mathcal{M}$. Then the normal and geodesic curvatures of $\boldsymbol{\alpha}$ are given (for $a < t < b$) by*

$$\boldsymbol{\kappa}_n[\boldsymbol{\alpha}](t) = \frac{\boldsymbol{\alpha}''(t) \cdot \mathbf{U}}{\|\boldsymbol{\alpha}'(t)\|^2} \tag{17.17}$$

and

$$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t) = \frac{[\boldsymbol{\alpha}'(t)\;\boldsymbol{\alpha}''(t)\;\mathbf{U}]}{\|\boldsymbol{\alpha}'(t)\|^3} = \frac{\boldsymbol{\alpha}''(t) \cdot J\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|^3}. \tag{17.18}$$

Proof. Since $\boldsymbol{\alpha}(t) = \boldsymbol{\beta}(s(t))$, we have

$$\boldsymbol{\alpha}'(t) = s'(t)\boldsymbol{\beta}'(s) \qquad \text{and} \qquad \boldsymbol{\alpha}''(t) = s''(t)\boldsymbol{\beta}'(s) + s'(t)^2\boldsymbol{\beta}''(s),$$

so that

$$\boldsymbol{\kappa}_n[\boldsymbol{\alpha}](t) = \boldsymbol{\kappa}_n[\boldsymbol{\beta}](s) = \boldsymbol{\beta}''(s) \cdot \mathbf{U} = \frac{\boldsymbol{\alpha}''(t) \cdot \mathbf{U}}{s'(t)^2}.$$

Since $s'(t) = \|\boldsymbol{\alpha}'(t)\|$, we get (17.17).

To prove (17.18) we first note that

$$\begin{aligned} \boldsymbol{\alpha}''(t) \times \boldsymbol{\alpha}'(t) \;&=\; \big(s''(t)\boldsymbol{\beta}'(s) + s'(t)^2\boldsymbol{\beta}''(s)\big) \times s'(t)\boldsymbol{\beta}'(s) \\ &=\; s'(t)^3\boldsymbol{\beta}''(s) \times \boldsymbol{\beta}'(s). \end{aligned}$$

Therefore,

$$[\boldsymbol{\alpha}''(t)\ \mathbf{U}\ \boldsymbol{\alpha}'(t)] = -[\mathbf{U}\ \boldsymbol{\alpha}''(t)\ \boldsymbol{\alpha}'(t)] = -s'(t)^3[\mathbf{U}\ \boldsymbol{\beta}''(s)\ \boldsymbol{\beta}'(s)].$$

Furthermore,

$$\begin{aligned} \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t) \;&=\; \boldsymbol{\kappa}_g[\boldsymbol{\beta}](s) = \boldsymbol{\beta}''(s) \cdot J\boldsymbol{\beta}'(s) \\ &=\; \boldsymbol{\beta}''(s) \cdot \big(\mathbf{U} \times \boldsymbol{\beta}'(s)\big) = \frac{[\boldsymbol{\alpha}'(t)\ \boldsymbol{\alpha}''(t)\ \mathbf{U}]}{s'(t)^3}, \end{aligned}$$

and so we get (17.18). ∎

Corollary 17.17. *In the case of a plane curve, the geodesic curvature $\boldsymbol{\kappa}_g$ coincides with the signed curvature $\boldsymbol{\kappa}\mathbf{2}$.*

Proof. Formulas (1.12) and (17.18) are the same. ∎

Thus the geodesic curvature is the generalization to an arbitrary surface of the signed curvature of a plane curve. In fact, we have the following geometric interpretation of geodesic curvature:

Theorem 17.18. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be an arbitrary-speed curve in an oriented regular surface $\mathcal{M}$ in $\mathbb{R}^3$, and let $\mathbf{p} = \boldsymbol{\alpha}(t_0)$, where $a < t_0 < b$. Then the geodesic curvature of $\boldsymbol{\alpha}$ at t_0 coincides with the signed curvature of the orthogonal projection of $\boldsymbol{\alpha}$ onto the tangent plane $\mathcal{M}_\mathbf{p}$ to $\mathcal{M}$ at $\mathbf{p}$.*

Proof. Let $\widehat{\boldsymbol{\alpha}}\colon (a,b) \to \mathcal{M}_\mathbf{p}$ denote the projection of $\boldsymbol{\alpha}$ in the tangent plane $\mathcal{M}_\mathbf{p}$. If $\mathbf{U}$ denotes the unit normal to $\mathcal{M}$, then $\mathbf{U}(\mathbf{p})$ is also the unit normal to $\mathcal{M}_\mathbf{p}$ at the origin of $\mathcal{M}_\mathbf{p}$ (where $\mathcal{M}_\mathbf{p}$ is considered as a surface in $\mathbb{R}^3$ in its own right). Clearly, for $a < t < b$ we have

$$\widehat{\boldsymbol{\alpha}}(t) = \boldsymbol{\alpha}(t) - \big(\boldsymbol{\alpha}(t) \cdot \mathbf{U}(\mathbf{p})\big)\mathbf{U}(\mathbf{p}) \qquad \text{and} \qquad \widehat{\boldsymbol{\alpha}}'(t_0) = \boldsymbol{\alpha}'(t_0).$$

This implies that

$$\begin{aligned}\boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t_0) &= \frac{[\boldsymbol{\alpha}'(t_0)\ \boldsymbol{\alpha}''(t_0)\ \mathbf{U}(\mathbf{p})]}{\|\boldsymbol{\alpha}'(t_0)\|^3} \\ &= \frac{[\widehat{\boldsymbol{\alpha}}'(t_0)\ \widehat{\boldsymbol{\alpha}}''(t_0)\ \mathbf{U}(\mathbf{p})]}{\|\widehat{\boldsymbol{\alpha}}'(t_0)\|^3} = \boldsymbol{\kappa}_g[\widehat{\boldsymbol{\alpha}}](t_0). \ \blacksquare\end{aligned}$$

Next, we show that the geodesic curvature $\boldsymbol{\kappa}_g$ is intrinsic to $\mathcal{M} \subset \mathbb{R}^3$ by expressing $\boldsymbol{\kappa}_g$ in terms of the coefficients E, F, G of the first fundamental form with respect to a patch. Since the computation is local, there is no loss of generality by assuming that $\mathcal{M}$ is oriented.

Theorem 17.19. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M} \subset \mathbb{R}^3$ be a curve whose trace is contained in $\mathbf{x}(\mathcal{U})$, where $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch. Write $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ for $a < t < b$. Then*

$$\begin{aligned}\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]\,\|\boldsymbol{\alpha}'\|^3 \;=\; -\sqrt{EG-F^2}\Big(&-\Gamma^2_{11}u'^3 + \Gamma^1_{22}v'^3 - (2\Gamma^2_{12} - \Gamma^1_{11})u'^2v' \\ &+(2\Gamma^1_{12} - \Gamma^2_{22})u'v'^2 + u''v' - v''u'\Big).\end{aligned}$$

Proof. We have $\boldsymbol{\alpha}' = \mathbf{x}_u u' + \mathbf{x}_v v'$, and also

$$\boldsymbol{\alpha}'' = \mathbf{x}_{uu}u'^2 + 2\mathbf{x}_{uv}u'v' + \mathbf{x}_{vv}v'^2 + \mathbf{x}_u u'' + \mathbf{x}_v v'',$$

so that

$$\begin{aligned}\boldsymbol{\alpha}'' \times \boldsymbol{\alpha}' \;=\; &(\mathbf{x}_{uu} \times \mathbf{x}_u)u'^3 + (\mathbf{x}_{vv} \times \mathbf{x}_v)v'^3 + (\mathbf{x}_{uu} \times \mathbf{x}_v + 2\mathbf{x}_{uv} \times \mathbf{x}_u)u'^2v' \\ &+(\mathbf{x}_{vv} \times \mathbf{x}_u + 2\mathbf{x}_{uv} \times \mathbf{x}_v)u'v'^2 + (u''v' - v''u')\mathbf{x}_u \times \mathbf{x}_v.\end{aligned}$$

Therefore,

$$\begin{aligned}(\boldsymbol{\alpha}'' \times \boldsymbol{\alpha}') \cdot (\mathbf{x}_u \times \mathbf{x}_v) \;=\; &(\mathbf{x}_{uu} \times \mathbf{x}_u) \cdot (\mathbf{x}_u \times \mathbf{x}_v)u'^3 \\ &+(\mathbf{x}_{vv} \times \mathbf{x}_v) \cdot (\mathbf{x}_u \times \mathbf{x}_v)v'^3 + (\mathbf{x}_{uu} \times \mathbf{x}_v + 2\mathbf{x}_{uv} \times \mathbf{x}_u) \cdot (\mathbf{x}_u \times \mathbf{x}_v)u'^2v' \\ &+(\mathbf{x}_{vv} \times \mathbf{x}_u + 2\mathbf{x}_{uv} \times \mathbf{x}_v) \cdot (\mathbf{x}_u \times \mathbf{x}_v)u'v'^2 + (u''v' - v''u')\|\mathbf{x}_u \times \mathbf{x}_v\|^2.\end{aligned} \tag{17.19}$$

We next show how to express the coefficients of u'^3, v'^3, u'^2v', $u'v'^2$ on the right-hand side of (17.19) in terms of E, F, G and the Christoffel symbols. First, we use the identity (7.1), page 193, and Theorem 17.7 to compute

$$\begin{aligned}(\mathbf{x}_{uu} \times \mathbf{x}_u) \cdot (\mathbf{x}_u \times \mathbf{x}_v) \;&=\; (\mathbf{x}_{uu} \cdot \mathbf{x}_u)(\mathbf{x}_u \cdot \mathbf{x}_v) - (\mathbf{x}_{uu} \cdot \mathbf{x}_v)\|\mathbf{x}_u\|^2 \\ &= \tfrac{1}{2}E_u F - \big(F_u - \tfrac{1}{2}E_v\big)E = -(EG - F^2)\Gamma^2_{11}.\end{aligned} \tag{17.20}$$

Similarly,

$$(\mathbf{x}_{vv} \times \mathbf{x}_v) \cdot (\mathbf{x}_u \times \mathbf{x}_v) = (EG - F^2)\Gamma^1_{22}. \tag{17.21}$$

Also,

$$
\begin{aligned}
&(\mathbf{x}_{uu} \times \mathbf{x}_v + 2\mathbf{x}_{uv} \times \mathbf{x}_u) \cdot (\mathbf{x}_u \times \mathbf{x}_v) \\
&\quad = (\mathbf{x}_{uu} \cdot \mathbf{x}_u)\|\mathbf{x}_v\|^2 - (\mathbf{x}_{uu} \cdot \mathbf{x}_v)(\mathbf{x}_u \cdot \mathbf{x}_v) \\
&\qquad\qquad +2(\mathbf{x}_{uv} \cdot \mathbf{x}_u)(\mathbf{x}_u \cdot \mathbf{x}_v) - 2(\mathbf{x}_{uv} \cdot \mathbf{x}_v)\|\mathbf{x}_u\|^2 \\
&\quad = \tfrac{1}{2}E_u G - \big(F_u - \tfrac{1}{2}E_v\big)F + E_v F - G_u E = (EG - F^2)(\Gamma^1_{11} - 2\Gamma^2_{12})
\end{aligned}
\tag{17.22}
$$

and

$$
(\mathbf{x}_{vv} \times \mathbf{x}_u + 2\mathbf{x}_{uv} \times \mathbf{x}_v) \cdot (\mathbf{x}_u \times \mathbf{x}_v) = -(EG - F^2)(\Gamma^2_{22} - 2\Gamma^1_{12}). \tag{17.23}
$$

When equations (17.20)–(17.23) are substituted into (17.19), we get the required result. ▮

17.5 Geodesic Torsion and Frenet Formulas

Just as the geodesic curvature $\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]$ of a curve $\boldsymbol{\alpha}$ is a variant of the ordinary curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$, there is a variant of the torsion of $\boldsymbol{\alpha}$.

Definition 17.20. *Let $\mathcal{M}$ be an oriented regular surface in $\mathbb{R}^3$. The* ***geodesic torsion*** $\boldsymbol{\tau}_g[\boldsymbol{\alpha}]$ *of a curve $\boldsymbol{\alpha}\colon (a, b) \to \mathcal{M}$ is given by*

$$
\boldsymbol{\tau}_g[\boldsymbol{\alpha}](t) = \frac{S\boldsymbol{\alpha}'(t) \cdot J\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|^2}. \tag{17.24}
$$

The following result is easily proved (see Exercise 3). Recall (Definition 13.10 and Section 15.2) that a curve on a surface is called ***principal*** if it is everywhere tangent to a principal direction.

Lemma 17.21. *The definition of geodesic torsion does not depend on the parametrization. Furthermore, a curve on a surface has geodesic torsion zero if and only if it is principal.*

Principal curves are studied further in Chapter 19.

There is also an analog of the Frenet formulas for a curve $\boldsymbol{\alpha}$ on an oriented surface $\mathcal{M} \subset \mathbb{R}^3$. Just as we did in Theorem 7.10, we put $\mathbf{T} = \boldsymbol{\alpha}'/\|\boldsymbol{\alpha}'\|$. However, since we want to study the relation between the geometry of $\boldsymbol{\alpha}$ and $\mathcal{M}$, we use the vector fields $J\mathbf{T}$ and $\mathbf{U}$ instead of $\mathbf{N}$ and $\mathbf{B}$. Note that $J\mathbf{T}$ is always tangent to $\mathcal{M}$, and $\mathbf{U}$ is always perpendicular to $\mathcal{M}$; in contrast, $\mathbf{N}$ and $\mathbf{B}$ in general will be neither tangent nor perpendicular to $\mathcal{M}$, and therefore are not useful to describe the geometry of $\mathcal{M}$.

Definition 17.22. *Let $\mathcal{M}$ be an oriented regular surface in $\mathbb{R}^n$ with unit normal $\mathbf{U}$, and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve on $\mathcal{M}$. Denote by J the complex structure of $\mathcal{M}$ determined by $\mathbf{U}$, and write $\mathbf{T} = \boldsymbol{\alpha}'/\|\boldsymbol{\alpha}'\|$. The* ***Darboux frame field*** *of $\boldsymbol{\alpha}$ with respect to $\mathcal{M}$ consists of the triple of vector fields $\{\mathbf{T}, J\mathbf{T}, \mathbf{U}\}$.*

Corresponding to the Frenet formulas are the ***Darboux formulas***; we first describe them for a unit-speed curve $\boldsymbol{\beta}$.

Theorem 17.23. *Let $\boldsymbol{\beta}\colon (c,d) \to \mathcal{M}$ be a unit-speed curve, and let $\{\mathbf{T}, J\mathbf{T}, \mathbf{U}\}$ be the Darboux frame field of $\boldsymbol{\beta}$ with respect to $\mathcal{M}$. Then*

$$\begin{cases} \mathbf{T}' &= & \boldsymbol{\kappa}_g\, J\mathbf{T} & +\boldsymbol{\kappa}_n \mathbf{U}, \\ (J\mathbf{T})' &= -\boldsymbol{\kappa}_g \mathbf{T} & & +\boldsymbol{\tau}_g \mathbf{U}, \\ \mathbf{U}' &= -\boldsymbol{\kappa}_n \mathbf{T} & -\boldsymbol{\tau}_g\, J\mathbf{T}. & \end{cases}$$

Proof. We have

$$\begin{aligned} \mathbf{T}' &= (\mathbf{T}' \cdot J\mathbf{T})J\mathbf{T} + (\mathbf{T}' \cdot \mathbf{U})\mathbf{U} \\ &= (\boldsymbol{\beta}'' \cdot J\mathbf{T})J\mathbf{T} + (\boldsymbol{\beta}'' \cdot \mathbf{U})\mathbf{U} = \boldsymbol{\kappa}_g\, J\mathbf{T} + \boldsymbol{\kappa}_n \mathbf{U}. \end{aligned}$$

The other formulas are proved in a similar fashion. ∎

The Darboux formulas for an arbitrary-speed curve can be derived in the same way that we derived the Frenet formulas for an arbitrary-speed curve on page 203. Here is the result.

Corollary 17.24. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve with speed $v = \|\boldsymbol{\alpha}'\|$, and let $\{\mathbf{T}, J\mathbf{T}, \mathbf{U}\}$ be the Darboux frame field of $\boldsymbol{\alpha}$ with respect to $\mathcal{M}$. Then*

$$\begin{cases} \mathbf{T}' &= & v\boldsymbol{\kappa}_g\, J\mathbf{T} & +v\boldsymbol{\kappa}_n \mathbf{U}, \\ (J\mathbf{T})' &= -v\boldsymbol{\kappa}_g \mathbf{T} & & +v\boldsymbol{\tau}_g \mathbf{U}, \\ \mathbf{U}' &= -v\boldsymbol{\kappa}_n \mathbf{T} & -v\boldsymbol{\tau}_g\, J\mathbf{T}. & \end{cases}$$

Frame fields constitute a very useful tool for studying curves and surfaces. See the paper [Cartan] of É. Cartan[4] for a general discussion of this subject. Cartan's work has been developed by S. S. Chern and others [BCGH] into the so-called *method of moving frames*, which is recognized as one of the basic techniques of research in modern differential geometry.



[4] Élie Cartan (1869–1951). French mathematician. One of the preeminent mathematicians of the 20th century. His important contributions include the classifications of compact Lie groups and symmetric spaces and the study of various geometries using moving frames and the structure equations.

17.6 Exercises

M **1.** Use Brioschi's formula to compute the Gaussian curvature of the following surfaces, and compare its effectiveness with the programs in Notebook 13.

(a) A sphere.

(b) A torus.

(c) A catenoid.

(d) Enneper's minimal surface.

M **2.** Compute the Christoffel symbols for each of the surfaces in Exercise 1.

3. Prove Lemma 17.21 and Corollary 17.24. Complete the proof of Theorem 17.23.

M **4.** Compute the normal and geodesic curvature of a spherical spiral, as defined in Exercise 14 of Chapter 7.

5. Show that the geodesic torsion of any curve on a sphere vanishes.

6. Compute the geodesic curvature and the geodesic torsion of the curve $t \mapsto (t, at, at^2)$ on the hyperbolic paraboloid $(u, v) \mapsto (u, v, uv)$, in terms of the constant a.

M **7.** Compute the geodesic torsion of a winding line on a torus, as discussed in connection with torus knots in Section 7.7.

8. Let $\mathbf{p} \in \mathcal{M}$ where $\mathcal{M}$ is a surface in $\mathbb{R}^3$, and let $\boldsymbol{\alpha}\colon (a, b) \to \mathcal{M}$ be a curve with $\boldsymbol{\alpha}(0) = \mathbf{p}$. Show that the geodesic torsion of $\boldsymbol{\alpha}$ at $\mathbf{p}$ is given by

$$\boldsymbol{\tau}_g[\boldsymbol{\alpha}](0) = \frac{1}{2}\big(k_2(\mathbf{p}) - k_1(\mathbf{p})\big)\sin 2\theta, \tag{17.25}$$

where θ denotes the oriented angle from $\mathbf{e}_1$ to $\mathbf{u}_\mathbf{p}$. Compare formulas (17.25) and (13.17).

9. Show, as claimed in (v) on page 537, that any rotation about the z-axis preserves the curvature of a monkey saddle, but only rotations of angles $\pm 2\pi/3$ are in fact isometries of the surface.

Notebook 17

■ N17.0

The main task of this notebook is to define the Christoffel symbols. As explained in Section 17.3, these are completely determined by the first fundamental form and its derivatives. We first convert the equations of Section 17.1 for the Gaussian curvature into *Mathematica*. The formula of Brioschi has the great advantage of being implicit, although the program gcurv defined in N13.5 is easier to use when the second fundamental form is available.

We shall in fact give two different programs for the Christoffel symbols that reflect the exposition in Chapter 17. The symbols defined by the two programs will be denoted by γ[i,j,k] in this notebook (but Γ^i_{jk} in Chapter 17). Upper case Greek letters are awkward to type, and in any case the notation $\boldsymbol{\gamma}^i_{jk}$ will appear in the text of Chapter 25. In the sequel, we shall only use the second definition, corresponding to Theorem 17.7, this being the more efficient. A third program for Christoffel symbols reflecting their role in tensor calculus will be given in Notebook 25.

The definitions will be applied to parametrizations of the elliptical torus of revolution and the sphere, which is really a special case of the former:

```
torus[a_,b_,c_][u_,v_]:=
  {(a+b Cos[v])Cos[u], (a+b Cos[v])Sin[u], c Sin[v]}
sphere[a_]:= torus[0,a,a]
```

We shall also experiment with the surfaces

```
helicoid[a_,b_][u_,v_]:= {a v Cos[u], a v Sin[u], b u}
monkeysaddle[u_,v_]:= {u, v, u^3-3 u v^2}
catenoid[c_][u_,v_]:= {c Cos[u]Cosh[v/c], c Sin[u]Cosh[v/c], v}
enneper[u_,v_]:= {u-u^3/3+u v^2, -v+v^3/3-v u^2, u^2-v^2}
```

The commands J, L, κ2, and the program metric are hidden in N17.0.1. We shall display other past programs as we need them.

■ N17.0.1

■ N17.1

First we recall, from Notebook 13, our preferred program for computing the Gaussian curvature. This has the theoretical disadvantage that it is expressed in terms of extrinsic quantities, namely the derivatives of the chart:

```
gcurv[x_][u_,v_]:=
Module[{xu,xv,xuu,xuv,xvv,U,V,A,B},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  B= Det[{xuu,xu,xv}]Det[{xvv,xu,xv}] - Det[{xuv,xu,xv}]^2;
  A= (xu.xu)(xv.xv) - (xu.xv)^2;
  Simplify[B/A^2]/.{U->u,V->v}
]
```

By contrast, the following program based on Theorem 17.3 is *intrinsic*, in the sense that it does not require to be fed a chart x, but only the first fundamental form. The latter is the first argument M and must be a *function* mapping a pair of variables to a symmetric 2 by 2 matrix, exactly like metric[x].

```
brioschi[M_][u_,v_]:=
Module[{U,V,A,B,ee,ff,gg},
  ee= M[U,V][[1,1]];
  ff= M[U,V][[1,2]];
  gg= M[U,V][[2,2]];
  A1={-D[ee,V,V]+2D[ff,U,V]-D[gg,U,U], D[ee,U], 2D[ff,U]-D[ee,V]};
  A2={D[ff,V]-D[gg,U]/2, ee, ff};
  A3={D[gg,V]/2, ff, gg};
  A= {A1/2, A2, A3};
  B1= {0, D[ee,V], D[gg,U]};
  B2= {D[ee,V]/2, ee, ff};
  B3= {D[gg,U]/2, ff, gg};
  B= {B1/2, B2, B3};
  AB= (Det[A]-Det[B])/Det[M[U,V]]^2;
  Simplify[AB]/.{U->u,V->v}
]
```

The above program defines ee,ff,gg to be the coefficients of M, and then computes the difference of the determinants of two matrices, which are denoted temporarily by A and B. Since the matrices are complicated, they are defined row by row, up to a factor of 2. In the examples of this notebook, the first argument M of brioschi will always assume the extrinsic form metric[x] for some chart x, though we shall see more general examples in Notebooks 21 and 26.

We next test the consistency of gcurv and brioschi on the sphere and monkey saddle. Of course, we are not questioning the theory, only our ability to copy out the various matrix entries in an accurate way! In its present implemention, brioschi is in general slower than gcurv.

```
sph:= sphere[a]
M:= metric[sph]
brioschi[M][u,v]//Simplify//Timing
gcurv[sph][u,v]//Simplify//Timing
%% - %
```

```
ms:= monkeysaddle
M:= metric[ms]
brioschi[M][u,v]//Simplify//Timing
gcurv[ms][u,v]//Simplify//Timing
%% - %//Simplify
```

■ N17.2

The following parametrization of the funnel surface is similar in form to that of the helicoid, defined in N17.0.

```
funnel[a_,b_,c_][u_,v_]:={a v Cos[u], b v Sin[u], c Log[v]}
```

Here are calculations that back up the statements in Section 17.2 (iv):

```
x:= funnel[a,a,b]
y:= helicoid[a,b]
brioschi[metric[x]][u,v]
brioschi[metric[y]][u,v]
```

The mapping Φ from one surface to the other cannot be an isometry because the respective metrics (expressed in the common coordinates u,v) are unequal:

```
metric[x][u,v]
metric[y][u,v]
```

■ N17.3

We begin with a program that computes the Christoffel symbols from their *definition* in Section 17.3 as the tangential components of the three second partial derivatives of the chart. Six linear equations arise by taking the scalar product of each of the three second derivatives with the two first partial derivatives of the chart, exactly as in the proof of Lemma 17.6. Using A[i,j,k] as temporary Christoffel symbols that are the unknowns, the linear system is

```
system:= {
  xuu.xu==A[1,1,1]ee+A[2,1,1]ff, xuu.xv==A[1,1,1]ff+A[2,1,1]gg,
  xuv.xu==A[1,1,2]ee+A[2,1,2]ff, xuv.xv==A[1,1,2]ff+A[2,1,2]gg,
  xvv.xu==A[1,2,2]ee+A[2,2,2]ff, xvv.xv==A[1,2,2]ff+A[2,2,2]gg
}
```

There is no point in using the command Module here, as there are no variables to be localized. We have given a definition out of context merely to insert it into the next program, a bad practice that is nonetheless convenient.

```
christoff1[x_][u_,v_]:= Module[{U,V,ta,so,si},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  ee= xu.xu;
  ff= xu.xv;
  gg=xv.xv;
  ta= Flatten[Table[A[i,j,k], {i,2}, {j,2}, {k,j,2}]];
  so= Solve[system, ta][[1]];
  si= Simplify[so]/.{U->u,V->v};
  γ[i_,j_,k_]:= A[i,j,k]/.si;
  ta/.si
]
```

We begin with the easy example of the sphere. The six Christoffel symbols (omitting γ[1,2,1], γ[2,2,1] that for the moment we do not bother to define as equal to γ[1,1,2], γ[2,1,2]) are listed in lexicographic order. Note that the symbols are independent of the sphere's radius a.

```
sph:= sphere[a]
christoff1[sph][u,v]//ColumnForm
```

An individual Christoffel symbol can be obtained as output in the same way that ee,ff,gg can be called after running metric. The result depends of course on the latest instance of christoffel.

```
γ[1,1,2]
tor:= torus[a,b,c]
christoff1[tor][u,v]
γ[1,1,2]
```

Below is a self-contained program that corresponds to Theorem 17.7. It is the one that we shall use in general. The Christoffel symbols that it produces are again labelled γ[i,j,k], and will in a given example be equal to the ones produced by the previous program. For future use, and to emphasize the implicit nature of the Christoffel symbols, the first fundamental form must be input as a matrix-valued function, in the same way as for brioschi.

```
christoff2[M_][u_,v_]:=
Module[{A},
  si= Simplify[#]/.{U->u,V->v} &;
  ee= M[U,V][[1,1]];
  ff=  M[U,V][[1,2]];
  gg= M[U,V][[2,2]];
  A[1,1,1]=  D[ee,V]ff - 2D[ff,U]ff + D[ee,U]gg;
  A[2,2,2]=  D[gg,V]ee - 2D[ff,V]ff + D[gg,U]ff;
  A[2,1,1]= -D[ee,V]ee + 2D[ff,U]ee - D[ee,U]ff;
  A[1,2,2]= -D[gg,V]ff + 2D[ff,V]gg - D[gg,U]gg;
  A[1,1,2]=  D[ee,V]gg - D[gg,U]ff;
  A[2,1,2]= -D[ee,V]ff + D[gg,U]ee;
  A[i_,2,1]= A[i,1,2];
  γ[i_,j_,k_]:= A[i,j,k]/(ee gg - ff^2)/2//si;
  Flatten[Table[γ[i,j,k], {i,2}, {j,2}, {k,j,2}]]
]
```

Here are three examples:

```
christoff1[sph][u,v]//Timing
christoff2[metric[sph]][u,v]//Timing
%% - %//Simplify
```

```
ms:= monkeysaddle
christoff1[ms][u,v]//Timing
christoff2[metric[ms]][u,v]//Timing
%% - %//Simplify
```

```
cat:= catenoid[a]
christoff1[cat][u,v]
christoff2[metric[cat]][u,v]
```

Finally, we apply our Christoffel symbol programs to an artificially defined chart designed to test their effectiveness. The resulting output is so complicated that it confirms their equivalence beyond doubt!

```
x[u_,v_]:= {u^5, Exp[v^3], Sin[u v]}
christoff1[x][u,v]//Timing
christoff2[metric[x]][u,v]//Timing
%% - %//FullSimplify
```

■ N17.4

Here is the unitnormal program from Notebook 11.

```
unitnormal[x_][u_,v_]:=
Module[{U,V,xu,xv},
  xu = D[x[U,V], U];
  xv = D[x[U,V], V];
  Nx = Cross[xu, xv]/.{U->u,V->v};
  Simplify[Nx/L[Nx]]
]
```

The following space curve was introduced in N7.8.

```
sphericalspiral[m_,n_][t_]=
  {Cos[m t]Cos[n t], Sin[m t]Cos[n t], Sin[n t]}
```

We can obtain the same curve by feeding a plane curve into the surface's chart. The string @@ is an abbreviation for Apply, and has the effect of replacing the curly brackets inherent in the definition of α[t] by square brackets, so that the *components* of α[t] become the *arguments* of sphere[1]. This device was already used in N8.6.

```
f[m_,n_][t_]:= {m t, n t}
α:= f[m,n]
σ[t_]= sphere[1]@@α[t]
```

This description of a curve on a surface is used in the following programs. The first two arguments are, respectively, a surface chart and a plane curve.

```
normalcurv[x_,α_][t_]:=
Module[{A,B,Nx,Q},
  Nx= unitnormal[x]@@α[tt];
  xt= x@@α[tt];
  A= D[xt,tt];
  B= D[xt,tt,tt];
  Q= PowerExpand[B.Nx/A.A];
  Simplify[Q]/.tt->t
]
```

The normal curvature of a spherical spiral is constant:

```
normalcurv[sphere[1], α][t]
```

The next program has a similar structure.

```
geodesiccurv[x_,α_][t_]:=
Module[{A,B,Nx,xt,Q},
  Nx= unitnormal[x]@@α[tt];
  xt= x@@α[tt];
  A= D[xt,tt];
  B= D[xt,tt,tt];
  Q= PowerExpand[Det[{A,B,Nx}]/(A.A)^(3/2)];
  Simplify[Q]/.tt->t
]
```

```
geodesiccurv[sphere[1], α][t]
```

The following calculations illustrate Theorem 17.18. The spherical spiral is denoted by σ, and passes through the south pole p when t = π/2. The tangent plane to p is the equatorial plane with opposite orientation, and ρ is therefore the projection of σ. The signed curvature of ρ equals that of σ at the point in question:

```
m= 1; n= 3;
geodesiccurv[sphere[1], α][π/2]
σ[π/2]
ρ[t_]:= {σ[t][[2]],σ[t][[1]]}
κ2[ρ][π/2]//Simplify
```

■ N17.5

The next program is completely analogous to geodesiccurv.

```
geodesictorsion[x_,α_][t_]:=
Module[{A,B,Nx,xt,Q},
  Nx= unitnormal[x]@@α[tt];
  xt= x@@α[tt];
  A= D[xt,tt];
  B= D[Nx,tt];
  Q= PowerExpand[Det[{A,Nx,B}]/(A.A)];
  Simplify[Q]/.tt->t
]
```

Any curve on the sphere is principal, so its geodesic torsion vanishes. This is verified by the computation in the next cell:

```
Clear[m,n]
β[t_]:= {u[t],v[t]}
σ[t_]= sphere[1]@@β[t]
geodesictorsion[sphere[1], β][t]
```

■ N17.6

These programs accomplish Exercise 1:

```
tor:= torus[a,b,b]
brioschi[metric[tor]][u,v]//Timing
gcurv[tor][u,v]//Timing
```

```
cat:= catenoid[a]
brioschi[metric[cat]][u,v]//Timing
gcurv[cat][u,v]//Timing
```

```
enn:= enneper
brioschi[metric[enn]][u,v]//Timing
gcurv[enn][u,v]//Timing
```

For Exercise 2, we compute

```
christoff2[metric[tor]][u,v]
```

```
christoff2[metric[cat]][u,v]
```

```
christoff2[metric[enn]][u,v]
```

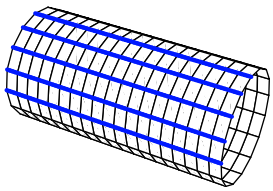
Exercise 4 has been done in N17.4, and Exercise 5 in N17.5. For Exercise 7, we define

```
h[t_]:= {t,2t}
w[t_]:= torus[2,1,2]@@f[t]
geodesictorsion[torus[2,1,1],h][t]
```

Chapter 18

Asymptotic Curves and Geodesics on Surfaces

In this chapter we begin a proper study of special curves lying on surfaces in $\mathbb{R}^3$. An *asymptotic curve* on a surface $\mathcal{M} \subset \mathbb{R}^3$ is a curve whose velocity always points in a direction in which the normal curvature of $\mathcal{M}$ vanishes. In some sense, $\mathcal{M}$ bends less along an asymptotic curve than it does along a general curve. As a simple example, illustrated below, the straight lines on the cylinder $(u, v) \mapsto (\cos u, \sin u, v)$ formed by setting u constant are asymptotic curves. If $\mathbf{p}$ is a hyperbolic point of $\mathcal{M}$ (meaning that the Gaussian curvature is negative at $\mathbf{p}$), there will be exactly two asymptotic curves passing through $\mathbf{p}$.



In Section 18.1, we derive the differential equation that must be satisfied in order that a curve be asymptotic. This is obtained by merely substituting the velocity vector of the curve into the second fundamental form, and setting the result equal to zero. We also prove Theorem 18.7, which relates the Gaussian curvature of $\mathcal{M}$ to the torsion (as defined in Section 7.2) of any asymptotic curve lying on $\mathcal{M}$. In Section 18.2 we identify families of asymptotic curves on various classes of surfaces. We also construct patches built out of asymptotic curves and illustrate the alternative surface plots that result.

Any straight line contained in a surface defines not only an asymptotic curve, but also a *geodesic*. A geodesic on a surface $\mathcal{M}$ is a curve with constant speed and vanishing geodesic curvature (as defined in Section 17.4). Roughly speaking, a geodesic is the shortest curve among all piecewise-differentiable curves on $\mathcal{M}$ connecting two points, just as for a straight line in the plane. However, this property of geodesics will not be demonstrated until Chapter 26.

Geodesics were first studied by Johann Bernoulli[1] in 1697, though the name "geodesic" is due to Liouville in 1844. In Section 18.3, we determine the differential equations for a geodesic. Their formulation allows one to prove that there is a geodesic through a given point $\mathbf{p}$ pointing in *any* direction tangent to $\mathcal{M}$ at $\mathbf{p}$. This of course contrasts with the case of asymptotic curves, but is what one expects intuitively from the distance-minimizing property of geodesics. Many examples of parametrizations of surfaces that we have already given in this book are so-called *Clairaut patches*, as defined in Section 18.5. These are patches for which the geodesics equations simplify and are more readily solvable. Applications are given in Section 18.6.

Since we shall be doing mostly local calculations in this chapter, we shall assume that all surfaces are orientable, unless explicitly stated otherwise. This means that we can choose once and for all a globally-defined unit normal $\mathbf{U}$ for any surface $\mathcal{M}$.

18.1 Asymptotic Curves

Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$. On page 390, we defined asymptotic directions and asymptotic curves on a surface $\mathcal{M}$. The former are defined by the vanishing of the normal curvature at a given point $\mathbf{p} \in \mathcal{M}$.

The following lemma is obvious from the definitions.

Lemma 18.1. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface.*

(i) *At an elliptic point of $\mathcal{M}$ there are no asymptotic directions.*

(ii) *At a hyperbolic point of $\mathcal{M}$ there are exactly two asymptotic directions.*

(iii) *At a parabolic point of $\mathcal{M}$ there is exactly one asymptotic direction.*

(iv) *At a planar point of $\mathcal{M}$ every direction is asymptotic.*

Hyperbolic points are considered in more detail in Theorem 18.4 below.

An asymptotic curve is a curve $\boldsymbol{\alpha}$ in $\mathcal{M}$ for which the normal curvature vanishes in the direction $\boldsymbol{\alpha}'$, that is

$$k\big(\boldsymbol{\alpha}'(t)\big) = 0$$



[1] Johann Bernoulli (1667–1748) occupied the chair of mathematics at Groningen from 1695 to 1705, and at Basel, where he succeeded his elder brother Jakob, from 1705 to 1748. Although Johann was tutored by Jakob, he became involved in controversies with him, and also with his own son Daniel (1700–1782). Discoveries of Johann Bernoulli include the exponential calculus, the treatment of trigonometry as a branch of analysis, the determination of orthogonal trajectories, and the solution of the brachistochrone problem. Other mathematician sons of Johann were Nicolaus (1695-1726) and Johann (II) (1710–1790).

for all t in the domain of definition of $\boldsymbol{\alpha}$. Here is an alternative description.

Lemma 18.2. *A curve $\boldsymbol{\alpha}$ in a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is asymptotic if and only if its acceleration $\boldsymbol{\alpha}''$ is always tangent to $\mathcal{M}$.*

Proof. Without loss of generality, we can assume that $\boldsymbol{\alpha}$ is a unit-speed curve. If $\mathbf{U}$ denotes the surface normal to $\mathcal{M}$, then differentiation of $\boldsymbol{\alpha}' \cdot \mathbf{U} = 0$ yields

$$0 = \boldsymbol{\alpha}' \cdot \mathbf{U}' + \boldsymbol{\alpha}'' \cdot \mathbf{U} = -k(\boldsymbol{\alpha}') + \boldsymbol{\alpha}'' \cdot \mathbf{U} \tag{18.1}$$

(see page 387). Hence $k(\boldsymbol{\alpha}')$ vanishes if and only if $\boldsymbol{\alpha}''$ is perpendicular to $\mathbf{U}$. ∎

Next, we derive the differential equation for the asymptotic curves.

Lemma 18.3. *Let $\boldsymbol{\alpha}$ be a curve that lies in the image of a patch $\mathbf{x}$. Write $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$. Then $\boldsymbol{\alpha}$ is an asymptotic curve if and only if one of the following equivalent conditions is satisfied for all t:*

(i) $e\big(\boldsymbol{\alpha}(t)\big)u'(t)^2 + 2f\big(\boldsymbol{\alpha}(t)\big)u'(t)v'(t) + g\big(\boldsymbol{\alpha}(t)\big)v'(t)^2 = 0$;

(ii) $\mathbf{II}\big(\boldsymbol{\alpha}'(t), \boldsymbol{\alpha}'(t)\big) = 0$, *where* $\mathbf{II}$ *is the second fundamental form defined on page* 401.

Proof. Corollary 10.15, page 294, implies that $\boldsymbol{\alpha}' = \mathbf{x}_u u' + \mathbf{x}_v v'$. Hence by Lemma 13.17 on page 395, we have

$$k\big(\boldsymbol{\alpha}'(t)\big) = \frac{e\big(\boldsymbol{\alpha}(t)\big)u'(t)^2 + 2f\big(\boldsymbol{\alpha}(t)\big)u'(t)v'(t) + g\big(\boldsymbol{\alpha}(t)\big)v'(t)^2}{E\big(\boldsymbol{\alpha}(t)\big)u'(t)^2 + 2F\big(\boldsymbol{\alpha}(t)\big)u'(t)v'(t) + G\big(\boldsymbol{\alpha}(t)\big)v'(t)^2}. \tag{18.2}$$

Then (i) follows from (18.2), and (ii) is a restatement of (i). ∎

The equation in (i) can be written more succinctly as

$$e\,u'^2 + 2f\,u'v' + g\,v'^2 = 0. \tag{18.3}$$

We call (18.3) the ***differential equation for the asymptotic curves*** of a surface. Using Lemma 13.31 on page 405, we see that (18.3) is equivalent to

$$\widetilde{e}u'^2 + 2\widetilde{f}u'v' + \widetilde{g}v'^2 = 0, \tag{18.4}$$

in terms of the triple products

$$\widetilde{e} = [\mathbf{x}_{uu}\ \mathbf{x}_u\ \mathbf{x}_v], \qquad \widetilde{f} = [\mathbf{x}_{uv}\ \mathbf{x}_u\ \mathbf{x}_v], \qquad \widetilde{g} = [\mathbf{x}_{vv}\ \mathbf{x}_u\ \mathbf{x}_v].$$

Theorem 18.4. *In a neighborhood of a hyperbolic point $\mathbf{p}$ of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ there exist two distinct families of asymptotic curves.*

Proof. In a neighborhood of a hyperbolic point $\mathbf{p}$, the equation

$$\widetilde{e}a^2 + 2\widetilde{f}ab + \widetilde{g}b^2 = 0$$

has real roots. Hence we have the factorization

$$\widetilde{e}a^2 + 2\widetilde{f}ab + \widetilde{g}b^2 = (Aa + Bb)(Ca + Db),$$

where A, B, C, D are real. Thus the differential equation for the asymptotic curves also factors:

$$(Au' + Bv')(Cu' + Dv') = 0, \tag{18.5}$$

where now A, B, C, D are real functions. One family consists of the solution curves to $A\,u' + B\,v' = 0$, and the other family consists of the solution curves to $C\,u' + D\,v' = 0$. ∎

Next, we give a characterization of an asymptotic curve $\boldsymbol{\alpha}$ in terms of the curvature of $\boldsymbol{\alpha}$ as a space curve.

Theorem 18.5. *Let $\boldsymbol{\alpha}$ be a regular curve on a regular surface $\mathcal{M} \subset \mathbb{R}^3$, and denote by $\{\mathbf{T}, \mathbf{B}, \mathbf{N}\}$ the Frenet frame and by $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ the curvature of $\boldsymbol{\alpha}$. Also, let $\mathbf{U}$ be the surface normal of $\mathcal{M}$. Then $\boldsymbol{\alpha}$ is an asymptotic curve if and only if at every point $\boldsymbol{\alpha}(t)$, either $\boldsymbol{\kappa}[\boldsymbol{\alpha}](t) = 0$ or $\mathbf{N}(t) \cdot \mathbf{U} = 0$.*

Proof. Without loss of generality, we can assume that $\boldsymbol{\alpha}$ has unit speed. The first Frenet formula (see page 197) in the case of nonzero curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ is

$$\mathbf{T}' = \boldsymbol{\kappa}[\boldsymbol{\alpha}]\mathbf{N}. \tag{18.6}$$

We can also make sense of (18.6) at points where $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ vanishes, by choosing the unit vector $\mathbf{N}$ arbitrarily. By (18.1), we have

$$k(\boldsymbol{\alpha}') = \boldsymbol{\alpha}'' \cdot \mathbf{U} = \mathbf{T}' \cdot \mathbf{U} = \boldsymbol{\kappa}[\boldsymbol{\alpha}]\mathbf{N} \cdot \mathbf{U}.$$

It follows that at each point $\boldsymbol{\alpha}(t)$, either $\boldsymbol{\kappa}[\boldsymbol{\alpha}] = 0$ or $\mathbf{N}$ is well defined and perpendicular to $\mathbf{U}$. The conclusion follows. ∎

The curvature of a straight line vanishes identically, whence

Corollary 18.6. *A straight line that is contained in a regular surface is necessarily an asymptotic curve.*

Now we establish an important relation between the torsion of an asymptotic curve and the Gaussian curvature of the surface containing it.

Theorem 18.7. (Beltrami[2]–Enneper) *Let $\boldsymbol{\alpha}$ be an asymptotic curve on a regular surface $\mathcal{M} \subset \mathbb{R}^3$, and assume the curvature $\boldsymbol{\kappa}[\boldsymbol{\alpha}]$ of $\boldsymbol{\alpha}$ does not vanish. Then the torsion $\boldsymbol{\tau}[\boldsymbol{\alpha}]$ of $\boldsymbol{\alpha}$ and the Gaussian curvature K of $\mathcal{M}$ are related* along $\boldsymbol{\alpha}$ *by*

$$K \circ \boldsymbol{\alpha} = -\boldsymbol{\tau}[\boldsymbol{\alpha}]^2.$$

Proof. Without loss of generality, we can assume that $\boldsymbol{\alpha}$ is a unit-speed curve. Since $\boldsymbol{\alpha}$ is an asymptotic curve with nonvanishing curvature, it follows from Theorem 18.5 that

$$\mathbf{T} \cdot \mathbf{U} = \mathbf{N} \cdot \mathbf{U} = 0.$$

Therefore, $\mathbf{U} = \pm\mathbf{B}$ along $\boldsymbol{\alpha}$. Then the third Frenet formula of (7.12) implies that

$$\boldsymbol{\tau}[\boldsymbol{\alpha}]^2 = \mathbf{B}' \cdot \mathbf{B}' = \frac{d\mathbf{U}}{ds} \cdot \frac{d\mathbf{U}}{ds}. \tag{18.7}$$

On the other hand, since $\mathbf{II}(\boldsymbol{\alpha}', \boldsymbol{\alpha}') = 0$, equation (13.23), page 402, for the third fundamental form reduces to

$$0 = (\mathbf{III} + K\,\mathbf{I})(\boldsymbol{\alpha}', \boldsymbol{\alpha}') = \frac{d\mathbf{U}}{ds} \cdot \frac{d\mathbf{U}}{ds} + K. \tag{18.8}$$

The theorem follows from (18.7) and (18.8). ∎

Corollary 18.8. *An asymptotic curve on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ with constant negative Gaussian curvature has constant torsion.*

This result applies in particular to the examples of surfaces given in Sections 15.4 and 15.6.

Next, we consider the problem of building patches from asymptotic curves.

Definition 18.9. *An* ***asymptotic patch*** *on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is a patch for which the u- and v-parameter curves are asymptotic curves.*

Theorem 18.10. *Let $\mathbf{x}$ be a patch for which f never vanishes. Then $\mathbf{x}$ is an asymptotic patch if and only if e and g vanish identically.*



[2] Eugenio Beltrami (1835–1900). Professor at the Universities of Bologna, Pisa, Padua and Rome. He found the first concrete model of non-Euclidean geometry, published in [Beltr] and in this paper showed how possible contradictions in non-Euclidean geometry would reveal themselves in the Euclidean geometry of surfaces. Beltrami also made known the work of the Jesuit mathematician Saccheri, whose 1733 publication *Euclides ab omni naevo vindicatus* foreshadowed non-Euclidean geometry.

Proof. If the curve $u \mapsto \mathbf{x}(u, v)$ is asymptotic, it follows from Lemma 18.2 that $e = \mathbf{x}_{uu} \cdot \mathbf{U} = 0$. Similarly, if $v \mapsto \mathbf{x}(u, v)$ is asymptotic, then $g = 0$. Conversely, if e and g vanish identically, the differential equation for the asymptotic curves becomes

$$f\, u' v' = 0.$$

Clearly, $u = u_0$ and $v = v_0$ are solutions of this differential equation. Otherwise said, the u- and v-parameter curves are asymptotic. ∎

18.2 Examples of Asymptotic Curves and Patches

Ruled surfaces

According to Corollary 18.6, ruled surfaces are constructed from asymptotic straight lines. For example, if we parametrize the hyperbolic paraboloid $z = x\,y$ by (10.12), page 296, it is easy to compute $e = g = 0$ and $f = (1 + u^2 + v^2)^{-1/2}$. Hence, by Theorem 18.10, $(u, v) \mapsto (u,\ v,\ uv)$ is an asymptotic patch. Of course the resulting curves

$$u \mapsto (u,\ v_0,\ u\,v_0) \qquad \text{and} \qquad v \mapsto (u_0,\ v,\ u_0 v)$$

are the straight-line rulings visible in Figure 14.3 on page 435.

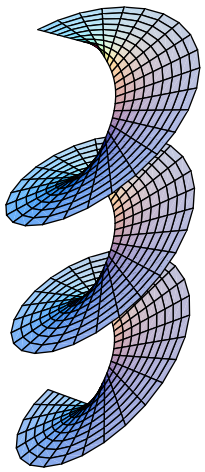


Figure 18.1: The elliptical helicoid

The helicoid defined on page 376 is a special case of the ***elliptical helicoid***

$$\mathsf{helicoid}[a, b, c](u, v) = \big(a\,v\cos u,\ b\,v\sin u,\ c\,u\big). \tag{18.9}$$

Like the previous example, $\mathsf{helicoid}[a, b, c]$ is an asymptotic patch, but this time only the v-parameter curves are straight lines. The u-parameter curves are, of course, elliptical helices. Theorem 18.7 tells us that the Gaussian curvature of the elliptical helicoid, when restricted to one of these helices, equals minus the square of the torsion of the helix. This fact is verified in Notebook 18.

The ***exponentially-twisted helicoid*** is the surface

$$\mathsf{exptwist}[a, c](u, v) = \big(av\cos u,\ av\sin u,\ a\,e^{cu}\big). \tag{18.10}$$

This is a helicoidal-like surface whose twisting varies exponentially and, unlike (18.9), $\mathsf{exptwist}[a, c]$ does not have $e = g = 0$. Instead, an asymptotic patch needs to be found by integrating (18.3); the result is

$$\mathsf{exptwistasym}[a, c](p, q) = \Big(a\,e^{c(p-q)/2}\cos p,\ a\,e^{c(p-q)/2}\sin p,\ a\,e^{cp}\Big).$$

We omit the details, which can be found in Notebook 18, though a similar technique is carried out for the funnel surface on the next page.

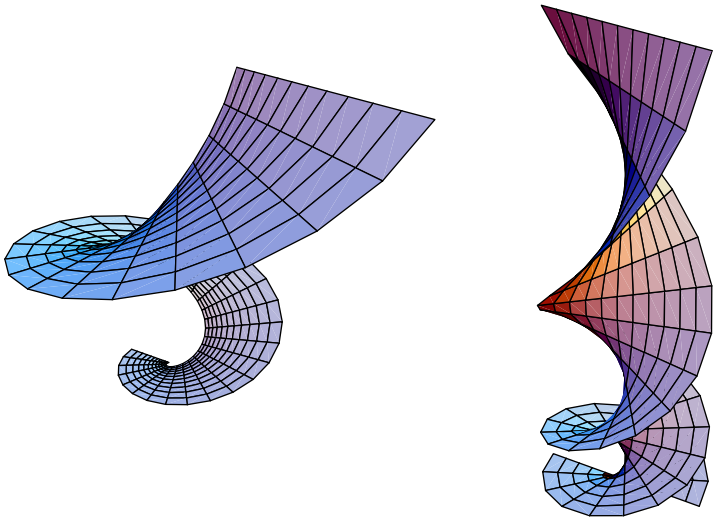


Figure 18.2: The patches $\mathsf{exptwistasym}[1, 0.3]$ (left) and $\mathsf{exptwist}[1, 0.3]$ (right)

The Monkey Saddle

We parametrize the monkey saddle

$$z = \mathfrak{Re}\,(x + iy)^3 = x^3 - 3xy^2$$

by the Monge patch (10.17), page 303. Because the point $(0, 0, 0)$ is planar, *every* direction at $(0, 0, 0)$ is asymptotic. This does not mean, however, that

there are asymptotic *curves* in every direction. But it is easy to find three straight lines passing through $(0,0,0)$ that lie entirely in the surface, namely

$$\begin{aligned} v &\mapsto (0,\, v,\, 0),\\ v &\mapsto (\sqrt{3}v,\, v,\, 0),\\ v &\mapsto (-\sqrt{3}v,\, v,\, 0). \end{aligned}$$

Independently, it can be checked that each of these curves satisfies the differential equation

$$u\,u'^2 - 2v\,u'v' - u\,v'^2 = 0$$

for asymptotic curves. These three lines are also examples of geodesics, and have been emphasized as such in Figure 18.4.

The Funnel Surface

Consider the regular surface which is the image of the patch

$$\begin{aligned} \mathbf{x}\colon (0,\infty)\times(0,2\pi) &\longrightarrow \mathbb{R}^3\\ \mathbf{x}(r,\theta) &= (r\cos\theta,\, r\sin\theta,\, \log r). \end{aligned} \tag{18.11}$$

This is a parametrization of the graph of the function

$$z = \frac{1}{2}\log(x^2+y^2).$$

We compute

$$e = \frac{-1}{r\sqrt{1+r^2}}, \qquad f = 0, \qquad g = \frac{r}{\sqrt{1+r^2}}.$$

Thus the differential equation for the asymptotic curves becomes

$$-\frac{r'^2}{r} + r\,\theta'^2 = 0, \qquad \text{or} \qquad \theta' = \pm\frac{r'}{r}. \tag{18.12}$$

The two equations $\theta' = \pm r'/r$ have respective solutions

$$\theta + 2u = \log r, \qquad \theta + 2v = -\log r,$$

where u and v are constants of integration. Solving for r and θ in terms of u and v gives

$$r = e^{u-v} \qquad \text{and} \qquad \theta = -u - v.$$

The patch

$$\begin{aligned} \mathbf{y}(u,v) &= \mathbf{x}\big(r(u,v),\, \theta(u,v)\big)\\ &= \Big(e^{u-v}\cos(u+v),\, -e^{u-v}\sin(u+v),\, u-v\Big) \end{aligned}$$

is therefore asymptotic. It is illustrated in Figure 18.3.

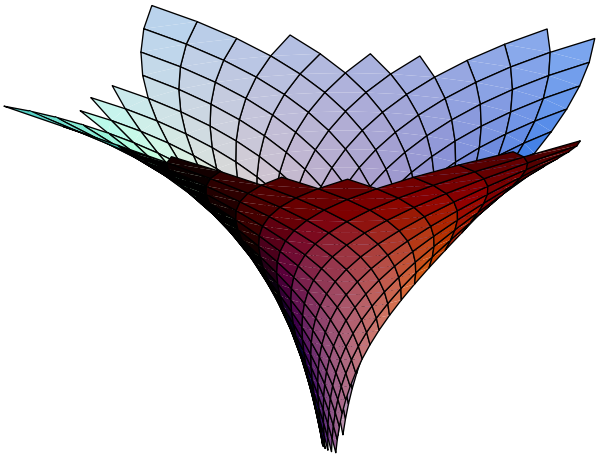


Figure 18.3: The funnel surface with asymptotic patch

18.3 The Geodesic Equations

Our starting point is the following definition of a geodesic on a surface in $\mathbb{R}^3$.

Definition 18.11. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface and $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ a curve. We say that $\boldsymbol{\alpha}$ is a **geodesic** on $\mathcal{M}$ if the tangential component $\boldsymbol{\alpha}''(t)^\top$ of the acceleration of $\boldsymbol{\alpha}$ vanishes.*

It is important to note that we do not assume that $\boldsymbol{\alpha}$ has unit speed in this definition. Nonetheless,

Lemma 18.12. *A geodesic $\boldsymbol{\alpha}$ necessarily has constant speed.*

Proof. We compute

$$\frac{d}{dt}\|\boldsymbol{\alpha}'(t)\|^2 = 2\boldsymbol{\alpha}''(t)\cdot\boldsymbol{\alpha}'(t).$$

Since $\boldsymbol{\alpha}'(t)$ is tangential to the surface, the right-hand side is zero, and

$$\frac{ds}{dt} = \|\boldsymbol{\alpha}'(t)\|$$

is constant. ∎

The definition of geodesic is in some sense complementary to that of asymptotic curve. Recall (see page 390) that $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M} \subset \mathbb{R}^3$ is an asymptotic curve provided the *normal* component of $\boldsymbol{\alpha}''$ vanishes. However, there is an important difference, since the definition of geodesic requires no knowledge of

normal directions, and immediately extends to a surface in $\mathbb{R}^n$. Indeed, we shall soon see that geodesics can be defined instrinsically, in terms of just the *first* fundamental form.

The following result is an immediate consequence of the definition on the previous page, and an analogue of Corollary 18.6.

Corollary 18.13. *A straight line in $\mathbb{R}^3$ that is contained in a surface $\mathcal{M}$ is a geodesic on $\mathcal{M}$, provided its parametrization has constant speed.*

As remarked on page 564, the monkey saddle has three such straight lines passing through its center. These form part of the plot in Figure 18.4 that shows geodesics emanating from the center; in fact all those visible are approximately straight lines.

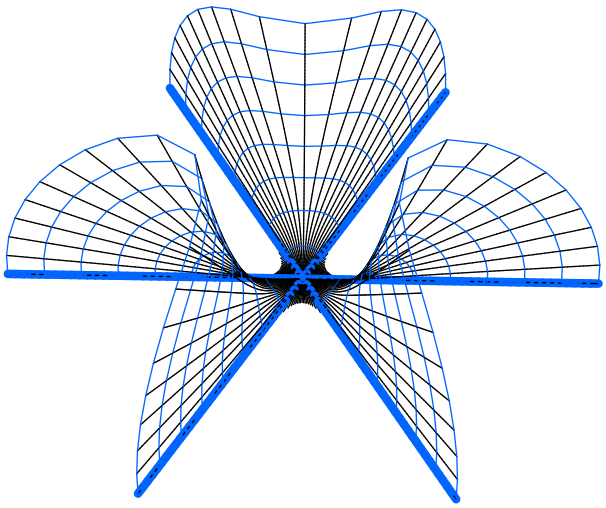


Figure 18.4: Geodesics on a monkey saddle $z = x^3 - 3xy^2$, plotted for $z > -0.05$ and viewed from above

Using the terminology of the previous chapter, we next determine the differential equations that any geodesic on a surface in $\mathbb{R}^3$ must satisfy.

Lemma 18.14. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface parametrized by a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$, where $\mathcal{U} \subset \mathbb{R}^2$. The geodesics on $\mathcal{M}$ are determined by the system of two second-order differential equations*:

$$\left\{ \begin{array}{rcl} u'' + \Gamma^1_{11} u'^2 + 2\Gamma^1_{12} u'v' + \Gamma^1_{22} v'^2 & = & 0, \\ v'' + \Gamma^2_{11} u'^2 + 2\Gamma^2_{12} u'v' + \Gamma^2_{22} v'^2 & = & 0, \end{array} \right. \tag{18.13}$$

where the Γ^i_{jk} are the Christoffel symbols of $\mathbf{x}$.

Proof. Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve. We can write

$$\boldsymbol{\alpha}(t) = \mathbf{x}\big(u(t), v(t)\big), \tag{18.14}$$

where $u, v\colon (a,b) \to \mathbb{R}$ are differentiable functions. When we differentiate (18.14) twice using the chain rule, we get

$$\boldsymbol{\alpha}'' = \mathbf{x}_{uu}u'^2 + \mathbf{x}_u u'' + 2\mathbf{x}_{uv}u'v' + \mathbf{x}_{vv}v'^2 + \mathbf{x}_v v''. \tag{18.15}$$

Then (17.8), page 538, and (18.15) imply that

$$\begin{aligned} \boldsymbol{\alpha}''(t) = &\big(u'' + \Gamma^1_{11}u'^2 + 2\Gamma^1_{12}u'v' + \Gamma^1_{22}v'^2\big)\,\mathbf{x}_u \\ &+ \big(v'' + \Gamma^2_{11}u'^2 + 2\Gamma^2_{12}u'v' + \Gamma^2_{22}v'^2\big)\,\mathbf{x}_v \\ &+ \big(e\,u'^2 + 2f\,u'v' + g\,v'^2\big)\,\mathbf{U}, \end{aligned} \tag{18.16}$$

where

$$\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}$$

is the unit normal vector field to $\mathcal{M}$. In order that $\boldsymbol{\alpha}$ be a geodesic, the coefficients of $\mathbf{x}_u$ and $\mathbf{x}_v$ in (18.16) must each vanish. Hence we get (18.13). ∎

Now we can obtain another important property of geodesics.

Corollary 18.15. Isometries and local isometries preserve geodesics.

Proof. Since the Christoffel symbols are expressible in terms of E, F and G, Corollary 12.8 implies that local isometries preserve equations (18.13). ∎

Next, we prove that from each point on a surface there is a geodesic in every direction. More precisely:

Theorem 18.16. Let $\mathbf{p}$ be a point on a surface $\mathcal{M}$ and $\mathbf{v}_\mathbf{p}$ a tangent vector to $\mathcal{M}$ at $\mathbf{p}$. Then there exists a geodesic $\boldsymbol{\gamma}$ parametrized on some interval containing 0 such that $\boldsymbol{\gamma}(0) = \mathbf{p}$ and $\boldsymbol{\gamma}'(0) = \mathbf{v}_\mathbf{p}$. Furthermore, any two geodesics with these initial conditions coincide on any interval containing 0 on which they are defined.

Proof. Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch with $\mathbf{x}(0,0) = \mathbf{p}$. Then (18.13) is the system of differential equations determining the geodesics on the trace of $\mathbf{x}$. Writing $\boldsymbol{\gamma}(t) = \mathbf{x}(u(t), v(t))$, the initial conditions translate into

$$\begin{cases} u(0) = 0, \\ v(0) = 0, \end{cases} \quad \text{and} \quad \begin{cases} u'(0) = v_1, \\ v'(0) = v_2, \end{cases} \tag{18.17}$$

where v_1, v_2 are the components of $\mathbf{v}_\mathbf{p}$. From the theory of ordinary differential equations, we know that the system (18.13) together with the four initial conditions (18.17) has a unique solution on some interval containing 0. (See, for example, [GrMePi].) Hence the theorem follows. ∎

Theorem 18.16 gives no information concerning the size of the interval on which the geodesic $\boldsymbol{\gamma}$ is defined. Let $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ be geodesics satisfying the same initial conditions. Since the two geodesics must coincide on some interval, it is clear that there is a geodesic $\boldsymbol{\gamma}_3$ whose domain of definition contains those of both $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ and coincides wherever possible with the two geodesics. Applying this consistency result, we obtain a single ***maximal geodesic***. The domain of a maximal geodesic may or may not be the whole real line, and we shall discuss this matter further in Chapter 26.

It is important to distinguish between a geodesic and the trace of a geodesic, because not every reparametrization of a geodesic has constant speed. Therefore, we introduce the following notion:

Definition 18.17. *A curve $\boldsymbol{\alpha}$ in a surface $\mathcal{M}$ is called a **pregeodesic** provided there is a reparametrization $\boldsymbol{\alpha} \circ h$ of $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha} \circ h$ is a geodesic.*

We can now characterize a pregeodesic by the vanishing of the geodesic curvature, defined in Section 17.4. See also Exercise 11.

Lemma 18.18. *Let $\mathcal{M}$ be a surface and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a regular curve. Then the following conditions are equivalent:*

(i) *$\boldsymbol{\alpha}$ is a pregeodesic;*

(ii) *there exists a function $f\colon (a,b) \to \mathbb{R}$ such that*

$$\boldsymbol{\alpha}''(t)^\top = f(t)\boldsymbol{\alpha}'(t)$$

for $a < t < b$;

(iii) *the geodesic curvature of $\boldsymbol{\alpha}$ vanishes.*

Proof. Suppose $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$ is a geodesic. Without loss of generality, $\boldsymbol{\beta}'$ never vanishes. Then $\boldsymbol{\beta}' = (\boldsymbol{\alpha}' \circ h)h'$, so that h' never vanishes. Furthermore, it follows that

$$0 = (\boldsymbol{\beta}'')^\top = h'^2(\boldsymbol{\alpha}'' \circ h)^\top + h''(\boldsymbol{\alpha}' \circ h),$$

so that

$$(\boldsymbol{\alpha}'')^\top = -\left(\frac{h''}{h'^2} \circ h^{-1}\right)\boldsymbol{\alpha}';$$

thus we can take $f = -(h''/h'^2) \circ h^{-1}$.

Conversely, assume that $(\boldsymbol{\alpha}'')^\top = f\boldsymbol{\alpha}'$ and let h be a nonzero solution of the differential equation $h'' + h'^2(f \circ h) = 0$. Put $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$; then

$$(\boldsymbol{\beta}'')^\top = h'^2(\boldsymbol{\alpha}'' \circ h)^\top + h''(\boldsymbol{\alpha}' \circ h) = \big((f \circ h)h'^2 + h''\big)\boldsymbol{\alpha}' \circ h = 0.$$

This establishes the equivalence of (i) and (ii).

The equivalence of (ii) and (iii) follows from equation (17.18) on page 542. For $\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]$ vanishes if and only if $\boldsymbol{\alpha}''(t) \cdot J\boldsymbol{\alpha}(t)' = 0$ for all t, which is equivalent to (ii). ∎

18.4 First Examples of Geodesics

We begin with a general result relating the shape operator of a surface in $\mathbb{R}^3$ and the Frenet frame of a curve on the surface.

Lemma 18.19. *Let $\mathcal{M}$ be an orientable surface in $\mathbb{R}^3$, and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a unit-speed geodesic on $\mathcal{M}$ with nonzero curvature. Denote by $\{\mathbf{T},\mathbf{N},\mathbf{B}\}$ the Frenet frame field of $\boldsymbol{\alpha}$ and by $\boldsymbol{\kappa}$ and $\boldsymbol{\tau}$ the curvature and torsion of $\boldsymbol{\alpha}$. Then it is possible to choose a unit normal vector field $\mathbf{U}$ to $\mathcal{M}$ such that*

$$S(\mathbf{T}) = \boldsymbol{\kappa}\mathbf{T} - \boldsymbol{\tau}\mathbf{B}, \tag{18.18}$$

where S denotes the shape operator of $\mathcal{M}$ with respect to $\mathbf{U}$.

Proof. Since $\boldsymbol{\alpha}$ has nonzero curvature, the vector field $\mathbf{N}$ is defined on all of (a,b). The assumption that $\boldsymbol{\alpha}$ is a geodesic implies that $\mathbf{N} = \boldsymbol{\alpha}''/\boldsymbol{\kappa}\mathbf{2}$ is everywhere perpendicular to $\mathcal{M}$. Thus it is possible to choose $\mathbf{U}$ so that

$$\mathbf{N}(t) = \mathbf{U}\big(\boldsymbol{\alpha}(t)\big)$$

for $a < t < b$. By the definition of the shape operator and by the Frenet Formulas (7.12), page 197, we have

$$S(\mathbf{T}) = -\frac{d}{dt}\mathbf{U}\big(\boldsymbol{\alpha}(t)\big) = -\mathbf{N}' = \boldsymbol{\kappa}\mathbf{T} - \boldsymbol{\tau}\mathbf{B}. \quad \blacksquare$$

Before proceeding further with the general theory, we find geodesics in three important cases.

Geodesics on a Sphere

We can use Lemma 18.19 to determine the geodesics of a sphere.

Lemma 18.20. *The geodesics on a sphere of radius $c > 0$ in $\mathbb{R}^3$ are parts of great circles.*

Proof. Let $\boldsymbol{\beta}$ be a geodesic on a sphere of radius c. By part (i) of Theorem 8.15, page 242, the curvature of $\boldsymbol{\beta}$ is nonzero, so that the Frenet frame $\{\mathbf{T},\mathbf{N},\mathbf{B}\}$ of $\boldsymbol{\beta}$ is well defined. Since the principal curvatures of a sphere are constant and equal, the shape operator of the sphere (with a proper choice of the unit normal) satisfies

$$S(\mathbf{T}) = \frac{1}{c}\mathbf{T}. \tag{18.19}$$

It follows from (18.18) and (18.19) that the curvature of $\boldsymbol{\beta}$ has the constant value $1/c$, and that the torsion vanishes. Moreover, part (vi) of Theorem 8.15 implies that $\boldsymbol{\beta}$ is part of a circle. Since the radius of the circle is the maximum possible, $\boldsymbol{\beta}$ must be part of a great circle. $\blacksquare$

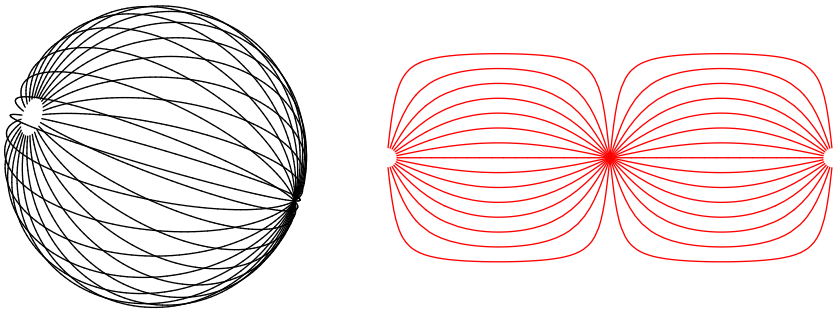


Figure 18.5: Geodesics on the sphere

Figure 18.5 (left) is an outline of the unit sphere $S^2(1)$, formed by a number of geodesics each of length 6 (rather than 2π), passing through a given point $\mathbf{p} \in S^2(1)$. The curves on the right are the same geodesics, but viewed in the rectangular domain of the standard chart (16.26). Both families of curves were plotted numerically using a program from Notebook 18.

Surfaces of Revolution

In this subsection, we limit ourselves to proving the following simple result. Recall Definition 15.3 on page 463.

Theorem 18.21. *Any meridian of a surface of revolution $\mathcal{M} \subset \mathbb{R}^3$ can be parametrized as a geodesic.*

Proof. Let $\boldsymbol{\beta}$ be a unit-speed parametrization of a meridian of $\mathcal{M}$, and let Π be a plane passing through $\boldsymbol{\beta}$ and the axis of revolution. Since $\boldsymbol{\beta}$ is a plane curve in Π, the unit normal $\mathbf{N} = J\boldsymbol{\beta}'$ also lies in Π. Theorem 15.26, page 485, implies that the plane Π meets $\mathcal{M}$ perpendicularly, and so $\mathbf{N}$ is also perpendicular to $\mathcal{M}$. But then $\boldsymbol{\beta}'' = \boldsymbol{\kappa 2}\mathbf{N}$ is perpendicular to $\mathcal{M}$ (see (1.12) on page 14). By definition, $\boldsymbol{\beta}$ is a geodesic. ∎

We shall give a criterion for a *parallel* to be a geodesic in Section 18.6. For a different generalization of Theorem 18.21, see Exercise 12.

Describing *all* the geodesics on a general surface of revolution is complicated. This complexity is illustrated by Figure 18.6, which results when the sphere of the previous figure is replaced by a torus — it shows a selection of geodesics emanating from a point $\mathbf{p}$ on the inside top left of the torus. For example, the one that appears horizontal relative to the chart (on the right) corresponds to the innermost circle of the torus. Figure 18.7 displays the boundary or "geodesic

circle" consisting of all those points a fixed distance r from $\mathbf{p}$, as measured along these geodesics. As r increases further, this closed curve will soon intersect itself in different places, and the "geodesic disk" of center $\mathbf{p}$ and radius r will no longer resemble a disk even topologically.

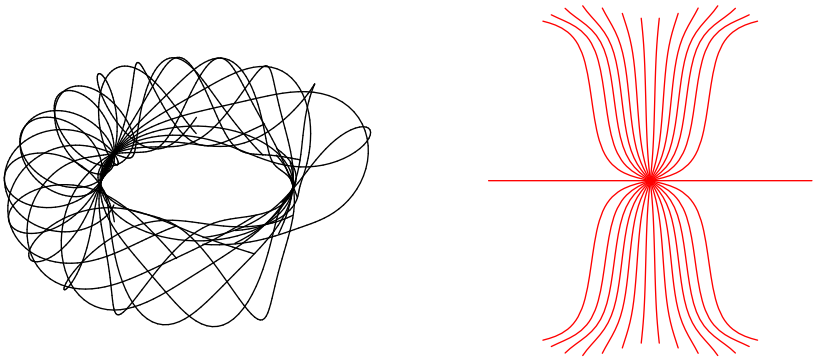


Figure 18.6: Geodesics on the torus of revolution

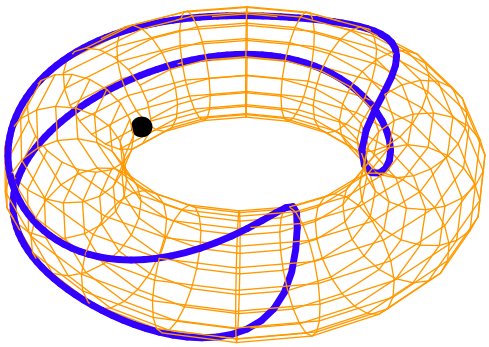


Figure 18.7: A geodesic circle on the torus

Geodesics on a Generalized Cylinder

Geodesics on a generalized cylinder (defined on page 438) are characterized by

Theorem 18.22. *Let $\mathcal{M}$ be the generalized cylinder whose base curve $\boldsymbol{\alpha}$ is a plane curve perpendicular to the rulings of $\mathcal{M}$, and denote by $\mathbf{u}$ a unit vector tangent to $\mathcal{M}$ and perpendicular to $\boldsymbol{\alpha}$, so that $\mathcal{M}$ is parametrized as*

$$\mathbf{x}(u, v) = \boldsymbol{\alpha}(u) + v\,\mathbf{u}.$$

Then the geodesics of $\mathcal{M}$ are precisely those constant speed curves on $\mathcal{M}$ that have constant slope with respect to $\mathbf{u}$.

Proof. Let $\boldsymbol{\gamma}\colon (a,b) \to \mathcal{M}$ be a curve on the generalized cylinder; we can write

$$\boldsymbol{\gamma}(t) = \boldsymbol{\alpha}\big(u(t)\big) + v(t)\mathbf{u} \tag{18.20}$$

for some functions $t \mapsto u(t)$ and $t \mapsto v(t)$. Then

$$\boldsymbol{\gamma}' = \boldsymbol{\alpha}' u' + v'\mathbf{u} \qquad \text{and} \qquad \boldsymbol{\gamma}'' = \boldsymbol{\alpha}'' u'^2 + \boldsymbol{\alpha}' u'' + v''\mathbf{u}. \tag{18.21}$$

Now assume that $\boldsymbol{\gamma}$ is a geodesic. Since any normal vector to $\mathcal{M}$ is always perpendicular to $\mathbf{u}$, it follows from the second equation of (18.21) that $v'' = 0$, so that v' is a constant. But $\mathbf{u} \cdot \boldsymbol{\alpha}' = 0$, so $\boldsymbol{\gamma}$ makes a constant angle with $\mathbf{u}$, and $\boldsymbol{\gamma}$ must have constant slope (see Section 8.5).

Conversely, let $\boldsymbol{\gamma}\colon (a,b) \to \mathcal{M}$ be a constant speed curve that has constant slope with respect to $\mathbf{u}$. Then (18.20) holds with v' constant, so that $\boldsymbol{\alpha} \circ u$ is the projection of $\boldsymbol{\gamma}$ onto the plane perpendicular to $\mathbf{u}$, and $\boldsymbol{\gamma}''$ is perpendicular to $\mathbf{u}$. Moreover, it follows from (18.20) that $\boldsymbol{\gamma}'' = (\boldsymbol{\alpha} \circ u)''$. Also, Lemma 8.21 on page 247 implies that $\boldsymbol{\alpha} \circ u$ has constant speed. Hence $\boldsymbol{\gamma}''$ is perpendicular to both $\mathbf{x}_u = \boldsymbol{\alpha}'$ and $\mathbf{u} = \mathbf{x}_v$, proving that $\boldsymbol{\gamma}$ is a geodesic. ∎

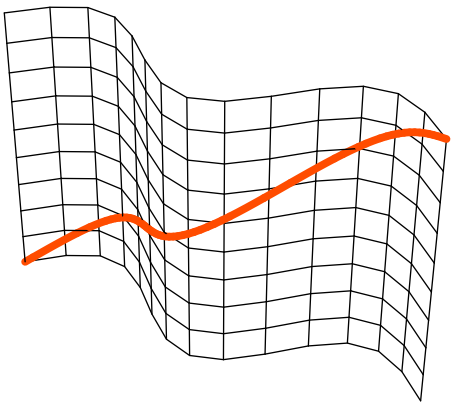


Figure 18.8: A geodesic on the generalized cylinder $(u,v) \mapsto (u, \sin u, v)$

We may deduce that the geodesics on a circular cylinder are circles, straight lines or helices. A more general example to illustrate the principle is shown in Figure 18.8.

18.5 Clairaut Patches

In general, the geodesic equations (18.13) are difficult to solve explicitly. However, there are two important cases where their solution can be reduced to computing integrals.

Definition 18.23. *Let $\mathcal{M}$ be a surface with metric*

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2.$$

(i) *A* ***u*-Clairaut patch** *on $\mathcal{M}$ is a patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ for which*

$$E_v = G_v = F = 0.$$

(ii) *A* ***v*-Clairaut patch** *on $\mathcal{M}$ is a patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ for which*

$$E_u = G_u = F = 0.$$

We work mainly with v-Clairaut patches, leaving the corresponding results for u-Clairaut patches to the reader. To cite two examples, consider

(i) the standard parametrization of the unit sphere $S^2(1)$, having $E = \cos^2 v$, $F = 0$ and $G = 1$.

(ii) the isothermal chart on $S^2(1)$ of Lemma 16.16 on page 519 (with v in place of w), having $E = G = \operatorname{sech} v$ and $F = 0$.

Many others can be found from earlier chapters.

The Christoffel symbols for a v-Clairaut patch are considerably simpler than those of a general patch. The following lemma is an immediate consequence of Theorem 17.7 on page 539.

Lemma 18.24. *For a v-Clairaut patch with $ds^2 = E\,du^2 + G\,dv^2$ we have*

$$\begin{cases} \Gamma^1_{11} = 0, & \Gamma^2_{11} = \dfrac{-E_v}{2G}, \\[2ex] \Gamma^1_{12} = \dfrac{E_v}{2E}, & \Gamma^2_{12} = 0, \\[2ex] \Gamma^1_{22} = 0, & \Gamma^2_{22} = \dfrac{G_v}{2G}, \end{cases} \tag{18.22}$$

so that the geodesic equations (18.13) *reduce to*

$$\begin{cases} u'' + \dfrac{E_v}{E} u'v' &= 0, \\[2ex] v'' - \dfrac{E_v}{2G} u'^2 + \dfrac{G_v}{2G} v'^2 &= 0. \end{cases} \tag{18.23}$$

Some geodesics on a Clairaut patch are easy to determine.

Lemma 18.25. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a v-Clairaut patch. Then*:

(i) *any v-parameter curve is a pregeodesic*;

(ii) *a u-parameter curve $u \mapsto \mathbf{x}(u, v_0)$ is a geodesic if and only if $E_v(u, v_0)$ vanishes along $u \mapsto \mathbf{x}(u, v_0)$.*

Proof. To prove (i), consider a v-parameter curve $\boldsymbol{\beta}$ defined by $\boldsymbol{\beta}(t) = \mathbf{x}(u_0, t)$. It suffices by Lemma 18.18 to show that the tangential component $\boldsymbol{\beta}''(t)^\top$ is a scalar multiple of $\mathbf{x}_v(u_0, t)$. Since $\mathbf{x}_u$ and $\mathbf{x}_v$ are orthogonal, this will be the case provided $\boldsymbol{\beta}''(t) \cdot \mathbf{x}_u(u_0, t)$ is zero. This dot product equals

$$\begin{aligned} \mathbf{x}_{vv}(u_0, t) \cdot \mathbf{x}_u(u_0, t) &= (\mathbf{x}_v \cdot \mathbf{x}_u)' - \mathbf{x}_v \cdot \mathbf{x}_{uv} \\ &= F_v - \tfrac{1}{2} G_u = 0, \end{aligned}$$

the prime indicating differentiation with respect to $t = v$.

To prove (ii), we observe that if the u-parameter curve $\boldsymbol{\alpha}(t) = \mathbf{x}(t, v_0)$ is a geodesic, then

$$\begin{aligned} E_v(u, v_0) = 2\mathbf{x}_{uv}(u, v_0) \cdot \mathbf{x}_u(u, v_0) &= 2(\mathbf{x}_v \cdot \mathbf{x}_u)' - 2\mathbf{x}_v \cdot \mathbf{x}_{uu} \\ &= 2F_u - 2\mathbf{x}_v \cdot \boldsymbol{\alpha}'' = 0, \end{aligned}$$

where this time the prime denotes d/du. Conversely, if $E_v(u, v_0) = 0$, then

$$\mathbf{x}_{uu} \cdot \mathbf{x}_v = (\mathbf{x}_u \cdot \mathbf{x}_v)_u - \mathbf{x}_u \cdot \mathbf{x}_{vu} = F_u - \tfrac{1}{2} E_v = -\tfrac{1}{2} E_v,$$

so that

$$(\boldsymbol{\alpha}'' \cdot \mathbf{x}_v)(u, v_0) = -\tfrac{1}{2} E_v(u, v_0) = 0.$$

But also,

$$\boldsymbol{\alpha}'' \cdot \mathbf{x}_u = \tfrac{1}{2} E_u = 0,$$

so that $\boldsymbol{\alpha}$ is a geodesic. ∎

We turn now to the problem of finding other geodesics on a Clairaut patch. The key result we need is:

Theorem 18.26. (Clairaut's relation) *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a v-Clairaut patch, and let $\boldsymbol{\beta}$ be a geodesic whose trace is contained in $\mathbf{x}(\mathcal{U})$. Let θ be the angle between $\boldsymbol{\beta}'$ and $\mathbf{x}_u$. Then*

$$\sqrt{E}\,\|\boldsymbol{\beta}'\| \cos\theta \quad \text{is constant along } \boldsymbol{\beta}. \tag{18.24}$$

Proof. We write $\boldsymbol{\beta}(t) = \mathbf{x}(u(t), v(t))$. From the assumption that $E_u = 0$ and the first equation of (18.23), it follows that

$$(E\,u')' = E'u' + E\,u'' = E_v v' u' + E\,u'' = 0.$$

Hence there is a constant c such that

$$E\,u' = c. \tag{18.25}$$

Furthermore,

$$\|\boldsymbol{\beta}'\| \cos\theta = \frac{\mathbf{x}_u \cdot \boldsymbol{\beta}'}{\|\mathbf{x}_u\|} = \frac{\mathbf{x}_u \cdot (\mathbf{x}_u u' + \mathbf{x}_v v')}{\|\mathbf{x}_u\|} = \|\mathbf{x}_u\| u' = \sqrt{E}\, u'.$$

When this is combined with (18.25), we obtain (18.24). ▌

Definition 18.27. *The constant c given by* (18.25) *is called the* ***slant*** *of the geodesic $\boldsymbol{\beta}$ in the v-Clairaut patch $\mathbf{x}$. Then the angle between the geodesic and $\mathbf{x}_u$ is given by*

$$\cos\theta = \frac{c}{\|\boldsymbol{\beta}'\| \sqrt{E}}.$$

We next obtain information about geodesics on a Clairaut patch other than those covered by Lemma 18.25.

Lemma 18.28. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a v-Clairaut patch, and let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}$ be a unit-speed curve whose trace is contained in $\mathbf{x}(\mathcal{U})$. Write $\boldsymbol{\beta}(s) = \mathbf{x}((u(s), v(s))$. If $\boldsymbol{\beta}$ is a geodesic, then there is a constant c such that*

$$\text{(18.26)} \qquad \begin{cases} u' \;=\; \dfrac{c}{E}, \\[2ex] v' \;=\; \pm\dfrac{\sqrt{E - c^2}}{\sqrt{EG}}. \end{cases}$$

Conversely, if (18.26) *holds, and if $v'(s) \neq 0$ for $a < s < b$, or if $v'(s) = 0$ for $a < s < b$, then $\boldsymbol{\beta}$ is a unit-speed geodesic.*

Proof. Suppose that $\boldsymbol{\beta}$ is a unit-speed geodesic. The first equation of (18.26) follows from (18.25). Furthermore,

$$\begin{aligned} 1 \;=\; \boldsymbol{\beta}' \cdot \boldsymbol{\beta}' \;&=\; (\mathbf{x}_u u' + \mathbf{x}_v v') \cdot (\mathbf{x}_u u' + \mathbf{x}_v v') \\ &=\; E u'^2 + G v'^2 \;=\; \frac{c^2}{E} + G v'^2, \end{aligned}$$

so that

$$v'^2 = \frac{1}{G}\left(1 - \frac{c^2}{E}\right) = \frac{E - c^2}{EG}.$$

Thus the second equation of (18.26) also holds.

Conversely, suppose that (18.26) holds. The first equation implies that

$$u'' = \left(\frac{c}{E}\right)' = -\frac{c E_v v'}{E^2} = -\frac{E_v u' v'}{E}.$$

Thus we see that the first equation of (18.23) is satisfied. Furthermore, $\boldsymbol{\beta}$ has unit speed because

$$\boldsymbol{\beta}' \cdot \boldsymbol{\beta}' = E u'^2 + G v'^2 = E\Big(\frac{c}{E}\Big)^2 + G\left(\frac{E - c^2}{EG}\right) \;=\; 1.$$

Next, when we differentiate the equation $Gv'^2 = 1 - Eu'^2$, we obtain

$$\begin{aligned} G_v v'^3 + 2Gv''v' &= -E_v v' u'^2 - 2Eu'u'' \\ &= -E_v v' u'^2 + 2Eu'\left(\frac{E_v}{E}u'v'\right) = E_v v' u'^2. \end{aligned} \tag{18.27}$$

Then (18.27) implies that the second equation of (18.23) is satisfied on any interval where v' is different from zero. On the other hand, on an interval where v' vanishes identically, the second equation of (18.26) implies that E is a constant on that interval. Again, the second equation of (18.23) is satisfied. ▮

Corollary 18.29. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a v-Clairaut patch. A curve $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ of the form*

$$\boldsymbol{\alpha}(v) = \mathbf{x}\big(u(v),\, v\big)$$

is a pregeodesic if and only if there is a constant c such that

$$\frac{du}{dv} = \pm c\sqrt{\frac{G}{E(E-c^2)}}. \tag{18.28}$$

Then c is the slant of $\boldsymbol{\alpha}$ with respect to $\mathbf{x}$.

Proof. There exists a unit-speed geodesic $\boldsymbol{\beta}$ reparametrizing $\boldsymbol{\alpha}$ such that

$$\boldsymbol{\beta}(s) = \boldsymbol{\alpha}\big(v(s)\big) = \mathbf{x}\big(u(v(s)), v(s)\big).$$

Then Lemma 18.28 implies that

$$u' = \frac{du}{ds} = \frac{c}{E} \qquad \text{and} \qquad v' = \frac{dv}{ds} = \pm\frac{\sqrt{E-c^2}}{\sqrt{EG}}.$$

Hence

$$\frac{du}{dv} = \frac{u'}{v'} = \frac{\dfrac{c}{E}}{\pm\dfrac{\sqrt{E-c^2}}{\sqrt{EG}}} = \pm c\sqrt{\frac{G}{E(E-c^2)}}.$$

Conversely, if (18.28) holds, we define u' and v' by

$$\frac{u'}{v'} = \pm c\sqrt{\frac{G}{E(E-c^2)}} \qquad \text{and} \qquad Eu'^2 + Gv'^2 = 1.$$

An easy calculation then shows that (18.26) holds, and so we get a pregeodesic. ▮

18.6 Use of Clairaut Patches

Let us consider several examples of Clairaut patches and find their geodesics.

The Euclidean Plane in Polar Coordinates

The polar-coordinate parametrization of the xy-plane is

$$\mathbf{x}(u,v) = (v\cos u,\ v\sin u,\ 0). \tag{18.29}$$

It is clear that (18.29) is a v-Clairaut patch; indeed

$$E = v^2, \qquad F = 0, \qquad G = 1.$$

Equation (18.28) becomes

$$\frac{du}{dv} = \pm\frac{c}{\sqrt{v^2(v^2-c^2)}}, \quad \text{or equivalently} \quad \frac{c\,dv}{\sqrt{v^2(v^2-c^2)}} = \pm du.$$

Carrying out the integration (done by computer in Notebook 18), we obtain

$$\arctan\left(\frac{c}{\sqrt{v^2-c^2}}\right) = \pm(u-u_0), \qquad \text{or} \qquad c = \pm v\sin(u-u_0)$$

for some constant u_0. This is merely the equation of a general straight line in polar coordinates.

Solving the geodesic equations in polar coordinates is harder without the techniques of the previous section. Finding the geodesics in the plane using Cartesian coordinates is of course much easier; see Exercise 13.

Surfaces of Revolution

The standard parametrization of a surface of revolution in $\mathbb{R}^3$ is

$$\mathbf{x}(u,v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big)$$

(see page 462). Let us assume that $\varphi(v) > 0$; then $\varphi(v)$ can be interpreted as the radius of the parallel $u \mapsto (\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v))$. Since

$$E = \varphi(v)^2, \qquad F = 0 \qquad \text{and} \qquad G = \varphi'(v)^2 + \psi'(v)^2,$$

we see that $E_u = G_u = F = 0$; thus $\mathbf{x}$ is again a v-Clairaut patch, and the equation for the geodesics is

$$\frac{du}{dv} = \pm\frac{\sqrt{\varphi'^2+\psi'^2}}{\varphi\sqrt{\varphi^2-c^2}}.$$

Since the v-parameter curves are meridians, we recover Lemma 18.21 from Lemma 18.25(i). Indeed, Theorem 18.26 is often quoted in the special case of a surface of revolution. From Lemma 18.25, we deduce

Corollary 18.30. *Let $\mathcal{M}$ be a surface of revolution in $\mathbb{R}^3$. Then a parallel is a geodesic if and only if $\varphi'(v_0) = 0$.*

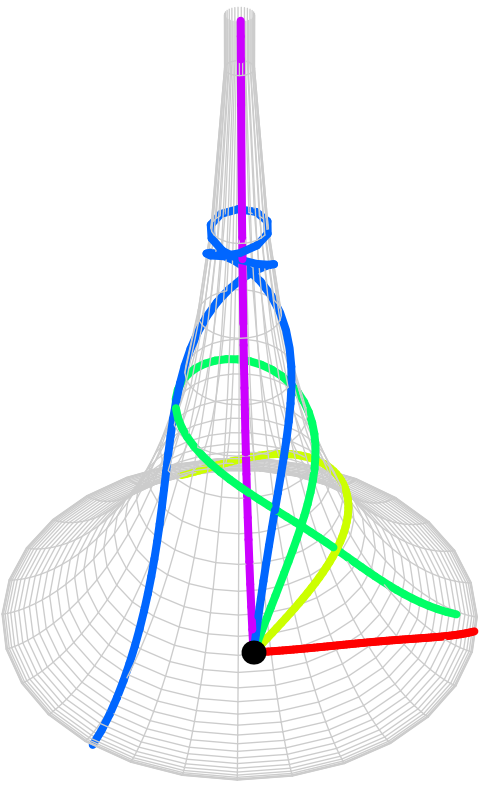


Figure 18.9: Geodesics winding around the pseudosphere

The Pseudosphere

Applying the method above to the pseudosphere, parametrized by (15.6) on page 480, allows us to determine explicitly its geodesics. The equations are

$$u = u_0 \pm \frac{\sqrt{a^2 - 2c^2 - a^2 \cos 2v}}{\sqrt{2}c \sin v},$$

with c, u_0 constants.[3] We plot u as a function of v in Notebook 18, so as to compare the results with the numerical solutions illustrated in Figure 18.9. For each geodesic on the pseudosphere, there are at most two points with equal "height," having the same value of v, but there may be many values of v with the same value of u modulo 2π.

[3]For more information on geodesics on the pseudosphere and interesting pictures of them, see F. Schilling's *Die Pseudosphäre und die nichteuklidische Geometrie* [Schill].

18.7 Exercises

1. Show that the differential equation for the asymptotic curves on a Monge patch $\mathbf{x}(u, v) = (u,\ v,\ h(u, v))$ is

$$h_{uu}u'^2 + 2h_{uv}u'v' + h_{vv}v'^2 = 0.$$

2. Show that the differential equation for the asymptotic curves on a polar patch of the form $\mathbf{x}(r, \theta) = (r\cos\theta,\ r\sin\theta,\ h(r))$ is

$$h''(r)r'^2 + h'(r)r\,\theta'^2 = 0.$$

M **3.** Show that an asymptotic parametrization of a catenoid is given by

$$\mathbf{y}(p, q) = \Big(\cos\frac{p+q}{2}\cosh\frac{-p+q}{2},\ \sin\frac{p+q}{2}\cosh\frac{-p+q}{2},\ \frac{-p+q}{2}\Big).$$

Plot this surface for $0 < p, q < 2\pi$.

4. Find the differential equation for the asymptotic curves on the torus

$$\mathbf{x}(u, v) = \big((a + b\cos v)\cos u,\ (a + b\cos v)\sin u,\ b\sin v\big).$$

M **5.** Find the asymptotic curves of the generalized hyperbolic paraboloid

$$z = x^{2n} - y^{2n},$$

where n is a positive integer. Display some particular cases.

M **6.** Find the asymptotic curves of the surface defined by

$$z = x^m y^n,$$

where m and n are positive integers. Display some particular cases.

7. Show that the Gaussian and mean curvatures of the (circular) helicoid defined on page 376 are given by

$$K = -\frac{c^2}{(c^2 + a^2v^2)^2}, \qquad H = 0.$$

Conclude that each u-parameter curve has constant torsion. Is this fact known from an earlier chapter?

8. Find an asymptotic patch for the surface

$$z = (x^2 + y^2)^\alpha,$$

where α is a real number with $0 \neq \alpha < 1/2$. Figure 18.10 is an ordinary plot of the case $\alpha = -1/2$ colored by Gaussian curvature.

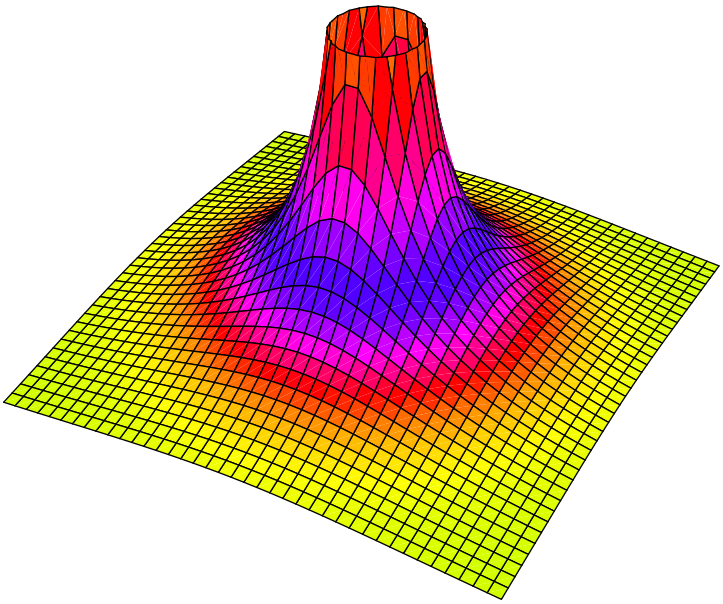


Figure 18.10: Part of the surface $z = (x^2 + y^2)^{-1/2}$ with $z < 3$

9. Find an asymptotic patch for the surface $z = e^{-\alpha(u^2+v^2)}$, where α is a real number with $0 \neq \alpha < 1$.

M **10.** The "shoe" surface is defined by the patch

$$\mathsf{shoe}[a,b](p,q) = \left(-\left(\frac{3b}{4a}\right)^{\frac{1}{3}}(p-q)^{\frac{2}{3}},\ -p-q,\ \frac{b}{4}\left(p^2+14pq+q^2\right)\right).$$

Plot this surface for $a = 1$, $b = -1$ and $-1 < p, q < 1$. Verify that the patch $\mathsf{shoe}[a,b]$ is asymptotic.

11. Show that a pregeodesic is a geodesic if and only if it has constant speed.

12. Generalize Theorem 18.21 to show that a meridian on a generalized helicoid can be parametrized as a geodesic. (Refer to Section 15.4.)

13. Use Lemma 18.28 to determine the geodesics on the plane parametrized by Cartesian coordinates.

M **14.** Find a parametrization of a general geodesic on a catenoid and draw several of them.

15. Let $\boldsymbol{\beta}$ be a geodesic on a catenoid other than the center circle, and suppose the initial velocity of $\boldsymbol{\beta}$ is perpendicular to the axis of revolution. Use Clairaut's relation (Theorem 18.26) to show that $\boldsymbol{\beta}$ approaches the center curve asymptotically, but never reaches it.

M **16.** Plot geodesics and geodesic circles on ellipsoids with two distinct axes and on ellipsoids with three distinct axes.

17. Let $\boldsymbol{\alpha}$ be an asymptotic curve defined by

$$\boldsymbol{\alpha}(t) = \mathsf{surfrev}[\boldsymbol{\alpha}]\big(u(t),\, v(t)\big),$$

where $\mathsf{surfrev}[\boldsymbol{\alpha}]$ is the standard parametrization of a surface of revolution given by (15.1). Show that u and v satisfy the differential equation

$$u'(t)^2 = \left(\frac{\varphi''(t)\psi'(t) - \varphi'(t)\psi''(t)}{\varphi(t)\psi'(t)}\right) v'(t)^2.$$

18. Fill in the proof of Lemma 18.24.

Notebook 18

■ N18.0

The main purpose of this notebook is to develop a simple program for solving numerically the geodesic equations on a surface. The program is the one that was adapted from the first edition of the book for inclusion in the tribute to the author contained in the volume [FeWo]. It enables geodesics and geodesic circles to be plotted with comparative ease. Before turning to geodesics, though, we study asymptotic curves in N18.1 and N18.2.

We shall need the following parametrizations, mostly investigated in Notebook 10.

```
ellipsoid[a_,b_,c_][u_,v_]:=
  {a Cos[v]Cos[u], b Cos[v]Sin[u], c Sin[v]}
sphere[a_]:= ellipsoid[a,a,a]
torus[a_,b_,c_][u_,v_]:=
  {(a+b Cos[v])Cos[u], (a+b Cos[v])Sin[u], c Sin[v]}
circularcone[a_,b_][u_,v_]:= {a v Cos[u], a v Sin[u], b v}
paraboloid[a_,b_][u_,v_]:= {u, v, a u^2 + b v^2}
```

We shall also investigate geodesics on the surfaces

```
monkeysaddle[u_,v_]:= {u, v, u^3-3 u v^2}
helicoid[a_,b_,c_][u_,v_]:= {a v Cos[u], b v Sin[u], c u}
pseudosphere[a_][u_,v_]:=
  a{Cos[u]Sin[v], Sin[u]Sin[v], Cos[v]+Log[Tan[v/2]]}
```

Apart from the package

```
<<Graphics`Colors`
```

options for displays, and other commands are loaded in N18.0.1. These include our usual programs unitnormal, metric, shape, for finding first and second fundamental forms, as well as gcurv, mcurv, for computing curvature. In addition, we load the program christoffel2 introduced in N17.3, and we plot the introductory figure showing the rulings on a cylinder.

■ N18.0.1

■ N18.1

The first new program produces the system of differential equations for the asymptotic curves. The relevant coefficients et,ft,gt are proportional to the second fundamental form coefficients eee, fff, ggg. In theory, the latter could have been generated using shape and used instead, but what follows has the advantage of being self-contained.

```
asymeq1[x_][u_,v_]:=
Module[{U,V,a},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  xuu= D[x[U,V],U,U];
  xuv= D[x[U,V],U,V];
  xvv= D[x[U,V],V,V];
  et:= Det[{xuu,xu,xv}];
  ft:= Det[{xuv,xu,xv}];
  gt:= Det[{xvv,xu,xv}];
  so:= Solve[et a^2 + 2ft a + gt == 0, a];
  Table[u'[v]==a/.so[[i]], {i,2}]/.{U->u[v],V->v}
]
```

The next output gives the system for a monkey saddle.

```
asymeq1[monkeysaddle][u,v]
```

Here is the program from Notebook 7 to compute the torsion of a space curve, used in the Beltrami-Enneper Theorem, and applied in N18.2.

```
τ[α_][t_]:=
Module[{tt,a,B},
  a= Det[{α'[tt], α''[tt], α'''[tt]}];
  B= Cross[α'[tt], α''[tt]];
  a/(B.B) /.tt->t
]
```

■ N18.2

The circular helix was introduced in N7.3. Here is an elliptical version:

```
helix[a_,b_,c_][t_]:={a Cos[t], b Sin[t], c t}
```

It is a coordinate curve on the elliptical helicoid:

```
helicoid[a,b,c][t,1]
```

The inner and outer edges of the elliptical helicoid shown in the next figure are also helices.

```
he:= helicoid[1,2,1]
S @ ParametricPlot3D[he[u,v], {u,-π,5π}, {v,1/3,3},
  PlotPoints->{80,10}]
```

Figure 18.1

The coefficients of the elliptical helicoid's second fundamental form are output below. Only the middle one, fff (standing for f), is nonzero.

```
shape[helicoid[a,b,c]][u,v];
eee
fff
ggg
```

Next, we compute the Gaussian curvature of the helicoid and the torsion of an associated helix. The output of the last line is 0, and this verifies an instance of the Beltrami-Enneper Theorem.

```
oid:= gcurv[helicoid[a,b,c]]
oid[u,v]
ix:= τ[helix[a v, b v, c]]
oid[u,v]+ix[u]^2//Simplify
```

The next few cells investigate asymptotic curves on the exponentially-twisted helicoid.

```
exptwist[a_,c_][u_,v_]:={a v Cos[u], a v Sin[u], a Exp[c u]}
```

```
ParametricPlot3D[exptwist[1,.3][u,v], {u,-π,2π}, {v,-1,1}];
```

```
x:= exptwist[a,c]
eqs:= asymeq1[x][u,v]
DSolve[eqs[[1]],u[v],v]/.C[1]->p
DSolve[eqs[[2]],u[v],v]/.C[1]->q
```

The two expressons for u[v] that appear in the output enable us to define an asymptotic patch, and compare it with the original parametrization.

```
exptwistasym[a_,c_][p_,q_]:=
  a{Exp[c(p-q)/2]Cos[p], Exp[c(p-q)/2]Sin[p], Exp[c p]}
```

```
Module[{x,y,z},
  x= exptwist[1,.3];
  y= exptwistasym[1,.3];
  z[2]= ParametricPlot3D[x[u,v], {u,-π,2π}, {v,-1,1},
    PlotPoints->{24,12}];
  z[1]= ParametricPlot3D[y[p,q], {p,-π,2π}, {q,-π,2π},
    PlotPoints->{36,12}, PlotRange->All];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->-.3]
]
```

Figure 18.2

Here are some more standard computations of second fundamental forms:

```
hp:= paraboloid[1,-1]
shape[hp][u,v]
```

```
shape[monkeysaddle][u,v];
eee
fff
ggg
```

Instructions to plot the funnel surface are given next, though a further examination is relegated to N18.2.1, in which we explain how Figure 18.3 is plotted using the program selectgraphics3d.

```
funnel[a_,b_,c_][u_,v_]:={a v Cos[u], b v Sin[u], c Log[v]}
```

```
fu:= funnel[1,1,1]
ParametricPlot3D[fu[u,v], {u,0,3π/2}, {v,.1,1}];
```

■ N18.2.1

■ N18.3

The following program solves numerically the equations for geodesics on a surface x[u,v] emanating from a given point x[u0,v0], with (arc length) parameter ranging from 0 to b. A geodesic is found for each of m directions that divide the circle (in the tangent space). The symbol e stands for the first order part of the equations, and eqic represents the geodesic equations themselves together with their initial conditions.

```
solvegeo[x_,{u0_,v0_},b_,m_]:=
Module[{su,e,so},
  christoff2[x][u,v];
  su= {u->u[s], v->v[s], p->u'[s], q->v'[s]};
  e[j_]:= e[j]= G[j,1,1]p^2 + 2G[j,1,2]p q + G[j,2,2]q^2 /.su;
  eqic:= {u''[s]+e[1]==0, v''[s]+e[2]==0,
    u[0]==u0, v[0]==v0, u'[0]==Cos[θ], v'[0]==Sin[θ]};
  so:= NDSolve[eqic, {u,v}, {s,0,b}];
  Flatten[Table[so, {θ, 2π/m, 2π, 2π/m}], 1]
]
```

The next figure illustrates a hyperbolic paraboloid, with saddle point (at the origin) marked by a red dot. First we solve for the geodesics. These emanate from a different point P, and "geodesic circles" join points of a given fixed arc length from P.

```
hp:= paraboloid[1,-1]
b:= 2
m:= 50
sg:= solvegeo[hp, {1,-1}/2, b, m]
```

Change Show to S to view the next plot:

```
z0:= Graphics3D[{Hue[0], PointSize[.05], Point[{0,0,0}]}]
z1:= Graphics3D[Table[Line[hp@@{u[t],v[t]}/.sg], {t,0,b,b/8}]]
z2:= ParametricPlot3D[Evaluate[hp@@{u[t],v[t]}/.sg], {t,0,b}]
Show[z0, z1, z2, AspectRatio->Automatic, Boxed->False]
```

N18.3.1

N18.4

We first apply the geodesic plotting program to the sphere. It produces a total of 10 functions of the form t->{u[t],v[t]}, with t running between 0 and 4, defined numerically; details are suppressed from the output.

```
x:= sphere[1]
solvegeo[x, {0,0}, 4, 10]
```

The next command constructs line segments joining a given sequence of points, including one to join the last to the first point (hence the name complete). The optional second argument will allow us to add colors, etc.

```
complete[seq_,opt___]:= {opt, Line[Append[seq, First[seq]]]}
```

The first picture shows the geodesics, as represented in the rectangular chart of the sphere. They would meet at the south pole with parameter value π, so stopping at 3 leaves a small gap. The situation is of course better when these are plotted in space:

```
Module[{x,z},
  x= sphere[1];
  sg= solvegeo[x, {0,0}, 3, 30];
  z[2]= ParametricPlot[Evaluate[{u[t],v[t]}/.sg], {t,0,3},
    PlotRange->{-2.5,2}, PlotStyle->Hue[0]];
  z[1]= ParametricPlot3D[Evaluate[x[u[t],v[t]]/.sg], {t,0,3}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->0]
]
```

Figure 18.5

The next picture shows the geodesic circles, though once these are plotted beyond $\pi/2$ (the equator), they are disconnected in the chart's domain, resulting in spurious lines that could in theory be eliminated using a trick like the one explained in N3.2.

```
Graphics[Table[complete[{u[t],v[t]}/.sg, Hue[0]], {t,0,3,1/2}],
  AspectRatio->Automatic];
```

Again, the situation is clearer when the chart x is applied to map the curves onto the surface in space. In the following code, one can try to see what happens if complete is replaced by Line:

```
Graphics3D[Table[complete[(x@@{u[t],v[t]})/.sg], {t,0,3,1/5}]];
```

We next pass to the torus:

```
to:= torus[3,1,1]
z0:= {PointSize[.05], Point[to@@{π,π}]}//Graphics3D
z1:= ParametricPlot3D[to[u,v], {u,-π,π}, {v,-π,π}, Shading->False]
Show[z1,z0];
```

```
Module[{x,P,z},
  b= 2π;
  x= torus[3,1,1];
  P= {π,π};
  sg= solvegeo[x, P, b, 30];
  z[2]= ParametricPlot[Evaluate[{u[t],v[t]}/.sg], {t,0,b},
    PlotRange->{-5,10}, PlotStyle->Hue[0]];
  z[1]= ParametricPlot3D[Evaluate[x[u[t],v[t]]/.sg], {t,0,b}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->0]
]
```

Figure 18.6

```
Show[Graphics[Table[complete[{u[t],v[t]}/.sg, Hue[0]],
   {t,0,b,1/2}]], AspectRatio->Automatic]
```

To draw the points of a given arc length from the point P with coordinates $\{\pi,\pi\}$, we need to increase the number m of geodesics:

```
b:= 2.5
m:= 200
sg:=solvegeo[to,{π,π},b,m]
ta:= Table[complete[to@@{u[t],v[t]}/.sg, Hue[.7]], {t,0,b,b/5}]
Show[ta//Graphics3D, AspectRatio->Automatic, Boxed->False]
```

The next figure duplicates the biggest closed curve, which is a geodesic circle with center P (the black dot). If the "radius" b were greater or equal to π, the geodesic disk it bounds would no longer be nonsingular (homeomorphic to an ordinary disk) as its boundary would have self-intersections. The geodesic disk is *not* the thin strip of torus represented by the "tongue" in Figure 18.7 but the curved "tube" lying to one side of it. The definition of z2 is an explicit variant of *Mathematica*'s WireFrame construction, and is recorded more formally in N18.4.1.

```
Module[{ta},
  ta= complete[to@@{u[b],v[b]}/.sg, Hue[.7], Thickness[.01]];
  z0= {PointSize[.03], Point[to[π,π]]}//Graphics3D;
  z1:= ParametricPlot3D[to[u,v], {u,-3,3}, {v,-π,π},
    PlotPoints->20];
  z2:= z1/.(Polygon[pts_]:>complete[pts, Hue[.1]]);
  S[z0, z2, ta//Graphics3D, Boxed->False, Shading->False]
]
```

Figure 18.7

■ N18.4.1

■ N18.5

Most of the examples in this notebook section are based on Corollary 18.29 that characterizes geodesics in a Clairaut v patch. The following program solves the resulting geodesic equation.

```
clairautv[pm_,c_,u0_][x_][u_,t_]:=
Module[{Q,R,tt},
  metric[x][u,tt];
  Q= Sqrt[c^2 gg/(ee(ee-c^2))];
  R= u0+Integrate[pm Q, tt]/.tt->t;
  PowerExpand[R]
]
```

Although the function clairautv has u as an argument, its value will not depend on u if x is a Clairaut v patch. The construction is first applied in the plane by means of

```
polarplane[u_,v_]:={v Cos[u], v Sin[u], 0}
```

The complicated looking solution is indeed a straight line.

```
f[t_]= clairautv[1,2,1][polarplane][1,t]
```

```
γ[t_]:= {t Cos[f[t]], t Sin[f[t]]}
ParametricPlot[γ[t], {t,5,9}, Axes->True];
```

The circular cone, studied in N12.2, is another surface with zero Gaussian curvature.

```
clairautv[pm,c,u0][circularcone[a,b]][u,t]
```

The next program parametrizes a geodesic on a Clairaut v patch in space, and is illustrated with some simple examples.

```
clairautgeodesic[x_,c_][t_]:=
Module[{f},
  f[v_]= clairautv[1,c,0][x][u,v];
  x[Re[f[t]], t]
]
```

```
γ:= clairautgeodesic[sphere[1], .5]
ParametricPlot3D[γ[t], {t,0,1}, Axes->True]
```

```
γ:= clairautgeodesic[paraboloid[1,1], 0]
ParametricPlot3D[γ[t], {t,-1,1}, Axes->True]
```

Here we compute the equations of geodesics on a pseudosphere:

```
Clear[a,c,u0,pm]
clairautv[pm,c,u0][pseudosphere[a]][u,t]//FullSimplify
clairautgeodesic[pseudosphere[a],u0][t]//FullSimplify
```

The simplified equation for u output first from the previous cell is recorded below; it will be inserted into a plotting program in N18.6.

```
psclairautv[a_,pm_,c_,u0_][t_]=
  u0 + pm Sqrt[a^2-2c^2-a^2 Cos[2t]]Csc[t]/(Sqrt[2] c)
```

Further examples of Clairaut patches are contained in the subsection that follows in the notebook's electronic version.

■ N18.5.1

■ N18.6

We repeat what we did above for the torus. First, we slightly alter the definition of the pseudosphere to avoid difficulties in applying the numerical method, and multiply by 0.99 to make the dot and geodesics visible.

```
psm[u_,v_]:=
  {Cos[u]Sin[v], Sin[u]Sin[v], Cos[v]+Log[.001+Tan[v/2]]}
P:= {5.3,2.5}
z0:= {PointSize[.05], Point[psm@@P]}//Graphics3D
z1:= ParametricPlot3D[.99psm[u,v], {u,0,2π}, {v,π/2,π},
  Shading->False]
{z1,z0}
```

The next cell customizes the program solvegeo above to cope with the pseudosphere. There are two main modifications: first we require geodesics spanning only the first quadrant, and we also wish each to have a different color:

```
sg:= Module[{su,e,so},
  christoff2[psm][u,v];
  su= {u->u[s], v->v[s], p->u'[s], q->v'[s]};
  e[j_]:= e[j]= G[j,1,1]p^2 + 2G[j,1,2]p q + G[j,2,2]q^2 /.su;
  eqic:= {u''[s]+e[1]==0, v''[s]+e[2]==0,
    u[0]==P[[1]], v[0]==P[[2]],
    u'[0]==Cos[θ], v'[0]==Sin[θ]};
  so:= NDSolve[eqic, {u,v}, {s,0,5}];
  Flatten[Table[so, {θ,0,π/2,π/8}], 1]
]
```

```
Module[{α,αα,z},
  z0:= {PointSize[.03], Point[psm@@P]}//Graphics3D;
  z1:= ParametricPlot3D[.99psm[u,v], {u,0,2π}, {v,π/2,π}];
  z2:= z1/.(Polygon[pts_]:>complete[pts, GrayLevel[.8]]);
  α[i_][t_]:= psm@@{u[t],v[t]}/.sg[[i+1]];
  αα[i_]:= highlight[α[i], i/5,1/100];
  z[i_]:= ParametricPlot3D[αα[i-1][t]//Evaluate, {t,0,.8i+1}];
  S[z2, z0, Array[z,5], ImageSize->400]
]
```

Figure 18.9

Changing the second coordinate of P from 2.5 to 2.7 as a dramatic effect. Using the explicit expression found in N18.5 for a geodesic on the pseudosphere, we plot u modulo 2π as a function of t to see how a geodesic winds its way up the neck.

```
psg[c_,t_]:= psclairautv[1,1,c,0][t]//N
f[c_,t_]:= 2π(psg[c,t]/(2π) - Floor[psg[c,t]/(2π)])
Plot[f[.02,t], {t, 3.10, 3.12}]
```

■ N18.7

The first command can (with appropriate sign choices) be used for both types of paraboloid:

```
Module[{h,x,z},
  h[u_,v_]= (u^2+v^2)^(-1/2);
  x[u_,v_]= {u, v, h[u,v]};
  y[u_,v_]= gcurv[x][u,v]/2;
  xx[u_,v_]= Append[x[u,v], FaceForm[Hue[.2+1.5y[u,v]]]];
  S @ ParametricPlot3D[xx[u,v], {u,-2,2}, {v,-2,2},
    PlotPoints->40, Lighting->False, ImageSize->350,
    PlotRange->{0,3}]
]
```

Figure 18.10

Chapter 19

Principal Curves and Umbilic Points

In this chapter we continue our study of special curves on surfaces. A *principal curve* on a surface is a curve whose velocity always points in a *principal direction*, that is, a direction in which the normal curvature is a maximum or a minimum. In Section 19.1, we derive the differential equation for the principal curves on a patch in $\mathbb{R}^3$ and give examples of its solution. As in Section 18.2, the idea then is to reparametrize a surface with specified coordinate curves.

An *umbilic point* on a surface is a point at which the principal curvatures are equal, so that at such a point it is not possible to distinguish principal directions. Every point of a sphere is an umbilic point. We discuss umbilic points in Section 19.2 and show, for example, that if a regular surface $\mathcal{M} \subset \mathbb{R}^3$ consists entirely of umbilic points, then it is (perhaps not surprisingly) part of a plane or sphere. For surfaces such as an ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

with a, b, c distinct, the umbilic points are isolated and can be considered to be degenerate principal curves. In fact, each looks very much like a navel, hence the name. The four umbilic points on an ellipsoid (with a, b, c distinct) are easy to locate visually, provided one draws the ellipsoid so that the principal curves can be seen, as in Figure 19.6.

The coefficients of the first and second fundamental forms of a regular surface $\mathcal{M}$ in $\mathbb{R}^3$ are not independent of one another. The Peterson–Mainardi–Codazzi equations consist of two equations that relate the derivatives of the coefficients of the second fundamental form to the coefficients of the first and second fundamental forms themselves. They simplify considerably when the coordinate curves are principal, and this justifies their inclusion in this chapter.

The derivation of these equations in Section 19.3 gives rise to formulas that foresee the definition of the Riemann curvature tensor in Chapter 25. In Section 19.4 we use the Peterson–Mainardi–Codazzi equations to prove Hilbert's Lemma (Lemma 19.12), which in turn is used to prove a so-called "rigidity result," namely Liebmann's Theorem.

For certain surfaces, such as the ellipsoid, there is a much more effective way to find principal curves than solving the differential equation discussed in Section 19.1. It is based on the notion of a *triply orthogonal system of surfaces*, something introduced by Lamé to study equations of mathematical physics (see, for example, [MoSp]). The corresponding notion in the plane, of an orthogonal system of curves, is discussed in the Exercises.

The connection between principal curves and triply orthogonal systems will become clear after we prove Dupin's Theorem, which states that surfaces from different families intersect in principal curves. Section 19.5 begins with some examples of triply orthogonal systems, and then introduces curvilinear patches in order to prove Dupin's Theorem 19.21. In Section 19.6, we define elliptic coordinates and use them to find and draw the principal curves on ellipsoids, hyperboloids of one sheet and hyperboloids of two sheets.

In Section 19.7, we succeed in using parabolic coordinates to find the principal curves of elliptic and hyperbolic paraboloids, and this raises the question of how widely applicable our method is. In fact, we show that there exists a triply orthogonal system containing any given surface. The construction uses the concept of *parallel surface*, and in Section 19.8 we show how to construct the parallel surface $\mathcal{M}(t)$ to a given surface $\mathcal{M} \subseteq \mathbb{R}^3$. We conclude the chapter with a study of the shape operator of parallel surfaces, and prove the important theorem of Bonnet on the subject.

19.1 The Differential Equation for Principal Curves

In this section we first determine when a tangent vector to a patch is a principal vector. Then we derive the differential equation for principal curves, examples of which have already been discussed in Section 15.2.

Lemma 19.1. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. A tangent vector $\mathbf{v_p} = v_1\mathbf{x}_u + v_2\mathbf{x}_v$ is a principal vector if and only if*

$$\det \begin{pmatrix} v_2^2 & -v_1v_2 & v_1^2 \\ E & F & G \\ e & f & g \end{pmatrix} = 0. \tag{19.1}$$

Proof. Using the notation of Theorem 13.16, page 394, we compute $S(\mathbf{v_p}) \times \mathbf{v_p}$ and apply Lemma 15.5, page 466:

$$\begin{aligned} S(\mathbf{v_p}) \times \mathbf{v_p} &= S(v_1\mathbf{x}_u + v_2\mathbf{x}_v) \times (v_1\mathbf{x}_u + v_2\mathbf{x}_v) \\ &= \big(a_{21}v_1^2 - (a_{11} - a_{22})v_1v_2 - a_{12}v_2^2\big)\mathbf{x}_u \times \mathbf{x}_v \\ &= \left(\left(\frac{eF - fE}{EG - F^2}\right)v_1^2 - \left(\frac{fF - eG}{EG - F^2} - \frac{fF - gE}{EG - F^2}\right)v_1v_2 \right. \\ &\qquad \left. - \left(\frac{gF - fG}{EG - F^2}\right)v_2^2\right)\mathbf{x}_u \times \mathbf{x}_v \\ &= -\left((fE - eF)v_1^2 + (gE - eG)v_1v_2 + (gF - fG)v_2^2\right)\frac{\mathbf{x}_u \times \mathbf{x}_v}{EG - F^2}. \end{aligned}$$

From Lemma 15.5, page 466, we know that a vector $\mathbf{v_p}$ is principal if and only if $S(\mathbf{v_p}) \times \mathbf{v_p} = 0$. It now follows that $\mathbf{v_p}$ is principal if and only if

$$(fE - eF)v_1^2 + (gE - eG)v_1v_2 + (gF - fG)v_2^2 = 0. \tag{19.2}$$

This is just another way of writing (19.1). ∎

The lemma and its proof enable us to write down the differential equation for the principal curves, in analogy to the differential equation on page 559 for the asymptotic curves. Substituting the velocity vector $(u'(t), v'(t))$ in place of (v_1, v_2) in (19.2) gives

Corollary 19.2. Let $\boldsymbol{\alpha}$ be a curve that lies on the trace of a patch $\mathbf{x}$. Write $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$. Then $\boldsymbol{\alpha}$ is a principal curve if and only if

$$\begin{aligned} (fE - eF)\big(\boldsymbol{\alpha}(t)\big)u'(t)^2 + (gE - eG)\big(\boldsymbol{\alpha}(t)\big)u'(t)v'(t) \\ +(gF - fG)\big(\boldsymbol{\alpha}(t)\big)v'(t)^2 = 0 \end{aligned}$$

for all t.

More informally, we may write the equation as

$$(fE - eF)u'^2 + (gE - eG)u'v' + (gF - fG)v'^2 = 0.$$

Let us use it to find the principal curves on the surface

$$\mathsf{helicoid}[a, b](u, v) = (a\,v\cos u,\ a\,v\sin u,\ b\,u),$$

introduced on page 376. Even better, we shall find a principal patch that reparametrizes the helicoid. The coefficients of the first and second fundamental forms of $\mathsf{helicoid}[a, b]$ are easily computed to be

$$F = e = g = 0, \qquad E = b^2 + a^2v^2, \qquad G = a^2, \qquad f = \frac{a\,b}{b^2 + a^2v^2}.$$

The differential equation reduces to $fEu'^2 - fGv'^2 = 0$, whence

$$\frac{du}{dv} = \frac{a}{\sqrt{a^2v^2 + b^2}}.$$

Integrating, we discover the two solutions

$$u - 2p = \operatorname{arcsinh}\frac{a\,v}{b} \qquad \text{and} \qquad u - 2q = -\operatorname{arcsinh}\frac{a\,v}{b},$$

where p and q are constants of integration. Hence to reparametrize $\mathsf{helicoid}[a, b]$, we set

$$u = p + q, \qquad v = \frac{b}{a}\sinh(-p + q),$$

so as to obtain

$$\mathbf{y}(p, q) = \big(-b\cos(p + q)\sinh(p - q),\ -b\sin(p + q)\sinh(p - q),\ b(p + q)\big).$$

The mapping $\mathbf{y}$ is called a ***principal patch***, since by construction the curves $p =$ constant and $q =$ constant are principal. The two families of such curves are visible in Figure 19.1.

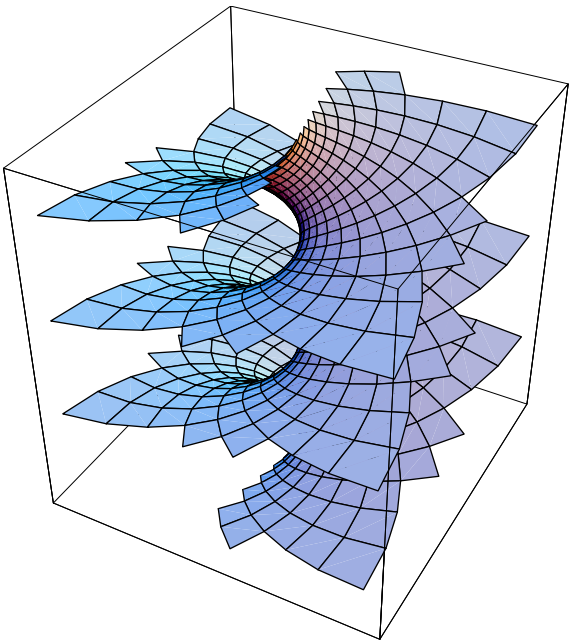


Figure 19.1: Principal patch on a helicoid

It is interesting to compare this plot of a helicoid by principal curves with that by asymptotic curves in Figure 18.1 on page 562.

19.2 Umbilic Points

In this section we consider those points on a surface in $\mathbb{R}^3$ at which all the normal curvatures are equal.

Definition 19.3. *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$. A point $\mathbf{p} \in \mathcal{M}$ is called an* ***umbilic point*** *provided the principal curvatures at $\mathbf{p}$ are equal: $k_1(\mathbf{p}) = k_2(\mathbf{p})$.*

It is clear that all points of a plane or a sphere are umbilic points.

The vertex $(0, 0, 0)$ of the paraboloid $z = x^2 + y^2$ is an umbilic point, since the symmetry there prevents us from choosing a principal direction. Such a symmetry does not apply to other points, and the vertex is in fact an isolated umbilic point. By contrast, an elliptical paraboloid $z = ax^2 + by^2$ with $a > b > 0$ has two umbilic points. Figure 19.2 is a rather uninformative plot of these two points made in Notebook 19, and it is the purpose of this chapter to better explain the existence of such umbilic points on paraboloids and similar surfaces.

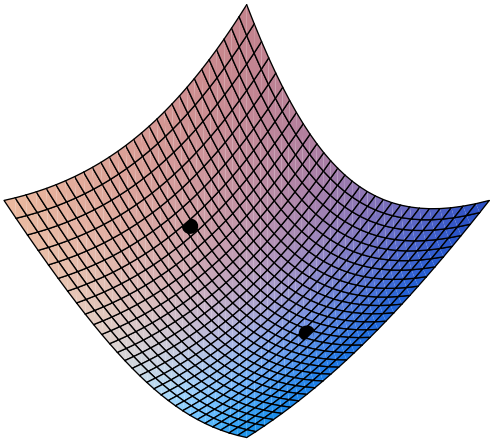


Figure 19.2: Two umbilic points on part of an elliptical paraboloid

Lemma 19.4. *A point $\mathbf{p}$ on a regular surface $\mathcal{M} \subset \mathbb{R}^3$ is an umbilic point if and only if the shape operator of $\mathcal{M}$ at $\mathbf{p}$ is a multiple of the identity.*

Proof. Let $S_\mathbf{p}$ denote the shape operator of $\mathcal{M}$ at $\mathbf{p}$. Then $S_\mathbf{p}$ is a multiple of the identity if and only if all of the normal curvatures of $\mathcal{M}$ at $\mathbf{p}$ coincide. This is true if and only if the maximum and minimum of the normal curvatures, namely, $k_1(\mathbf{p})$ and $k_2(\mathbf{p})$, coincide. Write $k(\mathbf{p}) = k_1(\mathbf{p}) = k_2(\mathbf{p})$; then $S_\mathbf{p} = k(\mathbf{p})I$, where I denotes the identity map on $\mathcal{M}_\mathbf{p}$. ∎

First, let us determine the regular surfaces that consist entirely of umbilic points.

Lemma 19.5. *If all of the points of a connected regular surface $\mathcal{M} \subset \mathbb{R}^3$ are umbilic, then $\mathcal{M}$ has constant Gaussian curvature $K \geqslant 0$.*

Proof. Let $k\colon \mathcal{M} \to \mathbb{R}$ be the function such that $k(\mathbf{q})$ is the common value of the principal curvatures at $\mathbf{q} \in \mathcal{M}$. Let $\mathbf{p} \in \mathcal{M}$, and let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch such that $\mathbf{p} \in \mathbf{x}(\mathcal{U})$. We can assume that $\mathcal{U}$ is connected. Then we have

$$\mathbf{U}_u = -S(\mathbf{x}_u) = -k\mathbf{x}_u, \tag{19.3}$$

$$\mathbf{U}_v = -S(\mathbf{x}_v) = -k\mathbf{x}_v. \tag{19.4}$$

We differentiate (19.3) with respect to v and (19.4) with respect to u, then subtract the results. Since $\mathbf{U}_{uv} = \mathbf{U}_{vu}$ and $\mathbf{x}_{uv} = \mathbf{x}_{vu}$, we obtain

$$k_v\mathbf{x}_u - k_u\mathbf{x}_v = 0.$$

The linear independence of $\mathbf{x}_u$ and $\mathbf{x}_v$ now implies that $k_v = 0 = k_u$, and k is constant on the neighborhood $\mathcal{U}$ of $\mathbf{p}$. Moreover, the Gaussian curvature K is also constant on $\mathcal{U}$, because it is given by $K = k^2 \geqslant 0$.

Let $\mathcal{V} = \{\, \mathbf{q} \in \mathcal{M} \mid K(\mathbf{q}) = K(\mathbf{p}) \,\}$. Then $\mathcal{V}$ is obviously closed, and by the previous argument it is also open. Since $\mathcal{M}$ is connected, it follows (see page 334) that $\mathcal{V} = \mathcal{M}$. Hence K is a constant nonnegative function on $\mathcal{M}$. ∎

To determine the connected regular surfaces consisting entirely of umbilic points, we imitate the proof of Theorem 1.22 on page 16.

Theorem 19.6. *Let $\mathcal{M}$ be a connected regular surface in $\mathbb{R}^3$ consisting entirely of umbilic points. Then $\mathcal{M}$ is part of a plane or sphere.*

Proof. We know from Lemma 19.5 that $\mathcal{M}$ has constant curvature $K \geqslant 0$, and that $K = k^2$, where k is the common value of the principal curvatures. If $K = 0$, then both k and S vanish identically. Hence $\mathbf{D}_\mathbf{v}\mathbf{U} = 0$ for any tangent vector $\mathbf{v}$ to $\mathcal{M}$ (see page 386). Therefore, for each $\mathbf{p} \in \mathcal{M}$ we have $\mathbf{U}(\mathbf{p}) = (n_1, n_2, n_3)_\mathbf{p}$, where the components n_i of the unit normal vector are constants. Then $\mathcal{M}$ is contained in the plane perpendicular to $(n_1, n_2, n_3)_\mathbf{p}$.

Next, suppose $K > 0$. Choose a point $\mathbf{p} \in \mathcal{M}$ and a unit normal vector $\mathbf{U}(\mathbf{p})$ to $\mathcal{M}$ at $\mathbf{p}$. We shall show that the point $\mathbf{c}$ defined by

$$\mathbf{c} = \mathbf{p} + \frac{1}{k}\mathbf{U}(\mathbf{p})$$

is equidistant from all points of $\mathcal{M}$. To this end, let $\mathbf{q}$ be any point of $\mathcal{M}$, and let $\boldsymbol{\alpha}\colon (a, b) \to \mathcal{M}$ be a curve with $a < 0 < 1 < b$ such that $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}(1) = \mathbf{q}$. We extend $\mathbf{U}(\mathbf{p})$ to a unit normal vector field $\mathbf{U} \circ \boldsymbol{\alpha}$ along $\boldsymbol{\alpha}$.

Define a new curve $\boldsymbol{\gamma}\colon (a, b) \to \mathbb{R}^3$ by

$$\boldsymbol{\gamma}(t) = \boldsymbol{\alpha}(t) + \frac{1}{k}\mathbf{U}\big(\boldsymbol{\alpha}(t)\big),$$

so that

$$\boldsymbol{\gamma}'(t) = \boldsymbol{\alpha}'(t) + \frac{1}{k}(\mathbf{U} \circ \boldsymbol{\alpha})'(t). \tag{19.5}$$

But

$$(\mathbf{U} \circ \boldsymbol{\alpha})'(t) = -S\big(\boldsymbol{\alpha}'(t)\big) = -k\boldsymbol{\alpha}'(t),$$

so that (19.5) reduces to $\boldsymbol{\gamma}' = 0$. Hence $\boldsymbol{\gamma}$ must be a constant, that is, a single point in $\mathbb{R}^3$, in fact, the point $\mathbf{c}$. Thus

$$\mathbf{c} = \boldsymbol{\gamma}(0) = \boldsymbol{\gamma}(1) = \mathbf{q} + \frac{1}{k}\mathbf{U}(\mathbf{q}). \tag{19.6}$$

From (19.6) it is immediate that

$$\|\mathbf{c} - \mathbf{q}\| = \frac{1}{|k|}.$$

Thus $\mathcal{M}$ is part of a sphere of radius $1/|k|$ with center $\mathbf{c}$. ∎

Next, we give a useful criterion for finding umbilic points.

Lemma 19.7. *A point* $\mathbf{p}$ *on a patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *is an umbilic point if and only if there is a number* k *such that at* $\mathbf{p}$ *we have*

$$E = k\,e, \qquad F = k\,f, \qquad G = k\,g. \tag{19.7}$$

Proof. The point $\mathbf{p}$ is umbilic if and only if equation (19.2) holds for all v_1, v_2. It is obvious that (19.7) implies this. For the converse, see Exercise 10. ∎

19.3 The Peterson–Mainardi–Codazzi Equations

The nine coefficients in the Gauss equations (17.8), page 538, are not independent. It turns out that there is a relation among e, f, g, their derivatives and the Christoffel symbols. These relations were proved by Gauss ([Gauss2, paragraph 11]) using obscure notation, then reproved successively by Peterson[1] in his thesis [Pson], Mainardi[2] [Main] and Codazzi[3] [Codaz]. See [Reich, page 303–305]

---

[1] 

Karl Mikhailovich Peterson (1828–1881). A student of Minding and Senff in Dorpat (Tartu). Peterson was one of the founders of the Moscow Mathematical Society. Peterson's derivation of the equations of Theorem 19.8 was not generally known in his lifetime.

[2] Gaspare Mainardi (1800–1879). Professor at the University of Padua.

[3] Delfino Codazzi (1824–1873). Professor at the University of Padua. The equations (19.8) were proved by Mainardi [Main] in 1856. However, Codazzi's formulation was simpler because he was careful that his expressions had geometric meaning, and his applications were wider. Since Codazzi submitted an early version of his work to the French Academy of Sciences for a prize competition in 1859, his contribution was known long before those of Peterson and Mainardi. Codazzi also published papers on geodesic triangles, equiareal mappings and the stability of floating bodies.

and [Cool2] for details on the history of these equations.

Theorem 19.8. (The Peterson–Mainardi–Codazzi Equations) *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a regular patch. Then*

$$\text{(19.8)} \qquad \begin{cases} \dfrac{\partial e}{\partial v} - \dfrac{\partial f}{\partial u} = e\Gamma^1_{12} + f\left(\Gamma^2_{12} - \Gamma^1_{11}\right) - g\Gamma^2_{11}, \\ \dfrac{\partial f}{\partial v} - \dfrac{\partial g}{\partial u} = e\Gamma^1_{22} + f\left(\Gamma^2_{22} - \Gamma^1_{12}\right) - g\Gamma^2_{12}. \end{cases}$$

Proof. The idea of the proof is to differentiate the equations of (17.8) and use the fact that $\mathbf{x}_{uuv} = \mathbf{x}_{uvu}$ and $\mathbf{x}_{uvv} = \mathbf{x}_{vvu}$. Thus from the first two equations of (17.8), it follows that

$$\begin{aligned} 0 &= \mathbf{x}_{uuv} - \mathbf{x}_{uvu} \\ &= \frac{\partial}{\partial v}\left(\Gamma^1_{11}\mathbf{x}_u + \Gamma^2_{11}\mathbf{x}_v + e\,\mathbf{U}\right) - \frac{\partial}{\partial u}\left(\Gamma^1_{12}\mathbf{x}_u + \Gamma^2_{12}\mathbf{x}_v + f\,\mathbf{U}\right) \\ &= \left(\frac{\partial\Gamma^1_{11}}{\partial v} - \frac{\partial\Gamma^1_{12}}{\partial u}\right)\mathbf{x}_u + \left(\frac{\partial\Gamma^2_{11}}{\partial v} - \frac{\partial\Gamma^2_{12}}{\partial u}\right)\mathbf{x}_v + (e_v - f_u)\mathbf{U} \\ &\qquad -\Gamma^1_{12}\mathbf{x}_{uu} + \left(\Gamma^1_{11} - \Gamma^2_{12}\right)\mathbf{x}_{uv} + \Gamma^2_{11}\mathbf{x}_{vv} + e\,\mathbf{U}_v - f\,\mathbf{U}_u. \end{aligned}$$

We use (17.8) again to expand $\mathbf{x}_{uu}, \mathbf{x}_{uv}, \mathbf{x}_{vv}$, and the Weingarten equations (13.10) on page 394 to expand $\mathbf{U}_v, \mathbf{U}_u$. Two coefficients of $\mathbf{x}_u$ cancel out, and collecting the remaining terms we obtain:

$$\begin{aligned} 0 &= \left(\frac{\partial\Gamma^1_{11}}{\partial v} - \frac{\partial\Gamma^1_{12}}{\partial u} - \Gamma^2_{12}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22} + e\frac{gF - fG}{EG - F^2} - f\frac{fF - eG}{EG - F^2}\right)\mathbf{x}_u \\ &\quad + \left(\frac{\partial\Gamma^2_{11}}{\partial v} - \frac{\partial\Gamma^2_{12}}{\partial u} + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \left(\Gamma^2_{12}\right)^2\right. \\ &\qquad\qquad \left. + e\frac{fF - gE}{EG - F^2} - f\frac{eF - fE}{EG - F^2}\right)\mathbf{x}_v \\ &\quad + \left(e_v - f_u - e\Gamma^1_{12} + f\left(\Gamma^1_{11} - \Gamma^2_{12}\right) + g\Gamma^2_{11}\right)\mathbf{U} \\ &= \left(\frac{\partial\Gamma^1_{11}}{\partial v} - \frac{\partial\Gamma^1_{12}}{\partial u} - \Gamma^2_{12}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22} + F\frac{eg - f^2}{EG - F^2}\right)\mathbf{x}_u \\ &+ \left(\frac{\partial\Gamma^2_{11}}{\partial v} - \frac{\partial\Gamma^2_{12}}{\partial u} + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \left(\Gamma^2_{12}\right)^2 - E\frac{eg - f^2}{EG - F^2}\right)\mathbf{x}_v \\ &\quad + \left(e_v - f_u - e\Gamma^1_{12} + f\left(\Gamma^1_{11} - \Gamma^2_{12}\right) + g\Gamma^2_{11}\right)\mathbf{U}. \end{aligned}$$

Taking the normal component in the last line yields the first equation of (19.8).

A similar argument exploiting $\mathbf{x}_{uvv} = \mathbf{x}_{uvu}$ and the second and third equations of (17.8) gives the second equation. ∎

More generally, let us consider the consequences of the equations

$$\text{(19.9)}\qquad \begin{cases} \mathbf{x}_{uuv} - \mathbf{x}_{uvu} = 0, \\ \mathbf{x}_{uvv} - \mathbf{x}_{vvu} = 0, \\ \mathbf{U}_{uv} - \mathbf{U}_{vu} = 0 \end{cases}$$

for a regular patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$. When the Gauss equations (17.8) are substituted into (19.9), the result is three equations of the form

$$\text{(19.10)}\qquad \begin{cases} A_1\mathbf{x}_u + B_1\mathbf{x}_v + C_1\mathbf{U} = 0, \\ A_2\mathbf{x}_u + B_2\mathbf{x}_v + C_2\mathbf{U} = 0, \\ A_3\mathbf{x}_u + B_3\mathbf{x}_v + C_3\mathbf{U} = 0. \end{cases}$$

Since $\mathbf{x}_u$, $\mathbf{x}_v$, $\mathbf{U}$ are linearly independent, we must have $A_j = B_j = C_j = 0$ for $j = 1, 2, 3$. In the proof of Theorem 19.8, we obtained the Peterson–Mainardi–Codazzi equations (19.8) from $C_1 = C_2 = 0$. In fact, $A_3 = B_3 = 0$ also yield (19.8). Furthermore, $C_3 = 0$ because when it is computed, all terms cancel. From $A_1 = B_1 = A_2 = B_2 = 0$ we get the four equations

$$\text{(19.11)}\qquad \begin{cases} EK = \dfrac{\partial\Gamma^2_{11}}{\partial v} - \dfrac{\partial\Gamma^2_{12}}{\partial u} + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \bigl(\Gamma^2_{12}\bigr)^2, \\ -FK = \dfrac{\partial\Gamma^1_{11}}{\partial v} - \dfrac{\partial\Gamma^1_{12}}{\partial u} - \Gamma^2_{12}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22}, \\ GK = \dfrac{\partial\Gamma^1_{22}}{\partial u} - \dfrac{\partial\Gamma^1_{12}}{\partial v} + \Gamma^1_{22}\Gamma^1_{11} + \Gamma^2_{22}\Gamma^1_{12} - \bigl(\Gamma^1_{12}\bigr)^2 - \Gamma^2_{12}\Gamma^1_{22}, \\ -FK = \dfrac{\partial\Gamma^2_{22}}{\partial u} - \dfrac{\partial\Gamma^2_{12}}{\partial v} + \Gamma^1_{22}\Gamma^2_{11} - \Gamma^1_{12}\Gamma^2_{12}. \end{cases}$$

The first two are visible on the previous page. But all four equations turn out to be equivalent to one another. Since each of them (when the Christoffel symbols are expanded) expresses the Gaussian curvature in terms of E, F, G, any one of them provides a new proof of Gauss's *Theorema Egregium* (Theorem 17.5 on page 536).

The Peterson–Mainardi–Codazzi equations simplify greatly in the case of principal patches, discussed on page 467.

Corollary 19.9. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a principal patch. Then*

$$\text{(19.12)}\qquad \begin{cases} \dfrac{\partial e}{\partial v} = \dfrac{E_v}{2}\Bigl(\dfrac{e}{E} + \dfrac{g}{G}\Bigr), \\ \dfrac{\partial g}{\partial u} = \dfrac{G_u}{2}\Bigl(\dfrac{e}{E} + \dfrac{g}{G}\Bigr). \end{cases}$$

Proof. Equation (19.12) follows from Lemma 17.8 and Theorem 19.8 ∎

We now state a formula linking principal curvatures and their derivatives:

Corollary 19.10. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular principal patch with principal curvatures k_1 and k_2. Then*

$$\begin{cases} k_{1v} = \dfrac{E_v}{2E}(k_2 - k_1), \\ k_{2u} = \dfrac{G_u}{2G}(k_1 - k_2). \end{cases} \tag{19.13}$$

For a proof of this result, see Exercise 3.

19.4 Hilbert's Lemma and Liebmann's Theorem

We need to start this section by quoting a lemma that can be found in [dC1, page 185]. It is actually a corollary of a more general result that implies that lines of curvature can be used as coordinate curves.

Lemma 19.11. *Let $\mathbf{p}$ be a nonumbilical point of a regular surface $\mathcal{M}$ in $\mathbb{R}^3$. Then there exists a principal patch that parametrizes a neighborhood of $\mathbf{p}$.*

We shall use this lemma to prove the next more celebrated one, that is a stepping stone to the theorem that follows.

Lemma 19.12. (Hilbert[4]) *Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$, and let $\mathbf{p} \in \mathcal{M}$ be a point such that*

(i) *k_1 has a local maximum at $\mathbf{p}$;*

(ii) *k_2 has a local minimum at $\mathbf{p}$;*

(iii) *$k_1(\mathbf{p}) > k_2(\mathbf{p})$.*

Then $K(\mathbf{p}) \leqslant 0$.

Proof. At $\mathbf{p}$ we have

$$k_{1v} = k_{2u} = 0, \qquad k_{1vv} \leqslant 0 \qquad \text{and} \qquad k_{2uu} \geqslant 0$$

because of (i) and (ii). Condition (iii) and Lemma 19.11 imply that there exists a principal patch $\mathbf{x}$ parametrizing a neighborhood of $\mathbf{p}$. Let E, F, G denote the

---



[4] David Hilbert (1862–1943). Professor at Göttingen, the leading German mathematician of his time. Hilbert contributed to many branches of mathematics, including invariants, algebraic number fields, functional analysis, integral equations, mathematical physics, and the calculus of variations. Hilbert's most famous contribution to differential geometry is [Hil], in which he proved that no regular surface of constant negative curvature in $\mathbb{R}^3$ can be complete.

coefficients of the first fundamental form of $\mathbf{x}$. From the Peterson–Codazzi–Mainardi equations for principal patches, that is (19.13), it follows that

$$E_v(\mathbf{p}) = G_u(\mathbf{p}) = 0. \tag{19.14}$$

Furthermore, if we differentiate (19.13) we obtain

$$\begin{cases} k_{1vv} &= \left(\dfrac{E\,E_{vv} - E_v^2}{2E^2}\right)(k_2 - k_1) + \dfrac{E_v}{2E}(k_2 - k_1)_v, \\ k_{2uu} &= \left(\dfrac{G\,G_{uu} - G_u^2}{2G^2}\right)(k_1 - k_2) + \dfrac{G_u}{2G}(k_1 - k_2)_u. \end{cases} \tag{19.15}$$

When we evaluate (19.15) at $\mathbf{p}$, we obtain

$$0 \geqslant k_{1vv}(\mathbf{p}) = \frac{E_{vv}}{2E}(k_2 - k_1)\bigg|_{\mathbf{p}} \quad \text{and} \quad 0 \leqslant k_{2uu}(\mathbf{p}) = \frac{G_{uu}}{2G}(k_1 - k_2)\bigg|_{\mathbf{p}}.$$

It now follows from (iii) that

$$E_{vv}(\mathbf{p}) \geqslant 0 \qquad \text{and} \qquad G_{uu}(\mathbf{p}) \geqslant 0.$$

Thus from (19.14) and (17.6), page 534, we get

$$K(\mathbf{p}) = -\frac{E_{vv} + G_{uu}}{2E\,G}\bigg|_{\mathbf{p}} \leqslant 0. \ \blacksquare$$

Now, we can prove the main result of this section, which is sometimes referred to as the "rigidity of the sphere."

Theorem 19.13. (Liebmann[5]) *Let $\mathcal{M}$ be a compact surface in $\mathbb{R}^3$ with constant Gaussian curvature K. Then $\mathcal{M}$ is a sphere of radius $1/\sqrt{K}$.*

Proof. We have

$$H^2 - K = \left(\frac{k_1 - k_2}{2}\right)^2. \tag{19.16}$$

Since $\mathcal{M}$ is compact, the function $H^2 - K$ assumes its maximum value (which must be nonnegative) at some point $\mathbf{p} \in \mathcal{M}$. Assume that this maximum is positive; we shall obtain a contradiction.

Choose an oriented neighborhood $\mathcal{U}$ containing $\mathbf{p}$ for which $H^2 - K$ is positive on $\mathcal{U}$. The choice of orientation gives rise to a unit normal vector field $\mathbf{U}$ defined on all of $\mathcal{U}$. Let k_1 and k_2 be the principal curvatures defined with respect to $\mathbf{U}$. Since $K > 0$ by Theorem 13.36 on page 410, we can arrange that $k_1 \geqslant k_2 > 0$

[5]Heinrich Liebmann (1874–1939). German mathematician, and professor at Munich and Heidelberg.

on $\mathcal{U}$. In fact, because of the identity (19.16), we have $k_1 > k_2 > 0$ on $\mathcal{U}$. Since $K = k_1 k_2$ is constant, it follows that k_1 has a maximum at $\mathbf{p}$ and k_2 has a minimum at $\mathbf{p}$. Hence Lemma 19.12 implies that $K(\mathbf{p}) \leqslant 0$, contradicting the assumption that $K(\mathbf{p}) > 0$.

Therefore, $H^2 - K$ is zero at $\mathbf{p}$. Equation (19.16) then implies that $H^2 - K$ vanishes identically on $\mathcal{M}$, so all points of $\mathcal{M}$ are umbilic. Now Theorem 19.6 implies that $\mathcal{M}$ is part of a sphere of radius $1/\sqrt{K}$; since $\mathcal{M}$ is compact, $\mathcal{M}$ must be an entire sphere. ∎

19.5 Triply Orthogonal Systems of Surfaces

This subject is based on the following notion.

Definition 19.14. *A* ***triply orthogonal system of surfaces*** *on an open set $\mathcal{U} \subseteq \mathbb{R}^3$ consists of three families $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$, of surfaces such that*

(i) *each point of $\mathcal{U}$ lies on one and only one member of each family;*

(ii) *each surface in each family meets every member of the other two families orthogonally.*

Three simple examples of such a system of surfaces are inherent in the use of Cartesian, cylindrical and spherical coordinates in space. Cartesian coordinates are effectively defined by three families of planes parallel to the yz-, zx- and xy-planes, as shown in Figure 19.3. To give the second example, let $\mathcal{U}$ be the complement of the vertical axis in $\mathbb{R}^3$; that is,

$$\mathcal{U} = \{\, (p_1, p_2, p_3) \in \mathbb{R}^3 \mid (p_1, p_2) \neq (0, 0) \,\}.$$

Then the system consisting of

$$\begin{aligned} \mathfrak{A} &= \text{planes through the } z\text{-axis}, \\ \mathfrak{B} &= \text{planes parallel to the } xy\text{-plane}, \\ \mathfrak{C} &= \text{circular cylinders around the } z\text{-axis} \end{aligned}$$

constitutes a triply orthogonal system of surfaces on $\mathcal{U}$. See Figure 19.4.

For the third case, set $\mathcal{V} = \mathbb{R}^3 \setminus \{(0, 0, 0)\}$, and

$$\begin{aligned} \mathfrak{A} &= \text{planes containing the } z\text{-axis}, \\ \mathfrak{B} &= \text{circular cones with vertex } (0, 0, 0) \text{ and common axis } x = 0 = y, \\ \mathfrak{C} &= \text{spheres centered at the origin } (0, 0, 0). \end{aligned}$$

This gives rise to a triply orthogonal system of surfaces on $\mathcal{V}$. See Figure 19.5.

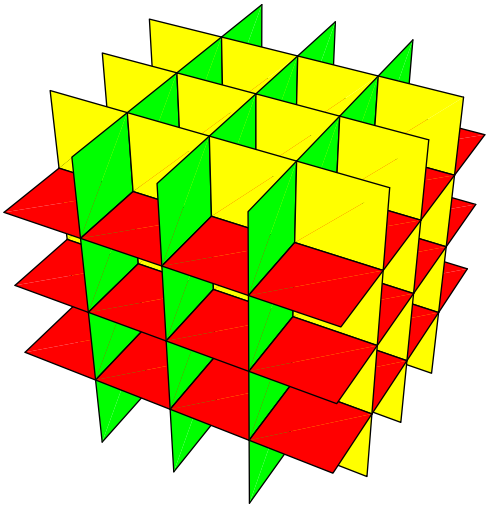


Figure 19.3: Three families of planes

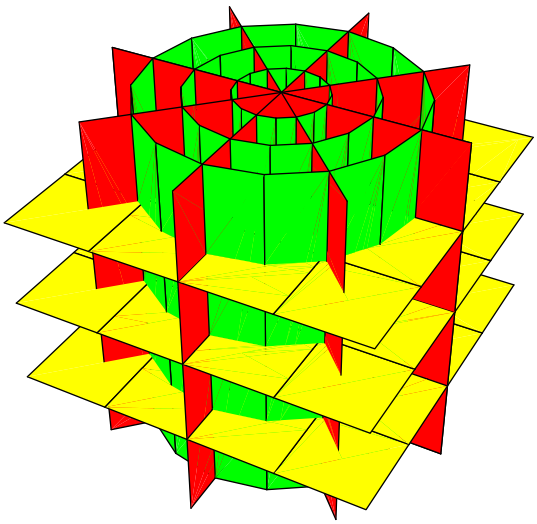


Figure 19.4: Two families of planes and one of cylinders

Many classical differential geometry books have a section on triply orthogonal systems, see for example [Bian], [Eisen1], and [Wea]. An extensive list of triply orthogonal systems is given in [MoSp]. We shall study more interesting examples in Sections 19.6 and 19.7.

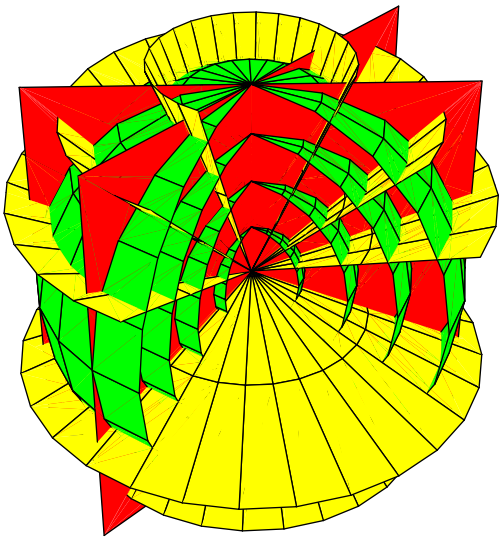


Figure 19.5: One family of planes, one of cones and one of spheres

A curve in $\mathbb{R}^3$ has one parameter, a surface two. It is also useful to consider a function of *three* parameters in order to generate a triply orthogonal system.

Definition 19.15. *A* ***curvilinear patch*** *for* $\mathbb{R}^3$ *is merely a differentiable function* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$*, where* $\mathcal{U}$ *is some open set of* $\mathbb{R}^3$*.*

If $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ is a curvilinear patch, we can write

$$\mathbf{x}(u, v, w) = \big(x_1(u, v, w),\ x_2(u, v, w),\ x_3(u, v, w)\big).$$

Differentiation of a curvilinear patch is defined in the same way as differentiation of a patch in $\mathbb{R}^3$. Thus

$$\mathbf{x}_u = \left(\frac{\partial x_1}{\partial u},\ \frac{\partial x_2}{\partial u},\ \frac{\partial x_3}{\partial u}\right),$$

and so forth.

Definition 19.16. *A curvilinear patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *is* ***regular*** *provided the vector fields* $\mathbf{x}_u, \mathbf{x}_v, \mathbf{x}_w$ *are linearly independent throughout* $\mathcal{U}$ *or (equivalently) the vector triple product* $[\mathbf{x}_u\ \mathbf{x}_v\ \mathbf{x}_w]$ *is nowhere zero. We shall say that* $\mathbf{x}$ *is* ***orientation preserving*** *if* $[\mathbf{x}_u\ \mathbf{x}_v\ \mathbf{x}_w] > 0$*, and* ***orientation reversing*** *if* $[\mathbf{x}_u\ \mathbf{x}_v\ \mathbf{x}_w] < 0$*.*

Note that if $\mathbf{x}$ is orientation reversing, then the curvilinear patch $\mathbf{y}$ defined by $\mathbf{y}(u, v, w) = \mathbf{x}(u, w, v)$ is orientation preserving.

Having introduced this terminology, a triply orthogonal system of surfaces can be considered to be a special kind of curvilinear patch.

Lemma 19.17. *A regular curvilinear patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *satisfying*

$$\mathbf{x}_v \cdot \mathbf{x}_w = \mathbf{x}_w \cdot \mathbf{x}_u = \mathbf{x}_u \cdot \mathbf{x}_v = 0 \tag{19.17}$$

determines a triply orthogonal system of surfaces, each of which is regular.

Proof. For fixed u, the mapping $(v, w) \mapsto \mathbf{x}(u, v, w)$ is a 2-dimensional patch. It is regular because $\mathbf{x}_v \times \mathbf{x}_w$ is everywhere nonzero. Each of its tangent spaces is spanned by $\mathbf{x}_v$ and $\mathbf{x}_w$. The analogous properties hold for the mappings $(w, u) \mapsto \mathbf{x}(u, v, w)$ and $(u, v) \mapsto \mathbf{x}(u, v, w)$. Then (19.17) implies that the tangent spaces to the traces of the three patches

$$(v, w) \mapsto \mathbf{x}(u, v, w), \quad (w, u) \mapsto \mathbf{x}(u, v, w), \quad (u, v) \mapsto \mathbf{x}(u, v, w) \tag{19.18}$$

are mutually perpendicular. Thus if we set

$$\begin{aligned} \mathfrak{A} &= \big\{ (v, w) \mapsto \mathbf{x}(u, v, w) \big\}, \\ \mathfrak{B} &= \big\{ (w, u) \mapsto \mathbf{x}(u, v, w) \big\}, \\ \mathfrak{C} &= \big\{ (u, v) \mapsto \mathbf{x}(u, v, w) \big\}, \end{aligned}$$

then $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ form a triply orthogonal system of surfaces.

That regularity of the curvilinear patch implies regularity of each of the surfaces in $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ results from the following identities, which are consequences of (7.1):

$$[\mathbf{x}_u\ \mathbf{x}_v\ \mathbf{x}_w]^2 = \|\mathbf{x}_u\|^2 \|\mathbf{x}_v\|^2 \|\mathbf{x}_w\|^2,$$

$$\|\mathbf{x}_v \times \mathbf{x}_w\| = \|\mathbf{x}_v\| \|\mathbf{x}_w\|, \quad \|\mathbf{x}_w \times \mathbf{x}_u\| = \|\mathbf{x}_w\| \|\mathbf{x}_u\|, \quad \|\mathbf{x}_u \times \mathbf{x}_v\| = \|\mathbf{x}_u\| \|\mathbf{x}_v\|. \qquad \blacksquare$$

Since we shall deal only with local questions concerning triply orthogonal systems of surfaces, we need only consider those systems determined by a curvilinear patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ satisfying (19.17). We shall not prohibit such a curvilinear patch from having isolated singularities and nonregular points.

Definition 19.18. *Let* $\mathcal{U}$ *be an open set in* $\mathbb{R}^3$. *A* ***triply orthogonal patch*** *is an orientation-preserving regular curvilinear patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *such that* (19.17) *holds. Given such a patch, set* $p = \|\mathbf{x}_u\|$, $q = \|\mathbf{x}_v\|$ *and* $r = \|\mathbf{x}_w\|$. *Then the* ***unit normals*** *determined by* $\mathbf{x}$ *are the vector fields*

$$\mathbf{A} = \frac{\mathbf{x}_u}{p}, \qquad \mathbf{B} = \frac{\mathbf{x}_v}{q}, \qquad \mathbf{C} = \frac{\mathbf{x}_w}{r}.$$

Using this definition, we can compute

$$\mathbf{x}_v \times \mathbf{x}_w = q\mathbf{B} \times r\,\mathbf{C} = qr\,\mathbf{A} = \frac{qr}{p}\mathbf{x}_u. \tag{19.19}$$

This enables us to state

Lemma 19.19. *A triply orthogonal patch* $\mathbf{x}$ *satisfies the following relations*:

$$\mathbf{x}_v \times \mathbf{x}_w = \frac{qr}{p}\mathbf{x}_u, \quad \mathbf{x}_w \times \mathbf{x}_u = \frac{rp}{q}\mathbf{x}_v, \quad \mathbf{x}_u \times \mathbf{x}_v = \frac{pq}{r}\mathbf{x}_w,$$

$$[\mathbf{x}_u\ \mathbf{x}_v\ \mathbf{x}_w] = pqr,$$

$$\mathbf{x}_u \cdot \mathbf{x}_{vw} = 0, \qquad \mathbf{x}_v \cdot \mathbf{x}_{wu} = 0, \qquad \mathbf{x}_w \cdot \mathbf{x}_{uv} = 0.$$

Proof. The triple product formula also follows from (19.19). For the dot products, we first differentiate (19.17), so as to obtain

$$\begin{cases} \mathbf{x}_{vu} \cdot \mathbf{x}_w + \mathbf{x}_v \cdot \mathbf{x}_{wu} &= 0, \\ \mathbf{x}_{wv} \cdot \mathbf{x}_u + \mathbf{x}_w \cdot \mathbf{x}_{uv} &= 0, \\ \mathbf{x}_{uw} \cdot \mathbf{x}_v + \mathbf{x}_u \cdot \mathbf{x}_{vw} &= 0. \end{cases} \tag{19.20}$$

The result follows by combining these equations with appropriate signs, exploiting symmetry of the mixed partial derivatives. ∎

Next, we show that the coefficients of the first and second fundamental forms of each of the surfaces determined by a triply orthogonal patch are especially simple.

Lemma 19.20. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a triply orthogonal patch. The coefficients* E, F, G *of the first fundamental form, the coefficients* e, f, g *of the second fundamental form, the principal curvatures* k_1, k_2*, the Gaussian curvature* K *and the mean curvature* H *of the patches* (19.18) *are given by the following table*:

Patch	E	F	G	e	f	g	k_1	k_2	K	H
$\mathbf{x}(v,w)$	q^2	0	r^2	$-\frac{qq_u}{p}$	0	$-\frac{rr_u}{p}$	$-\frac{q_u}{pq}$	$-\frac{r_u}{rp}$	$\frac{q_u r_u}{p^2 qr}$	$-\frac{(qr)_u}{2pqr}$
$\mathbf{x}(w,u)$	r^2	0	p^2	$-\frac{rr_v}{q}$	0	$-\frac{pp_v}{q}$	$-\frac{r_v}{qr}$	$-\frac{p_v}{pq}$	$\frac{r_v p_v}{pq^2 r}$	$-\frac{(rp)_v}{2pqr}$
$\mathbf{x}(u,v)$	p^2	0	q^2	$-\frac{pp_w}{r}$	0	$-\frac{qq_w}{r}$	$-\frac{p_w}{rp}$	$-\frac{q_w}{qr}$	$\frac{p_w q_w}{pqr^2}$	$-\frac{(pq)_w}{2pqr}$

Proof. That $F = 0$ for each of the surfaces is a consequence of the assumption that the surfaces are pairwise orthogonal. Furthermore, the formulas for E and G are just the definitions of p, q, r.

As for the second fundamental forms, we first note that (19.20) implies that $f = 0$ for each of the surfaces in (19.18). Next, we compute e for the surface $(v, w) \mapsto \mathbf{x}(u, v, w)$. Since the unit normal is $\mathbf{x}_u/\|\mathbf{x}_u\|$, we have

$$e = \mathbf{x}_{vv} \cdot \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|} = \frac{-\|\mathbf{x}_v\|_u^2}{2\|\mathbf{x}_u\|} = \frac{-q\,q_u}{p}.$$

The computation of the other coefficients is similar. ∎

Now we can give a simple proof of the following important theorem:

Theorem 19.21. (Dupin[6]) *The curves of intersection of the surfaces* (19.18) *of a triply orthogonal patch are principal curves on each.*

Proof. Lemma 19.20 implies that $F = f = 0$, the condition that appears in Lemma 15.5 on page 466. Thus, each of the patches given by (19.18) has the property that the parameter curves are principal curves. ∎

19.6 Elliptic Coordinates

In this section we show how to find triply orthogonal patches that give rise to principal patches on ellipsoids, hyperboloids of one sheet and hyperboloids of two sheets. Each of these surfaces can be rotated and translated so that it is described by a nonparametric equation of the form

$$\frac{x^2}{a} + \frac{y^2}{b} + \frac{z^2}{c} = 1 \tag{19.21}$$

for appropriate a, b, c. (In contrast to page 312, the constants are not squared, so that (19.21) can represent three different types of nondegenerate quadrics.) Fix a, b, c, and define

$$F[\lambda](x, y, z) = \frac{x^2}{a-\lambda} + \frac{y^2}{b-\lambda} + \frac{z^2}{c-\lambda} - 1. \tag{19.22}$$

The surfaces defined implicitly by the equations $F[\lambda] = 0$ are called the ***confocals*** of the surface defined by (19.21), which corresponds to $\lambda = 0$. (See [HC-V, 1.4].)



6 Baron Pierre-Charles-François Dupin (1784–1873). French mathematician, student of Monge. In addition to this theorem, Dupin's contributions to differential geometry include the theory of cyclides presented in Chapter 20, and the Dupin indicatrix that gives an indication of the local behavior of a surface in terms of its power series expansion. In 1830 Dupin was elected deputy for Tarn and continued in politics until 1870. He also wrote extensively on economics and industry.

For each fixed λ, the gradient of (19.22) is given by

$$\operatorname{grad} F[\lambda] = \Big(\frac{2x}{a-\lambda}, \frac{2y}{b-\lambda}, \frac{2z}{c-\lambda}\Big). \tag{19.23}$$

If we now choose to fix x, y, z, λ for which $F[\lambda](x, y, z) = 0$, then (19.23) is a normal vector to the confocal surface at (x, y, z) (see page 311). The link with triple orthogonal ststems is provided by

Lemma 19.22. *If λ, μ are distinct real numbers, then*

$$\frac{F[\lambda] - F[\mu]}{\lambda - \mu} = \tfrac{1}{4} \operatorname{grad} F[\lambda] \cdot \operatorname{grad} F[\mu].$$

Proof. It suffices to compute $F[\lambda] - F[\mu]$ and use (19.23). The formulation of the lemma is easy to remember, since if we vary λ and let it approach μ, the equation becomes the identity $F'[\lambda] = \frac{1}{4}\| \operatorname{grad} F[\lambda]\|^2$. ∎

In order to exploit Lemma 19.22, we fix a point $P = (x, y, z)$ and solve the equation

$$F[\lambda](x, y, z) = 0,$$

which becomes a cubic equation in λ by the time we have multiplied both sides by $(a-\lambda)(b-\lambda)(c-\lambda)$. Suppose that this equation has three real roots u, v, w all distinct from a, b, c, so that

$$\begin{aligned}(u-\lambda)(v-\lambda)(w-\lambda) &= (a-\lambda)(b-\lambda)(c-\lambda) \\ &\quad -x^2(b-\lambda)(c-\lambda) - y^2(a-\lambda)(c-\lambda) - z^2(a-\lambda)(b-\lambda).\end{aligned} \tag{19.24}$$

Not only does this mean that the confocal surfaces $F[u], F[v], F[w]$ all pass through P, but they are mutually perpendicular at P. For Lemma 19.22 tells us that their normal vectors are orthogonal. Carrying out this procedure at each point in $\mathbb{R}^3$ gives rise to elliptic coordinates, which we now describe.

To find the values of x, y, z, or rather x^2, y^2, z^2, such that (19.24) holds for all λ, we give λ the values a, b, c in succession. The result is:

$$\begin{cases} x^2 &= \dfrac{(a-u)(a-v)(a-w)}{(b-a)(c-a)}, \\ y^2 &= \dfrac{(b-u)(b-v)(b-w)}{(a-b)(c-b)}, \\ z^2 &= \dfrac{(c-u)(c-v)(c-w)}{(a-c)(b-c)}. \end{cases} \tag{19.25}$$

The solutions to equations (19.25) form eight patches.

Definition 19.23. *Let $a > b > c$. An* ***elliptic coordinate patch*** *is one of the eight curvilinear patches defined by*

$$\begin{aligned}(u,v,w) \quad \mapsto \quad & \left(\pm \sqrt{\frac{(a-u)(a-v)(a-w)}{(b-a)(c-a)}}, \right. \\ & \left. \pm\sqrt{\frac{(b-u)(b-v)(b-w)}{(c-b)(a-b)}},\ \pm\sqrt{\frac{(c-u)(c-v)(c-w)}{(a-c)(b-c)}} \right). \end{aligned} \tag{19.26}$$

The proof of the next result formalizes what we have already said, with $\mathbf{x}_u$ playing the role of $\mathsf{grad}\, F[\lambda]$.

Lemma 19.24. *For $a > b > c$, each mapping* (19.26) *is triply orthogonal. It is regular on the set*

$$\big\{ (p_1,p_2,p_3) \in \mathbb{R}^3 \mid p_1,p_2,p_3 \text{ are distinct, } p_1 \neq a,\ p_2 \neq b,\ p_3 \neq c \big\}.$$

Proof. If we indicate the mapping in question by $\mathbf{x}\colon (u,v,w) \mapsto (x,y,z)$, then it follows from (19.25) that

$$2x\,\frac{\partial x}{\partial u} = -\frac{(a-v)(a-w)}{(b-a)(c-a)} = \frac{x^2}{(u-a)}.$$

Similar formulas tell us that

$$\left\{ \begin{aligned} 2\mathbf{x}_u &= \left(\frac{x}{(u-a)}, \frac{y}{(u-b)}, \frac{z}{(u-c)} \right), \\ 2\mathbf{x}_v &= \left(\frac{x}{(v-a)}, \frac{y}{(v-b)}, \frac{z}{(v-c)} \right), \\ 2\mathbf{x}_w &= \left(\frac{x}{(w-a)}, \frac{y}{(w-b)}, \frac{z}{(w-c)} \right). \end{aligned} \right. \tag{19.27}$$

Then (19.25) and (19.27) imply that

$$\begin{aligned} 4\,\mathbf{x}_u \cdot \mathbf{x}_v &= \frac{x^2}{(u-a)(v-a)} + \frac{y^2}{(u-b)(v-b)} + \frac{z^2}{(u-c)(v-c)} \\ &= \frac{a-w}{(b-a)(c-a)} + \frac{b-w}{(c-b)(a-b)} + \frac{c-w}{(a-c)(b-c)} \\ &= \frac{-(a-w)(b-c) - (b-w)(c-a) - (c-w)(a-b)}{(a-b)(b-c)(c-a)} \\ &= 0. \end{aligned}$$

Similarly, $\mathbf{x}_w \cdot \mathbf{x}_u = 0 = \mathbf{x}_v \cdot \mathbf{x}_w$, and

$$p^2 = \|\mathbf{x}_u\|^2 = \frac{(u-v)(u-w)}{4(a-u)(b-u)(c-u)}, \tag{19.28}$$

$$q^2 = \|\mathbf{x}_v\|^2 = \frac{(v-w)(v-u)}{4(a-v)(b-v)(c-v)},$$

$$r^2 = \|\mathbf{x}_w\|^2 = \frac{(w-u)(w-v)}{4(a-w)(b-w)(c-w)}.$$

Hence the Jacobian matrix of $\mathbf{x}$ has rank 3 on the set stated. ▮

If we fix w, then (19.26) determines a patch $(u, v) \mapsto (x, y, z)$ as explained on page 607. In these circumstances, u and v are called ***elliptic coordinates***, and we next determine their possible ranges and associated surfaces.

Lemma 19.25. *Let $a > b > c > 0$.*

(i) *Suppose $c > w$ so that*

$$\frac{x^2}{a-w} + \frac{y^2}{b-w} + \frac{z^2}{c-w} = 1 \tag{19.29}$$

is an ellipsoid. In order that x, y, z in (19.25) *be real, either*

$$a \geqslant v \geqslant b \geqslant u \geqslant c \qquad \text{or} \qquad a \geqslant u \geqslant b \geqslant v \geqslant c.$$

(ii) *Suppose $b > w > c$, so that*

$$\frac{x^2}{a-w} + \frac{y^2}{b-w} + \frac{z^2}{c-w} = 1 \tag{19.30}$$

is a hyperboloid of one sheet. In order that x, y, z be real, either

$$a \geqslant v \geqslant b > c \geqslant u \qquad \text{or} \qquad a \geqslant u \geqslant b > c \geqslant v.$$

(iii) *Suppose $a > w > b$, so that*

$$\frac{x^2}{a-w} + \frac{y^2}{b-w} + \frac{z^2}{c-w} = 1 \tag{19.31}$$

is a hyperboloid of two sheets. In order that x, y, z be real, either

$$a > b \geqslant v \geqslant c \geqslant u \qquad \text{or} \qquad a > b \geqslant u \geqslant c \geqslant v.$$

Proof. If $w < c < b < a$, it follows from (19.25) that

$$\begin{cases} (a-u)(a-v) > 0, \\ (b-u)(b-v) < 0, \\ (c-u)(c-v) > 0. \end{cases} \tag{19.32}$$

Without loss of generality, $v \geqslant u$. Then (19.32) implies (i). The proofs of (ii) and (iii) are similar. ▮

It is complicated to generate an entire ellipsoid using elliptical coordinates, because each of the eight patches must be plotted separately as an octant of the ellipsoid. In Figure 19.6, the lower four octants have been translated down by -2 so that we can peek inside the egg. The new coordinates make it easy to spot the umbilics, and we can determine them analytically using methods from the previous section.

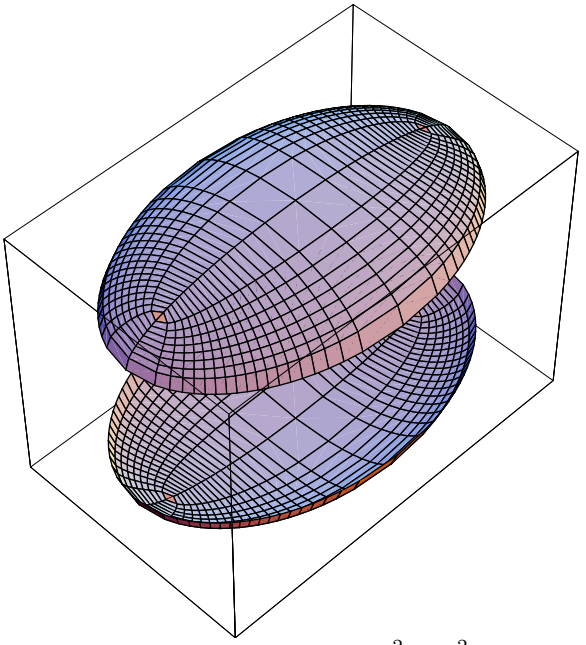


Figure 19.6: Principal curves on $\dfrac{x^2}{12} + \dfrac{y^2}{5} + z^2 = 1$

Theorem 19.26. *Suppose $a > b > c > 0$.*

(i) *If $c > w$, then the umbilic points of the ellipsoid* (19.29) *are the four points*

$$\left(\pm\sqrt{\frac{(a-b)(a-w)}{(a-c)}},\ 0,\ \pm\sqrt{\frac{(b-c)(c-w)}{(a-c)}} \right).$$

(ii) *If $b > w > c$, then the hyperboloid of one sheet* (19.30) *has no umbilic points.*

(iii) *If $a > w > b$, then the umbilic points of the hyperboloid of two sheets* (19.31) *are the four points*

$$\left(\pm\sqrt{\frac{(a-c)(a-w)}{(a-b)}},\ \pm\sqrt{\frac{(b-c)(b-w)}{(a-b)}},\ 0 \right).$$

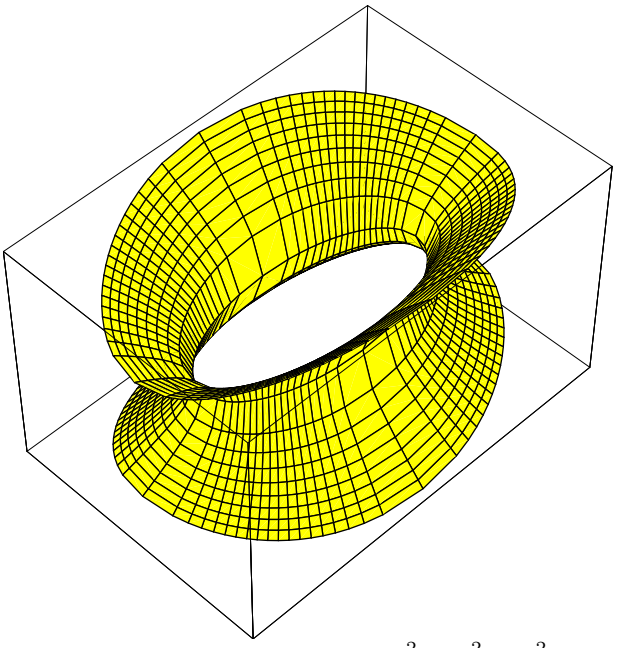


Figure 19.7: Principal curves on $\frac{x^2}{9} + \frac{y^2}{2} - \frac{z^2}{2} = 1$

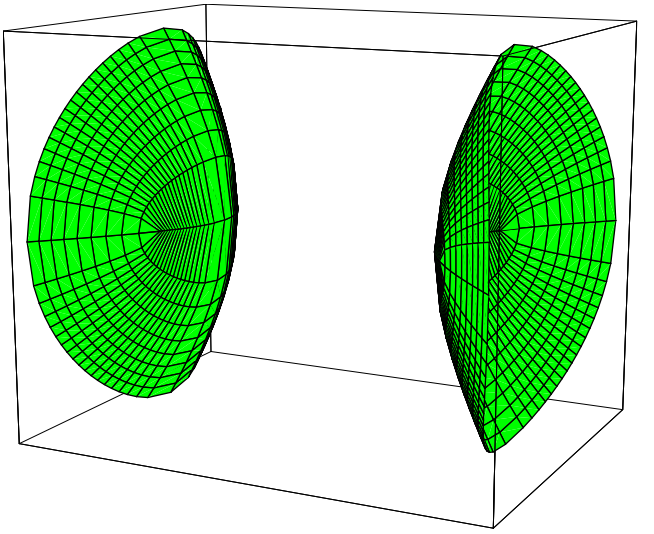


Figure 19.8: Principal curves on $\frac{x^2}{4} - \frac{y^2}{3} - \frac{z^2}{7} = 1$

Proof. From Lemma 19.20 and (19.28), it follows that the principal curvatures of the patch $(u, v) \mapsto (x, y, z)$ (defined by (19.26) with w fixed) are given by

$$k_1 = -\frac{1}{2r}\frac{\partial}{\partial w}\log(p^2) = \frac{1}{2(u-w)r},$$
$$k_2 = -\frac{1}{2r}\frac{\partial}{\partial w}\log(q^2) = \frac{1}{2(v-w)r}.$$

Then $k_1 = k_2$ implies that $u = v$. In case (i) we have $u = v = b$, and in case (iii) we have $u = v = c$, but in case (ii) no umbilic points are possible. Hence the umbilic points are as stated. ∎

Figure 19.8 shows the umbilics on the two-sheeted hyperboloid. Figure 19.9 illustrates how an ellipsoid, a hyperboloid of one sheet and a hyperboloid of two sheets intersect orthogonally in principal curves. Both these figures are computer-generated versions of sketches in [HC-V, chapter 4].

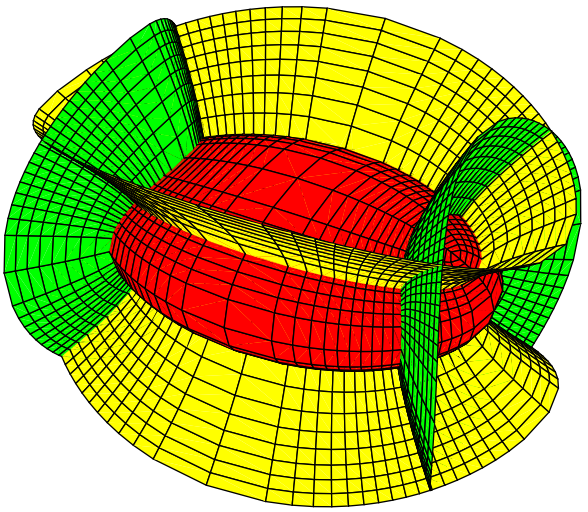


Figure 19.9: Triply orthogonal system formed by ellipsoids and hyperboloids

19.7 Parabolic Coordinates and a General Construction

Parabolic coordinates are defined in a similar way to elliptic coordinates, in order to furnish principal patches on elliptic paraboloids and hyperbolic paraboloids. We proceed as we did in Section 19.6, starting from the equation

$$\frac{x^2}{a} + \frac{y^2}{b} = 2z - c \tag{19.33}$$

of a general paraboloid.

This time, the confocals are defined as the zero sets of

$$G[\lambda] = \frac{x^2}{a-\lambda} + \frac{y^2}{b-\lambda} - 2z + c + \lambda. \tag{19.34}$$

This equation is chosen so that the analog of Lemma 19.22 remains valid, and it is also the case that the parabolas

$$\frac{y^2}{b-\lambda} - 2z + c + \lambda = 0$$

in the yz-plane have a common focus, namely the point $(y, z) = (0,\ (b+c)/2)$.

The confocal surfaces through a given point $(x, y, z) \in \mathbb{R}^3$ are given by the cubic equation

$$(a-\lambda)(b-\lambda)G[\lambda] = 0,$$

and the roots u, v, w satisfy

$$\begin{aligned}(u-\lambda)(v-\lambda)(w-\lambda)\\ = (a-\lambda)(b-\lambda)(2z-c-\lambda) - x^2(b-\lambda) - y^2(a-\lambda).\end{aligned} \tag{19.35}$$

In order that (19.35) hold for all λ, we must have

$$\begin{cases} x^2 &= \dfrac{(u-a)(v-a)(w-a)}{a-b}, \\ y^2 &= \dfrac{(u-b)(v-b)(w-b)}{b-a}, \\ z &= \dfrac{c-a-b+u+v+w}{2}; \end{cases} \tag{19.36}$$

the last equality was obtained by comparing coefficients of λ^2.

Expressing x, y, z in terms of u and v gives

Definition 19.27. *Suppose $a > b$. A* ***parabolic coordinate patch*** *is one of the four curvilinear patches defined by*

$$\begin{aligned}(u, v, w) \quad \mapsto \quad & \left(\pm \sqrt{\frac{(a-u)(a-v)(a-w)}{b-a}}, \right. \\ & \left. \pm\sqrt{\frac{(b-u)(b-v)(b-w)}{a-b}},\ \frac{c-a-b+u+v+w}{2} \right).\end{aligned} \tag{19.37}$$

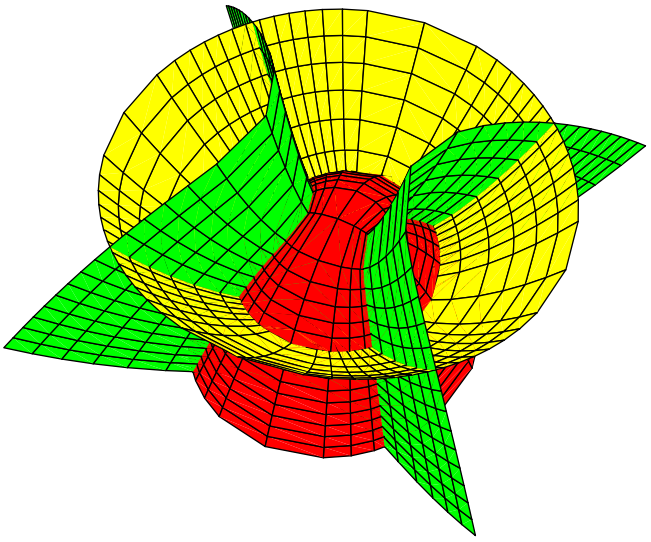


Figure 19.10: Triply orthogonal system formed by elliptic and hyperbolic paraboloids

The analog of Lemma 19.24 is illustrated in Figure 19.10, and the individual surfaces are plotted overleaf. With a different choice of parameters, it is possible to construct a triply orthogonal system involving two distinct families of hyperbolic paraboloids.

Lemma 19.28. *For $a > b$, the curvilinear patch* (19.37) *is triply orthogonal. It is regular on the set*

$$\big\{\,(p_1,p_2,p_3)\in\mathbb{R}^3 \;\big|\; p_1,p_2,p_3 \text{ are distinct, } p_1\neq a,\, p_2\neq b\,\big\}.$$

Proof. We skip some of the details, since the method is identical to that of (19.27). It follows from (19.36) that $\mathbf{x}_u\cdot\mathbf{x}_v=\mathbf{x}_w\cdot\mathbf{x}_u=\mathbf{x}_v\cdot\mathbf{x}_w=0$, and that

$$p^2=\|\mathbf{x}_u\|^2 \;=\; \frac{(u-v)(u-w)}{4(a-u)(b-u)},$$

$$q^2=\|\mathbf{x}_v\|^2 \;=\; \frac{(v-w)(v-u)}{4(a-v)(b-v)},$$

$$r^2=\|\mathbf{x}_w\|^2 \;=\; \frac{(w-u)(w-v)}{4(a-w)(b-w)}.$$

Hence the Jacobian matrix of $\mathbf{x}$ has rank 3 on the stated set. ∎

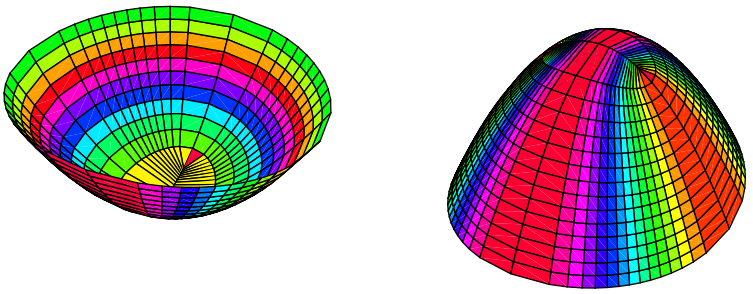


Figure 19.11: Principal curves on $\dfrac{x^2}{2} + \dfrac{y^2}{3} = -2z + 7$ and $\dfrac{x^2}{5} + \dfrac{y^2}{4} = 2z$

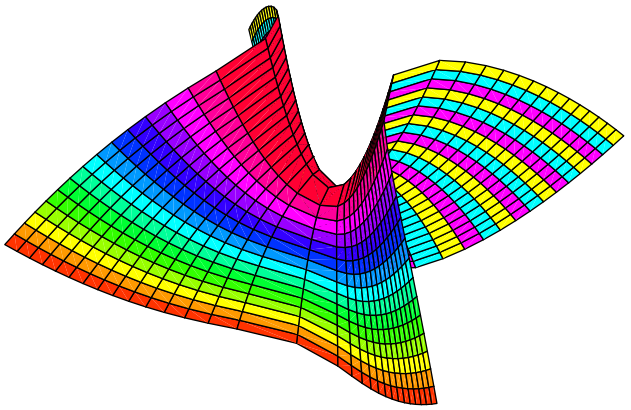


Figure 19.12: Principal curves on $x^2 - y^2 = 4z - 3$

We show that any principal patch in $\mathbb{R}^3$ is part of a triply orthogonal system of surfaces. We first prove

Theorem 19.29. *A necessary and sufficient condition that a curve on a surface be a principal curve is that the surface normals along the curve form a flat ruled surface.*

Proof. Let $\boldsymbol{\beta}\colon (a, b) \to \mathbb{R}^3$ be a unit-speed curve on a surface $\mathcal{M}$. Without loss of generality, we can assume that $\mathcal{M}$ is oriented with unit normal $\mathbf{U}$. Let $\mathbf{T}$ denote the unit-tangent vector of $\boldsymbol{\beta}$, and let $\mathcal{N}$ be the surface formed by the normals along $\boldsymbol{\beta}$. We parametrize $\mathcal{N}$ as

$$\mathbf{y}(s, v) = \boldsymbol{\beta}(s) + v\,\mathbf{U}(s),$$

where s is arc length along $\boldsymbol{\beta}$. Then

$$\mathbf{y}_s = \mathbf{T} + v\,\mathbf{U}_s, \qquad \mathbf{y}_v = \mathbf{U}, \qquad \mathbf{y}_{sv} = \mathbf{U}_s, \qquad \mathbf{y}_{vv} = 0. \tag{19.38}$$

According to Corollary 13.32 on page 405, and the last equation of (19.38), the Gaussian curvature of $\mathcal{N}$ is given by

$$K = \frac{[\mathbf{y}_{ss}\,\mathbf{y}_s\,\mathbf{y}_v][\mathbf{y}_{vv}\,\mathbf{y}_s\,\mathbf{y}_v] - [\mathbf{y}_{sv}\,\mathbf{y}_s\,\mathbf{y}_v]^2}{\big(\|\mathbf{y}_s\|^2\|\mathbf{y}_v\|^2 - (\mathbf{y}_s\cdot\mathbf{y}_v)^2\big)^2} = \frac{-[\mathbf{y}_{sv}\,\mathbf{y}_s\,\mathbf{y}_v]^2}{\big(\|\mathbf{y}_s\|^2\|\mathbf{y}_v\|^2 - (\mathbf{y}_s\cdot\mathbf{y}_v)^2\big)^2}.$$

Thus $K = 0$ if and only if the vector triple product $[\mathbf{y}_{sv}\,\mathbf{y}_s\,\mathbf{y}_v]$ is zero. From (19.38), we have

$$[\mathbf{y}_{sv}\,\mathbf{y}_s\,\mathbf{y}_v] = \big(\mathbf{U}_s \times (\mathbf{T} + v\,\mathbf{U}_s)\big)\cdot\mathbf{U} = \big(\mathbf{U}_s\times\mathbf{T}\big)\cdot\mathbf{U}.$$

Since both $\mathbf{T}$ and $\mathbf{U}_s$ are perpendicular to $\mathbf{U}$, this implies that $K = 0$ if and only if

$$0 = \mathbf{U}_s\times\mathbf{T} = -S(\mathbf{T})\times\mathbf{T}, \tag{19.39}$$

where S denotes the shape operator of $\mathcal{M}$. It follows from Lemma 15.5, page 466, that (19.39) is precisely the condition that $\boldsymbol{\beta}$ be a principal curve of $\mathcal{M}$. ∎

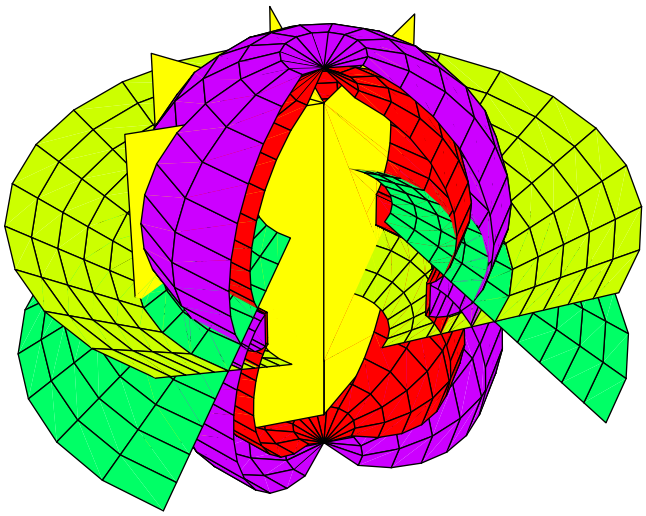


Figure 19.13: A triply orthogonal system constructed from surfaces parallel to a revolved nephroid, planes and cones

We are now able to construct the triply orthogonal system containing a given principal patch, such as that shown in Figure 19.13.

Theorem 19.30. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a principal patch. Define a curvilinear patch* $\mathbf{y}\colon \mathcal{U} \times \mathbb{R} \to \mathbb{R}^3$ *by*

$$\mathbf{y}(u, v, w) = \mathbf{x}(u, v) + w\,\mathbf{U}, \tag{19.40}$$

where $\mathbf{U}$ *is a unit normal to* $\mathbf{x}$*. Put*

$$\begin{aligned} \mathfrak{A} &= \big\{\, (u, v) \mapsto \mathbf{y}(u, v, w) \,\big\}, \\ \mathfrak{B} &= \big\{\, (w, u) \mapsto \mathbf{y}(u, v, w) \,\big\}, \\ \mathfrak{C} &= \big\{\, (v, w) \mapsto \mathbf{y}(u, v, w) \,\big\}; \end{aligned}$$

then $\mathfrak{A}$*,* $\mathfrak{B}$*,* $\mathfrak{C}$ *form a triply orthogonal system of surfaces. Each surface in* $\mathfrak{B}$ *and* $\mathfrak{C}$ *is a flat ruled surface.*

Proof. Let k_1 and k_2 be the principal curvatures corresponding to $\mathbf{x}_u$ and $\mathbf{x}_v$. Then we have

$$\begin{cases} \mathbf{y}_u &= \mathbf{x}_u + w\,\mathbf{U}_u = (1 - w\,k_1)\mathbf{x}_u, \\ \mathbf{y}_v &= \mathbf{x}_v + w\,\mathbf{U}_v = (1 - w\,k_2)\mathbf{x}_v, \\ \mathbf{y}_w &= \mathbf{U}. \end{cases}$$

Hence $\mathbf{y}_u, \mathbf{y}_v, \mathbf{y}_w$ are mutually orthogonal, so that $\mathbf{y}$ is a triply orthogonal patch. Now $\mathbf{y}$ determines a triply orthogonal system of surfaces by Lemma 19.17 on page 607. The last statement is a consequence of Theorem 19.29. ∎

19.8 Parallel Surfaces

In Section 4.5 we constructed a parallel curve to a given plane curve. There is a similar definition of parallel surface that was implicit in the preceding section.

Definition 19.31. *Let* $\mathcal{M} \subset \mathbb{R}^3$ *be a regular surface. The* ***surface parallel to*** $\mathcal{M}$ ***at a distance*** $t > 0$ *is the set*

$$\mathcal{M}(t) = \big\{\, \mathbf{q} \in \mathbb{R}^3 \;\big|\; \mathsf{distance}(\mathbf{q}, \mathcal{M}) = t \,\big\}.$$

Figure 11.10 on page 341 shows that a parallel surface to a nonorientable surface can itself be orientable. But if $\mathcal{M}$ is connected and orientable, then $\mathcal{M}(t)$ will consist of two components; see, for example, [Gray].

If t is large, $\mathcal{M}(t)$ may not be a regular surface, as can be seen in Figure 19.14 that extends Figure 4.13 on page 112. However, for t close to 0, we shall prove that $\mathcal{M}(t)$ *is* a regular surface. Figure 19.15 shows parallel surfaces to a catenoid for small t. The resulting components (for $t \neq 0$) are not themselves catenoids, since the "stem" can shrink to nothing, nor are they minimal surfaces. In the

next section we shall undertake a quantitive study of the curvature of a parallel surface.

First, let us define the notion of parallel patch. The idea is to move the patch a distance t along its normal $\mathbf{U}$. We want to be able to go in either direction along the normal, so we now allow t to be either positive or negative.

Definition 19.32. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. Then the* ***patch parallel to $\mathbf{x}$ at distance*** *t is the patch given by*

$$\mathbf{x}[t](u,v) = \mathbf{x}(u,v) + t\,\mathbf{U}(u,v), \tag{19.41}$$

where $\mathbf{U} = \mathbf{x}_u \times \mathbf{x}_v / \|\mathbf{x}_u \times \mathbf{x}_v\|$.

The similarity with (19.40) allows us to assert that each surface in the family $\mathfrak{A}$ in Theorem 19.30 is parallel to $\mathbf{x}(\mathcal{U})$. We next give conditions for the patch parallel to $\mathbf{x}$ at distance t to be regular.

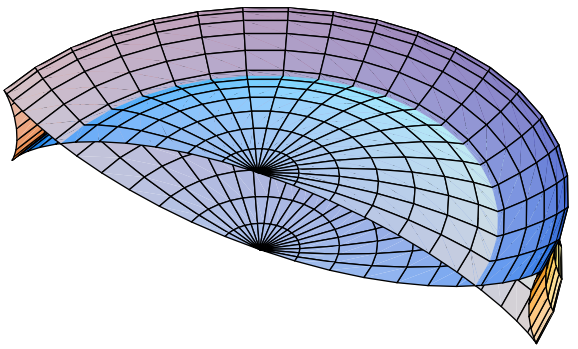


Figure 19.14: Parallel surface to the ellipsoid $\dfrac{4x^2}{9} + \dfrac{4y^2}{9} + z^2 = 1$

Lemma 19.33. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface for which there exists a single regular patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$. Assume that there is a number $t_0 > 0$ such that*

$$\det(I - tS) > 0 \qquad \text{for } |t| < t_0 \text{ on } \mathcal{M}, \tag{19.42}$$

where I denotes the identity transformation and S the shape operator. Then $\mathbf{x}[t]$ is a regular patch on $\mathcal{M}(t)$ for $|t| < t_0$.

Proof. We have

$$\mathbf{x}[t]_u = \mathbf{x}_u + t\,\mathbf{U}_u = (I - tS)\mathbf{x}_u; \tag{19.43}$$

similarly for $\mathbf{x}[t]_v$. Therefore,

$$\mathbf{x}[t]_u \times \mathbf{x}[t]_v = (I - tS)\mathbf{x}_u \times (I - tS)\mathbf{x}_v = \det(I - tS)\mathbf{x}_u \times \mathbf{x}_v. \tag{19.44}$$

Hence $\mathbf{x}[t]_u \times \mathbf{x}[t]_v$ is nonzero, and therefore $\mathbf{x}[t]$ is regular. Thus $\mathcal{M}(t)$ is a regular surface, since it is entirely covered by the regular patch $\mathbf{x}[t]$. ∎

Next, we consider parallel surfaces to arbitrary regular surfaces.

Lemma 19.34. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface and let $\mathcal{M}(t)$ denote the surface parallel to $\mathcal{M}$ at a distance t. Assume that there exists $t_0 > 0$ such that* (19.42) *holds. Then the parallel surface $\mathcal{M}(t)$ is a regular surface for $|t| < t_0$.*

Proof. Suppose that $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a regular injective patch on $\mathcal{M}$, and set $\mathcal{V} = \mathcal{U} \times \{\, t \mid |t| < t_0 \,\}$. Use (19.41) to define $\mathbf{y}\colon \mathcal{V} \to \mathbb{R}^3$ by merely writing $\mathbf{y}(u, v, t) = \mathbf{x}(u, v)[t]$. The resulting Jacobian matrix

$$\mathscr{J}(\mathbf{y}) = \Big((I - tS)\mathbf{x}_u,\ (I - tS)\mathbf{x}_v,\ \mathbf{U} \Big),$$

computed using (19.43), is obviously nonzero on $\mathcal{V}$. Hence $\mathbf{y}$ is injective on a neighborhood of each point of $\mathcal{V}$; in particular, $\mathbf{x}[t]$ is a regular injective patch on a neighborhood of each point of $\mathcal{U}$. Patches of the form $\mathbf{x}[t]$ cover $\mathcal{M}(t)$, so $\mathcal{M}(t)$ is a regular surface. ∎

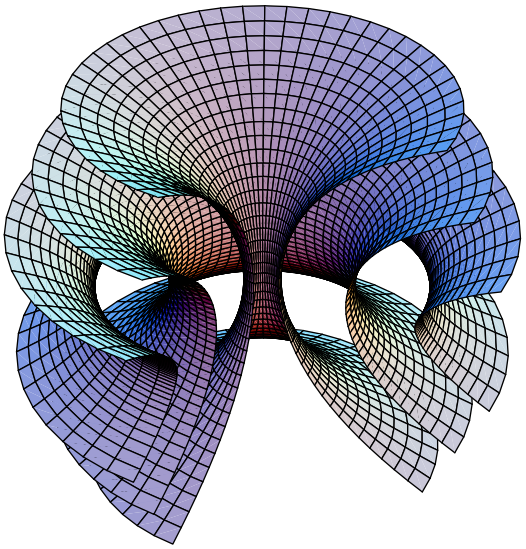


Figure 19.15: Parallel surfaces to a catenoid (in the middle)

19.9 The Shape Operator of a Parallel Surface

First, we show that the shape operator of a regular surface $\mathcal{M} \subseteq \mathbb{R}^3$ determines the shape operator of each of its parallel surfaces.

Lemma 19.35. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface and let $\mathcal{M}(t)$ denote the surface parallel to $\mathcal{M}$ at a distance t. Assume that condition* (19.42) *holds. Let S be the shape operator of $\mathcal{M}$ and $S(t)$ the shape operator of $\mathcal{M}(t)$. Then*

(i) *the matrix-valued function* $t \mapsto S(t)$ *satisfies* $S'(t) = S(t)^2$;

(ii) $S(t) = S(I - tS)^{-1}$;

(iii) *the principal curvatures of* $\mathcal{M}(t)$ *are given by*

$$k_i(t) = \frac{k_i}{1 - t\,k_i}, \tag{19.45}$$

for $i = 1, 2$, *where* k_1 *and* k_2 *are the principal curvatures of* $\mathcal{M}$.

(iv) *The Gaussian and mean curvatures of* $\mathcal{M}(t)$ *are given by*

$$K(t) = \frac{K}{1 - 2tH + t^2K} \quad \textit{and} \quad H(t) = \frac{H - tK}{1 - 2tH + t^2K}, \tag{19.46}$$

where K *and* H *denote the Gaussian and mean curvatures of* $\mathcal{M}$.

Proof. We may assume, without loss of generality, that $\mathcal{M}$ is parametrized by a single regular patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$. Then the patch $\mathbf{x}[t]$ on $\mathcal{M}(t)$ defined by (19.41) is regular, with unit normal

$$\mathbf{U}[t] = \frac{\mathbf{x}[t]_u \times \mathbf{x}[t]_v}{\big\|\mathbf{x}[t]_u \times \mathbf{x}[t]_v\big\|}.$$

But, given that $\det(I - tS) > 0$, (19.44) implies that $\mathbf{U}[t]$ coincides with $\mathbf{U}$. Therefore,

$$S(t)\mathbf{x}[t]_u = -\mathbf{U}[t]_u = -\mathbf{U}_u = S\mathbf{x}_u, \tag{19.47}$$

proving that $S(t)\mathbf{x}[t]_u$ does not depend on t. Hence

$$\big(S(t)\mathbf{x}[t]_u\big)' = 0, \tag{19.48}$$

where the prime denotes differentiation with respect to t. Also, from (19.41), we obtain

$$\mathbf{x}[t]'_u = \mathbf{U}_u = -S\mathbf{x}_u. \tag{19.49}$$

Then (19.47)–(19.49) imply that

$$S'(t)\mathbf{x}[t]_u = -S(t)\mathbf{x}[t]'_u = S(t)S\mathbf{x}_u = S(t)^2\mathbf{x}[t]_u.$$

Similarly, $S'(t)\mathbf{x}[t]_v = S(t)^2\mathbf{x}[t]_v$. Equation (19.44) implies that $\mathbf{x}[t]_u$ and $\mathbf{x}[t]_v$ are linearly independent. Hence $S'(t) = S(t)^2$, proving (i).

To prove (ii), we define $\widetilde{S}(t) = S(I - tS)^{-1}$, and use the binomial theorem to expand it as a convergent geometrical series for small t:

$$\widetilde{S}(t) = \sum_{k=0}^{\infty} t^k S^{k+1}. \tag{19.50}$$

It is clear from (19.50) that S commutes with $\widetilde{S}(t)$ for all t, and so

$$\widetilde{S}'(t) = S^2(I - tS)^{-2} = \widetilde{S}(t)^2.$$

Since $\widetilde{S}(0) = S$, we conclude that $\widetilde{S}$ coincides with S.

For (iii) we compute the eigenvalues of $S(I - tS)^{-1}$ and use (ii). To prove (iv), we use (19.45). In particular,

$$K(t) = k_1(t)k_2(t) = \frac{k_1 k_2}{(1 - t\,k_1)(1 - t\,k_2)} = \frac{K}{1 - 2tH + t^2K}$$

and calculation of $H(t)$ is similar. ∎

Now we can prove an important result due to Bonnet. As Chern has remarked (see [Chern3]), this theorem tells us that finding surfaces of constant mean curvature is equivalent to finding surfaces of constant positive Gaussian curvature.

Theorem 19.36. (Bonnet) *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface.*

(i) *If $\mathcal{M}$ has constant mean curvature $H \equiv 1/(2c)$, then the parallel surface $\mathcal{M}(c)$ has constant Gaussian curvature $1/c^2$, and the parallel surface $\mathcal{M}(2c)$ has constant mean curvature $-1/(2c)$.*

(ii) *If $\mathcal{M}$ has constant Gaussian curvature $K \equiv 1/c^2$, then the parallel surfaces $\mathcal{M}(\pm c)$ have constant mean curvature $\mp 1/(2c)$.*

Proof. If $H \equiv 1/(2c)$, then it follows from (19.46) that

$$K(c) = \frac{K}{1 - 2c/(2c) + K\,c^2} = \frac{1}{c^2}.$$

The rest of (i) and (ii) follow from (19.46) in a similar fashion (see Exercise 15). ∎

Finally, we prove a partial converse to Theorem 19.36.

Theorem 19.37. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface with constant positive Gaussian curvature $1/a^2$, where $a > 0$. Let $\mathcal{M}(t)$ denote the surface parallel to $\mathcal{M}$ at a distance t. Suppose that the umbilic points of $\mathcal{M}$ are isolated. If $\mathcal{M}(t)$ has constant mean curvature, then $t = \pm a$.*

Proof. Fix t, and suppose that $H(t)$ is constant on $\mathcal{M}(t)$. The second equation of (19.46) implies that

$$k_1(1 - t\,k_2) + k_2(1 - t\,k_1) = 2H(t)(1 - t\,k_1)(1 - t\,k_2),$$

or

$$k_1 + k_2 - \frac{2t}{a^2} = 2H(t)\Big(1 - t(k_1 + k_2) + \frac{t^2}{a^2}\Big).$$

Hence

$$(k_1 + k_2)\big(1 + 2tH(t)\big) = 2H(t) + 2H(t)\frac{t^2}{a^2} + \frac{2t}{a^2}. \tag{19.51}$$

By hypothesis, the right-hand of (19.51) is constant. But if the left-hand of (19.51) is constant, Lemma 19.12 forces it to vanish at the nonumbilic points of $\mathcal{M}$. Hence $1 + 2tH(t) = 0$ everywhere, and (19.51) now implies that

$$0 \;=\; 2H(t)\left(1 + \frac{t^2}{a^2}\right) + \frac{2t}{a^2} \;=\; -\frac{1}{t}\left(1 + \frac{t^2}{a^2}\right) + \frac{2t}{a^2} = -\frac{1}{t} + \frac{t}{a^2}.$$

Therefore, $t = \pm a$. ∎

We shall investigate surfaces with constant nonzero Gaussian curvature in Chapter 21.

19.10 Exercises

1. Prove the second equation of (19.8).

2. Verify that if $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ is a regular patch, then

$$\frac{\partial \Gamma^1_{11}}{\partial v} - \frac{\partial \Gamma^1_{12}}{\partial u} - \Gamma^2_{12}\Gamma^1_{12} + \Gamma^2_{11}\Gamma^1_{22} = -FK,$$

as stated in the proof of Theorem 19.8.

3. Prove Corollary 19.10.

4. The Peterson–Mainardi–Codazzi equations also simplify for an asymptotic patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ (see page 561). Show that

$$\frac{\partial \log f}{\partial u} = \Gamma^1_{11} - \Gamma^2_{12}, \qquad \frac{\partial \log f}{\partial v} = \Gamma^2_{22} - \Gamma^1_{12}.$$

5. Find a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ for which the coefficients of the first and second fundamental forms are

$$E = a^2\cos^2 v, \qquad F = 0, \qquad G = a^2,$$
$$e = -a\cos^2 v, \qquad f = 0, \qquad g = -a.$$

6. Find a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ for which the coefficients of the first and second fundamental forms are

$$E = a^2, \qquad F = 0, \qquad G = 1,$$
$$e = -a, \qquad f = 0, \qquad g = 0.$$

7. Find the most general patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ for which $\mathbf{x}_{uv} = 0$.

8. Show that there is no patch having $du^2 + dv^2$ and $du^2 - dv^2$ as its first and second fundamental forms.

9. Show that there is no patch having $du^2 + \cos^2 u\, dv^2$ and $\cos^2 u\, du^2 + dv^2$ as its first and second fundamental forms.

10. Let $\mathbf{p}$ be a point of an open set $\mathcal{U}$ in $\mathbb{R}^2$ at which a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ satisfies

$$f\,E = e\,F, \qquad g\,E = e\,G \qquad \text{and} \qquad g\,F = f\,G.$$

Show that there exists a real number k such that at $\mathbf{p}$ we have

$$e = k\,E \qquad f = k\,F \qquad \text{and} \qquad g = k\,G.$$

M **11.** Show that the surface of revolution generated by a parallel curve $\boldsymbol{\beta}$ to a curve $\boldsymbol{\alpha}$ is the same as a parallel surface $\mathbf{y}$ to the surface of revolution $\mathbf{x}$ generated by $\boldsymbol{\alpha}$.

12. The notion of a system of orthogonal curves in $\mathbb{R}^2$ corresponds to the notion of a system of triply orthogonal surfaces in $\mathbb{R}^3$:

Definition 19.38. *An* ***orthogonal system of curves*** *on an open set $\mathcal{U}$ of $\mathbb{R}^2$ consists of two families $\mathfrak{A}$, $\mathfrak{B}$, of curves such that*

(i) *each point of $\mathcal{U}$ lies on one and only one member of each family;*

(ii) *each curve in each family meets every member of the other family orthogonally.*

Let $\mathcal{U} \subseteq \mathbb{R}^2$ be an open subset and $u = u(x, y)$, $v = v(x, y)$ two functions defined on $\mathcal{U}$ so that $(x, y) \mapsto (u, v)$ is a regular patch $\mathcal{U} \to \mathbb{R}^2$. Suppose that the partial derivatives of u, v satisfy

$$u_x = v_y, \qquad u_y = -v_x. \tag{19.52}$$

Show that the images of the families $\mathfrak{H}$ and $\mathfrak{V}$ of horizontal and vertical lines in $\mathbb{R}^2$ constitute a system of orthogonal curves.

M **13.** In Exercise 12, (19.52) are the so-called Cauchy-Riemann equations that express the fact that $u + iv$ is a ***holomorphic*** function of $z = x + iy$ (such functions will be discussed in Section 22.1). Figure 19.16 illustrates the case $z \mapsto \tan z$. Use Notebook 19 to study instead $z \mapsto \cos z$ and $z \mapsto \log z$, and to construct corresponding surfaces of revolution.

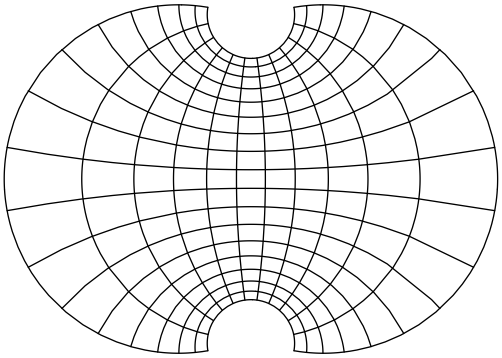


Figure 19.16: Image of horizontal and vertical lines by $z \mapsto \tan z$

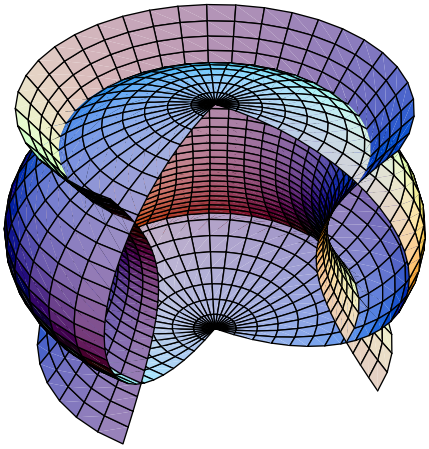


Figure 19.17: Orthogonal ellipsoid and hyperboloid of revolution

M **14.** Let $\mathfrak{A}$ and $\mathfrak{B}$ be a system of orthogonal curves in $\mathbb{R}^2$ that are symmetric with respect to reflection in a line ℓ. Let $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$ be the families of surfaces of revolution in $\mathbb{R}^3$ generated by $\mathfrak{A}$ and $\mathfrak{B}$, using ℓ as axis (see Figure 19.17). Show that $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$, together with the family $\mathfrak{C}$ of planes through ℓ, constitute a triply orthogonal system of surfaces.

15. Complete the proof of Theorem 19.36.

Notebook 19

■ N19.0

The aim of this notebook is to study principal curves, spot umbilic points and construct triply orthogonal systems. We base our investigation on some standard surfaces loaded in N19.0.1, though we shall also need the *nephroid*, which is a special case of the following curve defined in N5.5.

```
epicycloid[a_,b_][t_]:= {(a+b)Cos[t]-b Cos[((a+b)/b) t],
  (a+b) Sin[t]-b Sin[((a+b)/b) t]}
```

```
nephroid[a_][t_]:= a epicycloid[2,1][t]
```

```
nephroid[1][t]
```

We shall use the parametrization, first introduced in N15.1, of the surface formed by revolving a curve α around the z-axis.

```
surfrev[α_][u_,v_]:= {Cos[u]α[v][[1]], Sin[u]α[v][[1]], α[v][[2]]}
```

Because most of the 3-dimensional plots in this notebook are colored manually, we include Lighting->False as a default option.

```
SetOptions[ParametricPlot3D,
  Compiled->False, Axes->None, BoxRatios->Automatic,
  Boxed->False, Lighting->False
]
```

```
<<Graphics`Colors`
```

The Colors package enables one to use alphabetical names for coloring. Subsection N19.0.1 includes some surface definitions and the programs gcurv, mcurv.

■ N19.0.1

■ N19.1

We first reparametrize the helicoid manually, so that its coordinate curves are lines of curvature:

```
helicoidprin[b_][p_,q_]:=
  {b Cos[p+q]Sinh[q-p], b Sin[p+q]Sinh[q-p], b(p+q)}
```

```
hel:= helicoidprin[1]
hlp:= ParametricPlot3D[hel[p,q], {p,-2π,2π}, {q,-2π,2π},
  PlotPoints->{80,80}, PlotRange->All, Lighting->True];
hlp
```

At this point, one encounters a graphics problem that arises often. The figure is too squashed because the polygons from which it is formed have vertices too far apart. To improve the plot, we can manually give instructions to consider only those polygons with vertices less than a certain distance apart. This is the aim of the customized program selectgraphics3d, which we shall use again in the sequel.

```
selectgraphics3d[obj_,b_,opts___]:=
Show[Graphics3D[Select[obj,
Abs[#[[1,1,1]]]<b &&
Abs[#[[1,1,2]]]<b &&
Abs[#[[1,1,3]]]<b &&
Abs[#[[1,2,1]]]<b &&
Abs[#[[1,2,2]]]<b &&
Abs[#[[1,2,3]]]<b &&
Abs[#[[1,3,1]]]<b &&
Abs[#[[1,3,2]]]<b &&
Abs[#[[1,3,3]]]<b &&
Abs[#[[1,4,1]]]<b &&
Abs[#[[1,4,2]]]<b &&
Abs[#[[1,4,2]]]<b &]], opts]
```

Figure 19.1 sets the maximum distance at 5 units. Subsequent plots use different values; with the factory settings in N19.0.1, they will only be displayed after insertion of S@ at the start of each line.

```
S @ selectgraphics3d[hlp[[1]], 5]
```

Figure 19.1

```
selectgraphics3d[hlp[[1]], 10]
```

```
selectgraphics3d[hlp[[1]], 3]
```

■ N19.2

The next command could have been derived using paraboloid from N19.0, though we include it for the sake of completeness.

```
ellipticparaboloid[a_,b_][u_,v_]:= {u, v, u^2/a^2+v^2/b^2}
```

In the examples below, we shall take a=3/2 and b=1:

```
ellipticparaboloid[3/2,1][u,v]
ep:= paraboloid[4/9,1]
ep[u,v]
```

Umbilic points are characterized by the vanishing of $\delta = H^2 - K$. This quantity is always nonnegative on a surface (because the arithmetic mean of two positive numbers is no less than their geometric mean). In certain cases, umbilic points can be visualized by coloring a surface using δ, which it is convenient to define formally:

```
umbilic[x_][u_,v_]:= mcurv[x][u,v]^2 - gcurv[x][u,v] // Simplify
```

```
δ= umbilic[ep][u,v]
```

```
r:= 1
h:= If[δ<.001, 0, .5]
ParametricPlot3D[Append[ep[u,v], Hue[h]], {u,-r,r}, {v,-r,r},
  ViewPoint->{1,1,3}, PlotPoints->30]
```

In practice *Mathematica* cannot distinguish between real and complex values of u,v, so the vanishing of δ merely defines a complex curve in the complex uv-plane:

```
Solve[δ==0]
```

Each of the quantities under a square root sign in the output is nonpositive, and the value of v for which it is *zero* gives the umbilic points. A quicker way to find the umbilic points is to throw in one of the derivatives of δ and set this to zero. For the function δ has an absolute minimum of 0 on the surface, and its partial derivatives (computed in the real directions of the surface) must vanish.

```
Solve[{δ==0, D[δ,v]==0}]
```

This time, the output shows more clearly that there are two real solutions, and their position is consistent with the plot after the definition of δ above:

```
r:= 1
a:= Sqrt[5]/4
k:= {0,0,.01}
pts:= {PointSize[.03], Point[ep[0,a]+k],
Point[ep[0,-a]+k]}//Graphics3D
pa:= ParametricPlot3D[ep[u,v], {u,-r,r}, {v,-r,r},
  ViewPoint->{1,1,3}, PlotPoints->30, Lighting->True]
S[pa,pts]
Clear[a]
```

Figure 19.2

The calculation below shows that the upper half of a ellipsoid without any extra symmetry has two umbilic points. Thus, the entire ellipsoid will have four.

```
r:= 2
k:= {0,0,.01}
x[u_,v_]:= {u, v, Sqrt[1 - u^2/2 - v^2/4]}
Solve[{umbilic[x][u,v]==0, D[umbilic[x][u,v], u]==0}]
```

Next is the hyperboloid of one sheet. The real values of u,v giving solutions do not correspond to real points on the surface as the square root is negative. Thus, there are no umbilic pionts.

```
y[u_,v_]:= {u, v, Sqrt[u^2/2 + v^2/4 - 1]}
Solve[{umbilic[y][u,v]==0, D[umbilic[y][u,v], u]==0}]
```

Here are four umbilic points on a paraboloid-type surface. They are plotted automatically by finding all the solutions which become the coordinates of Point. Replace Show by S in the last line; the complex points give error messages which are suppressed.

```
Off[Graphics3D::nlist3]
r:= 1
k:={0,0,.1}
x[u_,v_]:={u, v, 1 - u^4 - v^2}
so:= Solve[{umbilic[x][u,v]==0, D[umbilic[x][u,v],u]==0}]
pts:= {PointSize[.03], Point/@(x@@{u,v}/.so)}//Graphics3D
pa:= ParametricPlot3D[x[u,v], {u,-r,r}, {v,-r,r},
  ViewPoint->{1,1,3}, Shading->False, PlotPoints->30]
S[pa, pts]
```

■ N19.5

The following commands allow us to plot three examples of triply orthogonal systems. The first parametrizes the plane that passes through the origin and the two chosen points, and the second specifies three mutually orthogonal planes, each with a specified color as fourth coordinate.

```
plane[a1_,a2_,b1_,b2_,c1_,c2_][u_,v_]:=
   {a1 u + a2 v, b1 u + b2 v, c1 u + c2 v}
```

```
pla= {
Table[Append[plane[1,0,0,1,0,0][u,v] + {0,0,z}, Red],
  {z,-.5, .5, .5}],
Table[Append[plane[1,0,0,0,0,1][u,v] + {0,y,0}, Yellow],
  {y,-.5, .5, .5}],
Table[Append[plane[0,0,1,0,0,1][u,v] + {x,0,0}, Green],
   {x,-.5, .5, .5}]
}
```

```
ta:= Flatten[pla,1]
S @ ParametricPlot3D[ta//Evaluate, {u,-1,1}, {v,-1,1},
  PlotPoints->{5,5}]
```

Figure 19.3

The next figures apply the same technique, though each family of surfaces is defined and colored separately. Patches for a cylinder and cone were loaded in N19.0.1.

```
t[1]:= Table[Append[plane[Cos[t],0,Sin[t],0,0,1][u,v], Red],
  {t,0,π,π/4}]
t[2]:= Table[Append[plane[1,0,0,1,0,0][u,v]+{0,0,h}, Yellow],
  {h,-.5,.5,.5}]
t[3]:= Table[Append[circularcylinder[a][u,v], Green],
  {a,.25,.75,.25}]
z[i_, a_,b_,A_]:= ParametricPlot3D[Evaluate[t[i]], {u,a,b}, {v,-1,1},
  PlotPoints->A]
S @ {z[1,-1,1,2], z[2,-1,1,5], z[3,0,2π,{17,2}]}
```

Figure 19.4

```
Clear[z]
t[1]:= Table[Append[plane[Cos[t],0,Sin[t],0,0,1][u,v], Red],
  {t, π/6, 5π/12, π/4}]
t[2]:= Table[Append[circularcone[Cos[ϕ],Sin[ϕ]][u,v], Yellow],
  {ϕ, π/8, 3π/8, π/8}]
t[3]:= Table[Append[sphere[a][θ,ϕ],Green], {a,.25,1,.25}]
z[i_, X__]:= ParametricPlot3D[Evaluate[t[i]], X]
S @ {z[1, {u,-1,1}, {v,-1,1}, PlotPoints->2],
  z[2, {u,0,3π/2},{v,-1.2,1.2}, PlotPoints->{22,5}],
  z[3, {θ,0,3π/2},{ϕ,-π/2, π/2}, PlotPoints->{19,11}]}
```

Figure 19.5

■ N19.6

To make use of elliptic coordinates in *Mathematica,* we need a function that defines each of the three triply orthogonal surfaces, and also allows for the choice of signs. This is accomplished by

```
ellipticcoor[d_,e_,f_][a_,b_,c_][w_][u_,v_]:=
  {d Sqrt[(a-u)(a-v)(a-w)/((a-b)(a-c))],
   e Sqrt[(b-u)(b-v)(b-w)/((b-a)(b-c))],
   f Sqrt[(c-u)(c-v)(c-w)/((c-a)(c-b))]}
```

```
ellipticcoor[1,1,1][12,5,1][a][u,v]
```

The following command will be used to combine eight different patches.

```
octants[x_]:=
Module[{w},
  w= {0,0,-3};
  {x[1,1,1], x[-1,1,1], x[1,-1,1], x[-1,-1,1],
   x[1,1,-1]+w, x[-1,1,-1]+w, x[1,-1,-1]+w, x[-1,-1,-1]+w}
]
```

```
Module[{x,xx},
  x[d_,e_,f_]= ellipticcoor[d,e,f][12,5,1][0][u,v];
  xx= octants[x];
  S @ ParametricPlot3D[xx//Evaluate, {u,1,5}, {v,5,12},
    PlotPoints->{10,20}, ViewPoint->{2.7,2.7,4.2},
    Boxed->True, Lighting->True]
]
```

Figure 19.6

■ N19.6.1

■ N19.7

Here are analogous definitions for parabolic coordinates and their combined patches:

```
paraboliccoor[d_,e_,f_][a_,b_,c_][w_][u_,v_]:=
  {d Sqrt[(a-u)(a-v)(a-w)/(b-a)], e Sqrt[(b-u)(b-v)(b-w)/(a-b)],
   f(c-a-b+u+v+w)/2}
```

```
octdown[x_]:= Flatten[Table[x[l,m,1], {l,-1,1,2}, {m,-1,1,2}], 1];
```

```
parabolicplot[a_,col_,opts___][p_,q_,r_,s_]:=
Module[{x,hi,xx},
  x[d_,e_,f_]= paraboliccoor[d,e,f][2,1,3][a][u,v];
  hi[x_]:= Append[x, FaceForm[col]];
  xx:= hi/@octdown[x];
  ParametricPlot3D[xx//Evaluate, {u,p,q}, {v,r,s},
    PlotPoints->{16,8}, Lighting->False, opts]
]
```

```
parabolicplot[4, Red][-9,1,1,2];
```

```
S[parabolicplot[4, Red, ViewPoint->{.8,-1.8,2.8}][-9,1,1,2],
  parabolicplot[1.5, Green][-9,1,2,12],
  parabolicplot[-3, Yellow][1,2,2,12]]
```

Figure 19.10

The following cells back up the equations leading to Definition 19.27.

```
A= (u-λ)(v-λ)(w-λ) - (a-λ)(b-λ)(2Z-c-λ) + X(b-λ) + Y(a-λ)
```

```
Solve[(A/.λ->a)==0, X]
Solve[(A/.λ->b)==0, Y]
```

```
su:= {X->-(a-u)(a-v)(-a+w)/(-a+b), Y->-(b-u)(b-v)(-b+w)/(a-b)}
AA= A/.su//Simplify
```

```
Solve[AA==0, Z]
A/.Z->(-a-b+c+u+v+w)/2 /.su //Simplify
```

The next example of a triply orthogonal system is manufactured in stages from planes and parallel surfaces to a catenoid.

```
xx[t_]:= Append[plane[Cos[t],0,Sin[t],0,0,1][u,v], Cyan]
yy1:= Table[xx[t], {t,-π/2,0,π/2}]
zz1:= ParametricPlot3D[yy1//Evaluate, {u,-3,0}, {v,-3,3},
  PlotPoints->3]
```

```
yy2:= Append[parallelsurf[catenoid[1], s][u,1], Purple]
zz2:= ParametricPlot3D[yy2, {u,0,3π/2}, {s,-2.3811,2},
  Lighting->False, PlotPoints->{21,11}]
```

```
yy3:= Append[catenoid[1][u,v], Orange]
zz3:= ParametricPlot3D[yy3, {u,0,3π/2},{v,-1.7,1.7},
  PlotPoints->{21,11}]
```

```
Show[zz1, zz2, zz3];
```

The next cell combines parabolic coordinate patches with color functions in each of the four upper octants, to form the right-hand half of Figure 19.11.

```
upplot:=
Module[{h,k,x,xx},
  h= FaceForm[Hue[(u+9)/10], Hue[v-1]];
  k= FaceForm[Hue[v-1], Hue[(u+9)/10]];
  x[d_,e_,f_]= paraboliccoor[d,e,f][2,1,3][4][u,v];
  xx= {Append[x[1,1,1], h], Append[x[-1,-1,1], h],
    Append[x[-1,1,1], k], Append[x[1,-1,1], k]};
  ParametricPlot3D[xx//Evaluate, {u,-9,1}, {v,1,2},
    PlotPoints->{16,16}]
]
```

```
downplot:=
Module[{h,k,x,xx},
  h= FaceForm[Hue[u-1], Hue[(v-1)/10]];
  k= FaceForm[Hue[(v-1)/10], Hue[u-1]];
  x[d_,e_,f_]= paraboliccoor[d,e,f][2,1,3][-3][u,v];
  xx= {Append[x[1,1,1], h], Append[x[-1,-1,1], h],
    Append[x[-1,1,1], k], Append[x[1,-1,1], k]};
  ParametricPlot3D[xx//Evaluate, {u,1,2},{v,2,15},
    PlotPoints->{12,12}, PlotRange->{0,10}]
]
S @ GraphicsArray[{downplot, upplot}, GraphicsSpacing->.1]
```

Figure 19.11

■ N19.7.1

■ N19.8

The next program gives us a patch that we use to plot parallel surfaces to ellipsoids and catenoids.

```
parallelsurf[x_,r_][u_,v_]:=
Module[{xu,xv,n},
  xu=D[x[uu,vv],uu];
  xv=D[x[uu,vv],vv];
  n[1]= Cross[xu,xv]//Simplify;
  n[2]= Factor[n[1].n[1]]//Simplify;
  n[3]= PowerExpand[Sqrt[n[2]]];
  n[4]= n[1]/n[3];
  Together[x[uu,vv]+r n[4]]/.{uu->u,vv->v}//Simplify
]
```

```
parallelsurf[ellipsoid[a,b,c],t][u,v]//FullSimplify//PowerExpand
```

```
Clear[x]
ell:= ellipsoid[3/2,3/2,1]
x:= parallelsurf[ell, .9]
S @ ParametricPlot3D[x[u,v]//Evaluate, {u,0,π},{v,-π/2,π/2},
  PlotPoints->{20,35}, Lighting->True]
```

Figure 19.14

```
parallelsurf[catenoid[a],t][u,v]//PowerExpand
```

```
Module[{y},
  y[0]= catenoid[1];
  y[-1]= parallelsurf[y[0], -.8];
  y[1]=  parallelsurf[y[0], .8];
  S @ ParametricPlot3D[Table[y[i][u,v], {i,-1,1}]//Evaluate,
    {u,0,3π/2}, {v,-1.5,1.5}, PlotPoints->{40,40}, Lighting->True]
]
```

Figure 19.15

■ N19.9.1

■ N19.10

A completely equivalent definition of the command ellipticcoor from N19.6 is given in the next cell, though the presence of parentheses allows us to distinguish the two definitions and check that their difference is zero.

```
ellipticcoor[{d,e,f}_][{a,b,c}_][w_][u_,v_]=
Module[{F},
  F[x_,s_]= Sqrt[(x-u)(x-v)(x-w)/s];
  {d F[a,(a-b)(a-c)], e F[b,(b-a)(b-c)], f F[c,(c-a)(c-b)]}
]
```

```
% - ellipticcoor[d,e,f][a,b,c][w][u,v]
```

The following miniprogram has the same functionality as Maeder's CartesianMap and PolarMap (see [Ma]), but draws curved instead of straight lines. It plots the images of horizontal and vertical lines under F.

```
plotpatch2d[F_, {x_,x1_,x2_}, {y_,y1_,y2_}, {nx_,ny_}]:=
Module[{t1,t2,z1,z2},
  reim[A_]:= ComplexExpand[{Re[A], Im[A]}];
  t1= Table[reim[F], {y, y1, y2, (y2-y1)/(ny-1)}];
  t2= Table[reim[F], {x, x1, x2, (x2-x1)/(nx-1)}];
  z1= ParametricPlot[t1//Evaluate, {x,x1,x2}, PlotPoints->nx];
  z2= ParametricPlot[t2//Evaluate, {y,y1,y2}, PlotPoints->ny];
  {z1, z2}
]
```

Exercise 13 suggests replacing Tan by Cos or Log in the plot

```
S @ plotpatch2d[Tan[x+I y], {x,-1,1}, {y,-1,1}, {12,18}]
```

Figure 19.16

The final subsection contains a 3-dimensional version of plotpatch2d that is used to draw Figure 19.17.

■ N19.10.1

Chapter 20

Canal Surfaces and Cyclides of Dupin

Let $\mathcal{M}$ be a regular surface in $\mathbb{R}^3$ and let $\mathcal{W} \subseteq \mathcal{M}$ be the image of a regular patch on which a unit normal $\mathbf{U}$ is defined and differentiable. Denote by k_1 and k_2 the principal curvatures of $\mathcal{M}$ with respect to $\mathbf{U}$, and suppose that they are ordered so that $k_1 \geqslant k_2$ on $\mathcal{W}$.

Definition 20.1. *The reciprocals*

$$\rho_1 = \frac{1}{k_1} \qquad \textit{and} \qquad \rho_2 = \frac{1}{k_2}$$

of the principal curvatures are called the ***principal radii of curvature*** *of* $\mathcal{M}$.

This chapter is concerned with the geometry arising from this definition, and analogs of the evolute of a curve, defined in Section 4.1.

For a point $\mathbf{q}$ in a regular surface $\mathcal{M}$, we denote by $\ell_{\mathbf{q}}$ the line normal to $\mathcal{M}$ at $\mathbf{q}$; then the surface normal $\mathbf{U}(\mathbf{q})$ is a vector in $\ell_{\mathbf{q}}$. Each normal section (see page 391) is a plane curve $\mathscr{C}$ in a plane $\mathit{\Pi}$ containing $\ell_{\mathbf{q}}$. Since $\mathbf{U}(\mathbf{q})$ is perpendicular to $\mathscr{C}$ at $\mathbf{q}$, the center of curvature (see page 99) of $\mathscr{C}$ lies on $\ell_{\mathbf{q}}$. Indeed, the centers of curvature at $\mathbf{q}$ fill out a connected subset F of $\ell_{\mathbf{q}}$ called the ***focal interval*** of $\mathcal{M}$ at $\mathbf{q}$.

The focal interval reduces to a point if $\mathbf{q}$ is itself an umbilic point (see page 597); otherwise, it is a line segment, or the complement of a line segment. The extremities of F are called the ***focal points*** of $\mathcal{M}$ at $\mathbf{q}$, and therefore coincide if and only if $\mathbf{q}$ is umbilic. If the Gaussian curvature of $\mathcal{M}$ vanishes at $\mathbf{q}$, then at least one focal point is at infinity. Figure 20.1 displays the typical behavior at an elliptic and a hyperbolic point; only in the latter case does the focal interval contain 0.

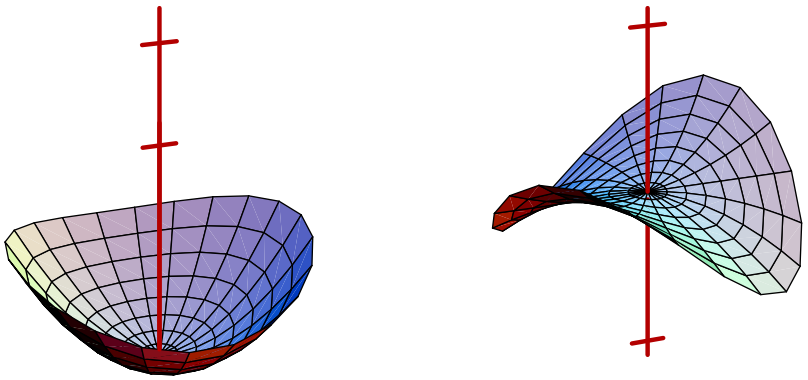


Figure 20.1: Focal interval at the vertex of both an elliptic and a hyperbolic paraboloid

It is clear from (13.22), page 400, that each k_i is continuous on $\mathcal{W}$. But it can happen that k_1 or k_2 is not differentiable at an umbilic point; this happens for the monkey saddle (see Figure 13.6 on page 404). But it *is* known that both k_1 and k_2 are differentiable provided $\mathcal{W}$ contains no umbilic points; for details see [Ryan, page 371], [CeRy, page 134]. In this chapter we assume that both k_1 and k_2 are differentiable on $\mathcal{W}$, except possibly at isolated points, and for convenience we take $\mathcal{W} = \mathcal{M}$.

Definition 20.2. *The* ***focal set*** *of a regular surface* $\mathcal{M} \subset \mathbb{R}^3$ *is*

$$\mathsf{focal}(\mathcal{M}) = \{\, \mathbf{p} \in \mathbb{R}^3 \mid \mathbf{p} \text{ is a focal point of some } \mathbf{q} \in \mathcal{M} \,\}.$$

In the case that $\mathcal{M} \subset \mathbb{R}^3$ is connected, orientable, and free of umbilic points, the set $\mathsf{focal}(\mathcal{M})$ has two components, one corresponding to each principal curvature. There are three possibilities:

Case 1. Each component of $\mathsf{focal}(\mathcal{M})$ is a surface.

Case 2. One component of $\mathsf{focal}(\mathcal{M})$ is a curve and the other a surface.

Case 3. Each component of $\mathsf{focal}(\mathcal{M})$ is a curve.

Sections 20.1, 20.2 and 20.3 deal successively with each of these cases. The generic situation is dealt with first, and we investigate properties of the focal surfaces. The second case characterizes envelopes of a 1-parameter family of

spheres, called *canal surfaces*, and the third gives rise to the *cyclides of Dupin*. A cyclide of Dupin can be considered to be a canal surface of either of the components of its focal set.

Perhaps the simplest type of mapping between open subsets of $\mathbb{R}^n$, other than an isometry or affine transformation, is an *inversion*. In Section 20.4 we give the precise definition of inversion and prove that it is a conformal map. Inversion of surfaces is studied in Section 20.5, the main aim being to exhibit cyclides of Dupin as inversions of tori.

20.1 Surfaces Whose Focal Sets Are 2-Dimensional

In this section, we consider Case 1. The first order of business is to make each component of $\mathsf{focal}(\mathcal{M})$ into a surface, possibly with singularities. This will be done by showing that each regular principal patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ gives rise to patches $\mathbf{z}_1\colon \mathcal{U} \to \mathsf{focal}(\mathcal{M})$ and $\mathbf{z}_2\colon \mathcal{U} \to \mathsf{focal}(\mathcal{M})$ parametrizing each of the components of $\mathsf{focal}(\mathcal{M})$.

Definition 20.3. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on a surface $\mathcal{M} \subset \mathbb{R}^3$. With the notation from* page 639*, the* ***focal patches*** *corresponding to $\mathbf{x}$ are defined by*

$$\mathbf{z}_1(u,v) = \mathbf{x}(u,v) + \rho_1\mathbf{U}(u,v) \quad \textit{and} \quad \mathbf{z}_2(u,v) = \mathbf{x}(u,v) + \rho_2\mathbf{U}(u,v). \tag{20.1}$$

By definition, the image of each focal patch $\mathbf{z}_i$ is contained in $\mathsf{focal}(\mathcal{M})$.

We want to compute the Gaussian curvatures of each component of the focal set of $\mathcal{M}$, and also the first and second fundamental forms. For this, we need to parametrize $\mathcal{M}$ by patches and then compute the first and second fundamental forms of the associated patches $\mathbf{z}_1, \mathbf{z}_2$. The calculations are considerably easier when we use principal patches, as defined on page 596.

Theorem 20.4. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular principal patch, let $\mathbf{z}_1$ and $\mathbf{z}_2$ be the corresponding focal patches, and let $\mathbf{U}_i$ be the unit normal of $\mathbf{z}_i$. Then*

$$\left\{\begin{array}{rcl} \mathbf{z}_{1u} &=& \rho_{1u}\mathbf{U}, \\ \mathbf{z}_{1v} &=& \left(1 - \dfrac{\rho_1}{\rho_2}\right)\mathbf{x}_v + \rho_{1v}\mathbf{U}, \\ \mathbf{U}_1 &=& -\operatorname{sign}\!\left(\rho_{1u}\left(1 - \dfrac{\rho_1}{\rho_2}\right)\right)\dfrac{\mathbf{x}_u}{\sqrt{E}}, \end{array}\right. \tag{20.2}$$

and

$$\left\{\begin{array}{rcl} \mathbf{z}_{2u} &=& \left(1 - \dfrac{\rho_2}{\rho_1}\right)\mathbf{x}_u + \rho_{2u}\mathbf{U}, \\ \mathbf{z}_{2v} &=& \rho_{2v}\mathbf{U}, \\ \mathbf{U}_2 &=& -\operatorname{sign}\!\left(\rho_{2v}\left(1 - \dfrac{\rho_2}{\rho_1}\right)\right)\dfrac{\mathbf{x}_v}{\sqrt{G}}. \end{array}\right. \tag{20.3}$$

It is a consequence of this theorem that $\mathbf{z}_1$ is regular on the set

$$\big\{(u,v)\in\mathcal{U} \mid \rho_1(u,v)\neq\rho_2(u,v) \text{ and } \rho_{1u}(u,v)\neq 0\big\},$$

and $\mathbf{z}_2$ is regular on

$$\big\{(u,v)\in\mathcal{U} \mid \rho_1(u,v)\neq\rho_2(u,v) \text{ and } \rho_{2v}(u,v)\neq 0\big\}.$$

Proof. We have

$$-\mathbf{U}_u = S(\mathbf{x}_u) = \frac{1}{\rho_1}\mathbf{x}_u \qquad \text{and} \qquad -\mathbf{U}_v = S(\mathbf{x}_v) = \frac{1}{\rho_2}\mathbf{x}_v. \tag{20.4}$$

Differentiating (20.1) and using (20.4), we get

$$\mathbf{z}_{1u} = \mathbf{x}_u + \rho_{1u}\mathbf{U} + \rho_1\mathbf{U}_u = \mathbf{x}_u + \rho_{1u}\mathbf{U} + \rho_1\Big(-\frac{1}{\rho_1}\mathbf{x}_u\Big) = \rho_{1u}\mathbf{U}. \tag{20.5}$$

Similarly,

$$\mathbf{z}_{1v} \;=\; \left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{x}_v + \rho_{1v}\mathbf{U}. \tag{20.6}$$

Using (20.5) and (20.6),

$$\mathbf{z}_{1u}\times\mathbf{z}_{1v} = \rho_{1u}\mathbf{U}\times\left(\left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{x}_v + \rho_{1v}\mathbf{U}\right) = \rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{U}\times\mathbf{x}_v.$$

Hence

$$\mathbf{U}_1 = \frac{\mathbf{z}_{1u}\times\mathbf{z}_{1v}}{\|\mathbf{z}_{1u}\times\mathbf{z}_{1v}\|} = \frac{\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{U}\times\mathbf{x}_v}{\sqrt{G}\left|\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right|} = -\operatorname{sign}\!\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\frac{\mathbf{x}_u}{\sqrt{E}}.$$

This proves (20.2). Similar calculations yield (20.3). ∎

A point of the principal curve corresponding to ρ_1 for which ρ_{1u} vanishes is called a ***ridge point*** of the principal curve on $\mathcal{M}$. A similar definition holds for the principal curve corresponding to ρ_2. From (20.2) and (20.3) we see that ridge points on $\mathcal{M}$ give rise to singular points on the focal set of $\mathcal{M}$.

The coefficients of the first fundamental forms of the focal patches are easily computed from Theorem 20.4. We omit the verification that is carried out in Notebook 20.

Lemma 20.5. *Let $\mathbf{x}\colon\mathcal{U}\to\mathbb{R}^3$ be a regular principal patch. Then the coefficients of the first fundamental form of $\mathbf{z}_1$ are given by*

$$E_1 = \rho_{1u}^2, \qquad F_1 = \rho_{1u}\rho_{1v}, \qquad G_1 = \left(1-\frac{\rho_1}{\rho_2}\right)^2 G + \rho_{1v}^2, \tag{20.7}$$

and the coefficients of the first fundamental form of $\mathbf{z}_2$ are given by

$$E_2 = \left(1-\frac{\rho_2}{\rho_1}\right)^2 E + \rho_{2u}^2, \qquad F_2 = \rho_{2u}\rho_{2v}, \qquad G_2 = \rho_{2v}^2. \tag{20.8}$$

Before computing the second fundamental forms of the focal patches, we note the following easy consequence of Corollary 19.10, page 602 (see Exercise 2).

Lemma 20.6. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular principal patch, and denote by ρ_1 and ρ_2 the principal radii of curvature of $\mathbf{x}$. Then*

$$\left\{\begin{array}{rcl} \dfrac{\rho_{1v}}{\rho_1^2} &=& \dfrac{E_v}{2E}\left(\dfrac{1}{\rho_1}-\dfrac{1}{\rho_2}\right), \\[2ex] \dfrac{\rho_{2u}}{\rho_2^2} &=& \dfrac{G_u}{2G}\left(\dfrac{1}{\rho_2}-\dfrac{1}{\rho_1}\right). \end{array}\right. \tag{20.9}$$

Now we can compute the second fundamental forms of the focal patches.

Lemma 20.7. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular principal patch such that the two branches of the focal set of $\mathbf{x}$ are parametrized by* (20.1)*. Then the coefficients of the second fundamental form of $\mathbf{z}_1$ are given by*

$$\left\{\begin{array}{rcl} e_1 &=& \operatorname{sign}\left(\rho_{1u}\left(1-\dfrac{\rho_1}{\rho_2}\right)\right)\dfrac{\sqrt{E}\,\rho_{1u}}{\rho_1}, \\[2ex] f_1 &=& 0, \\[2ex] g_1 &=& \operatorname{sign}\left(\rho_{1u}\left(1-\dfrac{\rho_1}{\rho_2}\right)\right)\left(1-\dfrac{\rho_1}{\rho_2}\right)\dfrac{G_u}{2\sqrt{E}}, \end{array}\right. \tag{20.10}$$

and the coefficients of the second fundamental form of $\mathbf{z}_2$ are given by

$$\left\{\begin{array}{rcl} e_2 &=& \operatorname{sign}\left(\rho_{2v}\left(1-\dfrac{\rho_2}{\rho_1}\right)\right)\left(1-\dfrac{\rho_2}{\rho_1}\right)\dfrac{E_v}{2\sqrt{G}}, \\[2ex] f_2 &=& 0, \\[2ex] g_2 &=& \operatorname{sign}\left(\rho_{2v}\left(1-\dfrac{\rho_2}{\rho_1}\right)\right)\dfrac{\sqrt{G}\,\rho_{2v}}{\rho_2}. \end{array}\right. \tag{20.11}$$

Furthermore, g_1 and e_2 are given by the alternate formulas

$$\left\{\begin{array}{rcl} g_1 &=& -\operatorname{sign}\left(\rho_{1u}\left(1-\dfrac{\rho_1}{\rho_2}\right)\right)\dfrac{G\,\rho_1\rho_{2u}}{\rho_2^2\sqrt{E}}, \\[2ex] e_2 &=& -\operatorname{sign}\left(\rho_{2v}\left(1-\dfrac{\rho_2}{\rho_1}\right)\right)\dfrac{E\,\rho_2\rho_{1v}}{\rho_1^2\sqrt{G}}. \end{array}\right. \tag{20.12}$$

Proof. From (20.2) we compute the second derivatives of $\mathbf{z}_1$:

$$\mathbf{z}_{1uu} = \rho_{1uu}\mathbf{U} + \rho_{1u}\mathbf{U}_u = \rho_{1uu}\mathbf{U} - \frac{\rho_{1u}}{\rho_1}\mathbf{x}_u, \tag{20.13}$$

$$\mathbf{z}_{1uv} = \rho_{1uv}\mathbf{U} + \rho_{1u}\mathbf{U}_v = \rho_{1uv}\mathbf{U} - \frac{\rho_{1u}}{\rho_2}\mathbf{x}_v. \tag{20.14}$$

Then (20.13) and (20.2) imply that

$$\begin{aligned} e_1 &= \mathbf{z}_{1uu} \cdot \mathbf{U}_1 = \left(\rho_{1uu}\mathbf{U} - \frac{\rho_{1u}}{\rho_1}\mathbf{x}_u\right) \cdot \left(-\operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\frac{\mathbf{x}_u}{\sqrt{E}}\right) \\ &= \operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\rho_{1u}\frac{\sqrt{E}}{\rho_1}. \end{aligned}$$

Similarly, (20.14) implies that

$$f_1 = \mathbf{z}_{1uv} \cdot \mathbf{U}_1 = \left(\rho_{1uv}\mathbf{U} - \frac{\rho_{1u}}{\rho_1}\mathbf{x}_v\right) \cdot \left(-\operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\frac{\mathbf{x}_u}{\sqrt{E}}\right) = 0,$$

because $F = 0$. To compute g_1, we use (20.2) and (20.14), together with the fact that $\mathbf{x}_u \cdot \mathbf{x}_v = \mathbf{x}_u \cdot \mathbf{U} = 0$; thus

$$\begin{aligned} g_1 &= \mathbf{z}_{1vv} \cdot \mathbf{U}_1 \\ &= \left(-\left(\frac{\rho_1}{\rho_2}\right)_v \mathbf{x}_v + \left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{x}_{vv} + \rho_{1vv}\mathbf{U} + \rho_{1v}\mathbf{U}_v\right) \\ &\qquad \cdot \left(-\operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\frac{\mathbf{x}_u}{\sqrt{E}}\right) \\ &= -\operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\left(1-\frac{\rho_1}{\rho_2}\right)\mathbf{x}_{vv} \cdot \frac{\mathbf{x}_u}{\sqrt{E}}. \end{aligned} \tag{20.15}$$

Since $\mathbf{x}_{vv} \cdot \mathbf{x}_u = -\mathbf{x}_v \cdot \mathbf{x}_{uv} = -G_u/2$, (20.15) can be written as

$$g_1 = \operatorname{sign}\left(\rho_{1u}\left(1-\frac{\rho_1}{\rho_2}\right)\right)\left(1-\frac{\rho_1}{\rho_2}\right)\frac{G_u}{2\sqrt{E}}. \tag{20.16}$$

Similar calculations yield (20.11). Finally, (20.12) is a consequence of (20.10), (20.11) and (20.9) (see Exercise 2 for details). ▮

Finally, we find concise formulas for the Gaussian curvatures of the focal patches.

Corollary 20.8. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular principal patch such that the two branches of the focal set of $\mathbf{x}$ are parametrized by* (20.1). *Then the Gaussian curvatures of $\mathbf{z}_1$ and $\mathbf{z}_2$ are given by*

$$\begin{cases} K_1 = -\dfrac{\rho_{2u}}{(\rho_1-\rho_2)^2\rho_{1u}}, \\ K_2 = -\dfrac{\rho_{1v}}{(\rho_1-\rho_2)^2\rho_{2v}}. \end{cases} \tag{20.17}$$

Proof. The first equation in (20.17) is an immediate consequence of formulas (20.7), (20.10), (20.12), and (13.20) on page 400. The second equation is found similarly. ▮

We conclude this section with an easy example, used in Notebook 20 to illustrate the theory. We merely state the results and plot the focal surfaces.

Focal Sets of a Hyperbolic Paraboloid

A hyperbolic paraboloid is parametrized, as on page 296, by $\mathbf{x}(u,v)=(u,\ v,\ uv)$. Its Gaussian and mean curvatures are given by

$$K = \frac{e\,g - f^2}{E\,G - F^2} = -\frac{1}{(1+u^2+v^2)^2}$$

and

$$H = \frac{e\,G - 2\,f\,F + g\,E}{2(E\,G - F^2)} = -\frac{u\,v}{(1+u^2+v^2)^{3/2}}.$$

Thus

$$H^2 - K = \frac{(1+u^2)(1+v^2)}{(1+u^2+v^2)^3},$$

so that

$$\begin{cases} k_1 = H + \sqrt{H^2-K} \ = \ \dfrac{-u\,v + \sqrt{(1+u^2)(1+v^2)}}{(1+u^2+v^2)^{3/2}}, \\ k_2 = H - \sqrt{H^2-K} \ = \ \dfrac{-u\,v - \sqrt{(1+u^2)(1+v^2)}}{(1+u^2+v^2)^{3/2}}. \end{cases}$$

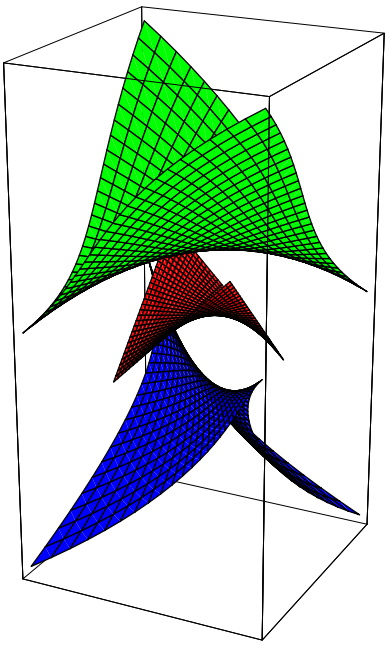


Figure 20.2: Hyperbolic paraboloid (in the center) and its focal set

Proposition 20.9. *The focal patches* $\mathbf{z}_1(u,v)$, $\mathbf{z}_2(u,v)$ *of the hyperbolic paraboloid* $\mathbf{x}$ *are given by*

$$\Big(-uv^2 \mp \sqrt{(u^2+1)(v^2+1)}\,v + u, \\ -vu^2 \mp \sqrt{(u^2+1)(v^2+1)}\,u + v,\ 2uv \pm \sqrt{(u^2+1)(v^2+1)}\Big),$$

with the respective choice of signs.

Their traces are illustrated in Figure 20.2. The principal curvatures of $\mathbf{x}$ satisfy

$$|k_1(u,v)|,\ |k_2(u,v)| \geqslant \frac{1}{(1+u^2+v^2)^{3/2}}.$$

20.2 Canal Surfaces

An ***envelope*** of a 1-parameter family of surfaces is constructed in a similar way to that of the envelope of a 1-parameter family of lines given on page 175. The family is described by a differentiable function $F(x,y,z,t) = 0$, where t is a parameter. When t can be eliminated from the equations

$$F(x,y,z,t) = 0 \quad \text{and} \quad \frac{\partial F(x,y,z,t)}{\partial t} = 0,$$

we get the envelope, as a surface described implicitly.

To give a concrete example, consider two first-order polynomials

$$F_1(x,y,z) = ax + by + cz + d, \qquad F_2(x,y,z) = a'x + b'y + c'z + d',$$

and define

$$F(x,y,z,t) = t^2 F_2(x,y,z) + t\,F_1(x,y,z) + f, \tag{20.18}$$

where f is another constant. Then (20.18) is a 1-parameter family of planes, whose envelope is obtained by solving the system

$$\begin{cases} t^2F_2 + tF_1 + f &= 0, \\ 2tF_2 + F_1 &= 0, \end{cases}$$

and consists of the parabolic cylinder

$$F_1{}^2 = 4fF_2,$$

assuming that f is nonzero and $(a,b,c),(a',b',c')$ are not proportional. For an extensive discussion of envelopes, see [BrGi, chapter 5] and [Bolt].

Definition 20.10. *The envelope of a* 1*-parameter family* $t \mapsto S^2(t)$ *of spheres in* $\mathbb{R}^3$ *is called a* ***canal surface****. The curve formed by the centers of the spheres is called the* ***center curve*** *of the canal surface. The* ***radius*** *of the canal surface is the function* r *such that* $r(t)$ *is the radius of the sphere* $S^2(t)$*.*

It is easy to find one principal curvature and its corresponding principal curve for a canal surface.

Lemma 20.11. *Let $t \mapsto S^2(t)$ be the 1-parameter family of spheres that defines a canal surface $\mathcal{M}$. Then for each t the intersection $S^2(t) \cap \mathcal{M}$ is a circle and a principal curve on $\mathcal{M}$.*

Proof. Since $S^2(t)$ and $\mathcal{M}$ are tangent along $S^2(t) \cap \mathcal{M}$, the angle between their normals is zero. Furthermore, any curve on $S^2(t)$ is a principal curve, in particular, $S^2(t) \cap \mathcal{M}$. Theorem 15.6 on page 466 then implies that $S^2(t) \cap \mathcal{M}$ is a principal curve on $\mathcal{M}$. ∎

Now we can characterize canal surfaces.

Theorem 20.12. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a regular surface without umbilic points. The following are equivalent*:

(i) *$\mathcal{M}$ is a canal surface*;

(ii) *one of the systems of principal curves of $\mathcal{M}$ consists of circles*;

(iii) *one of the components of* $\mathsf{focal}(\mathcal{M})$ *is a curve.*

Proof. That (i) implies (ii) is a consequence of Lemma 20.11. The normal lines passing through points on any principal curve that is a circle meet at a point, since the Gauss image of such a principal circle is a great circle on the sphere. As the principal circles move on $\mathcal{M}$, their centers generate a curve, which must be one of the components of $\mathsf{focal}(\mathcal{M})$. Thus (ii) implies (iii). Finally, if (iii) holds, we can use the component of $\mathsf{focal}(\mathcal{M})$ that is a curve to generate a canal surface using the distance between $\mathsf{focal}(\mathcal{M})$ and $\mathcal{M}$ as the radius of the canal surface. Evidently, this canal surface coincides with $\mathcal{M}$. Thus (iii) implies (i). ∎

We can find a parametrization of a canal surface, provided we make mild assumptions about the center curve.

Theorem 20.13. *Suppose the center curve of a canal surface is a unit-speed curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ with nonzero curvature. Then the associated canal surface can be parametrized by the mapping*

$$\begin{aligned}\mathbf{y}(t,\theta) \;=\; \boldsymbol{\gamma}(t) + r(t)\Big(&-\mathbf{T}\,r'(t)\\ &+\sqrt{1-r'(t)^2}\big(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta\big)\Big),\end{aligned} \tag{20.19}$$

where $\mathbf{T}, \mathbf{N}, \mathbf{B}$ denote the tangent, normal, binormal of $\boldsymbol{\gamma}$.

Proof. Ignoring (20.19) for the moment, let $\mathbf{y}$ denote a patch that parametrizes the envelope of spheres defining the canal surface. Since the curvature of $\boldsymbol{\gamma}$ is nonzero, the Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is well defined, and we can write

$$\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t) = a(t,\theta)\mathbf{T}(t) + b(t,\theta)\mathbf{N}(t) + c(t,\theta)\mathbf{B}(t), \tag{20.20}$$

where a, b and c are differentiable on the interval on which $\boldsymbol{\gamma}$ is defined. We must have

$$\big\|\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t)\big\|^2 = r(t)^2, \tag{20.21}$$

expressing analytically the fact that $\mathbf{y}(t,\theta)$ lies on a sphere $S^2(t)$ of radius $r(t)$ centered at $\boldsymbol{\gamma}(t)$. Furthermore, $\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t)$ is a normal vector to the canal surface, so

$$\big(\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t)\big) \cdot \mathbf{y}_t \;=\; 0, \tag{20.22}$$

$$\big(\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t)\big) \cdot \mathbf{y}_\theta \;=\; 0. \tag{20.23}$$

Equations (20.22) and (20.23) assert that the vectors $\mathbf{y}_t$ and $\mathbf{y}_\theta$ are tangent to the sphere $S^2(t)$.

From (20.20) and (20.21) we get

$$\begin{cases} a^2 + b^2 + c^2 \;=\; r^2, \\ a\,a_t + b\,b_t + c\,c_t \;=\; rr'. \end{cases} \tag{20.24}$$

On the other hand, when we differentiate (20.20) with respect to t and use the Frenet formulas (7.12) on page 197, we obtain

$$\mathbf{y}_t = (1 + a_t - b\boldsymbol{\kappa})\mathbf{T} + (a\boldsymbol{\kappa} - c\boldsymbol{\tau} + b_t)\mathbf{N} + (c_t + b\boldsymbol{\tau})\mathbf{B}. \tag{20.25}$$

Then (20.24), (20.25), (20.20) and (20.22) imply that

$$a + rr' = 0, \tag{20.26}$$

and from (20.24) and (20.26) we get $b^2 + c^2 = r^2(1 - r'^2)$. By Lemma 1.23, page 17, we are free to write

$$b = -r\sqrt{1 - r'^2}\cos\theta \quad \text{and} \quad c = r\sqrt{1 - r'^2}\sin\theta,$$

for a suitable parameter θ. Thus (20.20) becomes

$$\mathbf{y}(t,\theta) - \boldsymbol{\gamma}(t) = -rr'\,\mathbf{T} - r\sqrt{1 - r'^2}\,\mathbf{N}\cos\theta + r\sqrt{1 - r'^2}\,\mathbf{B}\sin\theta, \tag{20.27}$$

which is equivalent to (20.19). ∎

The notion of canal surface is a very general concept. In the next three subsections we consider three important subclasses of canal surfaces.

Tubes as Canal Surfaces

It is easy to see that when the radius function $r(t)$ is constant, the definition of canal surface reduces to the definition (7.36) of a tube given on page 209. In fact, we can characterize tubes among all canal surfaces.

Theorem 20.14. *Let $\mathcal{M}$ be a canal surface. Then the following conditions are equivalent*:

(i) *$\mathcal{M}$ is a tube parametrized by* (20.19);

(ii) *the radius of $\mathcal{M}$ is constant*;

(iii) *the radius vector of each sphere in the family that defines the canal surface $\mathcal{M}$ meets the center curve orthogonally.*

Proof. When we compare (20.19) and (7.36), we see that (i) is equivalent to (ii). The radius vector of each sphere is the left-hand side of (20.27), from where it follows that (ii) is equivalent to (iii). ▌

Let us compute the first and second fundamental forms of a tube parametrized by (20.19) with $r' = 0$, so that

$$\mathbf{y}(t,\theta) = \boldsymbol{\gamma}(t) + r(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta).$$

We have

$$\mathbf{y}_\theta = r\,(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta),$$

and

$$\begin{aligned}\mathbf{y}_t &= \mathbf{T} - r(-\boldsymbol{\kappa}\,\mathbf{T} + \boldsymbol{\tau}\,\mathbf{B})\cos\theta + r(-\boldsymbol{\tau}\,\mathbf{N})\sin\theta\\ &= (1 + r\,\boldsymbol{\kappa}\cos\theta)\mathbf{T} - r\,\boldsymbol{\tau}(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta)\\ &= (1 + r\,\boldsymbol{\kappa}\cos\theta)\mathbf{T} - \boldsymbol{\tau}\,\mathbf{y}_\theta,\end{aligned}$$

using the Frenet formulas (7.12). Hence

$$\begin{cases} E &= \mathbf{y}_t \cdot \mathbf{y}_t = (1 + r\boldsymbol{\kappa}\cos\theta)^2 + r^2\boldsymbol{\tau}^2,\\ F &= \mathbf{y}_t \cdot \mathbf{y}_\theta = -r^2\boldsymbol{\tau},\\ G &= \mathbf{y}_\theta \cdot \mathbf{y}_\theta = r^2. \end{cases}$$

Furthermore,

$$\begin{aligned}\mathbf{y}_t \times \mathbf{y}_\theta &= (1 + r\,\boldsymbol{\kappa}\cos\theta)\mathbf{T} \times r(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta)\\ &= r(1 + r\,\boldsymbol{\kappa}\cos\theta)(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta),\end{aligned}$$

so that

$$\|\mathbf{y}_t \times \mathbf{y}_\theta\|^2 = r^2(1 + r\,\boldsymbol{\kappa}\cos\theta)^2.$$

We can therefore define a unit normal vector to $\mathbf{y}$ by

$$\mathbf{U} = -\mathbf{N}\cos\theta + \mathbf{B}\sin\theta. \tag{20.28}$$

Next, we compute

$$\begin{aligned}
\mathbf{y}_{\theta\theta} &= r\,(\mathbf{N}\cos\theta - \mathbf{B}\sin\theta),\\
\mathbf{y}_{\theta t} &= r\,(-\boldsymbol{\kappa}\mathbf{T}\sin\theta - \boldsymbol{\tau}\mathbf{N}\cos\theta + \boldsymbol{\tau}\mathbf{B}\sin\theta),\\
\mathbf{y}_{tt} &= (r\boldsymbol{\kappa}'\cos\theta)\mathbf{T} + (1 + r\boldsymbol{\kappa}\cos\theta)(\boldsymbol{\kappa}\mathbf{N}) - r\boldsymbol{\tau}'(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta)\\
&\qquad -r\boldsymbol{\tau}\big(\sin\theta(-\boldsymbol{\kappa}\mathbf{T} + \boldsymbol{\tau}\mathbf{B}) + \cos\theta(-\boldsymbol{\tau}\mathbf{N})\big)\\
&= \big(r\boldsymbol{\kappa}'\cos\theta - r\boldsymbol{\kappa}\boldsymbol{\tau}\sin\theta\big)\mathbf{T}\\
&\qquad +\big(\boldsymbol{\kappa} + r(\boldsymbol{\kappa}^2 + \boldsymbol{\tau}^2)\cos\theta - r\boldsymbol{\tau}'\sin\theta\big)\mathbf{N}\\
&\qquad -\big(r\boldsymbol{\tau}'\cos\theta + r\boldsymbol{\tau}^2\sin\theta\big)\mathbf{B}.
\end{aligned}$$

Therefore, relative to the choice (20.28),

$$\begin{cases}
e &= -\boldsymbol{\kappa}\cos\theta - r\,\boldsymbol{\kappa}^2\cos^2\theta - r\boldsymbol{\tau}^2,\\
f &= r\,\boldsymbol{\tau},\\
g &= -r.
\end{cases}$$

From equation (13.20) on page 400, we obtain

$$K = \frac{r\big(\boldsymbol{\kappa}\cos\theta + r\,\boldsymbol{\kappa}^2\cos^2\theta + r\,\boldsymbol{\tau}^2\big) - (r\,\boldsymbol{\tau})^2}{r^2(1 + r\,\boldsymbol{\kappa}\cos\theta)^2} = \frac{\boldsymbol{\kappa}\cos\theta}{r(1 + r\,\boldsymbol{\kappa}\cos\theta)}$$

and

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)} = -\frac{1}{2}\left(\frac{1}{r} + K\,r\right).$$

Hence

$$H^2 - K = \frac{1}{4}\left(r\,K - \frac{1}{r}\right)^2,$$

so that

$$\begin{cases}
k_1 &= H + \sqrt{H^2 - K} \;=\; -\dfrac{1}{r},\\[2ex]
k_2 &= H - \sqrt{H^2 - K} \;=\; \dfrac{-\boldsymbol{\kappa}\cos\theta}{1 + r\,\boldsymbol{\kappa}\cos\theta}.
\end{cases}$$

Surfaces of Revolution as Canal Surfaces

Almost any surface of revolution is a canal surface. The following theorem characterizes surfaces of revolution among all canal surfaces.

Theorem 20.15. *The center curve of a canal surface $\mathcal{M}$ is a straight line if and only if $\mathcal{M}$ is a surface of revolution for which no normal line to the surface is parallel to the axis of revolution. Furthermore, the parametrization* (20.19) *reduces to the standard parametrization* (15.1) *of a surface of revolution given on page* 462.

Proof. Let $\mathcal{M}$ be a surface of revolution and let $\mathbf{p} \in \mathcal{M}$. The normal line $\ell_{\mathbf{p}}$ perpendicular to $\mathcal{M}$ at $\mathbf{p}$ lies in the plane passing through the meridian of $\mathcal{M}$ containing $\mathbf{p}$; this plane contains the axis of revolution α of $\mathcal{M}$. Hence either $\ell_{\mathbf{p}}$ intersects α or is parallel to it. By hypothesis, the latter possibility is excluded. Let $\ell_{\mathbf{p}}$ intersect α at $\mathbf{q}$. The sphere $S^2(\mathbf{q})$ of radius $r(\mathbf{q}) = \|\mathbf{q} - \mathbf{p}\|$ and center $\mathbf{q}$ meets $\ell_{\mathbf{p}}$ perpendicularly at $\mathbf{p}$, and so it is tangent to $\mathcal{M}$ at $\mathbf{p}$. Hence $\mathcal{M}$ is the canal surface with center curve α and variable radius $\mathbf{q} \mapsto r(\mathbf{q})$.

Conversely, suppose that $\mathcal{M}$ is a canal surface whose center curve is a straight line $\boldsymbol{\gamma}$. Instead of a Frenet Frame field along $\boldsymbol{\gamma}$, we choose a constant frame field $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$. (See Exercise 5 of Chapter 7.) We can proceed exactly as in the proof of Theorem 20.13; the only difference is that $\boldsymbol{\kappa} = \boldsymbol{\tau} = 0$ for the center curve. In particular, (20.19) holds.

To show that $\mathcal{M}$ is a surface of revolution, we use a Euclidean motion to move $\mathcal{M}$ so that the center curve is the z-axis and so that

$$\mathbf{T} = (0, 0, 1), \qquad \mathbf{N} = (1, 0, 0), \qquad \mathbf{B} = (0, 1, 0).$$

We may also assume that $\boldsymbol{\gamma}(t) = (0, 0, t)$. Then (20.19) becomes

$$\big(- r(t)\sqrt{1 - r'(t)^2}\cos\theta,\ r(t)\sqrt{1 - r'(t)^2}\sin\theta,\ t - r(t)r'(t)\big).$$

This coincides with (15.1) when we take $\theta = -u$, $t = v$,

$$\varphi(v) = -r(v)\sqrt{1 - r'(v)^2} \qquad \text{and} \qquad \psi(v) = v - r(v)r'(v).$$

Hence $\mathcal{M}$ is a surface of revolution. ∎

There is a close connection between evolutes of plane curves and focal sets of surfaces of revolution:

Theorem 20.16. *Let $\mathscr{C}$ be a plane curve that is neither a straight line nor part of a circle. Then one of the components of the focal set of a surface of revolution $\mathcal{M}$ generated by $\mathscr{C}$ is the surface of revolution generated by the evolute of $\mathscr{C}$.*

Proof. Let $\mathscr{C}$ be parametrized by a unit-speed curve $\boldsymbol{\alpha}$, and write $\boldsymbol{\alpha} = (\varphi, \psi)$. The standard parametrization of $\mathcal{M}$ is

$$\mathbf{x}(u, v) = \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big).$$

For simplicity we assume that $\varphi > 0$. According to Corollary 15.14, page 470, the principal curvatures of $\mathcal{M}$ are given by

$$k_m = (\varphi''\psi' - \varphi'\psi'') \qquad \text{and} \qquad k_p = \frac{-\psi'}{\varphi}.$$

By hypothesis, k_m is not identically zero. By Lemma 15.10 on page 468, the unit normal is given by

$$\mathbf{U} = (\psi'\cos u,\ \psi'\sin u,\ -\varphi').$$

Hence the formula for the nondegenerate focal patch of $\mathcal{M}$ is

$$\begin{aligned}
\mathbf{z}_1(u, v) \ &= \ \mathbf{x}(u, v) + \rho_1 \mathbf{U}(u, v) \\
&= \big(\varphi(v)\cos u,\ \varphi(v)\sin u,\ \psi(v)\big) + \frac{\big(\psi'(v)\cos u,\ \psi'(v)\sin u,\ -\varphi'(v)\big)}{\varphi''(v)\psi'(v) - \varphi'(v)\psi''(v)} \\
&= \Bigg(\Big(\varphi(v) + \frac{\psi'(v)}{\varphi''(v)\psi'(v) - \varphi'(v)\psi''(v)}\Big)\cos u, \\
&\qquad \Big(\varphi(v) + \frac{\psi'(v)}{\varphi''(v)\psi'(v) - \varphi'(v)\psi''(v)}\Big)\sin u, \\
&\qquad\quad \psi(v) - \frac{\varphi'(v)}{\varphi''(v)\psi'(v) - \varphi'(v)\psi''(v)}\Bigg).
\end{aligned}$$

This is the surface of revolution generated by the curve

$$\boldsymbol{\gamma}(t) = \left(\varphi(t) + \frac{\psi'(t)}{\varphi''(v)\psi'(t) - \varphi'(t)\psi''(t)},\ \psi(t) - \frac{\varphi'(t)}{\varphi''(t)\psi'(t) - \varphi'(t)\psi''(t)}\right).$$

This curve is precisely the evolute of $\boldsymbol{\alpha}$, as we see by noting that

$$J\boldsymbol{\alpha}'(t) = \big(-\psi'(t),\ \varphi(t)\big) \qquad \text{and} \qquad \|\boldsymbol{\alpha}'(t)\| = 1,$$

and making use of formulas on pages 15 and 99. ∎

For example, we know from Section 4.2 that the evolute of a tractrix is a catenary curve. Hence the 2-dimensional component of the focal set of a pseudosphere is a catenoid. Figure 20.3 displays the pseudosphere together with this 2-dimensional component, corresponding to the plot of a tractrix and its evolute on page 103.

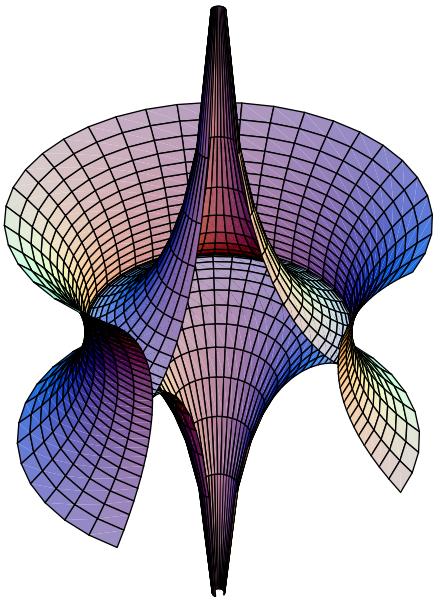


Figure 20.3: Pseudosphere and the catenoid component of its focal set

Canal Surfaces with a Planar Center Curve

The parametrization (20.19) may or may not be principal. We have already seen (Lemma 20.11) that any canal surface has a system of circles as principal curves. Let us prove this fact by direct calculation.

Lemma 20.17. *Let $\mathcal{M}$ be a canal surface parametrized by* (20.19)*. For fixed t, the circle $\theta \mapsto \mathbf{y}(t,\theta)$ is a principal curve with principal curvature $-1/r(t)$.*

Proof. Define $\mathbf{y}$ by equation (20.19). We can take for the unit normal of the canal surface the vector field

$$\mathbf{U} = \frac{\mathbf{y}-\boldsymbol{\gamma}}{\|\mathbf{y}-\boldsymbol{\gamma}\|} = \frac{1}{r}(\mathbf{y}-\boldsymbol{\gamma}). \tag{20.29}$$

Thus

$$\mathbf{y} = \boldsymbol{\gamma} + r\,\mathbf{U}. \tag{20.30}$$

When we differentiate (20.30) and use the fact that neither $\boldsymbol{\gamma}$ nor r depends on θ, we get

$$\mathbf{y}_\theta = r\,\mathbf{U}_\theta = -r\,S(\mathbf{y}_\theta), \tag{20.31}$$

where S is the shape operator of the canal surface. The lemma follows. ∎

Next, we characterize canal surfaces for which (20.19) *is* principal.

Theorem 20.18. *Let $\mathcal{M}$ be a canal surface parametrized by* (20.19). *Then the following conditions are equivalent.*

(i) *For each θ, the curve $t \mapsto \mathbf{y}(t,\theta)$ is a principal curve.*

(ii) *Either r is constant and the torsion of the center curve $\boldsymbol{\gamma}$ vanishes, or both the curvature and torsion of $\boldsymbol{\gamma}$ vanish.*

(iii) *The center curve $\boldsymbol{\gamma}$ is a plane curve, and either r is constant or $\boldsymbol{\gamma}$ is a straight line.*

Proof. Assume that (i) holds. We first use (20.27) and (20.29) to write unit normal $\mathbf{U}$ to the canal surface as

$$\mathbf{U} = -r'\mathbf{T} + \sqrt{1-r'^2}\,(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta), \tag{20.32}$$

so that

$$\mathbf{U}_\theta = \sqrt{1-r'^2}\,\big(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta\big). \tag{20.33}$$

Also, (20.32) and the Frenet formulas (7.12) imply that

$$\begin{aligned}\mathbf{U}_t \;&=\; -r''\mathbf{T} - r'(\boldsymbol{\kappa}\mathbf{N}) + \big(\sqrt{1-r'^2}\big)'\big(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta\big)\\ &\qquad +\sqrt{1-r'^2}\,\big(-(-\boldsymbol{\kappa}\mathbf{T} + \boldsymbol{\tau}\mathbf{B})\cos\theta + (-\boldsymbol{\tau}\mathbf{N})\sin\theta\big)\\ &=\; \big(-r'' + \boldsymbol{\kappa}\cos\theta\sqrt{1-r'^2}\big)\mathbf{T} - r'\boldsymbol{\kappa}\mathbf{N}\\ &\qquad +\big(\sqrt{1-r'^2}\big)'\big(-\mathbf{N}\cos\theta + \mathbf{B}\sin\theta\big) - \boldsymbol{\tau}\sqrt{1-r'^2}\big(\mathbf{N}\sin\theta + \mathbf{B}\cos\theta\big).\end{aligned}$$

Combined with (20.33), we get

$$\mathbf{U}_\theta \cdot \mathbf{U}_t = -r'\boldsymbol{\kappa}\sin\theta\sqrt{1-r'^2} - \boldsymbol{\tau}(1-r'^2). \tag{20.34}$$

On the other hand, when we differentiate (20.30) with respect to t, we find that

$$\mathbf{y}_t = \mathbf{T} + r'\mathbf{U} + r\,\mathbf{U}_t.$$

Then (20.30), (20.31), (20.33), (20.34) imply that

$$\begin{aligned}\mathbf{y}_t \cdot \mathbf{y}_\theta \;&=\; \big(\mathbf{T} + r'\mathbf{U} + r\,\mathbf{U}_t\big)\cdot(r\,\mathbf{U}_\theta) = r^2\mathbf{U}_t\cdot\mathbf{U}_\theta\\ &=\; -r^2\big(r'\boldsymbol{\kappa}\sin\theta\sqrt{1-r'^2} + \boldsymbol{\tau}(1-r'^2)\big).\end{aligned} \tag{20.35}$$

If $\mathbf{y}_t$ is principal, then it must be perpendicular to the principal vector $\mathbf{y}_\theta$. If this happens, both terms on the right-hand side of (20.35) vanish, since the first term depends on θ and the second does not. Thus $\boldsymbol{\tau} = 0$ and either $r' = 0$ or $\boldsymbol{\kappa} = 0$. Thus (i) implies (ii). Conversely, if (ii) holds, then (20.35) implies that $\mathbf{y}_t$ is perpendicular to $\mathbf{y}_\theta$, and so itself principal. Finally, the equivalence of (ii) and (iii) follows from results in Section 7.2. ∎

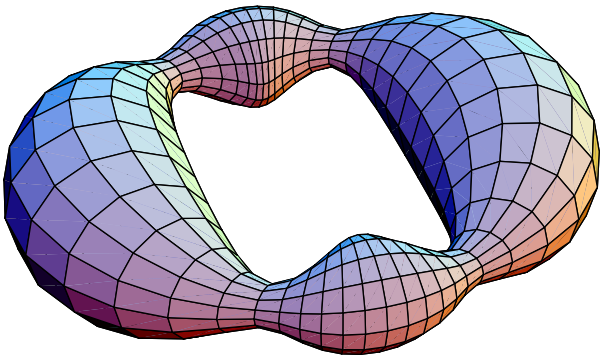


Figure 20.4: The canal surface generated by a sphere of radius $1.8 + 0.48\sin(2t)$ moving on a circle of radius 5

Figure 20.4 shows the sort of surface that is formed as the envelope of spheres with varying radius. Since its center curve is not a straight line, the parametrization (20.19) is not principal in this case.

20.3 Cyclides of Dupin via Focal Sets

In the early 1800s, a class of surfaces all of whose lines of curvature are circles or lines was studied by Charles Dupin. Not only does this class include cones, cylinders and tori, but also distinct new surfaces. Dupin called such surfaces "cyclides" in his book [Dupin2], and the name seems to have been coined by him. Since there is now a more general notion of cyclide, those studied by Dupin are referred to as ***cyclides of Dupin***. Three methods are available for their study:

(i) Dupin's original method consisted of describing a cyclide as an envelope of spheres tangent to three given spheres.

(ii) Liouville [Liou] showed that any cyclide of Dupin is the image under an *inversion* (see page 662) of a torus, circular cylinder or circular cone.

(iii) Maxwell[1] [Max] gave a direct construction of a cyclide of Dupin as the envelope of spheres whose centers move along ellipses, hyperbolas or parabolas, generalizing the torus of revolution.



[1] James Clerk Maxwell (1831–1879). A leading scientist of the 19th century, celebrated for his *Treatise on Electricity and Magnetism*. The paper [Max] is informative and elegant. Before giving parametrizations of cyclides, Maxwell characterized them by a geometrical argument using almost no algebra. The paper also includes images viewable by a stereoscope together with Maxwell's recommendations for constructing the stereoscope.

In this section we follow the approach of Maxwell, as modified by Darboux. Our treatment is based on [Darb4, pages 405–406], though cyclides are also discussed in [CeRy, page 150–166], [Bomp], [Sny], and [Cool1]. The theory of inversion is developed in Sections 20.4 and 20.5.

Firstly, we characterize focal sets consisting of two curves.

Theorem 20.19. *Let $\mathcal{M}$ be a surface for which the focal set* $\mathsf{focal}(\mathcal{M})$ *consists of two curves. Then each curve is a conic section (an ellipse, hyperbola, parabola, or straight line), and the planes of each component are perpendicular to one another.*

Proof. We denote the two curves by $\mathsf{focal}(\mathcal{M})_1$ and $\mathsf{focal}(\mathcal{M})_2$, and start by considering a point $\mathbf{p} \in \mathsf{focal}(\mathcal{M})_1$. The set of normal lines to $\mathcal{M}$ that pass through $\mathbf{p}$ forms a cone, which we call $\mathsf{cone}(\mathbf{p}, \mathcal{M})$, with $\mathbf{p}$ as its vertex. By definition of focal set, each line in this cone also passes through a point of $\mathsf{focal}(\mathcal{M})_2$; furthermore, for each point $\mathbf{q} \in \mathsf{focal}(\mathcal{M})_2$, the line containing $\mathbf{p}$ and $\mathbf{q}$ is in $\mathsf{cone}(\mathbf{p}, \mathcal{M})$. It follows that the whole curve $\mathsf{focal}(\mathcal{M})_2$ must lie on $\mathsf{cone}(\mathbf{p}, \mathcal{M})$.

This construction creates a cone for any $\mathbf{p} \in \mathsf{focal}(\mathcal{M})_1$. In particular, it works for a point that is nearest to $\mathsf{focal}(\mathcal{M})_2$. For such a point, the corresponding cone is part of a plane. It follows that $\mathsf{focal}(\mathcal{M})_2$ is a plane curve.

Clearly, the same argument shows that $\mathsf{focal}(\mathcal{M})_1$ is also a plane curve lying on a cone. Hence both $\mathsf{focal}(\mathcal{M})_1$ and $\mathsf{focal}(\mathcal{M})_2$ are conic sections. Indeed, $\mathsf{cone}(\mathbf{p}, \mathcal{M})$ must be a circular cone for $\mathbf{p} \in \mathsf{focal}(\mathcal{M})_1$, since it contains a principal curve of $\mathcal{M}$, which is a circle. ∎

We are now in a position to define the cyclides of Dupin and give explicit parametrizations of them.

Elliptic-Hyperbolic Cyclides

Let ell_{xy} be the ellipse in the xy-plane given by

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1,$$

and let hyp_{xz} be the hyperbola in the xz-plane given by

$$\frac{x^2}{c^2} - \frac{z^2}{a^2 - c^2} = 1.$$

We assume that $a > c > 0$. Notice that one focus of the ellipse is $(c, 0, 0)$, which lies on the hyperbola, whereas one focus of the hyperbola is $(a, 0, 0)$, which lies on the ellipse.

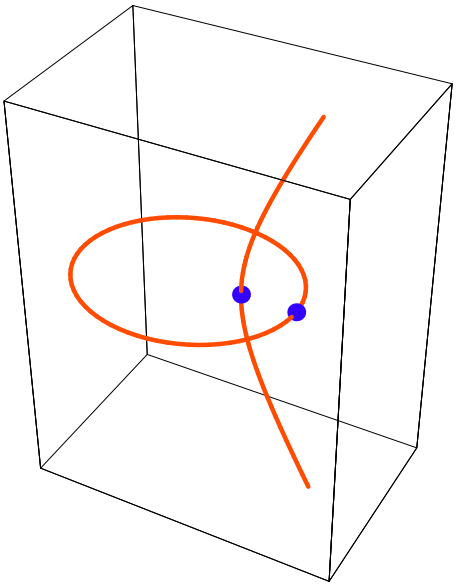


Figure 20.5: Ellipse and hyperbola in orthogonal planes

We first derive a simple formula for the distance between a point on the ellipse and a point on the hyperbola.

Lemma 20.20. *The distance between a point* $(x_1, y_1, 0)$ *on* ell_{xy} *and a point* $(x_2, 0, z_2)$ *on the* hyp_{xz} *is given by the formula*

$$\mathsf{distance}\big((x_1,y_1,0),(x_2,0,z_2)\big) = \Big|\frac{c}{a}x_1 - \frac{a}{c}x_2\Big|. \tag{20.36}$$

Proof. We compute

$$\begin{aligned}\mathsf{distance}^2\big((x_1,y_1,0),(x_2,0,z_2)\big) &= (x_1-x_2)^2 + y_1^2 + z_2^2\\ &= (x_1-x_2)^2 + (a^2-c^2)\left(1 - \frac{x_1^2}{a^2} + \frac{x_2^2}{c^2} - 1\right)\\ &= x_1^2 - 2x_1x_2 + x_2^2 - x_1^2 + \frac{c^2}{a^2}x_1^2 + \frac{a^2}{c^2}x_2^2 - x_2^2 = \left(\frac{c}{a}x_1 - \frac{a}{c}x_2\right)^2,\end{aligned}$$

as required. ∎

Next, we parametrize ell_{xy} and hyp_{xz} by

$$\begin{cases}\boldsymbol{\alpha}(u) &= \big(a\cos u,\ \sqrt{a^2-c^2}\sin u,\ 0\big),\\ \boldsymbol{\beta}(v) &= \big(c\sec v,\ 0,\ \sqrt{a^2-c^2}\tan v\big),\end{cases} \tag{20.37}$$

and we restrict to the branch of hyp_{xz} for which $-\pi/2 < v < \pi/2$, so as to state

Lemma 20.21. *Let* $\mathsf{sph}_1(u)$ *be a sphere with radius* $\rho_1(u)$ *centered at* $\boldsymbol{\alpha}(u)$ *and let* $\mathsf{sph}_2(v)$ *be a sphere with radius* $\rho_2(v)$ *centered at* $\boldsymbol{\beta}(v)$. *Suppose that* $\mathsf{sph}_1(u)$ *and* $\mathsf{sph}_2(v)$ *are tangent to each other. Then there exists* k, *which is constant with respect to* u *and* v, *such that*

$$\rho_1(u) = -c\cos u + k \qquad \textit{and} \qquad \rho_2(v) = a\sec v - k. \tag{20.38}$$

Proof. By hypothesis, the two spheres have a unique point in common, say (x_3, y_3, z_3). Then

$$\begin{cases} \mathsf{distance}\,\big((x_1, y_1, 0), (x_3, y_3, z_3)\big) \;=\; \rho_1(u), \\ \mathsf{distance}\,\big((x_2, 0, z_2), (x_3, y_3, z_3)\big) \;=\; \rho_2(v). \end{cases} \tag{20.39}$$

Since the spheres are tangent at (x_3, y_3, z_3),

$$\rho_1(u) + \rho_2(v) = \mathsf{distance}\,\big((x_1, y_1, 0), (x_2, 0, z_2)\big),$$

and Lemma 20.20 implies that this distance equals

$$\left|\frac{c}{a}x_1 - \frac{a}{c}x_2\right| = |c\cos u - a\sec v| = a\sec v - c\cos u.$$

Thus,

$$\rho_1(u) + c\cos u = -\rho_2(v) + a\sec v. \tag{20.40}$$

Since the left-hand side of (20.40) depends only on u, while the right-hand side of (20.40) depends only on v, we let k be the common value, and obtain (20.38). ∎

Let us now prove the converse of Lemma 20.21.

Lemma 20.22. *Let* $\mathsf{sph}_1(u)$ *be a sphere with radius* $\rho_1(u)$ *centered at* $\boldsymbol{\alpha}(u)$, *and let* $\mathsf{sph}_2(v)$ *be a sphere with radius* $\rho_2(v)$ *centered at* $\boldsymbol{\beta}(v)$. *Suppose that* $\rho_1(u)$ *and* $\rho_2(v)$ *are given by* (20.38) *for some* k. *Let* (x_3, y_3, z_3) *be the point on the line connecting the points* $\boldsymbol{\alpha}(u)$ *and* $\boldsymbol{\beta}(v)$ *at a distance* $\rho_1(u)$ *from* $\boldsymbol{\alpha}(u)$. *Then the distance between the points* (x_3, y_3, z_3) *and* $\boldsymbol{\beta}(v)$ *is* $\rho_2(v)$. *Furthermore,* $\mathsf{sph}_1(u)$ *and* $\mathsf{sph}_2(v)$ *are mutually tangent at* (x_3, y_3, z_3).

Proof. By Lemma 20.20, the distance from $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}(v)$ is

$$\begin{aligned} \mathsf{distance}\,\big((x_1, y_1, 0), (x_2, 0, z_2)\big) \;=\; \left|\frac{c}{a}x_1 - \frac{a}{c}x_2\right| \;&=\; |c\cos u - a\sec v| \\ &=\; a\sec v - c\cos u, \end{aligned}$$

which is precisely $\rho_1(u) + \rho_2(v)$. Therefore, (20.39) holds. This can only happen if $\mathsf{sph}_1(u)$ and $\mathsf{sph}_2(v)$ are tangent at (x_3, y_3, z_3). ∎

As $\mathsf{sph}_1(u)$ moves along the ellipse $\boldsymbol{\alpha}(u)$ it traces out a surface, the envelope of the spheres. This envelope must coincide with the envelope formed by $\mathsf{sph}_2(v)$ as it moves along a branch of the hyperbola $\boldsymbol{\beta}(v)$. The resulting surface is called a ***cyclide of Dupin of elliptic-hyperbolic type***. It depends on three parameters a, c and k. In order to parametrize the cyclide, we define

$$\mathbf{x}(u,v) = \boldsymbol{\alpha}(u) + \rho_1(u)V(u,v) = \boldsymbol{\beta}(v) - \rho_2(v)\mathbf{V}(u,v),$$

where $\mathbf{V}(u,v)$ is a unit vector. We have

$$0 = \boldsymbol{\alpha}(u) - \boldsymbol{\beta}(v) + \big(\rho_1(u) + \rho_2(v)\big)\mathbf{V}(u,v),$$

whence

$$\mathbf{x}(u,v) \;=\; \boldsymbol{\alpha}(u) - \rho_1(u)\left(\frac{\boldsymbol{\alpha}(u) - \boldsymbol{\beta}(v)}{\rho_1(u) + \rho_2(v)}\right) \;=\; \frac{\rho_2(v)\boldsymbol{\alpha}(u) + \rho_1(u)\boldsymbol{\beta}(v)}{\rho_1(u) + \rho_2(v)}.$$

Substituting in (20.37) and (20.38), and multiplying top and bottom by $\cos v$ gives

$$\mathbf{x}(u,v) = \left(\frac{c(k - c\cos u) + a\cos u(a - k\cos v)}{a - c\cos u\cos v}, \right.$$
$$\left. \frac{\sqrt{a^2 - c^2}(a - k\cos v)\sin u}{a - c\cos u\cos v}, \; \frac{\sqrt{a^2 - c^2}(k - c\cos u)\sin v}{a - c\cos u\cos v}\right).$$

Figures 20.6–20.8 display one of each of the three classes designated by

Definition 20.23. *An elliptic-hyperbolic cyclide $\mathcal{M}$ is called a*

ring cyclide	*if it has*	*no self-intersections,*
spindle cyclide	*if it has*	1 *self-intersection,*
horn cyclide	*if it has*	2 *self-intersections.*

Notice also that generically an elliptic-hyperbolic cyclide has two planes of symmetry. If there are additional planes of symmetry, an elliptic-hyperbolic cyclide becomes a torus.

Parabolic Cyclides

There is another type of cyclide whose focal set consist of two parabolas. Let par_{xz} be the parabola in the xz-plane given by

$$z = -\frac{x^2}{8a} + a,$$

and let par_{yz} be the parabola in the yz-plane given by

$$z = \frac{y^2}{8a} - a.$$

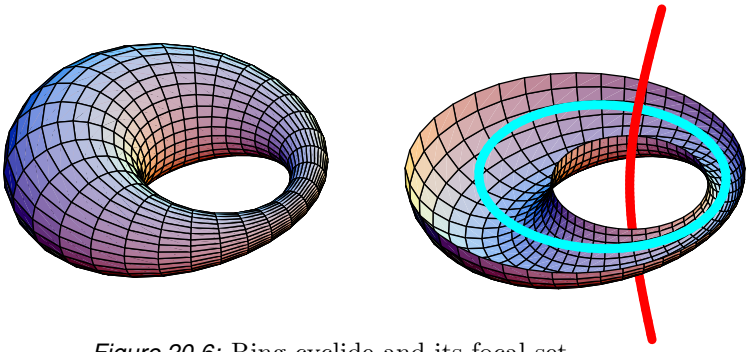


Figure 20.6: Ring cyclide and its focal set

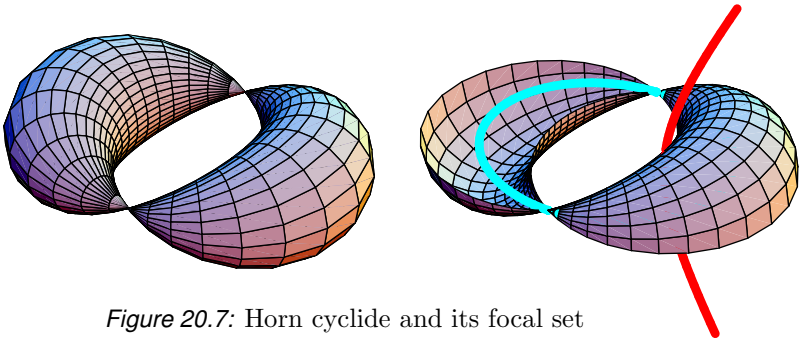


Figure 20.7: Horn cyclide and its focal set

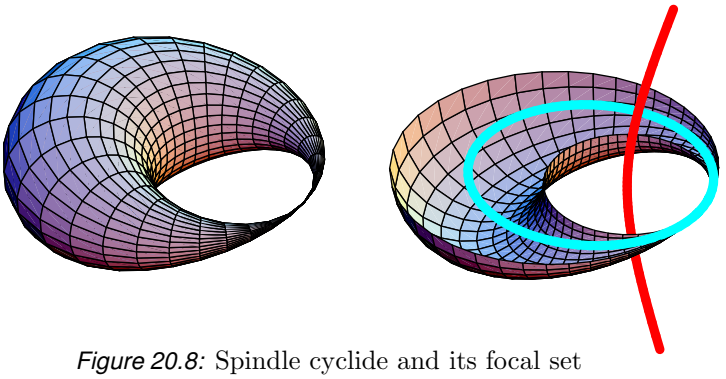


Figure 20.8: Spindle cyclide and its focal set

We parametrize par_{xz} and par_{yz} by

$$\text{(20.41)} \qquad \begin{cases} \boldsymbol{\alpha}(u) \;=\; \Big(u,\, 0,\, -\dfrac{u^2}{8a} + a\Big), \\ \boldsymbol{\beta}(v) \;=\; \Big(0,\, -v,\, \dfrac{v^2}{8a} - a\Big). \end{cases}$$

Note that $\boldsymbol{\alpha}$ passes through the focus of $\boldsymbol{\beta}$ (namely $(0,0,a)$) and $\boldsymbol{\beta}$ passes through the focus of $\boldsymbol{\alpha}$ (namely $(0,0,-a)$). The analog of Lemma 20.21 is

Lemma 20.24. *Let* $\mathsf{sph}_1(u)$ *be a sphere with radius* $\rho_1(u)$ *centered at* $\boldsymbol{\alpha}(u)$*, and let* $\mathsf{sph}_2(v)$ *be a sphere with radius* $\rho_2(v)$ *centered at* $\boldsymbol{\beta}(v)$*. Suppose that* $\mathsf{sph}_1(u)$ *and* $\mathsf{sph}_2(v)$ *are tangent to each other. Then there exists* k*, which is constant with respect to* u *and* v*, such that*

$$\rho_1(u) = \frac{u^2}{8a} + a + k \qquad \textit{and} \qquad \rho_2(v) = \frac{v^2}{8a} + a - k.$$

Proof. This is very similar to that of Lemma 20.21. Let (x_3, y_3, z_3) be the point of intersection of the two spheres. Then

$$\begin{aligned} \big\|\boldsymbol{\alpha}(u) - \boldsymbol{\beta}(v)\big\|^2 \;&=\; \left\|\left(u,\, v,\, 2a - \frac{u^2}{8a} - \frac{v^2}{8a}\right)\right\|^2 \\ &=\; \frac{1}{2}(u^2+v^2) + \frac{1}{64a^2}(u^2+v^2)^2 + 4a^2 \\ &=\; \left(\frac{u^2+v^2+16a^2}{8a}\right)^2. \end{aligned}$$

Furthermore,

$$\rho_1(u) + \rho_2(v) = \mathsf{distance}\,\big((x_1,0,z_1),(0,y_2,z_2)\big) = \big\|\boldsymbol{\alpha}(u) - \boldsymbol{\beta}(v)\big\|,$$

so that

$$\rho_1(u) - \frac{u^2}{8a} - a = -\rho_2(v) + \frac{v^2}{8a} + a.$$

Just as with (20.40), both sides must equal a constant k. ∎

The equation of a parabolic cyclide can be found in Notebook 20.

20.4 The Definition of Inversion

We want to invert $\mathbb{R}^n$ with respect to a point $\mathbf{q} \in \mathbb{R}^n$; this means that points near to $\mathbf{q}$ are mapped far away in a reasonable way, and vice versa.

Definition 20.25. *The* ***inversion*** *of* $\mathbb{R}^n$ *with respect to a point* $\mathbf{q} \in \mathbb{R}^n$ *with* ***inversion radius*** ρ *is the map* $\Psi\colon \mathbb{R}^n \setminus \{\mathbf{q}\} \to \mathbb{R}^n \setminus \{\mathbf{q}\}$ *given by*

$$\Psi(\mathbf{p}) = \mathbf{q} + \rho^2 \frac{\mathbf{p}-\mathbf{q}}{\|\mathbf{p}-\mathbf{q}\|^2}. \tag{20.42}$$

We call $\mathbf{q}$ *the* ***inversion center*** *or* ***pole****.*

In the special case $\mathbf{q} = 0$, we see that $\Psi(\mathbf{p}) = \rho^2\mathbf{p}/\|\mathbf{p}\|^2$.

It is easy to check that two applications of an inversion lead back to the same point; that is, $\Psi \circ \Psi$ is the identity map of $\mathbb{R}^n \setminus \mathbf{q}$. This has the important consequence that (20.42) is a *bijective* mapping. Here is another important property:

Lemma 20.26. *Let* Ψ *be an inversion with pole* $\mathbf{q} \in \mathbb{R}^n$*. Then* Ψ *maps the sphere*

$$S^{n-1}(\mathbf{p}, a) = \{\mathbf{v} \in \mathbb{R}^n \mid \|\mathbf{v}-\mathbf{p}\| = a\}$$

of dimension $n-1$*, radius* a *and center* $\mathbf{p}$ *to another sphere, or a plane.*

Proof. Without loss of generality, we may suppose that $\mathbf{q} = \mathbf{0}$ is the origin. In an attempt to show that Ψ maps the sphere onto a sphere whose center lies on the line straight line joining $\mathbf{0}$ to $\mathbf{p}$, let $\mathbf{v} \in S^{n-1}(\mathbf{p}, a)$ and consider

$$\begin{aligned} \|\Psi(\mathbf{v}) - c\mathbf{p}\|^2 &= \left\| \frac{\rho^2}{\|\mathbf{v}\|^2}\mathbf{v} - c\mathbf{p} \right\|^2 \\ &= \rho^4 \frac{1}{\|\mathbf{v}\|^2} - \frac{2\rho^2 c}{\|\mathbf{v}\|^2}\mathbf{v}\cdot\mathbf{p} + c^2\|\mathbf{p}\|^2, \end{aligned} \tag{20.43}$$

where c is to be determined. Since

$$a^2 = \|\mathbf{v}-\mathbf{p}\|^2 = \|\mathbf{v}\|^2 - 2\mathbf{v}\cdot\mathbf{p} + \|\mathbf{p}\|^2, \tag{20.44}$$

we may eliminate $\mathbf{v}\cdot\mathbf{p}$ to find that (20.43) equals

$$\frac{\rho^2}{\|\mathbf{v}\|^2}\left(\rho^2 - c\|\mathbf{p}\|^2 + ca^2\right) - c\rho^2 + c^2\|\mathbf{p}\|^2.$$

To make the coefficient of $1/\|\mathbf{v}\|^2$ vanish, we set

$$c = \frac{\rho^2}{\|\mathbf{p}\|^2 - a^2},$$

assuming for the moment that $\|\mathbf{p}\| \neq a$. In this case, Ψ maps the sphere onto a sphere with center $c\mathbf{p}$.

If $\|\mathbf{p}\| = a$ then the sphere passes through the inversion center $\mathbf{0}$. In this case, (20.44) implies that $\|\mathbf{v}\|^2 = 2\mathbf{p}\cdot\mathbf{v}$, whence

$$\Psi(\mathbf{v})\cdot\mathbf{p} = \frac{\rho^2}{2},$$

which is the equation of a plane. ∎

An inversion is characterized by two properties.

Lemma 20.27. *Let $\mathbf{q} \in \mathbb{R}^n$. A map $\Phi\colon \mathbb{R}^n \setminus \{\mathbf{q}\} \to \mathbb{R}^n \setminus \{\mathbf{q}\}$ is an inversion with inversion radius ρ if and only if there is a number $\lambda > 0$ such that*

$$\|\Phi(\mathbf{p}) - \mathbf{q}\|\,\|\mathbf{p} - \mathbf{q}\| = \rho^2 \quad \textit{and} \quad \Phi(\mathbf{p}) - \mathbf{q} = \lambda(\mathbf{p} - \mathbf{q}) \tag{20.45}$$

for all $\mathbf{p}$.

Proof. It is easy to prove that an inversion satisfies (20.45). Conversely, suppose that Ψ satisfies (20.45). Then $\lambda\|\mathbf{p} - \mathbf{q}\|^2 = \rho^2$, and so

$$\Psi(\mathbf{p}) - \mathbf{q} = \rho^2 \frac{\mathbf{p} - \mathbf{q}}{\|\mathbf{p} - \mathbf{q}\|^2}. \quad \blacksquare$$

Because of Lemma 20.27, an inversion of $\mathbb{R}^n$ is sometimes called a ***transformation by reciprocal radii*** in the classical literature.

Next, we show that an inversion has the important property that its tangent map preserves angles between tangent vectors. For this, we make the following definition, which corresponds to the notion of conformality for surface mappings given on page 370.

Definition 20.28. *Let $\Phi\colon \mathcal{U} \to \Phi(\mathcal{U}) \subset \mathbb{R}^n$ be a map, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. We say that Φ is a* ***conformal map*** *of $\mathcal{U}$ onto $\Phi(\mathcal{U})$ provided there is a positive differentiable function $\lambda\colon \mathcal{U} \to \mathbb{R}$ such that*

$$\|\Phi_*(\mathbf{v_p})\| = \lambda(\mathbf{p})\|\mathbf{v_p}\| \tag{20.46}$$

for all $\mathbf{p} \in \mathcal{U}$.

Proposition 20.29. *An inversion is a conformal map.*

Proof. Let Ψ be the inversion defined by (20.42), and let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^n$ be a curve passing through $\mathbf{q}$. Then

$$(\Psi \circ \boldsymbol{\alpha})(t) = \mathbf{q} + \rho^2 \frac{\boldsymbol{\alpha}(t) - \mathbf{q}}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^2}$$

for $a < t < b$. Differentiating the last equation gives

$$(\Psi \circ \boldsymbol{\alpha})'(t) = \frac{\rho^2 \boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^2} - \frac{2\rho^2\big((\boldsymbol{\alpha}(t) - \mathbf{q}) \cdot \boldsymbol{\alpha}'(t)\big)\big(\boldsymbol{\alpha}(t) - \mathbf{q}\big)}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^4}.$$

It follows that

$$\begin{aligned} \big\|(\Psi \circ \boldsymbol{\alpha})'(t)\big\|^2 &= \left\| \frac{\rho^2 \boldsymbol{\alpha}'(t)}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^2} - \frac{2\rho^2\big((\boldsymbol{\alpha}(t) - \mathbf{q}) \cdot \boldsymbol{\alpha}'(t)\big)\big(\boldsymbol{\alpha}(t) - \mathbf{q}\big)}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^4} \right\|^2 \\ &= \frac{\rho^4 \big\|\boldsymbol{\alpha}'(t)\big\|^2}{\big\|\boldsymbol{\alpha}(t) - \mathbf{q}\big\|^4}. \end{aligned} \tag{20.47}$$

A geometrical explanation of why the norm squared of $(\Psi \circ \boldsymbol{\alpha})'(t)$ simplifies dramatically is given on page 666. Since any tangent vector can be realized as the velocity vector of a curve, we can replace $\boldsymbol{\alpha}(t)$ by $\mathbf{p}$ and $\boldsymbol{\alpha}'(t)$ by $\mathbf{v_p}$, so that

$$\|\Psi_*(\mathbf{v_p})\| = \frac{\rho^2 \|\mathbf{v_p}\|}{\|\mathbf{p}-\mathbf{q}\|^2}$$

for all tangent vectors $\mathbf{v_p}$. Hence Ψ distorts lengths of all tangent vectors at $\mathbf{p}$ by the same factor

$$\lambda(p) = \frac{\rho^2}{\|\mathbf{p}-\mathbf{q}\|^2},$$

in accordance with (20.46). ▌

We shall mainly be interested in finding inversions of *surfaces* in $\mathbb{R}^3$, but we conclude this section by inverting some plane curves.

Inversion provides a useful way to visualize what happens to a curve at infinity. For example, Figure 20.9 compares inversions (with respect to the origin) of a catenary and one branch of a hyperbola. Although the shape of a central portion of a hyperbola is not so different from that of a catenary, the "inverse curves" are themselves quite different near the origin.

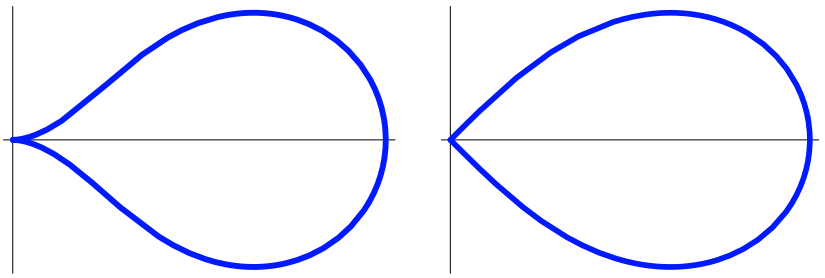


Figure 20.9: Inverse curve of a catenary (left) and a hyperbola (right)

Similar principles apply to space curves. Figure 20.10 shows an inversion of the twisted cubic, defined on page 202, with pole indicated by the black bead. The fact that the inverse curve rapidly "closes up" indicates that the twisted cubic can be extended to a closed curve in projective space (see forward to page 793). Similar remarks apply to many other space curves that we have studied, including the helix whose inversion is shown in Figure 20.15 on page 671. That figure also serves to outline the result of inverting a cylinder.

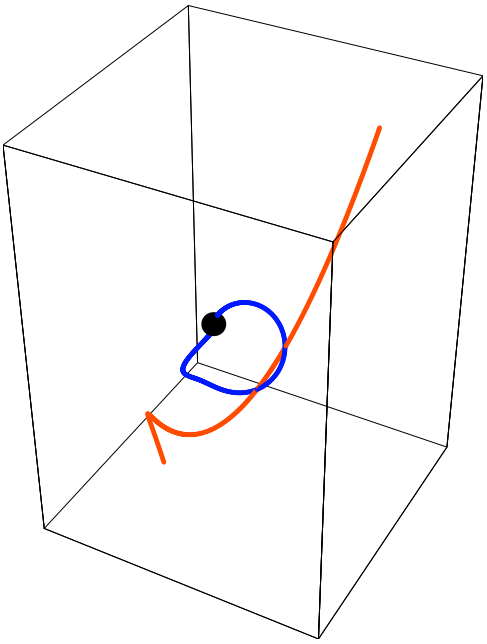


Figure 20.10: Inverse of the twisted cubic with pole $(0, 1, 1)$

20.5 Inversion of Surfaces

The relevance of the present topic to our recent study of surfaces is perhaps best illustrated by the next result.

Proposition 20.30. *Inversion maps principal curves onto principal curves.*

Proof. The mapping Ψ in (20.42) is conformal by Proposition 20.29. It therefore preserves orthogonality, and maps a triply orthogonal system of surfaces onto another triply orthogonal system of surfaces. The proposition is now a consequence of Dupin's Theorem on page 609 and the general construction of Theorem 19.30. ∎

This fact can be used to construct many interesting parametrizations of surfaces. For example, Figure 20.11 shows the inverse of an ellipsoid with respect to a point whose position is approximately represented by the midpoint of the top edge of the book's page.

Definition 20.31. *Let* $\mathbf{x}:\mathcal{U} \to \mathbb{R}^3$*, and let* Ψ *be the mapping* (20.42)*. The* ***inverse patch*** *of* $\mathbf{x}$ *with respect to* $\mathbf{q} \in \mathbb{R}^n$ *with radius of inversion* ρ *is simply the patch* $\Psi \circ \mathbf{x}$*.*

In other words, the inverse patch is the image of the original one by inversion.

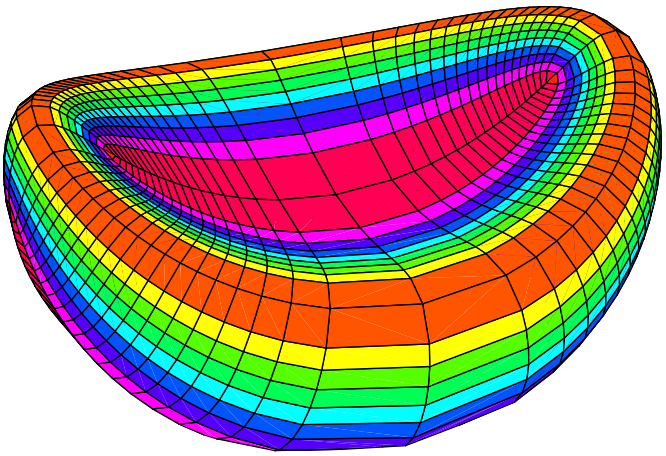


Figure 20.11: Inverse of an ellipsoid with lines of curvature

We proceed to obtain formulas for the first and second fundamental forms of an inverse patch. The notion of inversion is such a natural one that these formulas can be obtained without excessive computation, in a similar spirit to that of Section 19.9. We first observe that the tangent vectors of an inverse patch are obtained by *reflection* in a mirror perpendicular to $\mathbf{x} - \mathbf{q}$.

Let $\mathbf{y} = \Psi \circ \mathbf{x}$ be the inverse patch. Differentiating (20.42), we get

$$\mathbf{y}_u = \frac{\rho^2 \mathbf{x}_u}{\|\mathbf{x}-\mathbf{q}\|^2} - \frac{2\rho^2\big(\mathbf{x}_u \cdot (\mathbf{x}-\mathbf{q})\big)(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4}, \tag{20.48}$$

with a similar formula for $\mathbf{x}_v$. Using the notation of (5.1) on page 128,

$$\mathbf{y}_u = \mathsf{refl}_{\mathbf{x}-\mathbf{q}}\Big(\frac{\rho^2 \mathbf{x}_u}{\|\mathbf{x}-\mathbf{q}\|^2}\Big), \quad \mathbf{y}_v = \mathsf{refl}_{\mathbf{x}-\mathbf{q}}\Big(\frac{\rho^2 \mathbf{x}_v}{\|\mathbf{x}-\mathbf{q}\|^2}\Big).$$

We give two immediate corollaries of these formulas.

Corollary 20.32. *Let $\mathbf{x}$ be a patch in $\mathbb{R}^n$, and let $\mathbf{y}$ be the inverse patch of $\mathbf{x}$ with respect to an inversion with center $\mathbf{q} \in \mathbb{R}^n$ and radius ρ. Then the coefficients of the first fundamental form of $\mathbf{x}$ are related to those of $\mathbf{y}$ by*

$$E(\mathbf{y}) = \frac{\rho^4}{\|\mathbf{x}-\mathbf{q}\|^4} E(\mathbf{x}), \quad F(\mathbf{y}) = \frac{\rho^4}{\|\mathbf{x}-\mathbf{q}\|^4} F(\mathbf{x}), \quad G(\mathbf{y}) = \frac{\rho^4}{\|\mathbf{x}-\mathbf{q}\|^4} G(\mathbf{x}).$$

Proof. These equations are consequences of the fact that a reflection preserves the norm of a vector. ∎

The second is a formula for the unit normal to an inverse patch.

Corollary 20.33. *Let* $\mathbf{x}$ *be a patch in* $\mathbb{R}^3$*, and let* $\mathbf{y}$ *be the inverse patch of* $\mathbf{x}$ *with respect to* $\mathbf{q} \in \mathbb{R}^3$ *with radius of inversion* ρ*. Denote by* $\mathbf{U_x}$ *and* $\mathbf{U_y}$ *the respective normals. Then*

$$\mathbf{U_y} = -\mathsf{refl}_{\mathbf{x}-\mathbf{q}}\big(\mathbf{U_x}\big) = -\mathbf{U_x} + \frac{2\big(\mathbf{U_x}\cdot(\mathbf{x}-\mathbf{q})\big)(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^2}. \tag{20.49}$$

Proof. Let $\mathsf{refl} = \mathsf{refl}_{\mathbf{x}-\mathbf{q}}$ denote the reflection that figured in Corollary 20.32. Since reflections in $\mathbb{R}^3$ are orientation-reversing,

$$\mathsf{refl}(\mathbf{x}_u \times \mathbf{x}_v) = -\mathsf{refl}(\mathbf{x}_u) \times \mathsf{refl}(\mathbf{x}_v),$$

and more to the point,

$$\mathbf{y}_u \times \mathbf{y}_v = -\mathsf{refl}\Big(\frac{\rho^2\,\mathbf{x}_u}{\|\mathbf{x}_u-\mathbf{q}\|^2} \times \frac{\rho^2\,\mathbf{x}_v}{\|\mathbf{x}_v-\mathbf{q}\|^2}\Big).$$

The result follows from the fact that $\mathbf{U_x}$ and $\mathbf{U_y}$ are proportional to $\mathbf{x}_u \times \mathbf{x}_v$ and $\mathbf{y}_u \times \mathbf{y}_v$ respectively, and that a reflection preserves norm. ∎

We are now in a position to find the coefficients of the second fundamental form of an inverse patch.

Lemma 20.34. *Let* $\mathbf{x}$ *be a patch in* $\mathbb{R}^3$*, and let* $\mathbf{y}$ *be the inverse patch of* $\mathbf{x}$ *with respect to* $\mathbf{q} \in \mathbb{R}^3$ *with radius of inversion* ρ*. Then the coefficients of the second fundamental form of* $\mathbf{y}$ *are related to those of* $\mathbf{x}$ *by*

$$\begin{cases} e(\mathbf{y}) \;=\; \dfrac{-\rho^2 e(\mathbf{x})}{\|\mathbf{x}-\mathbf{q}\|^2} - \dfrac{2\rho^2 E(\mathbf{x})\,\mathbf{U_x}\cdot(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4}, \\[2ex] f(\mathbf{y}) \;=\; \dfrac{-\rho^2 f(\mathbf{x})}{\|\mathbf{x}-\mathbf{q}\|^2} - \dfrac{2\rho^2 F(\mathbf{x})\,\mathbf{U_x}\cdot(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4}, \\[2ex] g(\mathbf{y}) \;=\; \dfrac{-\rho^2 g(\mathbf{x})}{\|\mathbf{x}-\mathbf{q}\|^2} - \dfrac{2\rho^2 G(\mathbf{x})\,\mathbf{U_x}\cdot(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4}. \end{cases} \tag{20.50}$$

Proof. We first use (20.48) to compute

$$\begin{aligned} \mathbf{y}_{uu} \;=\; & \frac{\rho^2\mathbf{x}_{uu}}{\|\mathbf{x}-\mathbf{q}\|^2} - \frac{4\rho^2\big(\mathbf{x}_u\cdot(\mathbf{x}-\mathbf{q})\big)\mathbf{x}_u}{\|\mathbf{x}-\mathbf{q}\|^4} - \frac{2\rho^2\big(\mathbf{x}_{uu}\cdot(\mathbf{x}-\mathbf{q})\big)(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4} \\ & - \frac{2\rho^2\|\mathbf{x}_u\|^2(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^4} + \frac{8\rho^2\big(\mathbf{x}_u\cdot(\mathbf{x}-\mathbf{q})\big)^2(\mathbf{x}-\mathbf{q})}{\|\mathbf{x}-\mathbf{q}\|^6}. \end{aligned}$$

Taking the dot product with (20.49) and cancelling the various terms gives the first equation of (20.50); proofs of the other equations are similar. ∎

Next, we determine the relation between the principal curvatures of a surface $\mathcal{M}$ and those of the image of $\mathcal{M}$ under an inversion Ψ. Let $\mathbf{x}$ be a principal patch on $\mathcal{M}$. By Corollary 13.33 on page 405, the principal curvatures on the trace of $\mathbf{x}$ are

$$k_1(\mathbf{x}) = \frac{e(\mathbf{x})}{E(\mathbf{x})} \qquad \text{and} \qquad k_2(\mathbf{x}) = \frac{g(\mathbf{x})}{G(\mathbf{x})}.$$

Proposition 20.30 implies that $\Psi \circ \mathbf{y}$ is a principal patch, and so the same equations hold with $\mathbf{y}$ in place of $\mathbf{x}$.

Lemma 20.35. *Let $\mathbf{x}$ be a principal patch in $\mathbb{R}^3$, and let $\mathbf{y}$ be the inverse patch of $\mathbf{x}$ with respect to $\mathbf{q} \in \mathbb{R}^3$ with radius of inversion ρ. Then*

$$\left\{ \begin{array}{rcl} k_1(\mathbf{y}) & = & \dfrac{-\|\mathbf{x}-\mathbf{q}\|^2}{\rho^2} k_1(\mathbf{x}) - \dfrac{2\mathbf{U}_\mathbf{x} \cdot (\mathbf{x}-\mathbf{q})}{\rho^2}, \\ k_2(\mathbf{y}) & = & \dfrac{-\|\mathbf{x}-\mathbf{q}\|^2}{\rho^2} k_2(\mathbf{x}) - \dfrac{2\mathbf{U}_\mathbf{x} \cdot (\mathbf{x}-\mathbf{q})}{\rho^2}. \end{array} \right. \tag{20.51}$$

Proof. This is now a consequence of Corollary 20.32 and (20.50). ∎

As a corollary, an inversion of $\mathbb{R}^3$ maps umbilic points into umbilic points, a statement already implicit in Proposition 20.30.

Combining (20.1), (20.49) and (20.51), we deduce that the inverse patch $\mathbf{y}$ itself has focal patches

$$\mathbf{z}_i = \mathbf{y} - \rho^2 \frac{-\|\tilde{\mathbf{x}}\|^2 \mathbf{U}_\mathbf{x} + 2(\mathbf{U}_\mathbf{x} \cdot \tilde{\mathbf{x}})\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|^2 \big(\|\tilde{\mathbf{x}}\|^2 k_i + 2\mathbf{U}_\mathbf{x} \cdot \tilde{\mathbf{x}}\big)}, \tag{20.52}$$

where $\tilde{\mathbf{x}}$ is shorthand for $\mathbf{x} - \mathbf{q}$. This formula can be used to prove

Proposition 20.36. *Inversion maps a torus of revolution onto a cyclide of Dupin.*

Proof. Let $\mathbf{x} = \mathsf{torus}[a, b, b]$ be a torus of revolution, as defined on page 210. Let $\mathbf{y} = \Psi \circ \mathbf{x}$ be an inverse patch, depending on the center $\mathbf{q}$ and radius ρ of inversion. It suffices to show that *both* the associated patches (20.52) are curves, a fact that is readily verified by computer. To cite a simple but typical example, let $\mathbf{q} = (m, 0, 0)$ and $\rho = 1$. The following formulas for $\mathbf{z}_1$ and $\mathbf{z}_2$ were copied from output from Notebook 20:

$$\left(\frac{m\left(a^2 - b^2 + m^2 - 1\right) + \left(a - 2am^2\right)\cos u}{a^2 - 2am\cos u - b^2 + m^2}, \ \frac{a \sin u}{a^2 - 2am\cos u - b^2 + m^2}, \ 0 \right),$$

$$\left(-\frac{m\left(-2ab - \left(a^2 + b^2 - m^2 + 1\right)\cos v\right)}{2ab + \left(a^2 + b^2 - m^2\right)\cos v}, \ 0, \ \frac{a \sin v}{2ab + \left(a^2 + b^2 - m^2\right)\cos v} \right).$$

Each one is certainly a curve, indeed a plane conic. ∎

The last result enables us to reconcile the types of cyclides in Definition 20.23 with those of tori on page 306. For example, Figure 20.12 shows the effect of inverting the spindle torus, parametrized by torus$[4, 4, 4]$ in Figure 10.11 (right) on page 306, with center $\mathbf{q} = (0, 5, 0)$. It evidently corresponds to the cyclide shown in Figure 20.8.

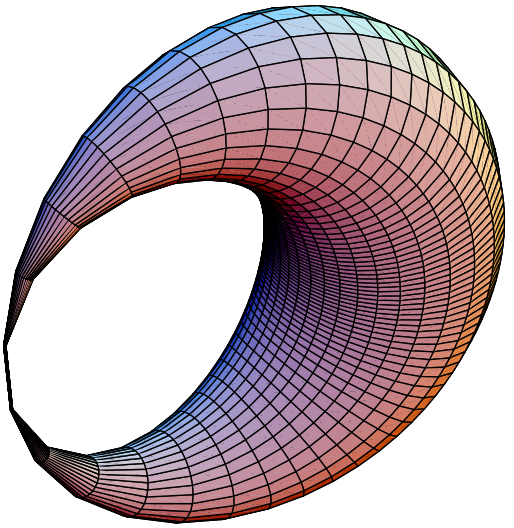


Figure 20.12: Inverse of the inside-out torus

20.6 Exercises

1. Complete the proofs of Theorem 20.4, Lemma 20.5 and Lemma 20.7.

2. Prove Lemma 20.6 and give the details of equations (20.9) and (20.17).

3. Let $\mathcal{M}$ be a surface with the property that $\rho_1 - \rho_2 = a$, where a is a constant. Assume that each component of the focal set of $\mathcal{M}$ is a regular surface. Show that each component of the focal set of $\mathcal{M}$ has constant negative curvature $-a^{-2}$.

4. Describe the envelope of the 1-parameter of spheres

$$(x - 2\cos t)^2 + (y - 2\sin t)^2 + z^2 = 1,$$

for $0 \leqslant t < 2\pi$.

5. Find the equations of the two focal hypersurfaces of an elliptical paraboloid, all shown together in Figure 20.13.

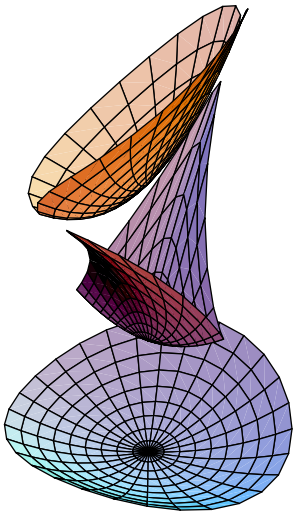


Figure 20.13: Elliptical paraboloid and its focal set

M **6.** Find the two components of the focal set of a circular helicoid defined on page 376, referring also to Exercise 3 of Chapter 12.

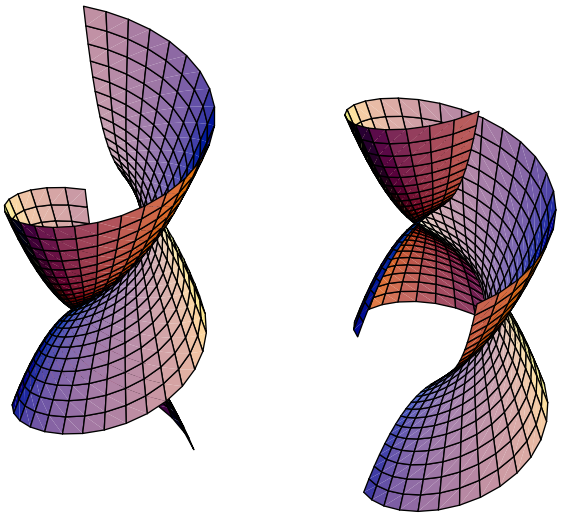


Figure 20.14: The two components of the focal set of a helicoid

M **7.** Use the technique that led to Definition 20.23 to find the equation of a parabolic cyclide. Plot several examples.

M **8.** Compute the metric and second fundamental form of a general elliptic-hyperbolic cyclide.

M **9.** Determine analytically the inverse of a helix. The one in Figure 20.15 also allows one to imagine the inverse of a cylinder.

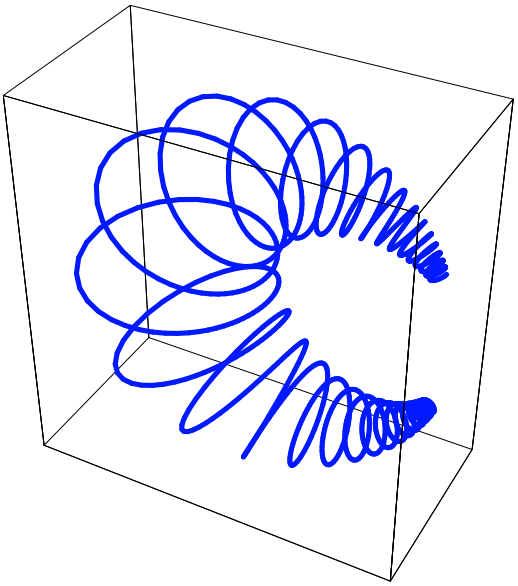


Figure 20.15: Inverse curve of a helix

10. Show that the inverse curve with respect to the origin of an equilateral hyperbola is a lemniscate. The parametrization of an equilateral hyperbola that corresponds to the lemniscate is $t \mapsto \big(a \sec t,\ a \tan t\big)$.

M **11.** Plot the inverse curves of an ellipse with respect to several points.

12. Use Lemma 20.35 to obtain the following formula for the mean curvature of an inverse patch:

$$H(\mathbf{y}) = -\frac{\|\mathbf{x}-\mathbf{q}\|^2 H(\mathbf{x})}{\rho^2} - \frac{2\mathbf{U}_\mathbf{x} \cdot (\mathbf{x}-\mathbf{q})}{\rho^2}.$$

Find an analogous formula for the Gaussian curvature.

M **13.** Find the image of the circular cone

$$\mathbf{x}(u, v) = \big(v \cos u,\ v \sin u,\ v\big)$$

under an inversion about an arbitrary point (p_1, p_2, p_3) and radius $\rho = 1$. Plot several inverted circular cones, like the one in Figure 20.16.

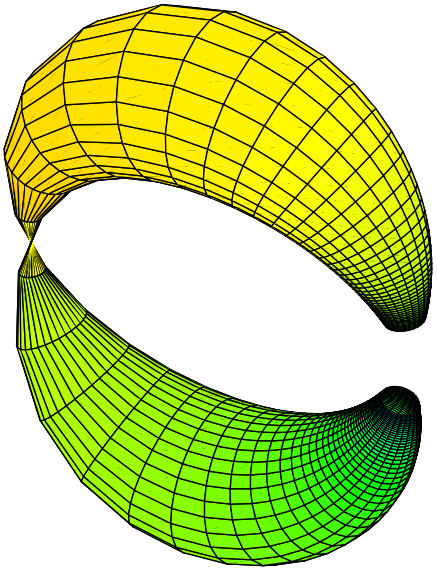


Figure 20.16: Inverted circular cone

14. Show that the focal set of each surface parallel to a surface $\mathcal{M}$ coincides with the focal set of $\mathcal{M}$.

Notebook 20

■ N20.0

The aim of this notebook is to compute principal curvatures, plot focal surfaces, investigate canal surfaces, compute cyclides of Dupin, and define the inversion map for curves and surfaces.

We shall make use of the following plane curves and surfaces. The Notebook Index at the end of the book can be used to find the notebook in which each was first defined.

```
hyperbola[a_,b_][t_]:= {a Cosh[t], b Sinh[t]}
catenary[c_][t_]:= {c Cosh[t/c], t}
helix[a_,b_][t_]:= {a Cos[t], a Sin[t], b t}
```

```
circularcone[a_,b_][u_,v_]:= {a v Cos[u], a v Sin[u], b v}
circularcylinder[a_][u_,v_]:= {a Cos[u], a Sin[u], v}
monkeysaddle[u_,v_]:={u, v, u^3 - 3u v^2}
torus[a_,b_,c_][u_,v_]:=
  {(a+b Cos[v])Cos[u], (a+b Cos[v])Sin[u], c Sin[v]}
pseudosphere[a_][u_,v_]:=
  a{Cos[u]Sin[v], Sin[u]Sin[v], Cos[v]+Log[Tan[v/2]]}
```

```
ellipticparaboloid[a_,b_][u_,v_]:= {u, v, u^2/a^2 + v^2/b^2}
hyperbolicparaboloid[u_,v_]:= {u, v, u v}
helicoid[a_,c_][u_,v_]:= {a v Cos[u], a v Sin[u], c u}
catenoid[c_][u_,v_]:= {c Cos[u]Cosh[v/c], c Sin[u]Cosh[v/c], v}
```

The definition of elliptic coordinates, from N19.5, is required in N20.5:

```
ellipticcoor[d_,e_,f_][a_,b_,c_][w_][u_,v_]:=
  {d Sqrt[(a-u)(a-v)(a-w)/((a-b)(a-c))],
   e Sqrt[(b-u)(b-v)(b-w)/((b-a)(b-c))],
   f Sqrt[(c-u)(c-v)(c-w)/((c-a)(c-b))]}
```

N20.0.1 contains further definitions, including evolute from N4.1. It also contains the code for Figure 20.1, showing focal intervals.

■ N20.0.1

■ N20.1

The following program, from N13.5, computes the principal curvatures of a surface.

```
princurv[x_][u_,v_]:=
Module[{K,H},
  K= gcurv[x][u,v];
  H= mcurv[x][u,v];
  k1= H + Sqrt[Simplify[H^2-K]];
  k2= H - Sqrt[Simplify[H^2-K]];
  {k1,k2}
]
```

Using it, we can easily parametrize the two focal sets:

```
focal[i_,x_][u_,v_]:=
Module[{U,V,k,y},
  k:= princurv[x][U,V];
  y:= unitnormal[x];
  x[U,V]+(1/k[[i]])y[U,V]/.{U->u,V->v}
]
```

The equations in Proposition 20.9 for the focal sets of a hyperbolic paraboloid were extracted from the next output. Here, only instructions for the second set are shown.

```
hp:= hyperbolicparaboloid
K= gcurv[hp][u,v]
H= mcurv[hp][u,v]
H^2-K//Simplify
princurv[hp][u,v]//Factor//PowerExpand
```

```
fhp:= focal[2,hp]
fhp[u,v]//PowerExpand//Together;
fhp[u,v]//PowerExpand//FullSimplify
%//TeXForm
```

The metric of the second focal set is given by

```
metric[fhp][u,v];
ee//PowerExpand//FullSimplify
```

The same technique is used to plot the next figure.

```
Module[{f},
  f[0][u,v]= {u, v, u v, Hue[0]};
  f[i_][u_,v_]:= Append[focal[i,hp][u,v], Hue[i/3]];
  S @ ParametricPlot3D[Table[f[i][u,v], {i,0,2}]//Evaluate,
    {u,-1,1}, {v,-1,1}, ViewPoint->{3,1.5,1.3},
    Boxed->True, Lighting->False]
]
```

Figure 20.2

■ N20.2

The following calculations refer to Theorem 20.16. The first definition is the standard parametrization of a surface of revolution, copied from N15.1.

```
surfrev[α_][u_,v_]:= {Cos[u]α[v][[1]], Sin[u]α[v][[1]], α[v][[2]]}
```

```
γ[t_]:= {ϕ[t], ψ[t]}
metric[surfrev[γ]][u,v]
```

Here is an alternative way of obtaining the principal curvatures that works when both ff and fff are zero. We use it for a surface of revolution.

```
prcurv[x_][u_,v_]:=
Module[{},
  metric[x][u,v];
  shape[x][u,v];
  kp:= eee/ee;
  km:= ggg/gg;
  {kp,km}
]
```

The command usrule that follows is a valid substitution for a unit-speed curve.

```
usrule= ϕ'[t_]^2+ ψ'[t_]^2->1;
prcurv[surfrev[γ]][u,v];
kp/.usrule//PowerExpand
km/.usrule//PowerExpand
```

An analog of the program focal uses the alternative definition of principal curvatures, and is then used to examine the focal set of a surface of revolution.

```
foc[i_,x_][u_,v_]:=
Module[{U,V,y},
  prcurv[x][U,V];
  y:= unitnormal[x];
  x[U,V] + (1/If[i==1,kp,km])y[U,V]/.{U->u,V->v}
]
```

```
Collect[foc[1,surfrev[γ]][u,v]/.usrule, {Cos[u],Sin[u]}]
foc[2, surfrev[γ]][u,v]/.usrule
```

Using the definition of evolute from N20.0.1, we can compare the surface of revolution of the evolute with the output from above, so as to verify Theorem 20.16.

```
ϵ:= evolute[α]
ϵ[t]/.usrule
surfrev[ϵ][u,v]/.usrule
```

- N20.2.1

■ N20.3

The following definitions of spatial conics are used for Figure 20.5:

```
α[a_,c_][u_]:= {a Cos[u], Sqrt[a^2-c^2]Sin[u], 0}
β[a_,c_][v_]:= {c Sec[v], 0, Sqrt[a^2-c^2]Tan[v]}
```

When a,c are constant, the first is an ellipse while the second is a hyperbola.

```
Module[{z},
  αα= highlight[α[1,.5]];
  ββ= highlight[β[1,.5]];
  z[1]= ParametricPlot3D[αα[t], {t,-π,π}];
  z[2]= ParametricPlot3D[ββ[t], {t, -.499π, .499π}];
  pts= {Hue[.7], PointSize[.04], Point[α[1,.5][0]], Point[β[1,.5][0]]};
  z[3]= Graphics3D[pts];
  S[Array[z,3], PlotRange->{{-1,1.8}, Automatic, Automatic},
    Boxed->True]
]
```

Figure 20.5

Here is a version with asymptotes:

```
ParametricPlot3D[{α[1,.5][t], β[1,.5][t]}//Evaluate, {t,-π,π},
  PlotRange->{{-1.5,1.5},{-1,1},{-1.8,1.8}}]
```

Using results from Section 20.3, we can define an elliptic-hyperbolic cyclide directly:

```
ellhypcyclide[a_,c_,k_][u_,v_]:=
Module[{A,B,s},
  A= k - c Cos[u];
  B= a - k Cos[v];
  s= Sqrt[a^2-c^2];
  {c A + a Cos[u]B, s B Sin[u], s A Sin[v]}/(a - c Cos[u]Cos[v])
]
```

The following cyclide has been colored by the difference u-v of its two coordinates. Add S@ at the start of the first or fourth line to display it.

```
Module[{x,xx},
  x= ellhypcyclide[12,4,4];
  xx[u_,v_]:= Append[x[u,v], Hue[u-v]];
  ParametricPlot3D[xx[u,v], {u,0,2π}, {v,0,2π},
    PlotPoints->{80,40}, PlotRange->All, Lighting->False]
]
```

We can now check that the definitions of α and β given at the start of N20.3 are compatible with this cyclide.

```
ehc:= ellhypcyclide[a,c,k]
α[a,c][u]
(ehc[u,v]/.Cos[u]->k/c)/.k->c Cos[u]
```

When this cell is evaluated, the last two lines produce identical outputs, confirming that α gives rise to a parametrization of the elliptic-hyperbolic cyclide. The same is true when we use the chart β associated to the hyperbola.

```
β[a,c][v]
(ehc[u,v]/.Cos[v]->a/k)/.k->a Sec[v]//Simplify
```

A parametrization of the parabolic cyclide is

```
paraboliccyclide[a_,k_][u_,v_]:=
  {(u(8a^2+k+v^2))/(16a^2+u^2+v^2),
  ((8a^2-k+u^2)v)/(16a^2+u^2+v^2),
  (16a^2k-16a^2u^2-k u^2+16a^2v^2-k v^2)
    /(8a(16a^2+u^2+v^2))}
```

```
pc:= paraboliccyclide[2,1]
ParametricPlot3D[pc[u,v], {u,-20,20}, {v,-20,20}];
```

Next, we construct a generalized cone, as parametrized in N14.2, over a focal curve.

```
cone[p_, γ_][u_,v_]:= (1-v)p + v γ[u]
```

```
Module[{x,y},
  x= ellhypcyclide[12,2,4];
  y= cone[β[12,2][0.35π], α[12,3]];
  ParametricPlot3D[{x[u,v], y[u,v/(2π)]}//Evaluate, {u,0,π},
  {v,0,2π}]
];
```

The next program is used to generate a cyclide together with its focal curves, for use in Figures 20.6-20.8.

```
cycgen[a_,c_,k_,rau_,rav_,opts___]:=
Module[{x,z,α,β},
  x= ellhypcyclide[a,c,k];
  z[1]:= ParametricPlot3D[x[u,v], rau, rav, opts];
  α[t_]:= {a Cos[t], (a^2-c^2)^(1/2) Sin[t], 0};
  β[t_]:= {c Sec[t], 0, (a^2-c^2)^(1/2) Tan[t]};
  αα:= highlight[α, .5, .015];
  ββ:= highlight[β, 0, .015];
  z[2]= ParametricPlot3D[αα[t], {t,-π,π}];
  z[3]= ParametricPlot3D[ββ[t], {t, -.499π, .499π}];
  Show[Array[z,3],
    PlotRange->{{-a-c-k,a+c+k},{-a-c-k,a+c+k},{-a-c-k,a+c+k}}]
]
```

```
Module[{x,z},
  x= ellhypcyclide[12,3,4.5];
  z[1]:= ParametricPlot3D[{x[u,v], 9{u,0,.35u}}//Evaluate,
    {u,0,2π}, {v,0,2π}];
  z[2]= cycgen[12, 3, 4.5, {u,0,2π}, {v,-π,0}, PlotPoints->{45,15}];
  S[GraphicsArray[Array[z,2]], GraphicsSpacing->-.75,
    ImageSize->500]
]
```

Figure 20.6

Similar instructions for Figures 20.7 and 20.8 are hidden in the next subsection.

■ N20.3.1

■ N20.4

Here are the basic inversion formulas. The first definition requires q to be a vector. The second deals with the special case in which q=0.

```
inversion[q_List,ρ_][p_]:= q + (ρ^2/((p-q).(p-q)))(p-q)
inversion[ρ_][p_]:= (ρ^2/p.p)p
```

The clothoid curve is generated by the next cell, copied from N2.7.

```
clothoidprime[n_,a_][t_]:=
  a{Sin[t^(n+1)/(n+1)], Cos[t^(n+1)/(n+1)]}
clothoid[n_,a_][t_]:= Integrate[clothoidprime[n,a][tt], {tt,0,t}]
```

The next two definitions will allow us one to define "inverse" curves and surfaces without an explicit parameter. This is convenient for plotting purposes.

```
invertedcurve[q_, ρ_][α_][t_]:= inversion[q,ρ][α[t]]
invertedsurf[q_,ρ_][surf_][u_,v_]:= inversion[q,ρ][surf[u,v]]
```

```
invertedcurve[{q1,q2}, ρ][clothoid[1,a]][t]
```

We can now display the effect of inversion on a clothoid.

```
ParametricPlot[invertedcurve[{2,2}, 1][clothoid[1,1]][t],
  {t,-10,10}, AspectRatio->Automatic]
```

N20.4.1

N20.5

The inverse of an ellipsoid needs to be constructed in eight separate parts.

```
Module[{h,k,u,v,x,y,ta},
  h:= Hue[(u-1)/6];
  k:= Hue[(v-7)/9];
  x[a_,b_,c_]:= ellipticcoor[a,b,c][16,7,1][0];
  y[a_,b_,c_,h_,k_]:= Append[inversion[{0,0,5}, 1][x[a,b,c][u,v]],
    FaceForm[h,k]];
  ta:= {y[1,1,1,h,k], y[-1,1,1,k,h], y[1,-1,1,k,h],
    y[-1,-1,1,h,k], y[1,1,-1,k,h], y[-1,1,-1,h,k],
    y[1,-1,-1,h,k], y[-1,-1,-1,k,h]};
  S @ ParametricPlot3D[ta//Evaluate, {u,1,7}, {v,7,16},
    Lighting->False, PlotPoints->{10,20}, ViewPoint->{1,3,2}]
]
```

Figure 20.11

Using definitions from N20.1 and N20.4, we may define the inverse of a torus and compute its principal curvatures.

```
invertedtorus[{p1_,p2_,p3_},ρ_,a_,b_,c_][u_,v_]:=
Module[{x},
  x= torus[a,b,c];
  invertedsurf[{p1,p2,p3}, ρ][x][u,v]
]
```

```
princurv[invertedtorus[{2,0,0}, 1, 8, 3, 3]][1,2]//N
```

Next we use Lemma 20.35 to compute the same principal curvatures to check consistency.

```
invertedprincurv[q_,ρ_][x_][u_,v_]:=
Module[{xq,un},
  princurv[x][u,v];
  xq= x[u,v] - q;
  un= unitnormal[x][u,v];
  kk[i_]= (-L[xq]^2 k[i] - 2un.xq)/ρ^2;
  Array[kk,2]
]
```

```
invertedprincurv[{2,0,0},1][torus[8,3,3]][1,2]//N
```

Next we compute focal sets of the inverse of a torus of revolution.

```
invertedfocal[i_][u_,v_]:=
Module[{x,q,w,un,A,B},
  x= torus[a,b,b];
  q= {m,0,0};
  princurv[x][u,v];
  w= x[u,v] - q;
  un= unitnormal[x][u,v];
  A= -un + 2(un.w/w.w)w;
  B= -w.w k[i] - 2un.w;
  invertedsurf[q,1][x][u,v] + A/B
]
```

The next two cells give the answers analytically.

```
invertedfocal[1][u,v]
%//PowerExpand
%//Simplify
```

```
invertedfocal[2][u,v]//PowerExpand//Simplify;
%//TeXForm
```

We can easily compute the first fundamental form of the inverse of a torus:

```
itor:= invertedtorus[{1,2,0},1,4,4,4]
metric[itor][u,v];
ee
ff
gg
```

■ N20.5.1

■ N20.6

Here is the first figure in the Exercises. Others are included in N20.6.1.

```
Module[{y,z},
  ep:= ellipticparaboloid[1.5,1];
  y[0]:= ep;
  y[i_]:= focal[i, ep];
  z[i_]:= ParametricPlot3D[y[i][r Cos[θ], r Sin[θ]], {r,0,.5}, {θ,0,2π},
    PlotPoints->{10,30}, PlotRange->All];
  S @ Array[z,3,0]
 ]
```

Figure 20.13

■ N20.6.1

Chapter 21

The Theory of Surfaces of Constant Negative Curvature

In 1879, Luigi Bianchi gave for the first time in his thesis a method for deriving from one surface of constant negative Gaussian curvature an infinite number of other surfaces with the same curvature. His work created great interest and was followed by papers by Darboux, Lie, and Bäcklund. In [Bck], Bäcklund gave a more general construction that led to the *Bäcklund transform*. In spite of the fact that the paper [Bck] dealt exclusively with surfaces of constant negative curvature, the notion of Bäcklund transform has come to have enormous importance in soliton theory (see for example [Lamb2, Remo]).

Our treatment is based on the concept of a Tchebyshef[1] net. Patches of two interrelated types are introduced in Section 21.1 and are characterized by properties of the first fundamental form and its coefficients E, F, G. We first define a *Tchebyshef patch* for which E and G are equal and constant, and second one for which $E + G$ is constant and $F = 0$. At this stage, the theory proceeds without reference to the second fundamental form or knowledge of the way that $\mathcal{M}$ is contained in $\mathbb{R}^3$ or $\mathbb{R}^n$, so the principal curvatures k_1, k_2 are not yet defined. The results can therefore be applied to the more abstract types of surfaces to be studied in Chapter 26.

In Section 21.2, we turn attention to surfaces of constant negative curvature



[1] Pafnuty Lvovich Tchebyshef (1821–1894). A leading Russian mathematician of his time. He worked in probability theory, number theory, and differential geometry.

in $\mathbb{R}^3$. We show that principal patches on such surfaces have special properties, and the two types of Tchebyshef patches can be realized as asymptotic and principal patches respectively in which k_1, k_2 are determined by the metric. Section 21.3 introduces the sine–Gordon equation, solutions of which determine angle functions of Tchebyshef patches with constant curvature. Special solutions are used to treat surfaces of revolution in Section 21.4, culminating in explicit parametrizations for the pseudosphere.

Section 21.5 is devoted to the Bianchi transform of surfaces of constant negative curvature. Moving frames on surfaces in $\mathbb{R}^3$ are discussed in Section 21.6, as well as an analog of the Frenet formulas of Chapter 7. In Section 21.7 we describe Kuen's surface as the Bianchi transform of the pseudosphere, itself the transform of a degenerate patch. The Bäcklund transform generalizes that of Bianchi; it is defined and applied to the construction of Dini's surface in Section 21.8.

Throughout this chapter, we shall be considering surfaces of constant negative Gaussian curvature $-1/a^2$. We shall express this with the statement "$K \equiv -a^{-2} < 0$," in which it is implicitly understood that a is a positive constant.

21.1 Intrinsic Tchebyshef Patches

We begin by defining a special kind of patch that will give rise (in the next section) to an asymptotic patch on a surface with constant negative Gaussian curvature.

Definition 21.1. *Let a be a positive constant. A* ***Tchebyshef patch*** *or* ***net*** *of* ***radius*** *a is a patch $\mathbf{y}\colon \mathcal{U} \to \mathbb{R}^3$ whose metric $ds^2 = E\,dp^2 + 2F\,dp\,dq + G\,dq^2$ has the property that $E = G = a^2$.*

Like that of isothermal patches on page 518, this definition is *intrinsic*, in that it imposes conditions merely on the coefficients of the first fundamental form. Sometimes it is required in the definition that $a = 1$, so that the coordinate curves are parametrized by arc length, but we prefer to allow an arbitrary positive constant. The resulting functions p, q on the surface $\mathbf{y}(\mathcal{U})$ are called ***Tchebyshef coordinates***.

The Cauchy–Schwarz inequality implies that

$$F^2 = \langle \mathbf{y}_p, \mathbf{y}_q \rangle^2 \leqslant \|\mathbf{y}_p\|^2 \|\mathbf{y}_q\|^2 = EG = a^4,$$

so we can write $F = a^2 \cos\omega$ for some ω. Thus the metric of a Tchebyshef patch of radius a can be written as

$$ds^2 = a^2\big(dp^2 + 2\cos\omega\, dp\,dq + dq^2\big). \tag{21.1}$$

We say that the metric given by (21.1) has ***angle function*** ω. Note that

$$\cos\omega = \frac{\langle \mathbf{y}_p, \mathbf{y}_q \rangle}{\|\mathbf{y}_p\| \, \|\mathbf{y}_q\|},$$

so that geometrically, ω represents an unoriented angle between the coordinate curves $p \mapsto \mathbf{y}(p,q)$ and $q \mapsto \mathbf{y}(p,q)$.

Before giving examples of such patches, we compute their Gaussian curvature and Christoffel symbols.

Lemma 21.2. *The Gaussian curvature of a Tchebyshef patch of radius $a > 0$ is given by*

$$K = -\frac{\omega_{pq}}{a^2 \sin\omega},$$

where the subscripts denote partial derivatives.

Proof. Apply Brioschi's formula (17.3), page 533, to the metric (21.1). Since E and G are constant,

$$\begin{aligned} K &= \frac{1}{(EG-F^2)^2} \det \begin{pmatrix} F_{pq} & 0 & F_p \\ F_q & a^2 & F \\ 0 & F & a^2 \end{pmatrix} \\ &= \frac{F_{pq}(a^4 - F^2) + F_p F_q F}{\left(a^4 - a^4\cos^2\omega\right)^2} \\ &= \frac{a^6\big((-\omega_p\omega_q\cos\omega - \omega_{pq}\sin\omega)\sin^2\omega + \omega_p\omega_q\sin^2\omega\cos\omega\big)}{(a^2\sin\omega)^4} \\ &= \frac{-\omega_{pq}}{a^2\sin\omega}. \quad \blacksquare \end{aligned}$$

Easy calculations using Theorem 17.7, page 539, yield the following formulas for the Christoffel symbols (see Exercise 1 and Notebook 21).

Lemma 21.3. *The Christoffel symbols of a Tchebyshef patch are given by*

$$\begin{aligned} \Gamma^1_{11} &= (\cot\omega)\omega_p, & \Gamma^2_{11} &= -(\csc\omega)\omega_p, \\ \Gamma^1_{12} &= 0, & \Gamma^2_{12} &= 0, \\ \Gamma^1_{22} &= -(\csc\omega)\omega_q, & \Gamma^2_{22} &= (\cot\omega)\omega_q. \end{aligned}$$

A key feature is the vanishing of the "mixed" Christoffel symbols, and this provides a situation in which it is relatively easy to find Tchebyshef coordinates.

Lemma 21.4. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a regular patch. A sufficient condition that there exist a Tchebyshef patch that reparametrizes $\mathbf{x}$ is that*

$$\Gamma^1_{12} = \Gamma^2_{12} = 0. \tag{21.2}$$

Proof. Let $\mathbf{x}$ be a patch with $ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$ for which $\Gamma^1_{12} = \Gamma^2_{12} = 0$. From (17.7), we get

$$0 = GE_v - FG_u = EG_u - FE_v. \tag{21.3}$$

When we regard (21.3) as two algebraic equations in the unknowns E_v and G_u, we see that the unique solution of (21.3) is $E_v = G_u = 0$. Thus E is a function of u alone and G is a function of v alone, so we can define functions p, q by

$$p(u) = \frac{1}{a}\int\sqrt{E}\,du \qquad \text{and} \qquad q(v) = \frac{1}{a}\int\sqrt{G}\,dv. \tag{21.4}$$

The constants of integration are immaterial. Since p' and q' never vanish, the inverse function theorem assures us of the existence of inverse functions $u = u(p)$ and $v = v(q)$. We define a patch $\mathbf{y}$ by

$$\mathbf{y}(p, q) = \mathbf{x}\big(u(p), v(q)\big).$$

From (21.4) we have

$$\frac{du}{dp} = \frac{a}{\sqrt{E}}, \qquad \frac{dv}{dq} = \frac{a}{\sqrt{G}} \qquad \text{and} \qquad \frac{du}{dq} = \frac{dv}{dp} = 0.$$

Now Lemma 12.4 on page 365 implies that $E_{\mathbf{y}} = G_{\mathbf{y}} = a^2$. ∎

A closely-related type of patch is defined by

Lemma 21.5. *Let $\mathbf{y}$ be a Tchebyshef patch, and set*

$$\mathbf{x}(u, v) = \mathbf{y}\left(\frac{u+v}{2},\ \frac{u-v}{2}\right).$$

Then the first fundamental form of $\mathbf{x}$ is given by

$$ds^2_{\mathbf{x}} = (a\cos\theta)^2\,du^2 + (a\sin\theta)^2\,dv^2, \tag{21.5}$$

where $\theta = \omega/2$. Conversely, if the coefficients of the first fundamental form of $\mathbf{x}$ satisfy $F = 0$ and $E + G$ constant, then $\mathbf{y}(p, q) = \mathbf{x}(p + q,\ p - q)$ defines a Tchebyshef patch.

Proof. Use subscripts to specify the patch in question. Since $\mathbf{x}_u = \frac{1}{2}(\mathbf{y}_p + \mathbf{y}_q)$ and $\mathbf{x}_v = \frac{1}{2}(\mathbf{y}_p - \mathbf{y}_q)$, we have

$$E_{\mathbf{x}} = \tfrac{1}{4}(\mathbf{y}_p + \mathbf{y}_q)\cdot(\mathbf{y}_p + \mathbf{y}_q) = \tfrac{1}{4}(E_{\mathbf{y}} + 2F_{\mathbf{y}} + G_{\mathbf{y}}).$$

Thus if $E_{\mathbf{y}} = G_{\mathbf{y}} = a^2$ and $F_{\mathbf{y}} = a^2\cos\omega$, then

$$E_{\mathbf{x}} = \tfrac{1}{2}a^2(1 + \cos\omega) = a^2\cos^2\frac{\omega}{2} = a^2\cos^2\theta,$$

as stated; $G_{\mathbf{x}}$ similarly.

Conversely, suppose that $F = 0$ and that $E + G$ is constant. Since E and G must both be positive, there exists $\theta \in \mathbb{R}$ for which $E\,du^2 + G\,dv^2$ equals (21.5), where $a^2 = E + G$. The rest then follows by setting $\omega = 2\theta$. ∎

In the light of Lemma 21.5, we call a patch $\mathbf{x}$ whose metric is given by (21.5) a ***Tchebyshef patch of the second kind*** with ***angle function*** θ. The latter is determined up to sign and addition of integer multiples of π.

With this terminology, we have

Lemma 21.6. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef patch of the second kind with angle function θ. Then the Gaussian curvature of $\mathbf{x}$ is given by*

$$K = \frac{-\theta_{uu} + \theta_{vv}}{a^2 \sin\theta \cos\theta}.$$

Proof. Note that the expression for K is unaffected by the ambiguity in θ. We have

$$\begin{cases} \dfrac{1}{\sqrt{E}} \dfrac{\partial \sqrt{G}}{\partial u} = \dfrac{1}{a\cos\theta} \dfrac{\partial (a \sin\theta)}{\partial u} = \theta_u, \\ \dfrac{1}{\sqrt{G}} \dfrac{\partial \sqrt{E}}{\partial v} = \dfrac{1}{a\sin\theta} \dfrac{\partial (a \cos\theta)}{\partial v} = -\theta_v. \end{cases}$$

Hence by Corollary 17.4, page 534,

$$K = \frac{-1}{\sqrt{EG}} \left\{ \frac{\partial}{\partial u} \left(\frac{1}{\sqrt{E}} \frac{\partial \sqrt{G}}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{1}{\sqrt{G}} \frac{\partial \sqrt{E}}{\partial v} \right) \right\} = \frac{-\theta_{uu} + \theta_{vv}}{a^2 \sin\theta \cos\theta}. \quad \blacksquare$$

Once again, the Christoffel symbols can readily be found from (17.7) (see Exercise 2 and Notebook 21).

Lemma 21.7. *The Christoffel symbols of the metric* (21.5) *are given by*

$$\begin{aligned} \Gamma^1_{11} &= -(\tan\theta)\theta_u = \Gamma^1_{22}, & \Gamma^2_{11} &= (\cot\theta)\theta_v = \Gamma^2_{22}, \\ \Gamma^1_{12} &= -(\tan\theta)\theta_v, & \Gamma^2_{12} &= (\cot\theta)\theta_u. \end{aligned}$$

In the next section, we proceed to construct a Tchebyshef patch of the second kind, starting from a principal patch.

21.2 Patches on Surfaces of Constant Negative Curvature

We can establish some important facts about the principal curvatures of surfaces of constant negative curvature in $\mathbb{R}^3$, using the Peterson-Mainardi-Codazzi equations (Theorem 19.8 on page 600).

Lemma 21.8. *Let $\mathbf{y}\colon \mathcal{U} \to \mathbb{R}^3$ be a principal patch whose metric is given by $ds^2 = E\,dp^2 + G\,dq^2$, and let k_1, k_2 denote the principal curvatures. Suppose that $\mathbf{y}$ has constant Gaussian curvature $-a^{-2}$. Then*

$$\begin{cases} E(1 + a^2 k_1^2) & \text{is a function of } p \text{ alone}, \\ G(1 + a^2 k_2^2) & \text{is a function of } q \text{ alone}. \end{cases}$$

Proof. The first equation of (19.13), page 602, together with the assumption that $k_1 k_2 = -a^{-2}$ imply that

$$\frac{E_q}{2E} = \frac{k_{1q}}{k_2 - k_1} = \frac{k_{1q}}{-\dfrac{1}{a^2 k_1} - k_1} = \frac{-a^2 k_1 k_{1q}}{1 + a^2 k_1^2}. \tag{21.6}$$

We integrate (21.6) with respect to q, obtaining

$$-\tfrac{1}{2}\log\bigl(1 + a^2 k_1^2\bigr) = \tfrac{1}{2}\log E + A(p),$$

where $A(p)$ denotes the constant of integration, a function of p. Exponentiation yields

$$(1 + a^2 k_1^2)^{-1} = e^{2A(p)} E, \tag{21.7}$$

and the lemma's first conclusion. The second equation of (19.13) yields the second conclusion in the same way. ∎

The next important existence result shows that one can choose the two functions of Lemma 21.8 to be constant and equal to a^2.

Theorem 21.9. *Suppose that* $\mathbf{y}\colon \mathcal{U} \to \mathbb{R}^3$ *is a principal patch with metric* $ds^2 = E\,dp^2 + G\,dq^2$, *and* $K \equiv -a^{-2} < 0$. *Then there is a principal patch* $\mathbf{x}$ *reparametrizing* $\mathbf{y}$ *of the form*

$$\mathbf{x}(u, v) = \mathbf{y}\bigl(p(u), q(v)\bigr) \tag{21.8}$$

such that the principal curvatures of $\mathbf{x}$ *satisfy*

$$\begin{cases} k_1^2 \;=\; \dfrac{a^2 - E}{a^2 E}, \\[2ex] k_2^2 \;=\; \dfrac{a^2 - G}{a^2 G}. \end{cases} \tag{21.9}$$

Proof. We define $\mathbf{x}$ by (21.8) and determine functions p and q so that (21.9) holds. Since $\mathbf{x}_u = p'(u)\mathbf{y}_p$ and $\mathbf{x}_v = q'(v)\mathbf{y}_q$, the coefficients $E = E_{\mathbf{x}}$, $F = F_{\mathbf{x}}$, $G = G_{\mathbf{x}}$ of the first fundamental form of $\mathbf{x}$ are related to those of $\mathbf{y}$ by

$$\begin{cases} E(u, v) \;=\; p'(u)^2 E_{\mathbf{y}}\bigl(p(u), q(v)\bigr), \\ F(u, v) \;=\; 0, \\ G(u, v) \;=\; q'(v)^2 G_{\mathbf{y}}\bigl(p(u), q(v)\bigr). \end{cases} \tag{21.10}$$

Similarly, the coefficients of the second fundamental form of $\mathbf{x}$ are

$$\mathbf{x}_{uu} = p''(u)\mathbf{y}_p + p'(u)^2 \mathbf{y}_{pp}, \quad \mathbf{x}_{uv} = p'(u)q'(v)\mathbf{y}_{pq}, \quad \mathbf{x}_{vv} = q''(v)\mathbf{y}_q + q'(v)^2 \mathbf{y}_{qq},$$

and

$$\left\{\begin{array}{rcl} e(u,v) &=& p'(u)^2 e_{\mathbf{y}}\big(p(u),q(v)\big), \\ f(u,v) &=& f_{\mathbf{y}}\big(p(u),q(v)\big) = 0, \\ g(u,v) &=& q'(v)^2 g_{\mathbf{y}}\big(p(u),q(v)\big) \end{array}\right. \tag{21.11}$$

(see equations (13.41), page 419). Denote by $\widetilde{k}_1, \widetilde{k}_2$ the principal curvatures of $\mathbf{y}$; these are patch-independent, so

$$k_1(u,v) = \widetilde{k}_1\big(p(u),q(v)\big) \quad \text{and} \quad k_2(u,v) = \widetilde{k}_2\big(p(u),q(v)\big), \tag{21.12}$$

equations that also follow directly from (21.10) and (21.11).

From (21.7), (21.10) and (21.12), we get

$$\left\{\begin{array}{rcl} k_1(u,v)^2 &=& \dfrac{1}{a^2}\left(\dfrac{p'(u)^2}{E(u,v)e^{A(p(u))}} - 1\right), \\ k_2(u,v)^2 &=& \dfrac{1}{a^2}\left(\dfrac{q'(v)^2}{G(u,v)e^{B(q(v))}} - 1\right). \end{array}\right.$$

To finish the proof, it suffices to choose p and q so that

$$p'(u)^2 e^{-A(p(u))} = q'(v)^2 e^{-B(q(v))} = a^2.$$

Then (21.9) follows. ∎

The time has come to give a concrete realization of the angle functions of the previous section. Recall the discussion in the proof on page 401. Let $\mathcal{M}$ be a surface in $\mathbb{R}^3$, and fix a point $\mathbf{p} \in \mathcal{M}$. Let k_1, k_2 be the principal curvatures at $\mathbf{p}$, and let $\mathbf{e}_1, \mathbf{e}_2$ be corresponding unit principal vectors. If $\mathbf{u}_1$ is a unit asymptotic vector, there exists an angle θ such that

$$\mathbf{u}_1 = \mathbf{e}_1 \cos\theta + \mathbf{e}_2 \sin\theta. \tag{21.13}$$

By Lemma 13.27,

$$0 = k(\mathbf{u}_1) = k_1 \cos^2\theta + k_2 \sin^2\theta, \tag{21.14}$$

which shows that

$$\mathbf{u}_2 = \mathbf{e}_1 \cos\theta - \mathbf{e}_2 \sin\theta \tag{21.15}$$

is also an asymptotic vector. It follows that $\omega = 2\theta$ is the (or rather, an) angle between the two asymptotic directions.

Theorem 21.10. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface with $K \equiv -a^{-2} < 0$. Denote by 2θ the angle between asymptotic vectors. In a neighborhood of each point of $\mathcal{M}$ there exists a principal patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ such that*

(i) *the metric of* $\mathbf{x}$ *is given by* (21.5) *on page* 686.

(ii) *The principal curvatures satisfy*

$$a^2k_1^2 = \tan^2\theta, \qquad \textit{and} \qquad a^2k_2^2 = \cot^2\theta. \tag{21.16}$$

(iii) *The mean curvature is* $H = \pm\dfrac{1}{a\sin 2\theta}$.

(iv) *The second fundamental form of* $\mathbf{x}$ *is given by*

$$\mathbf{II} = \pm a\sin\theta\cos\theta(du^2 - dv^2).$$

Proof. Choose a principal patch and normalize it according to Theorem 21.9. Choose θ so that $\tan^2\theta = a^2k_1^2$; then

$$E\tan^2\theta = a^2Ek_1^2 = a^2 - E,$$

and so $E\sec^2\theta = a^2$, or $E = a^2\cos^2\theta$. The second part of (ii) is a consequence of the equation $k_1k_2 = -a^{-2}$ (see (21.14)). It also follows that $G = a^2\sin^2\theta$. Hence (i) holds.

For (iii), observe that

$$2H = k_1 + k_2 = \pm\Big(\frac{\tan\theta}{a} + \frac{\cot\theta}{a}\Big) = \pm\frac{2}{a\sin(2\theta)}.$$

Let e, f, g be the coefficients of the second fundamental form of $\mathbf{x}$. Lemma 15.5 on page 466 implies that $f = 0$. To prove (iii), we use Corollary 13.33 on page 405 to compute

$$a\,e = ak_1E = \pm a^2\tan\theta\cos^2\theta = \pm a\sin\theta\cos\theta.$$

Similarly, $g = \mp a\sin\theta\cos\theta$. ∎

Part (i) of the above theorem asserts that $\mathbf{x}$ is a Tchebyshef patch of the second kind, and we record

Definition 21.11. *A* ***Tchebyshef principal patch*** *of* ***radius*** $a > 0$ *and* ***angle function*** θ *is patch satisfying* (i) *and* (ii).

Ordinarily, the principal curvatures of a patch $\mathbf{x}$ depend on the choice of immersion of the patch into $\mathbb{R}^3$. But for a Tchebyshef principal patch of constant Gaussian curvature $-a^{-2}$, the metric determines the principal curvatures.

Given a Tchebyshef principal patch $\mathbf{x}$, the associated map $\mathbf{y}$ in Lemma 21.5 will be an asymptotic patch. For (21.13) and (21.15) are unit vectors tangent to the coordinate curves defining the metric (21.1). We call $\mathbf{y}$ a ***Tchebyshef asymptotic patch***. Thus, one feature of Tchebyshef patches is the easy way in which an asymptotic patch can be converted to a principal patch, and vice versa; this feature is not shared by general patches. When each parametrization is plotted, the resulting images give us distinct visual information.

21.3 The Sine–Gordon Equation

Finding surfaces of constant negative curvature in $\mathbb{R}^3$ is closely related to solving a famous partial differential equation, called the ***sine–Gordon equation***[2]. Let $\theta = \theta(u, v)$ be a function of two variables. There are two versions. The first,

$$\theta_{uu} - \theta_{vv} = \sin\theta\cos\theta, \tag{21.17}$$

is related by Lemma 21.6 to finding principal patch parametrizations of constant negative curvature.

Let $u = p + q$ and $v = p - q$; then

$$\frac{\partial}{\partial p} = \frac{\partial}{\partial u} + \frac{\partial}{\partial v} \qquad \text{and} \qquad \frac{\partial}{\partial q} = \frac{\partial}{\partial u} - \frac{\partial}{\partial v},$$

so that

$$\frac{\partial^2}{\partial p\,\partial q} = \frac{\partial^2}{\partial u^2} - \frac{\partial^2}{\partial v^2}.$$

Hence, if $\omega = 2\theta$, we have $\theta_{pq} = \theta_{uu} - \theta_{vv}$, and

$$\omega_{pq} = \sin\omega. \tag{21.18}$$

This is the version of the sine–Gordon equation that is related, by Lemma 21.2, to finding asymptotic patch parametrizations of constant negative curvature. Clearly, a change of coordinates converts solutions of one version into solutions of the other version, as in Lemma 21.5.

Here is a basic existence theorem for the sine–Gordon equation. The proof ([Bian, volume 1, pages 660–664]) resembles the method of Picard iteration used to establish the existence of solutions to certain first-order differential equations.

Theorem 21.12. *Let $\phi, \psi\colon \mathbb{R} \to \mathbb{R}$ be functions possessing first derivatives such that $\phi(0) = \psi(0)$. Then the sine–Gordon equation* (21.18) *has a unique solution ω satisfying $\omega(p, 0) = \phi(p)$ and $\omega(0, q) = \psi(q)$.*

It is natural to look for solutions of the sine–Gordon equation (21.17) that depend on only one variable.

Lemma 21.13. *Let θ be a solution of* (21.17)*, depending on u alone. Then there exists a constant b such that*

$$\theta_u^2 = b^2 - \cos^2\theta. \tag{21.19}$$

Proof. Since θ is a function of u alone, (21.17) can be multiplied by θ_u and written as

$$2\theta_{uu}\theta_u = -(-2\cos\theta\sin\theta)\theta_u. \tag{21.20}$$

Integrating (21.20) yields (21.19). ∎

[2] A variant of the *Klein*–Gordon equation $\theta_{uu} - \theta_{vv} = \theta$

To assist in subsequent applications of this lemma, we record some useful trigonometric identities.

Lemma 21.14. *Suppose* $u = \log\tan(\theta/2)$. *Then*

$$\sin\theta = \operatorname{sech} u \qquad \textit{and} \qquad \cos\theta = -\tanh u.$$

Proof. We have

$$\operatorname{sech} u = \frac{2}{e^u + e^{-u}} = \frac{2}{\tan(\theta/2) + \cot(\theta/2)} = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} = \sin\theta,$$

and

$$\tanh u = \frac{e^u - e^{-u}}{e^u + e^{-u}} = \frac{\tan(\theta/2) - \cot(\theta/2)}{\tan(\theta/2) + \cot(\theta/2)} = \sin^2\frac{\theta}{2} - \cos^2\frac{\theta}{2} = -\cos\theta. \quad \blacksquare$$

21.4 Tchebyshef Patches on Surfaces of Revolution

We first determine the general form of a Tchebyshef principal parametrization of a surface of revolution with constant negative Gaussian curvature.

Theorem 21.15. *Let* $\mathcal{M}$ *be a surface of revolution with* $K \equiv -a^{-2} < 0$, *and let* θ *be an angle function satisfying* $\theta_v = 0$ *and* (21.19). *A Tchebyshef principal patch on* $\mathcal{M}$ *is given by*

$$\mathbf{x}(u,v) = \Big(\frac{a}{b}\sin(\theta(u))\cos(bv),\ \frac{a}{b}\sin(\theta(u))\sin(bv),\ \psi(u)\Big), \tag{21.21}$$

where

$$\psi'(u) = \pm\frac{a}{b}\cos^2\big(\theta(u)\big). \tag{21.22}$$

Proof. The surface of revolution $\mathcal{M}$ can be parametrized locally as

$$\mathbf{x}(u,v) = \big(\alpha\sin(\theta(u))\cos(\beta v),\ \alpha\sin(\theta(u))\sin(\beta v),\ \psi(u)\big), \tag{21.23}$$

for some constants α, β and another function ψ. For this is equivalent to the standard parametrization on page 462, with u and v reversed. We compute

$$\begin{cases} \mathbf{x}_u(u,v) \ = \ \big(\alpha\cos\theta\cos(\beta v)\theta_u,\ \alpha\cos\theta\sin(\beta v)\theta_u,\ \psi'(u)\big), \\ \mathbf{x}_v(u,v) \ = \ \big(-\alpha\beta\sin\theta\sin(\beta v),\ \alpha\beta\sin\theta\cos(\beta v),\ 0\big). \end{cases}$$

Next we impose the condition that $\mathbf{x}$ be a Tchebyshef principal patch. Firstly, $a^2\sin^2\theta = G = \alpha^2\beta^2\sin^2\theta$ implies that $\alpha^2\beta^2 = a^2$; similarly,

$$a^2\cos^2\theta = E = \alpha^2(\cos^2\theta)\theta_u^2 + \psi'(u)^2.$$

The condition that $\mathcal{M}$ have constant negative curvature is (21.19), and we get

$$\psi'(u)^2 = \cos^2\theta\big(a^2 - \alpha^2 b^2 + \alpha^2\cos^2\theta\big). \tag{21.24}$$

When we choose $\alpha = a/b$ and $\beta = b$, we get (21.22). $\blacksquare$

Next, we examine the special case of the preceding theorem in which $b^2 = 1$.

Theorem 21.16. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch of radius $a > 0$ whose angle function θ satisfies the partial differential equations*

$$\theta_u^2 = \sin^2\theta \qquad \text{and} \qquad \theta_v = 0, \tag{21.25}$$

and for which

$$\mathbf{x}(u,v) = \big(\varphi(u)\cos v,\ \varphi(u)\sin v,\ \psi(u)\big). \tag{21.26}$$

Then

(i) $\mathbf{x}$ *has constant negative curvature* $-a^{-2}$.

(ii) φ *and* ψ *are given by*

$$\varphi(u) = \alpha\, a\,\operatorname{sech}(u+C) \quad \text{and} \quad \psi(u) = \beta\, a\big(u + B - \tanh(u+C)\big),$$

where α *and* β *are* ± 1*, and* B, C *are constants.*

(iii) *Up to a Euclidean motion,* $\mathbf{x}$ *is a reparametrization of a pseudosphere of radius* a.

(iv) *Any angle function is given by* $\theta = 2\delta \arctan(e^{\gamma(u+C)})$*, where* γ *and* δ *are* ± 1.

Proof. From (21.25), we get $\theta_{uu} = \sin\theta\cos\theta$; Lemma 21.6 implies that $\mathbf{x}$ has $K \equiv -a^{-2} < 0$, proving (i). Furthermore, the first equation of (21.25) can be rewritten as

$$\theta_u = \gamma \sin\theta, \tag{21.27}$$

where $\gamma = \pm 1$, and the solution to 21.25 is

$$\tan\left|\frac{\theta}{2}\right| = \frac{1}{|\csc\theta + \cot\theta|} = e^{\gamma(u+C)},$$

where C is a constant. Whence (iv).

From (21.27), (iv) and Lemma 21.14 we get

$$\gamma\,\theta_u = \sin\theta = \delta\,\sin(\delta\theta) = \delta\,\operatorname{sech}\big(\gamma(u+C)\big) = \delta\,\operatorname{sech}(u+C) \tag{21.28}$$

and

$$\cos\theta = \cos(\delta\theta) = -\tanh\big(\gamma(u+C)\big) = -\gamma\tanh(u+C). \tag{21.29}$$

Now suppose that $\mathbf{x}$ is given by (21.26), so that

$$G = \varphi(u)^2 \qquad \text{and} \qquad E = \varphi'(u)^2 + \psi'(u)^2.$$

For $\mathbf{x}$ to be a Tchebyshef principal patch, we must have $\varphi(u)^2 = G = a^2 \sin^2\theta$, so that

$$\varphi(u) = \alpha\, a \sin\theta, \tag{21.30}$$

where $\alpha = \pm 1$. Then (21.28) and (21.30) imply that

$$\varphi(u) = \alpha\, \delta\, a \operatorname{sech}(u + C). \tag{21.31}$$

Without loss of generality, we can assume $\delta = 1$ in (21.31), and so we obtain the first equation of (ii).

Similarly, the fact that $\varphi'(u)^2 + \psi'(u)^2 = E = a^2 \cos^2\theta$, combined with (21.30) and (21.27), implies that

$$\begin{aligned} \psi'(u)^2 \;=\; E - \varphi'(u)^2 \;&=\; a^2\cos^2\theta - a^2(\cos^2\theta)\theta_u^2 \\ &=\; a^2(\cos^2\theta)\big(1 - \theta_u^2\big) = a^2\cos^4\theta. \end{aligned} \tag{21.32}$$

From (21.32) and (21.29) we get $\psi'(u)^2 = a^2\big(\tanh(u+C)\big)^4$, so that

$$\psi'(u) = \beta\, a\big(\tanh(u+C)\big)^2, \tag{21.33}$$

where $\beta = \pm 1$. Integrating (21.33), we find that

$$\psi(u) = \beta\, a\big(u + B - \tanh(u+C)\big), \tag{21.34}$$

where B is the constant of integration, completing the proof of (ii).

Written out, (21.26) now becomes

$$\mathbf{x}(u,v) = a\left(\frac{\alpha\cos v}{\cosh(u+C)},\ \frac{\alpha\sin v}{\cosh(u+C)},\ \beta\big(u + B - \tanh(u+C)\big)\right),$$

where both α and β are ± 1. A translation along the z-axis combined with the change of variable $u \mapsto u - C$ transforms this into

$$\mathbf{x}(u,v) = a\left(\frac{\alpha\cos v}{\cosh u},\ \frac{\alpha\sin v}{\cosh u},\ \beta\,(u - \tanh u)\right). \tag{21.35}$$

Let $u = \log\tan(q/2)$ and $v = p + \pi(\alpha - 1)/2$. Lemma 21.14 gives

$$\operatorname{sech} u = \sin q \qquad \text{and} \qquad u - \tanh u = \cos q + \log\big(\tan\tfrac{q}{2}\big).$$

Furthermore, $\alpha\cos v = \cos p$ and $\alpha\sin v = \sin p$. It follows that

$$\mathbf{x}(u,v) = a\left(\cos p\sin q,\ \sin p\sin q,\ \beta\Big(\cos q + \log\big(\tan\tfrac{q}{2}\big)\Big)\right). \tag{21.36}$$

When $\beta = 1$, the right-hand side of (21.36) is equivalent to that of the equation displayed on page 480. The choice $\beta = -1$ leads to the same pseudosphere reflected in the xy-plane. Thus (iii) is proved. ∎

Choosing $\alpha = \beta = 1$ in (21.35) leads us to make

Definition 21.17. *Let $a > 0$. The* ***Tchebyshef principal parametrization*** *of a pseudosphere of radius a is the mapping from $\mathbb{R}^2 \to \mathbb{R}^3$ given by*

$$(u, v) \;\mapsto\; a\left(\frac{\cos v}{\cosh u},\; \frac{\sin v}{\cosh u},\; u - \tanh u\right). \tag{21.37}$$

Any angle function of a pseudosphere is given by $\theta(u) = \pm 2\arctan(e^{\pm u})$.

The image produced by the Tchebyshef parametrization in Figure 21.1 looks similar to that of the standard parametrization in Figure 15.10 on page 480, but the parameter lines are more evenly spaced.

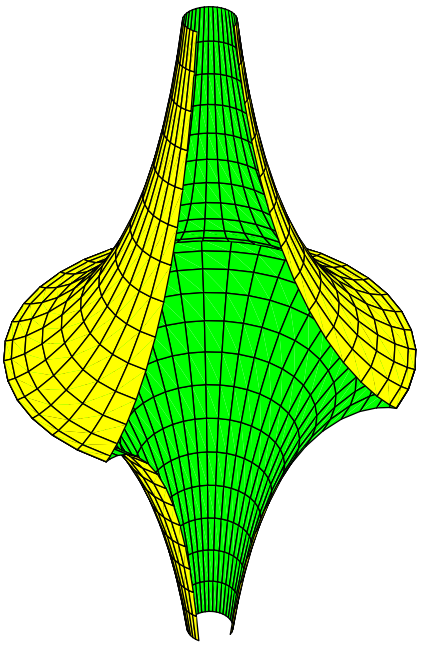


Figure 21.1: Tchebyshef principal parametrization of a pseudosphere

The asymptotic Tchebyshef parametrization of a pseudosphere can be found by substituting $u = p + q$ and $v = p - q$ into (21.37). It is

$$(p, q) \mapsto a\left(\frac{\cos(p-q)}{\cosh(p+q)},\; \frac{\sin(p-q)}{\cosh(p+q)},\; p + q - \tanh(p+q)\right). \tag{21.38}$$

The trace with $-\pi/2 \leqslant p, q \leqslant \pi/2$ is shown in Figure 21.2. In spite of the fact that a pseudosphere has singularities along the rim, the asymptotic curves are smooth across the rim.

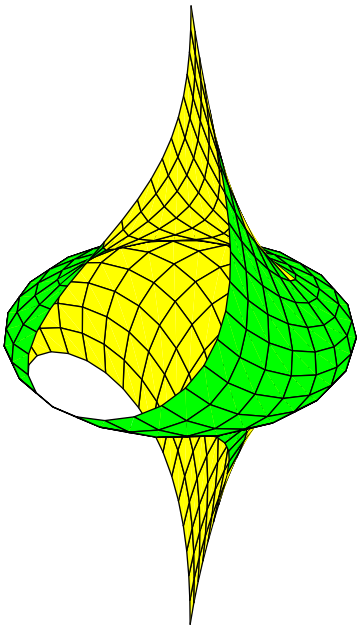


Figure 21.2: Tchebyshef asymptotic parametrization of a pseudosphere

Next we turn to finding Tchebyshef principal patches on the other constant negative curvature surfaces of revolution. We need the Jacobi elliptic functions $\mathrm{cn}, \mathrm{am}, \mathrm{dn}$, in addition to the elliptic integral of the second kind discussed in Section 15.5. The next result follows from Theorem 21.15, after verifying that the relevant functions satisfy equation (21.22).

Theorem 21.18. *A Tchebyshef principal patch on a constant negative curvature surface of revolution of conic type is given by mapping* (u, v) *to*

$$\left(\frac{a}{b}\,\mathrm{cn}\Big(bu, \frac{1}{b^2}\Big)\cos(bv),\ \ \frac{a}{b}\,\mathrm{cn}\Big(bu, \frac{1}{b^2}\Big)\sin(bv),\ \ \pm a\Big(bu - \mathrm{E}\Big(\mathrm{am}\Big(bu, \frac{1}{b^2}\Big)\Big|\frac{1}{b^2}\Big)\Big)\right),$$

where $b > 1$.

Similarly, for $0 < b < 1$, we can use Notebook 21 to prove

Theorem 21.19. *A Tchebyshef principal patch on a constant negative curvature surface of revolution of hyperbolic type is given by mapping* (u, v) *to*

$$\frac{a}{b}\Big(\mathrm{dn}(u, b^2)\cos(bv),\ \ \mathrm{dn}(u, b^2)\sin(bv),\ \ \pm\Big(u - \mathrm{E}\big(\mathrm{am}(u, b^2)\ \big|\ b^2\big)\Big)\Big).$$

The combined results are shown in Figures 21.3 and 21.4.

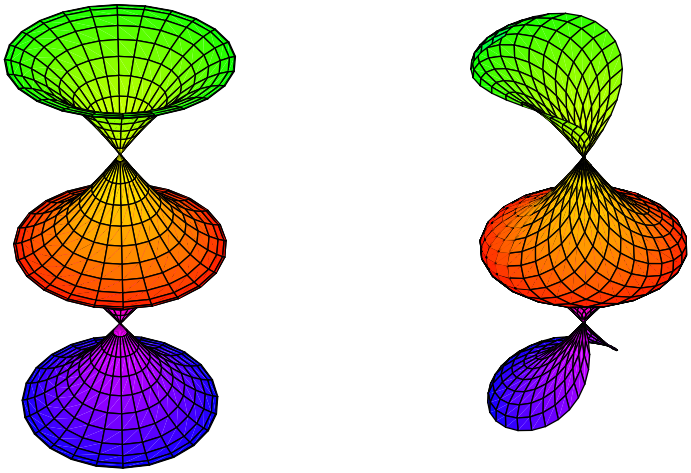


Figure 21.3: Principal and asymptotic parametrizations of a constant negative curvature surface of revolution of conic type

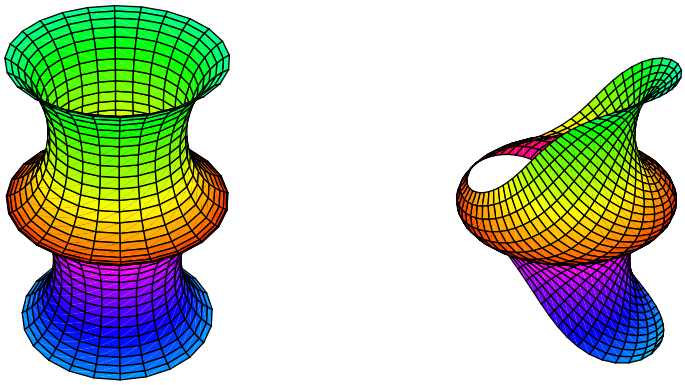


Figure 21.4: Principal and asymptotic parametrizations of a constant negative curvature surface of revolution of hyperbolic type

21.5 The Bianchi Transform

There are several constructions of new surfaces of constant negative curvature from old: the transforms of Bianchi, Lie and Bäcklund. Although the Bianchi and Lie transforms are special cases of a Bäcklund transformation, we treat them first because they are simpler.

The basic set-up is as follows.

Definition 21.20. *Let $\mathcal{M}$ be a surface in $\mathbb{R}^3$ whose Gaussian curvature has the constant value $-a^{-2}$, where $a > 0$. We say that a surface $\mathcal{N}$ with unit normal vector field $\mathbf{U}_{\mathcal{N}}$ is a* ***Bianchi transform*** *of $\mathcal{M}$ provided there is a map $\Phi\colon \mathcal{M} \to \mathcal{N}$ such that for each $\mathbf{p} \in \mathcal{M}$ the following relations hold:*

(21.39) $\|\Phi(\mathbf{p}) - \mathbf{p}\| = a$;

(21.40) $\Phi(\mathbf{p}) - \mathbf{p}$ *is parallel to a tangent vector to $\mathcal{M}$ at $\mathbf{p}$;*

(21.41) $\mathbf{U}_{\mathcal{N}}\big(\Phi(\mathbf{p})\big)$ *is parallel to a tangent vector to $\mathcal{M}$ at $\mathbf{p}$ and perpendicular to $\Phi(\mathbf{p}) - \mathbf{p}$.*

Bianchi in [Bian] calls $\mathcal{N}$ a ***complementary surface*** of $\mathcal{M}$.

Figure 21.5 shows the geometrical idea underlying the Bianchi tranform. In this case, the initial surface $\mathcal{M}$ is part of a circular cone and therefore has zero Gaussian curvature, so (21.39) fails. Nonetheless, the construction illustrates the meaning of (21.40). The new surface is formed by translating a point $\mathbf{p}$ of the cone in a tangent direction at $\mathbf{p}$, which we have indicated by a line segment. Condition (21.41) implies that $\Phi(\mathbf{p}) - \mathbf{p}$ is also tangent to the new surface at $\Phi(\mathbf{p})$, which is approximately true for the point we have chosen.

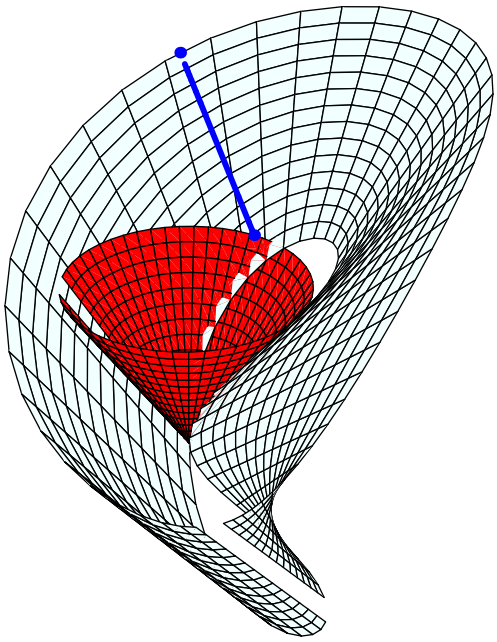


Figure 21.5: A surface formed by translating points of a circular cone

Lemma 21.21. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ with $K \equiv -a^{-2}$, and let $\theta\colon \mathcal{M} \to \mathbb{R}^3$ be any function such that $ak_1 = -\tan\theta$, so that*

$$E = a^2\cos^2\theta \qquad \text{and} \qquad G = a^2\sin^2\theta,$$

in agreement with (21.16). *Let $\widetilde{\mathbf{x}} = \Phi \circ \mathbf{x}$ be a patch that satisfies* (21.39) *and* (21.40). *Then there exists a function $\tilde{\theta}\colon \mathcal{U} \to \mathbb{R}^3$ such that*

$$\widetilde{\mathbf{x}} = \mathbf{x} + \frac{\cos\tilde{\theta}}{\cos\theta}\mathbf{x}_u + \frac{\sin\tilde{\theta}}{\sin\theta}\mathbf{x}_v. \tag{21.42}$$

We call $\tilde{\theta}$ an ***angle function*** *of the Bianchi transform relative to θ.*

Proof. Equation (21.40) implies that there exist functions A and B such that $\widetilde{\mathbf{x}} = \mathbf{x} + A\mathbf{x}_u + B\mathbf{x}_v$. Then from (21.39) we get

$$a^2 = \|\widetilde{\mathbf{x}} - \mathbf{x}\|^2 = A^2\|\mathbf{x}_u\|^2 + B^2\|\mathbf{x}_v\|^2 = a^2(A^2\cos^2\theta + B^2\sin^2\theta).$$

We can choose A and B, so that $A = \cos\tilde{\theta}/\cos\theta$ and $B = \sin\tilde{\theta}/\sin\theta$, and so we obtain (21.42). ∎

Next, we obtain a set (21.43) of partial differential equations (called the ***Bianchi–Darboux equations***) relating θ and $\tilde{\theta}$. They derive from (21.41).

Theorem 21.22. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ of constant negative curvature $-a^{-2}$. Let $\widetilde{\mathbf{x}}$ be a patch given by* (21.42). *Then $\widetilde{\mathbf{x}}$ is a Bianchi transform of $\mathbf{x}$ if and only if*

$$\begin{cases} \tilde{\theta}_u + \theta_v \;=\; \sin\tilde{\theta}\cos\theta, \\ \theta_u + \tilde{\theta}_v \;=\; -\cos\tilde{\theta}\sin\theta. \end{cases} \tag{21.43}$$

Proof. Suppose that $\widetilde{\mathbf{x}}$ is a Bianchi transform of $\mathbf{x}$, so that $\widetilde{\mathbf{x}}$ is given by (21.42). We differentiate $\widetilde{\mathbf{x}}$ with respect to u and obtain

$$\widetilde{\mathbf{x}}_u = \mathbf{x}_u + \Big(\frac{\cos\tilde{\theta}}{\cos\theta}\Big)_u\mathbf{x}_u + \frac{\cos\tilde{\theta}}{\cos\theta}\mathbf{x}_{uu} + \Big(\frac{\sin\tilde{\theta}}{\sin\theta}\Big)_u\mathbf{x}_v + \frac{\sin\tilde{\theta}}{\sin\theta}\mathbf{x}_{uv}.$$

By (17.8) on page 538 and Lemma 21.7, we have

$$\begin{aligned} \mathbf{x}_{uu} \;&=\; \Gamma^1_{11}\mathbf{x}_u + \Gamma^2_{11}\mathbf{x}_v + e\,\mathbf{U} \\ &=\; -(\tan\theta)\theta_u\mathbf{x}_u + (\cot\theta)\theta_v\mathbf{x}_v - a\sin\theta\cos\theta\,\mathbf{U}, \end{aligned}$$

and

$$\mathbf{x}_{uv} = \Gamma^1_{12}\mathbf{x}_u + \Gamma^2_{12}\mathbf{x}_v = -(\tan\theta)\theta_v\mathbf{x}_u + (\cot\theta)\theta_u\mathbf{x}_v.$$

Hence

$$\widetilde{\mathbf{x}}_u = A\mathbf{x}_u + B\mathbf{x}_v - a\sin\theta\cos\tilde{\theta}\,\mathbf{U}, \tag{21.44}$$

where

$$A = 1 + \Big(\frac{\cos\tilde{\theta}}{\cos\theta}\Big)_u - \frac{(\sin\theta\cos\tilde{\theta})\theta_u}{\cos^2\theta} - \frac{(\sin\tilde{\theta})\theta_v}{\cos\theta} = \frac{\cos\theta - (\tilde{\theta}_u + \theta_v)\sin\tilde{\theta}}{\cos\theta},$$

and

$$B = \Big(\frac{\sin\tilde{\theta}}{\sin\theta}\Big)_u + \frac{(\cos\tilde{\theta})\theta_v}{\sin\theta} + \frac{(\sin\tilde{\theta}\cos\theta)\theta_u}{\sin^2\theta} = \frac{(\tilde{\theta}_u + \theta_v)\cos\tilde{\theta}}{\sin\theta}.$$

Referring to (21.44), and carrying out the analogous procedure for $\widetilde{\mathbf{x}}_v$, we obtain the pair of equations

$$\begin{cases} \widetilde{\mathbf{x}}_u = \dfrac{\cos\theta - (\tilde{\theta}_u + \theta_v)\sin\tilde{\theta}}{\cos\theta}\mathbf{x}_u + \dfrac{(\tilde{\theta}_u + \theta_v)\cos\tilde{\theta}}{\sin\theta}\mathbf{x}_v - a\sin\theta\cos\tilde{\theta}\,\mathbf{U}, \\ \widetilde{\mathbf{x}}_v = -\dfrac{(\tilde{\theta}_v + \theta_u)\sin\tilde{\theta}}{\cos\theta}\mathbf{x}_u + \dfrac{\sin\theta + (\tilde{\theta}_v + \theta_u)\cos\tilde{\theta}}{\sin\theta}\mathbf{x}_v + a\sin\tilde{\theta}\cos\theta\,\mathbf{U}. \end{cases}$$

Let $\widetilde{\mathbf{U}}$ denote the unit normal vector field of $\widetilde{\mathbf{x}}$. Then (21.41) and (21.42) imply that up to sign

$$\widetilde{\mathbf{U}} = -\frac{\sin\tilde{\theta}}{a\cos\theta}\mathbf{x}_u + \frac{\cos\tilde{\theta}}{a\sin\theta}\mathbf{x}_v. \tag{21.45}$$

It now follows that $\widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{U}}$ equals

$$\begin{aligned} &\left(\frac{\cos\theta - (\tilde{\theta}_u + \theta_v)\sin\tilde{\theta}}{\cos\theta}\mathbf{x}_u + \frac{(\tilde{\theta}_u + \theta_v)\cos\tilde{\theta}}{\sin\theta}\mathbf{x}_v\right) \cdot \left(-\frac{\sin\tilde{\theta}}{a\cos\theta}\mathbf{x}_u + \frac{\cos\tilde{\theta}}{a\sin\theta}\mathbf{x}_v\right) \\ &\quad = \left(\frac{\cos\theta - (\tilde{\theta}_u + \theta_v)\sin\tilde{\theta}}{\cos\theta}\right)\Big(-\frac{\sin\tilde{\theta}}{a\cos\theta}\Big)a^2\cos^2\theta \\ &\qquad\qquad + \left(\frac{(\tilde{\theta}_u + \theta_v)\cos\tilde{\theta}}{\sin\theta}\right)\Big(\frac{\cos\tilde{\theta}}{a\sin\theta}\Big)a^2\sin^2\theta \\ &\quad = a\big(-\cos\theta\sin\tilde{\theta} + (\tilde{\theta}_u + \theta_v)\sin^2\tilde{\theta} + (\tilde{\theta}_u + \theta_v)\cos^2\tilde{\theta}\big) \\ &\quad = a\big(-\sin\tilde{\theta}\cos\theta + \tilde{\theta}_u + \theta_v\big). \end{aligned}$$

A similar calculation gives

$$\widetilde{\mathbf{x}}_v \cdot \widetilde{\mathbf{U}} = a\big(\tilde{\theta}_v + \theta_u + \cos\tilde{\theta}\sin\theta\big).$$

The result (21.43) now follows from the fact that $\widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{U}} = \widetilde{\mathbf{x}}_v \cdot \widetilde{\mathbf{U}} = 0$.

Conversely, suppose that (21.43) holds. Then the pair of expressions for $\widetilde{\mathbf{x}}_u, \widetilde{\mathbf{x}}_v$ can be rewritten as

$$\begin{cases} \widetilde{\mathbf{x}}_u = \cos^2\tilde{\theta}\,\mathbf{x}_u + \cos\tilde{\theta}\sin\tilde{\theta}\cot\theta\,\mathbf{x}_v - a\cos\tilde{\theta}\sin\theta\,\mathbf{U}, \\ \widetilde{\mathbf{x}}_v = \sin\tilde{\theta}\cos\tilde{\theta}\tan\theta\,\mathbf{x}_u + \sin^2\tilde{\theta}\,\mathbf{x}_v + a\sin\tilde{\theta}\cos\theta\,\mathbf{U}. \end{cases}$$

Defining $\widetilde{\mathbf{U}}$ by (21.45), we see that $\widetilde{\mathbf{U}}$ is a unit normal vector field of $\widetilde{\mathbf{x}}$, and is also perpendicular to $\widetilde{\mathbf{x}} - \mathbf{x}$. Hence, $\widetilde{\mathbf{x}}$ is a Bianchi transform of $\mathbf{x}$. ∎

The equations (21.43) are analogous to the Cauchy–Riemann equations on page 626 that we shall use in the next chapter.

Theorem 21.23. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ with $K \equiv -a^{-2}$, where $a > 0$. Let $\tilde{\theta}$ denote an angle function of the Bianchi transform $\widetilde{\mathbf{x}}$ of $\mathbf{x}$. Then $\widetilde{\mathbf{x}}$ is a Tchebyshef principal patch of constant negative curvature $-a^{-2}$, and $\tilde{\theta}$ is the angle function of $\widetilde{\mathbf{x}}$.*

Proof. From (21.43), we get

$$\begin{aligned} \tilde{\theta}_{uu} - \tilde{\theta}_{vv} &= \big(-\theta_v + \sin\tilde{\theta}\cos\theta\big)_u - \big(-\theta_u - \cos\tilde{\theta}\sin\theta\big)_v \\ &= \tilde{\theta}_u\cos\tilde{\theta}\cos\theta - \theta_u\sin\tilde{\theta}\sin\theta - \tilde{\theta}_v\sin\tilde{\theta}\sin\theta + \theta_v\cos\tilde{\theta}\cos\theta \\ &= (\tilde{\theta}_u + \theta_v)\cos\tilde{\theta}\cos\theta - (\tilde{\theta}_v + \theta_u)\sin\tilde{\theta}\sin\theta \\ &= \sin\tilde{\theta}\cos\tilde{\theta}. \end{aligned}$$

Now Lemma 21.6 implies that $\widetilde{\mathbf{x}}$ has constant negative curvature $-a^{-2}$.

Next, it follows from (21.5) that

$$\begin{aligned} \widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{x}}_u &= a^2\big(\cos^2\theta(\cos^4\tilde{\theta} + \cos^2\tilde{\theta}\sin^2\tilde{\theta}) + \cos^2\tilde{\theta}\sin^2\theta\big) \\ &= a^2\cos^2\tilde{\theta}. \end{aligned}$$

Similarly, $\widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{x}}_v = 0$ and $\widetilde{\mathbf{x}}_v \cdot \widetilde{\mathbf{x}}_v = a^2\sin^2\tilde{\theta}$, as in (21.5). ∎

21.6 Moving Frames on Surfaces in $\mathbb{R}^3$

In Chapters 7 and 8, we made extensive use of the Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$, defined on page 196, to describe the geometry of a space curve. One of the main advantages of such a frame is that it is orthonormal; that is, $\mathbf{T}, \mathbf{N}, \mathbf{B}$ are mutually perpendicular and each has unit length. In this section, we shall use a frame field analogous to $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ to study the geometry of a principal patch. It is really a variant of the Darboux frame field, defined in Section 17.4 to study curves on a surface in $\mathbb{R}^3$.

Definition 21.24. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a principal patch whose metric is* $ds^2 = E\,du^2 + G\,dv^2$. *Put*

$$\mathbf{E}_1 = \frac{\mathbf{x}_u}{\sqrt{E}} \qquad \text{and} \qquad \mathbf{E}_2 = \frac{\mathbf{x}_v}{\sqrt{G}}.$$

Then the triple $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{U}\}$ (*where* $\mathbf{U}$ *denotes the unit normal of* $\mathbf{x}$) *is called the* ***moving orthonormal frame*** *of* $\mathbf{x}$.

The Frenet formulas (Theorem 7.10 on page 197) express the derivatives $\mathbf{T}', \mathbf{N}', \mathbf{B}'$ in terms of $\mathbf{T}, \mathbf{N}, \mathbf{B}$. The analog for surfaces is

Theorem 21.25. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a principal patch whose metric is* $ds^2 = E\,du^2 + G\,dv^2$. *The derivatives of* $\mathbf{E}_1, \mathbf{E}_2, \mathbf{U}$ *satisfy*

$$\begin{cases} \mathbf{E}_{1u} = -\dfrac{(\sqrt{E})_v}{\sqrt{G}}\mathbf{E}_2 + \dfrac{e}{\sqrt{E}}\mathbf{U}, & \mathbf{E}_{1v} = \dfrac{(\sqrt{G})_u}{\sqrt{E}}\mathbf{E}_2, \\ \mathbf{E}_{2u} = \dfrac{(\sqrt{E})_v}{\sqrt{G}}\mathbf{E}_1, & \mathbf{E}_{2v} = -\dfrac{(\sqrt{G})_u}{\sqrt{E}}\mathbf{E}_1 + \dfrac{g}{\sqrt{G}}\mathbf{U}, \\ \mathbf{U}_u = -\dfrac{e}{\sqrt{E}}\mathbf{E}_1, & \mathbf{U}_v = -\dfrac{g}{\sqrt{G}}\mathbf{E}_2. \end{cases} \tag{21.46}$$

Proof. Since $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{U}\}$ constitutes an orthonormal basis at each point on the image of $\mathbf{x}$, we may write

$$\mathbf{E}_{1u} = A_1\mathbf{E}_1 + A_2\mathbf{E}_2 + A_3\mathbf{U}, \tag{21.47}$$

where $A_i = \mathbf{E}_{1u} \cdot \mathbf{E}_i$ (and $\mathbf{E}_3$ temporarily denotes $\mathbf{U}$). Since the norm of $\mathbf{E}_1$ is constant, we have $A_1 = 0$. As for A_2, A_3 these can easily be found using the equations $\mathbf{x}_{uu} \cdot \mathbf{x}_v = -\frac{1}{2}E_v$ (that follows because $\mathbf{x}_u \cdot \mathbf{x}_v = 0$) and $\mathbf{x}_{uu} \cdot \mathbf{U} = e$. The derivatives of $\mathbf{E}_2$ and $\mathbf{U}$ are computed similarly (see Exercise 4). ∎

Corollary 21.26. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ *be a Tchebyshef principal patch whose metric is given by* (21.5). *Assume that* $0 < \theta < \pi/2$. *Then*

$$\begin{cases} \mathbf{E}_{1u} = \theta_v\,\mathbf{E}_2 - \sin\theta\,\mathbf{U}, & \mathbf{E}_{1v} = \theta_u\,\mathbf{E}_2, \\ \mathbf{E}_{2u} = -\theta_v\,\mathbf{E}_1, & \mathbf{E}_{2v} = -\theta_u\,\mathbf{E}_1 + \cos\theta\,\mathbf{U}, \\ \mathbf{U}_u = \sin\theta\,\mathbf{E}_1, & \mathbf{U}_v = -\cos\theta\,\mathbf{E}_2. \end{cases} \tag{21.48}$$

For a proof, see Exercise 5.

The Pseudosphere via Bianchi Transform

We shall apply the Bianchi transform in an appropriate way to the singular patch

$$\mathbf{x}(u, v) = (0,\ 0,\ \epsilon\, a\, u), \tag{21.49}$$

where $a > 0$ is a constant and $\epsilon = \pm 1$. Formally, this satisfies (21.5) with $\cos\theta = \epsilon$ and $\sin\theta = 0$. From (21.43), we get

$$\tilde{\theta}_u = \epsilon \sin\tilde{\theta} \qquad \text{and} \qquad \tilde{\theta}_v = 0.$$

Integrating the first equation, and translating to eliminate any constant of integration,

$$\log\tan\Big(\frac{\delta\tilde{\theta}}{2}\Big) = \epsilon\, u,$$

where $\delta = \pm 1$. Thus, there are four possible angle functions for the Bianchi transform; by Lemma 21.14 we obtain

$$\begin{cases} \cos\tilde{\theta} = \cos(\delta\tilde{\theta}) = -\tanh(\epsilon u) = -\epsilon\tanh u, \\ \sin\tilde{\theta} = \delta\sin(\delta\tilde{\theta}) = \delta\,\operatorname{sech}(\epsilon u) = \delta\,\operatorname{sech} u. \end{cases} \tag{21.50}$$

Strictly speaking, the moving orthonormal frame of $\mathbf{x}$ is not defined because its trace is 1-dimensional, but we are at liberty to declare

$$\mathbf{E}_1 = (0,0,1), \qquad \mathbf{E}_2 = (\cos v, \sin v, 0), \qquad \mathbf{U} = (-\sin v, \cos v, 0).$$

The resulting Bianchi transform of (21.49) is then given by

$$\begin{aligned} \widetilde{\mathbf{x}}(u,v) &= \mathbf{x}(u,v) + a\big(\cos\tilde{\theta}\,\mathbf{E}_1 + \sin\tilde{\theta}\,\mathbf{E}_2\big) \\ &= a\epsilon\,(0,0,u) - a\epsilon\tanh u(0,0,1) + a\delta\,\operatorname{sech} u(\cos v, \sin v, 0) \\ &= a\big(\delta\cos v\operatorname{sech} u,\ \delta\sin v\operatorname{sech} u,\ \epsilon(u - \tanh u)\big). \end{aligned}$$

Hence $\widetilde{\mathbf{x}}$ coincides with the patch

$$(u,v) \mapsto \left(\frac{a\delta\cos v}{\cosh u},\ \frac{a\delta\sin v}{\cosh u},\ a\epsilon(u - \tanh u)\right). \tag{21.51}$$

This parametrizes a pseudosphere, and coincides with (21.37) if $\delta = \epsilon = 1$.

Lemma 21.27. *Under the hypothesis of* Theorem 21.23*, the moving frame of the Bianchi transform* $\widetilde{\mathbf{x}}$ *of* $\mathbf{x}$ *is given by* $\{\widetilde{\mathbf{E}}_1, \widetilde{\mathbf{E}}_2, \widetilde{\mathbf{U}}\}$*, where*

$$\begin{cases} \widetilde{\mathbf{E}}_1 &= \cos\theta\cos\tilde{\theta}\,\mathbf{E}_1 + \cos\theta\sin\tilde{\theta}\,\mathbf{E}_2 - \sin\theta\,\mathbf{U}, \\ \widetilde{\mathbf{E}}_2 &= \sin\theta\cos\tilde{\theta}\,\mathbf{E}_1 + \sin\theta\sin\tilde{\theta}\,\mathbf{E}_2 + \cos\theta\,\mathbf{U}, \\ \widetilde{\mathbf{U}} &= -\sin\tilde{\theta}\,\mathbf{E}_1 + \cos\tilde{\theta}\,\mathbf{E}_2. \end{cases}$$

21.7 Kuen's Surface as Bianchi Transform of the Pseudosphere

Lemma 21.28. *Let $\widetilde{\mathbf{x}}$ be a Bianchi transform of the Tchebyshef principal patch* (21.37) *of a pseudosphere. Take $\theta = 2\arctan(e^u)$, and let $\tilde{\theta}$ be an angle function of $\widetilde{\mathbf{x}}$. Then*

$$\tilde{\theta}(u,v) = 2\arctan\Big(-\frac{v}{\cosh u}\Big). \tag{21.52}$$

Proof. It follows from Lemma 21.14 that

$$\cos\theta = -\tanh u \qquad \text{and} \qquad \sin\theta = \operatorname{sech} u. \tag{21.53}$$

Also,

$$\theta_u = \frac{2e^u}{1+e^{2u}} = \operatorname{sech} u, \quad \text{and} \quad \theta_v = 0. \tag{21.54}$$

From (21.43), (21.53) and (21.54) we get

$$\tilde{\theta}_u = -\sin\tilde{\theta}\tanh u, \tag{21.55}$$

and

$$\tilde{\theta}_v + \operatorname{sech} u = -\cos\tilde{\theta}\operatorname{sech} u. \tag{21.56}$$

The solution of the differential equation (21.55) is

$$\log\tan\Big|\frac{\tilde{\theta}}{2}\Big| = -\int \tanh u\, du = -\log\cosh u + \log A(v),$$

or

$$\tan\Big(\frac{\gamma\tilde{\theta}}{2}\Big) = A(v)\operatorname{sech} u, \tag{21.57}$$

where $\gamma = \pm 1$ and $A(v)$ is a constant of integration. Equation (21.56) can be rewritten as

$$\tilde{\theta}_v = -2\cos^2\frac{\theta}{2}\operatorname{sech} u,$$

and has solution

$$\tan\frac{\tilde{\theta}}{2} = -v\operatorname{sech} u + B(u),$$

where $B(u)$ is a constant of integration. From (21.57), we now see that

$$\gamma A(v) = -v, \qquad \text{and} \qquad B(u) = 0.$$

Hence we get (21.52). ∎

The calculations of the next lemma can be checked in Notebook 21.

Lemma 21.29. *If $\tilde{\theta}$ is given by* (21.52), *then*

$$\begin{cases} \cos\tilde{\theta} \;=\; -\dfrac{v^2-\cosh^2 u}{v^2+\cosh^2 u}, \\[2ex] \sin\tilde{\theta} \;=\; \dfrac{-2v\cosh u}{v^2+\cosh^2 u}. \end{cases} \tag{21.58}$$

Theorem 21.30. *Let $a = 1$. The Bianchi transform of the pseudosphere* (21.37) *is given by setting $\widetilde{\mathbf{x}}(u, v)$ equal to*

$$\left(\frac{2(\cos v + v \sin v)\cosh u}{v^2 + \cosh^2 u},\ \frac{2(\sin v - v \cos v)\cosh u}{v^2 + \cosh^2 u},\ u - \frac{\sinh 2u}{v^2 + \cosh^2 u}\right).$$

Proof. With the same notation as in Lemma 21.28, we have

$$\theta = 2\arctan(e^u), \qquad \cos\theta = -\tanh u, \qquad \sin\theta = \operatorname{sech} u. \tag{21.59}$$

Since

$$\begin{cases} \mathbf{x} &= \big(\cos v \operatorname{sech} u,\ \sin v \operatorname{sech} u,\ u - \tanh u\big), \\ \mathbf{x}_u &= \big(-\cos v \operatorname{sech} u \tanh u,\ -\sin v \operatorname{sech} u \tanh u,\ \tanh^2 u\big), \\ \mathbf{x}_v &= \big(-\sin v \operatorname{sech} u,\ \cos v \operatorname{sech} u,\ 0\big), \end{cases}$$

it follows from (21.59) that

$$\begin{cases} \dfrac{\mathbf{x}_u}{\cos\theta} = \big(\cos v \operatorname{sech} u,\ \sin v \operatorname{sech} u,\ -\tanh u\big), \\ \dfrac{\mathbf{x}_v}{\sin\theta} = \big(-\sin v,\ \cos v,\ 0\big). \end{cases} \tag{21.60}$$

From (21.52) and (21.58) we have that

$$\begin{cases} \tilde\theta = 2\arctan\Big(-\dfrac{v}{\cosh u}\Big), \\ \cos\tilde\theta = -\dfrac{v^2 - \cosh^2 u}{v^2 + \cosh^2 u}, \\ \sin\tilde\theta = -\dfrac{2v\cosh u}{v^2 + \cosh^2 u}. \end{cases} \tag{21.61}$$

The second equation of (21.61) implies that

$$1 + \cos\tilde\theta = 1 - \frac{v^2 - \cosh^2 u}{v^2 + \cosh^2 u} = \frac{2\cosh^2 u}{v^2 + \cosh^2 u}.$$

Therefore,

$$\begin{aligned} \widetilde{\mathbf{x}} &= \mathbf{x} + \frac{\cos\tilde\theta}{\cos\theta}\mathbf{x}_u + \frac{\sin\tilde\theta}{\sin\theta}\mathbf{x}_v \\ &= \big((1 + \cos\tilde\theta)\cos v \operatorname{sech} u - \sin\tilde\theta \sin v, \\ &\qquad (1 + \cos\tilde\theta)\sin v \operatorname{sech} u + \sin\tilde\theta\cos v,\ u - (1 + \cos\tilde\theta)\tanh u\big), \end{aligned}$$

which gives the required result. ∎

The principal patch $\widetilde{\mathbf{x}}$ defined by Theorem 21.30 and its asymptotic counterpart are shown in Figure 21.6, which should be compared to Figure 15.14 on page 484.

The equivalence of these surfaces is the subject of Exercise 8.

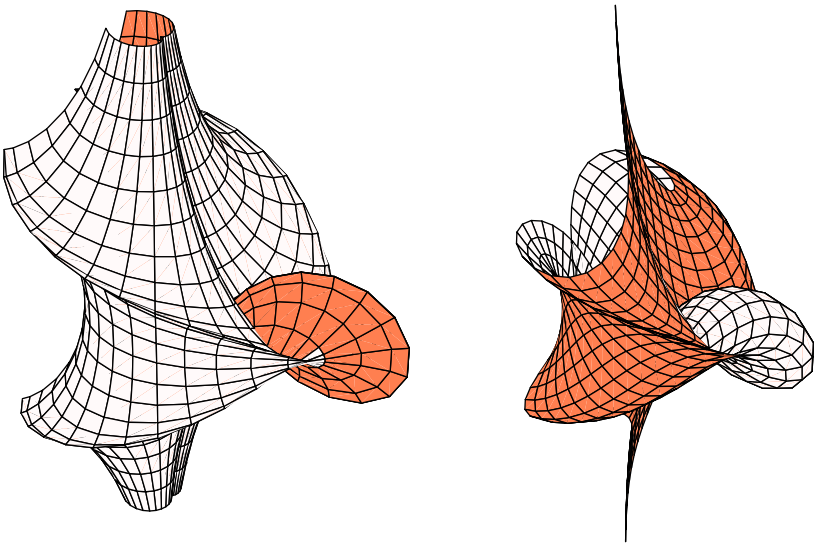


Figure 21.6: Principal and asymptotic parametrizations of Kuen's surface

21.8 The Bäcklund Transform

In a famous paper [Bck] (originally written in Swedish, but translated into English) Bäcklund[3] gave an important generalization of Bianchi's construction.

Definition 21.31. *Let $\mathcal{M}$ be a surface in $\mathbb{R}^3$ whose Gausian curvature has the constant value $-a^{-2}$. We say that a surface $\mathcal{N}$ with unit normal vector field $\mathbf{U}_{\mathcal{N}}$ is a* ***Bäcklund transform*** *of $\mathcal{M}$ of* ***inclination*** *σ if there is a map $\Phi\colon \mathcal{M} \to \mathcal{N}$ such that for each $\mathbf{p} \in \mathcal{M}$ the following relations hold:*

(21.62) $\|\Phi(\mathbf{p}) - \mathbf{p}\| = a\cos\sigma$;

(21.63) $\Phi(\mathbf{p}) - \mathbf{p}$ *is parallel to a tangent vector to $\mathcal{M}$ at $\mathbf{p}$;*

(21.64) $\mathbf{U}_{\mathcal{N}}\big(\Phi(\mathbf{p})\big)$ *is perpendicular to $\Phi(\mathbf{p}) - \mathbf{p}$;*

(21.65) *the angle between the normals $\mathbf{U}_{\mathcal{M}}(\mathbf{p})$ and $\mathbf{U}_{\mathcal{N}}\big(\Phi(\mathbf{p})\big)$ is $\pi/2 - \sigma$.*

---

[3] 

Albert Victor Bäcklund (1845–1922). Swedish mathematician. Rector of the University of Lund (1907–1909).

We can now present some generalizations of the previous section, though we omit some of the proofs in doing so. The analogue of Lemma 21.21 is

Lemma 21.32. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ with $K \equiv -a^{-2} < 0$, and let $\theta\colon \mathcal{M} \to \mathbb{R}^3$ be any function such that $ak_1 = -\tan\theta$, so that*

$$E = a^2\cos^2\theta \qquad \text{and} \qquad G = a^2\sin^2\theta.$$

Let $\widetilde{\mathbf{x}} = \Phi \circ \mathbf{x}$ be a patch that satisfies (21.62) *and* (21.63). *Then there exists a function $\tilde{\theta}\colon \mathcal{U} \to \mathbb{R}^3$ such that*

$$\widetilde{\mathbf{x}} = \mathbf{x} + \cos\sigma\left(\frac{\cos\tilde{\theta}}{\cos\theta}\mathbf{x}_u + \frac{\sin\tilde{\theta}}{\sin\theta}\mathbf{x}_v\right). \tag{21.66}$$

We call $\tilde{\theta}$ the ***angle function*** *of the Bäcklund transform of inclination σ relative to θ.*

The Bianchi–Darboux equations (21.43) are generalized in the next theorem to the ***Bäcklund–Darboux equations***.

Theorem 21.33. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ with $K \equiv -a^{-2} < 0$. Let $\widetilde{\mathbf{x}}$ be a patch given by* (21.66). *Then $\widetilde{\mathbf{x}}$ is a Bäcklund transform of inclination σ of $\mathbf{x}$ if and only if*

$$\begin{cases} \tilde{\theta}_u + \theta_v \;=\; \dfrac{\sin\tilde{\theta}\cos\theta + \sin\sigma\cos\tilde{\theta}\sin\theta}{\cos\sigma}, \\[2ex] \theta_u + \tilde{\theta}_v \;=\; -\dfrac{\cos\tilde{\theta}\sin\theta + \sin\sigma\sin\tilde{\theta}\cos\theta}{\cos\sigma}. \end{cases} \tag{21.67}$$

Proof. Suppose that $\widetilde{\mathbf{x}}$ is a Bäcklund transform of inclination σ of $\mathbf{x}$; then $\widetilde{\mathbf{x}}$ is given by (21.66). We differentiate $\widetilde{\mathbf{x}}$ with respect to u and obtain

$$\widetilde{\mathbf{x}}_u = \mathbf{x}_u + \cos\sigma\left(\Big(\frac{\cos\tilde{\theta}}{\cos\theta}\Big)_u\mathbf{x}_u + \frac{\cos\tilde{\theta}}{\cos\theta}\mathbf{x}_{uu} + \Big(\frac{\sin\tilde{\theta}}{\sin\theta}\Big)_u\mathbf{x}_v + \frac{\sin\tilde{\theta}}{\sin\theta}\mathbf{x}_{uv}\right). \tag{21.68}$$

Just as in the proof of Theorem 21.22, we have

$$\mathbf{x}_{uu} = -(\tan\theta)\theta_u\,\mathbf{x}_u + (\cot\theta)\theta_v\,\mathbf{x}_v - a\sin\theta\cos\theta\,\mathbf{U}$$

and

$$\mathbf{x}_{uv} = -(\tan\theta)\theta_v\mathbf{x}_u + (\cot\theta)\theta_u\mathbf{x}_v.$$

It follows that

$$\widetilde{\mathbf{x}}_u = A\,\mathbf{x}_u + B\,\mathbf{x}_v - a\cos\sigma\sin\theta\cos\tilde{\theta}\,\mathbf{U}, \tag{21.69}$$

where

$$A = \frac{\cos\theta - (\tilde\theta_u + \theta_v)\cos\sigma\sin\tilde\theta}{\cos\theta}, \quad B = \frac{(\tilde\theta_u + \theta_v)\cos\sigma\cos\tilde\theta}{\sin\theta}.$$

By (21.64), (21.65), (21.66), the unit normal vector field of $\widetilde{\mathbf{x}}$ is given up to sign by

$$\widetilde{\mathbf{U}} = -\frac{\cos\sigma\sin\tilde\theta}{a\cos\theta}\,\mathbf{x}_u + \frac{\cos\sigma\cos\tilde\theta}{a\sin\theta}\,\mathbf{x}_v + \sin\sigma\,\mathbf{U}. \tag{21.70}$$

Calculations using (21.69), an analogous formula for $\widetilde{\mathbf{x}}_v$, and (21.70) yield

$$\begin{cases} \widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{U}} = a\cos\sigma\big(-\sin\tilde\theta\cos\theta + (\tilde\theta_u + \theta_v)\cos\sigma - \sin\sigma\sin\theta\cos\tilde\theta\big), \\ \widetilde{\mathbf{x}}_v \cdot \widetilde{\mathbf{U}} = a\cos\sigma\big((\tilde\theta_v + \theta_u)\cos\sigma + \sin\theta\cos\tilde\theta + \sin\sigma\sin\tilde\theta\cos\theta\big). \end{cases}$$

Since $\widetilde{\mathbf{x}}_u \cdot \widetilde{\mathbf{U}} = \widetilde{\mathbf{x}}_v \cdot \widetilde{\mathbf{U}} = 0$, these equations imply (21.67).

Conversely, suppose that (21.67) holds. Referring to (21.69), we can write

$$\begin{aligned} \widetilde{\mathbf{x}}_u &= (\cos^2\tilde\theta - \sin\sigma\sin\tilde\theta\cos\tilde\theta\tan\theta)\mathbf{x}_u \\ &\quad + (\cos\tilde\theta\sin\tilde\theta\cot\theta + \sin\sigma\cos^2\tilde\theta)\mathbf{x}_v - a\cos\sigma\cos\tilde\theta\sin\theta\,\mathbf{U}, \\ \widetilde{\mathbf{x}}_v &= (\sin\tilde\theta\cos\tilde\theta\tan\theta + \sin\sigma\sin^2\tilde\theta)\mathbf{x}_u \\ &\quad + (\sin^2\tilde\theta - \sin\sigma\cos\tilde\theta\sin\tilde\theta\cot\theta)\mathbf{x}_v + a\cos\sigma\sin\tilde\theta\cos\theta\,\mathbf{U}. \end{aligned}$$

Define $\widetilde{\mathbf{U}}$ by (21.70). The equations above for $\widetilde{\mathbf{x}}_u \cdot \mathbf{U}$ and $\widetilde{\mathbf{x}}_v \cdot \mathbf{U}$ imply that $\widetilde{\mathbf{U}}$ is (plus or minus) the unit normal vector field of $\widetilde{\mathbf{x}}$. Clearly, $\widetilde{\mathbf{U}}$ is perpendicular to $\widetilde{\mathbf{x}} - \mathbf{x}$. Hence $\widetilde{\mathbf{x}}$ is a Bäcklund transform of $\mathbf{x}$. ∎

Theorem 21.34. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch that parametrizes a surface $\mathcal{M}$ in $\mathbb{R}^3$ with $K \equiv -a^{-2} < 0$. Let $\tilde\theta$ denote an angle function of the Bäcklund transform $\widetilde{\mathbf{x}}$ with inclination σ of $\mathbf{x}$. Then $\widetilde{\mathbf{x}}$ is a Tchebyshef principal patch with $K \equiv -a^{-2}$, and $\tilde\theta$ is an angle function of $\widetilde{\mathbf{x}}$.*

Dini's Surface via Bäcklund Transform

We begin with the singular patch $\mathbf{x}$ and moving frame $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{U}\}$ defined on pages 702–703. As for the angle function of $\mathbf{x}$, we again take $\theta = 0$ so that (21.50) holds. The Bäcklund transform of inclination σ of $\mathbf{x}$ is the patch $\widetilde{\mathbf{x}}$ given by

$$\begin{aligned} \widetilde{\mathbf{x}}(u,v) &= \mathbf{x}(u,v) + a\cos\sigma\big(\cos\tilde\theta\,\mathbf{E}_1 + \sin\tilde\theta\,\mathbf{E}_2\big) \\ &= a(0,0,u) + a\cos\sigma\big(\cos\tilde\theta\,(0,0,1) + \sin\tilde\theta\,(\cos u, \sin u, 0)\big) \\ &= a\big(\cos\sigma\cos u\sin\tilde\theta,\ \cos\sigma\sin u\sin\tilde\theta,\ u + \cos\sigma\cos\tilde\theta\big). \end{aligned} \tag{21.71}$$

From (21.67) we get

$$\tilde{\theta}_u = \frac{\sin\tilde{\theta}}{\cos\sigma} \qquad \text{and} \qquad \tilde{\theta}_v = -\frac{\sin\sigma\sin\tilde{\theta}}{\cos\sigma}.$$

Hence

$$\Big(\log\tan\frac{\tilde{\theta}}{2}\Big)_u = \frac{1}{\cos\sigma} \qquad \text{and} \qquad \Big(\log\tan\frac{\tilde{\theta}}{2}\Big)_v = -\frac{\sin\sigma}{\cos\sigma}.$$

In terms of differentials, we may write

$$d\Big(\log\tan\frac{\tilde{\theta}}{2}\Big) = \frac{du - \sin\sigma\, dv}{\cos\sigma}.$$

The solution of the differential equation is

$$\tan\Big|\frac{\tilde{\theta}}{2}\Big| = c\exp\Big(\frac{u - v\sin\sigma}{\cos\sigma} + c\Big),$$

where c is a constant of integration, which, without loss of generality, we may take to be 0.

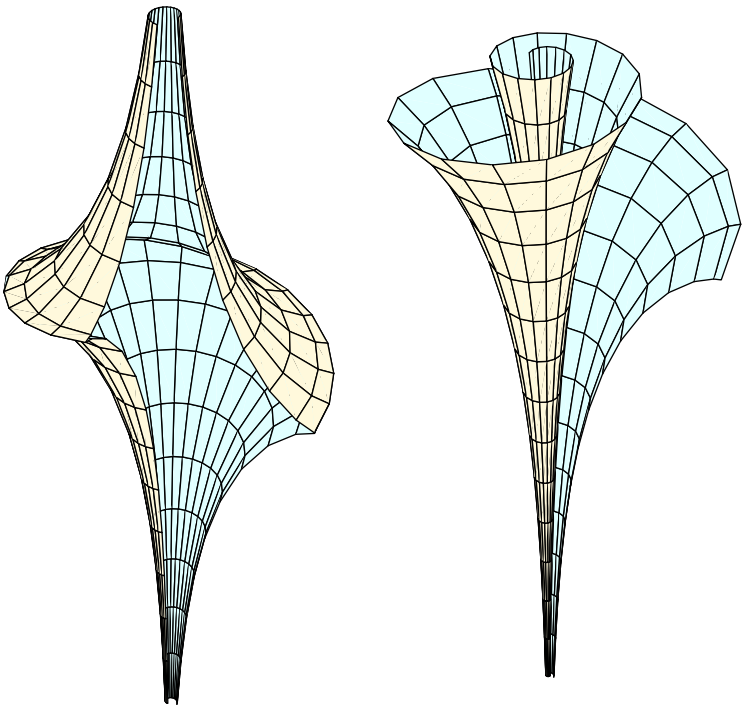


Figure 21.7: Two representations of the Tchebyshef principal parametrization of Dini's surface

By Lemma 21.14,

$$\cos\tilde{\theta} = -\tanh\Big(\frac{u - v\sin\sigma}{\cos\sigma}\Big) \qquad \text{and} \qquad \sin\tilde{\theta} = \pm\operatorname{sech}\Big(\frac{u - v\sin\sigma}{\cos\sigma}\Big).$$

From (21.71), we conclude (choosing the plus sign) that the transformed patch $\widetilde{\mathbf{x}}$ maps (u, v) to

$$a\left(\frac{\delta\cos\sigma\cos v}{\cosh\Big(\frac{u - v\sin\sigma}{\cos\sigma}\Big)},\ \frac{\cos\sigma\sin v}{\cosh\Big(\frac{u - v\sin\sigma}{\cos\sigma}\Big)},\ u - \cos\sigma\tanh\Big(\frac{u - v\sin\sigma}{\cos\sigma}\Big)\right).$$

This is therefore a Tchebyshef principal patch of a surface with constant negative Gaussian curvature, with an associated asymptotic patch given by Lemma 21.5.

In Notebook 21, we show that the last patch is in fact a parametrization of Dini's surface shown on page 482. This is visually evident from comparing Figures 15.13 and Figure 21.7.

21.9 Exercises

M **1.** Prove Lemma 21.3 by hand. Then use Notebook 21 to compute the Christoffel symbols of a Tchebyshef patch, thus checking the hand calculations. Compute the Gaussian curvature.

M **2.** Prove Lemma 21.7 by hand. Then compute the Christoffel symbols of a Tchebyshef patch of the second type, thus checking the hand calculations. Compute the Gaussian curvature.

3. Starting from a Tchebyshef principal parametrization of a pseudosphere with angle function $\theta = 2\arctan(e^u)$ (see (21.5) and Definition 21.17), verify that the pseudosphere's metric can be expressed in the form

$$ds^2 = a^2\big(dt^2 + e^{-2t}dv^2\big),$$

where $\cosh u = e^t$.

4. Complete the proof of Theorem 21.25.

5. Prove Corollary 21.26.

6. Use the results of this chapter to deduce that a Bianchi transform maps asymptotic curves to asymptotic curves.

7. Verify that the patch $\widetilde{\mathbf{x}}$ of Theorem 21.30 has

$$\begin{cases} E = \left(\dfrac{v^2 - \cosh^2 u}{v^2 + \cosh^2 u}\right)^2, \\ F = 0, \\ G = \left(\dfrac{2v \cosh u}{v^2 + \cosh^2 u}\right)^2. \end{cases}$$

8. Rewrite the parametrization of Theorem 21.30 in terms of v and the angle function θ of (21.53), by substituting $u = u(\theta)$. Show that the resulting chart $(\theta, v) \mapsto \widetilde{\mathbf{x}}(\theta, v)$ is (apart from a renaming of the variables and an overall factor of 2) equal to the one defining Kuen's surface on page 484.

9. Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef principal patch. Show that the Gauss equations (17.8), page 538, for $\mathbf{x}$ reduce to

$$\begin{cases} \mathbf{x}_{uu} &= -(\tan\theta)\theta_u \mathbf{x}_u + (\cot\theta)\theta_v \mathbf{x}_v \pm a\cos\theta\sin\theta\,\mathbf{U}, \\ \mathbf{x}_{uv} &= -(\tan\theta)\theta_u \mathbf{x}_u + (\tan\theta)\theta_u \mathbf{x}_v, \\ \mathbf{x}_{vv} &= -(\tan\theta)\theta_u \mathbf{x}_u + (\cot\theta)\theta_v \mathbf{x}_v \mp a\cos\theta\sin\theta\,\mathbf{U}, \end{cases}$$

and that the Weingarten equations (13.10), page 394, for $\mathbf{x}$ become

$$\begin{cases} S(\mathbf{x}_u) &= -\mathbf{U}_u &= \pm\dfrac{\tan\theta}{a}\mathbf{x}_u, \\ S(\mathbf{x}_v) &= -\mathbf{U}_v &= \mp\dfrac{\cot\theta}{a}\mathbf{x}_v. \end{cases}$$

10. Let $\mathbf{y}\colon \mathcal{U} \to \mathbb{R}^3$ be a Tchebyshef asymptotic patch. Show that the Gauss equations for $\mathbf{y}$ reduce to

$$\begin{cases} \mathbf{y}_{pp} &= (\cot\omega)\omega_p \mathbf{y}_p - (\csc\omega)\omega_p \mathbf{y}_q, \\ \mathbf{y}_{pq} &= \mp a(\sin\omega)\mathbf{U}, \\ \mathbf{y}_{qq} &= -(\csc\omega)\omega_q \mathbf{y}_p + (\cot\omega)\omega_q \mathbf{y}_q, \end{cases}$$

and find the corresponding Weingarten equations for $\mathbf{y}$.

Notebook 21

■ N21.0

In this notebook, we investigate Tchebyshef patches both abstractly and for surfaces in space. We then construct and parametrize surfaces of constant negative curvature by transform methods in N21.5, N21.7 and N21.8.

We shall color surfaces with the help of the following package and subsequent options.

```
<<Graphics`Colors`
```

```
SetOptions[ParametricPlot3D,
  Compiled->False, Axes->None, BoxRatios->Automatic,
  Boxed->False, Lighting->False
]
```

In order to verify that the surfaces we construct do have constant Gaussian curvature, we reload the programs gcurv and brioschi in N21.0.1. We also include christoff2 that will be used in N21.2 to determine the Christoffel symbols of Tchebyshef patches, and the definition circularcone for Figure 21.5.

■ N21.0.1

■ N21.1

To verify the statement of Lemma 21.2, we need to convert the first fundamental form of a Tschebyshef patch into a symmetric 2 by 2 matrix of functions, so that (as explained in N17.1) it can be input into brioschi.

```
M[u_,v_]:= a^2{{1,Cos[ω[u,v]]},{Cos[ω[u,v]],1}}
brioschi[M][u,v]
```

To verify that the Christoffel symbols are as stated in Lemma 21.3, we now use the program christoff2.

```
christoff2[M][u,v]
```

The same techniques can be used to check the analogous statements for a Tchebyshef patch of the second kind, namely Lemmas 21.6 and 21.7 (see N21.9).

■ N21.3

We define a "sine-Gordon operator" that characterizes the differential equation related to asymptotic patches:

```
singordon[f_][u_,v_]:= D[f[u,v],u,u] - D[f[u,v],v,v] - Sin[2f[u,v]]/2
```

Armed with this, we may check to see if a particular function is a solution of the sine-Gordon equation. Here is one:

```
f[u_,v_]:= 2ArcTan[-v/Cosh[u]]
singordon[f][u,v]
%//FullSimplify
```

■ N21.4

The Tchebyshev principal patch of Definition 21.17 can be used to plot principal and asymptotic curves on the pseudosphere.

```
pseudospheretcheb[a_][u_,v_]:=
  a{Cos[v]/Cosh[u], Sin[v]/Cosh[u], u - Tanh[u]}
```

```
Module[{x,xx},
  x= pseudospheretcheb[1];
  xx[u_,v_]= Append[x[u,v], FaceForm[Green,Yellow]];
  S @ ParametricPlot3D[xx[u,v], {u,-2.8, 2.8}, {v,0,3π/2},
    PlotPoints->{30,30}, PlotRange->All]
]
```

Figure 21.1

```
Module[{x,xx},
  x[p_,q_]= pseudospheretcheb[1][p+q, p-q];
  xx[p_,q_]= Append[x[p,q], FaceForm[Green,Yellow]];
  S @ ParametricPlot3D[xx[p,q], {p,-π/2,π/2}, {q,-π/2,π/2},
    PlotPoints->{20,20}, PlotRange->All, ViewPoint->{-4,-4,4}]
]
```

Figure 21.2

```
ParametricPlot3D[pseudospheretcheb[1][p+q,p-q], {p,-2,2},
  {q,-1,1}, PlotPoints->{50,5}, Lighting->True]
```

The functions JacobiCN, JacobiDN, JacobiAmplitude are used in Theorems 21.18 and 21.19. The following angle function gives rise to $\psi(u)$ in Theorem 21.15.

```
θ[b_][u_]:= π/2 + JacobiAmplitude[b u, 1/b^2]
```

```
Simplify[Integrate[(a/b)Cos[θ[b][u]]^2,u] /.
JacobiDN[b u,1/b^2]->Sqrt[1 - JacobiSN[b u,1/b^2]^2/b^2]]
```

The resulting parametrizations are used in N21.4.1 for Figures 21.3 and 21.4.

```
cnccotcheb[a_,b_][u_,v_]:= {(a/b)JacobiCN[b u,1/b^2]Cos[b v],
  (a/b)JacobiCN[b u,1/b^2]Sin[b v],
  a(b u - EllipticE[JacobiAmplitude[b u,1/b^2],1/b^2])}
```

```
cnchytcheb[a_,b_][u_,v_]:= (a/b){JacobiDN[u,b^2]Cos[b v],
  JacobiDN[u,b^2]Sin[b v],
  (u - EllipticE[JacobiAmplitude[u,b^2],b^2])}
```

■ N21.4.1

■ N21.5

The following program mimics the definition of the Bianchi transform, in which ϕ plays the role of the angle function "θ tilde."

```
bianchi[x_,θ_,ϕ_][u_,v_]:=
Module[{xu,xv,A},
  xu= D[x[uu,vv],uu];
  xv= D[x[uu,vv],vv];
  A= Cos[ϕ[u,v]]xu/Cos[θ[u,v]] + Sin[ϕ[u,v]]xv/Sin[θ[u,v]];
  x[u,v] + A/.{uu->u,vv->v}
]
```

The next figure provides an idea of how a new surface is obtained from a circular cone by applying the Bianchi transform to translate each point in a tangent direction. The string @@ is short for the command Apply; it was first used in N8.6 and has the effect of replacing curly brackets around the expression it precedes by square brackets.

```
Module[{θ,ϕ,x,xx,z},
  θ[u_,v_]= π/4;
  ϕ[u_,v_]= Cos[u];
  p= {1,1};
  x[1]= circularcone[1,1.5];
  x[2]= bianchi[x[1], θ, ϕ];
  xx[i_][u_,v_]:= Append[x[i][u,v], If[i==1,Red,Azure]];
  z[i_]:= ParametricPlot3D[xx[i][u,v], {u,-3,3}, {v,0,1},
    PlotPoints->{50,15}];
  li= Line[{x[1]@@p, x[2]@@p}];
  po= Table[Point[x[i]@@p], {i,2}];
  z[3]= {Blue, PointSize[.02],Thickness[.01],li,po}//Graphics3D;
  S[Array[z,3], ViewPoint->{0,-2.5,1.5}]
]
```

Figure 21.5

It can be modified by changing the angle functions θ, ϕ, although in these simple cases, the new unit normal vector is *not* tangent to the original surface.

■ N21.6

The next program, which constructs the moving frame associated to a patch x, is a modification of unitnormal. We are free to use the protected symbol E (usually the exponential constant) within Module, but the simplified components are denoted by mf[i] for external use.

```
movingframe[x_][u_,v_]:=
Module[{U,V,xu,xv,E},
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  E[1]= xu/L[xu];
  E[2]= xv/L[xv];
  E[3]= Cross[E[1],E[2]]//Simplify;
  mf[i_]:= E[i]/.{U->u, V->v};
  Array[mf,3]
]
```

In practice, this may need to be followed by appropriate simplification commands:

```
movingframe[pseudospheretcheb[a]][u,v]
%//PowerExpand
%//Simplify
```

For the singular patch considered in Section 21.6, we need to specify the moving frame components manually.

```
Module[{x,E},
  x[u_,v_]={0, 0, a u};
  E[1]= {0, 0, 1};
  E[2]= {Cos[v], Sin[v], 0};
  E[3]= {-Sin[v], Cos[v], 0};
  Array[E,3]
]
```

Here is a relevant transformation rule:

```
ExpToTrig[Exp[u]]
```

■ N21.7

To compute the Bianchi transform of the pseudosphere, the latter must be inserted in the form of a principal patch:

```
bianchips=
Module[{x,θ,ϕ},
  x= pseudospheretcheb[a];
  θ[u_,v_]= 2ArcTan[Exp[u]];
  ϕ[u_,v_]= 2ArcTan[-v/Cosh[u]];
  bianchi[x,θ,ϕ][u,v]//FullSimplify
]
```

A principal patch for Kuen's surface depending on two parameters is

```
kuentcheb[a_,b_][u_,v_]:=
  {2a Cosh[u](Cos[v] + v Sin[v])/(v^2 + Cosh[u]^2),
   2a Cosh[u](Sin[v] - v Cos[v])/(v^2 + Cosh[u]^2),
   b(u - Sinh[2u]/(v^2 + Cosh[u]^2))}
```

We verify that, when a=b, this surface coincides with the previous Bianchi transform.

```
bianchips - kuentcheb[a,a][u,v] //FullSimplify
```

Here is another way to compute the same thing.

```
Module[{b},
  x= pseudospheretcheb[a];
  movingframe[x][u,v];
  b[1]= Tanh[Log[-v/Cosh[u]]];
  b[2]= Sech[Log[-v/Cosh[u]]];
  x[u,v] + a Sum[b[i]mf[i], {i,2}]
]//PowerExpand//Simplify
```

The first fundamental form of Kuen's surface is given by

```
metric[kuentcheb[a,b]][u,v];
ee//Simplify
ff//Simplify
```

This analysis is completed in an electronic subsection, which contains the plot of kuentcheb[1,1] and its associated asymptotic patch visible in Figure 21.6.

N21.7.1

N21.8

A principal patch for a generalization of Dini's surface is

```
dinitcheb[a_,b_,c_][u_,v_]:=
  {a Cos[c]Cos[v]/Cosh[(u - v Sin[c])/Cos[c]],
   a Cos[c]Sin[v]/Cosh[(u - v Sin[c])/Cos[c]],
   b(u - Cos[c]Tanh[(u - v Sin[c])/Cos[c]])}
```

We are mainly interested in the case `a=b`, as otherwise the Gaussian curvature is not constant.

```
dinitcheb[a_,c_]:= dinitcheb[a,a,c]
```

```
dinitcheb[a,c][u,v]//StandardForm
```

```
gcurv[dinitcheb[a,c]][u,v]
```

```
metric[dinitcheb[a,c]][u,v]
```

N21.8.1

N21.9

The computation of Gaussian curvature and Christoffel symbols in N21.1 can be equally well carried out for a Tchebyshef patch of the second kind:

```
M[u_,v_]:= a^2{{Cos[θ[u,v]]^2, 0},{0, Sin[θ[u,v]]^2}}
brioschi[M][u,v]
christoff2[M][u,v]
```

Chapter 22

Minimal Surfaces via Complex Variables

We undertake a more detailed study of minimal surfaces in this chapter. It builds on the theory of complex functions of one variable, the basics of which we take for granted. The concept of minimal isothermal patch, previously introduced in Section 16.7, allows us to start the present chapter by associating to a conjugate harmonic pair of patches an *isometric deformation* of minimal surfaces. The terminology of complex derivatives in Section 22.2 leads to the central notion of *minimal curve* in Section 22.3, and a simple algebraic method for constructing an isometric deformation from a minimal isothermal patch in Section 22.4.

The above methods are effective if one has a minimal surface to start with, but do not help otherwise. In 1866, Weierstrass[1] achieved a major breakthrough in this regard. He gave a formula using complex variables that allows the easy generation of new minimal surfaces, and permits their investigation in a general manner. Weierstrass's formula for generating minimal surfaces is described in Section 22.5. It is applied to some familiar examples, and also to generate a new class of minimal surfaces with *one planar end*.

Section 22.6 adresses the problem, first introduced on page 510, of constructing a minimal surface that incorporates a given plane curve as geodesic. This is achieved by means of a simple formula that first associates to a plane curve



1 Karl Theodor Wilhelm Weierstrass (1815–1897). German mathematician. Weierstrass began his mathematical career as a high school teacher, but his work became recognized after he published a major paper on Abelian functions. This led to an appointment at the University of Berlin. His lectures in mathematics attracted students from all over the world. Topics included mathematical physics, elliptic functions, and applications to problems in geometry and mechanics. His introduction of the so-called Weierstrass $\wp$-function revolutionized the theory of elliptic functions.

a minimal curve, which is then used to construct the required minimal surface. Examples of new minimal surfaces formed in this way include a generalization of the catenoid that has an ellipse instead of a circle as its center curve. The whole process can be effectively automated, as we shall see in Notebook 22.

Few complete minimal surfaces are *embedded*, meaning "without points of self-intersection." For example, Enneper's minimal surface has self-intersections in spite of its simple definition (see page 509). Moreover, Weierstrass's formula usually produces self-intersecting minimal surfaces. The four complete minimal surfaces without self-intersections that we have seen are: the catenoid, the helicoid, Scherk's minimal surface and Scherk's fifth minimal surface. Another much more complicated complete minimal surface with no self-intersections is Costa's minimal surface, the combined beauty and simplicity of which has been the subject of attention from the artistic world. It is defined using the Weierstrass $\wp$ and ζ functions for a square lattice, rapidly introduced in the chapter's final section. The plotting programs from Notebook 22 provide three images of Costa's surface.

22.1 Isometric Deformations of Minimal Surfaces

Minimal surfaces and complex analytic functions have many properties in common. We quickly recall some pertinent facts regarding complex analytic functions of one variable. For more details see, for example, [Ahlf] or [Hille].

Let $\mathcal{U}$ be an open subset of $\mathbb{C}$. A function $f : \mathcal{U} \to \mathbb{C}$ is called ***complex analytic*** or ***holomorphic*** if f has a derivative in the complex sense at each point $\mathbf{p}$ of $\mathcal{U}$. This means that the limit

$$\lim_{\mathbf{q}\to\mathbf{p}} \frac{f(\mathbf{q}) - f(\mathbf{p})}{\mathbf{q} - \mathbf{p}}$$

exists no matter how $\mathbf{q}$ approaches the fixed point $\mathbf{p}$ in the complex plane. In this case, if we write $g = \mathfrak{Re}\, f$ and $h = \mathfrak{Im}\, f$, then g and h are related by the Cauchy[2]–Riemann equations

$$g_u = h_v, \qquad g_v = -h_u. \tag{22.1}$$

These are deduced by first letting $\mathbf{q}$ tend to $\mathbf{p}$ parallel to the real x-axis, and comparing the result with that when $\mathbf{q}$ tends to $\mathbf{p}$ parallel to the imaginary y-axis. Basic theorems from introductory complex analysis assert that a function

---



[2] Augustin Louis Cauchy (1789–1857). One of the leading French mathematicians of the first half of the 19th century. He introduced rigor into calculus, including precise definitions of continuity and integration. His best known results include the Cauchy integral formula and Cauchy–Riemann equations in complex analysis, and the Cauchy-Kovalevskaya existence theorem for the solution of partial differential equations. Cauchy's scientific output was enormous, second only to that of Euler.

f satisfying (22.1) on an open set $\mathcal{U} \subseteq \mathbb{C}$ not only has derivatives of all orders on $\mathcal{U}$, but has a power series expansion on any disk in $\mathcal{U}$.

Using the symmetry of the mixed second partial derivatives, we see that (22.1) implies that

$$g_{uu} + g_{vv} = h_{uu} + h_{vv} = 0,$$

so that that g and h are *harmonic* (see page 280). For this reason, one can say that a pair g, h of real-valued functions of two variables are ***conjugate harmonic*** when (22.1) holds. These notions extend in a coordinate-wise way to n-tuples of functions:

Definition 22.1. *Let $\mathcal{U}$ be an open subset of $\mathbb{R}^2$, and let $\mathbf{x}, \mathbf{y} \colon \mathcal{U} \to \mathbb{R}^n$ be patches. We say that*

(i) $\mathbf{x}$ *is* ***harmonic*** *if $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$.*

(ii) $\mathbf{x}$ *and* $\mathbf{y}$ *are* ***conjugate harmonic*** *if*

$$\mathbf{x}_u = \mathbf{y}_v \qquad \text{and} \qquad \mathbf{x}_v = -\mathbf{y}_u. \tag{22.2}$$

It again follows that (ii) implies (i). We shall discuss the problem of determining the conjugate $\mathbf{y}$ of a given harmonic patch $\mathbf{x}$ in Section 22.4.

As an immediate consequence of Lemma 16.18, page 520, we obtain

Corollary 22.2. *A regular isothermal patch $\mathbf{x} \colon \mathcal{U} \to \mathbb{R}^3$ is a minimal surface if and only if it is harmonic.*

So far we have only defined the notion of minimal surface in $\mathbb{R}^3$. Corollary 22.2 provides one way of extending this definition so as to talk of a minimal surface in $\mathbb{R}^n$ for any n (another way is given in Exercise 11):

Definition 22.3. *A* ***minimal isothermal patch*** *$\mathbf{x} \colon \mathcal{U} \to \mathbb{R}^n$ is a patch that is both isothermal and harmonic.*

Next, we describe a method for obtaining a family of isometric minimal patches. This construction generalizes the deformation between the helicoid and the catenoid, and (16.8) on page 505 will assist in understanding the general case below.

Definition 22.4. *Let $\mathbf{x}, \mathbf{y} \colon \mathcal{U} \to \mathbb{R}^n$ be conjugate harmonic isothermal patches. Then the* ***associated family*** *of $\mathbf{x}$ and $\mathbf{y}$ is the 1-parameter family of patches $\mathbf{z}[t] \colon \mathcal{U} \to \mathbb{R}^n$ given by*

$$\mathbf{z}[t] = \mathfrak{Re}\left(e^{-it}(\mathbf{x} + i\,\mathbf{y})\right) = \mathbf{x}\cos t + \mathbf{y}\sin t. \tag{22.3}$$

The last expression is of course equal to $(\cos t)\mathbf{x} + (\sin t)\mathbf{y}$, but writing scalars on the right will save us from inserting lots of parentheses below. First, we compute the derivatives of $\mathbf{z}[t]$.

Lemma 22.5. *Let $\mathbf{x}, \mathbf{y} \colon \mathcal{U} \to \mathbb{R}^n$ be conjugate harmonic isothermal patches. Then the associated family $t \mapsto \mathbf{z}[t]$ of $\mathbf{x}$ and $\mathbf{y}$ satisfies*

$$\text{(22.4)} \qquad \begin{cases} \mathbf{z}[t]_u &= \mathbf{x}_u \cos t - \mathbf{x}_v \sin t, \\ \mathbf{z}[t]_v &= \mathbf{x}_u \sin t + \mathbf{x}_v \cos t, \\ \mathbf{z}[t]_{uu} &= \mathbf{x}_{uu} \cos t - \mathbf{x}_{uv} \sin t = -\mathbf{z}[t]_{vv}, \\ \mathbf{z}[t]_{uv} &= \mathbf{x}_{uu} \sin t + \mathbf{x}_{uv} \cos t. \end{cases}$$

Furthermore, $\mathbf{z}[t]$ and $\mathbf{z}\big[t + \frac{\pi}{2}\big]$ are conjugate harmonic for any t.

Proof. Equations (22.4) are a consequence of 22.3 and the Cauchy–Riemann equations. The last statement of the lemma can be proved as follows:

$$\begin{aligned} \mathbf{z}[t]_u \;=\; \mathfrak{Re}\,\big(e^{-it}(\mathbf{x}_u + i\,\mathbf{y}_u)\big) \;&=\; \mathfrak{Re}\,\big(e^{-it}(\mathbf{y}_v - i\,\mathbf{x}_v)\big) \\ &= \mathfrak{Re}\,\big(e^{-i(t+\pi/2)}(\mathbf{x}_v + i\,\mathbf{y}_v)\big) \;=\; \mathbf{z}\big[t + \tfrac{\pi}{2}\big]_v. \end{aligned}$$

Similarly, $\mathbf{z}[t]_v = -\mathbf{z}\big[t + \frac{\pi}{2}\big]_u$. ∎

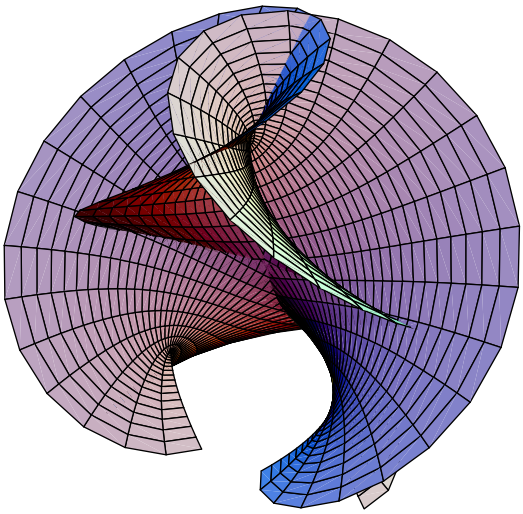


Figure 22.1: A deformation of Catalan's surface

Figure 22.1 illustrates the patch (22.3), in which $\mathbf{x}$ is Catalan's surface shown in Figure 16.4 on page 510, and $t = 3\pi/8$. The snag is that the techniques of this section are not yet sufficient to find the equation of this deformation. To

make the plot, it is necessary to find the conjugate patch $\mathbf{y}$, and the resulting *complexification* $\mathbf{x} + i\mathbf{y}$ of $\mathbf{x}$; this was done automatically in Notebook 22 using the recipe given in Section 22.4 below.

Lemma 22.5 is a useful tool for obtaining information about the first and second fundamental forms of an associated family.

Theorem 22.6. *Let $\mathbf{x}, \mathbf{y} : \mathcal{U} \to \mathbb{R}^n$ be conjugate harmonic isothermal patches, and let $t \mapsto \mathbf{z}[t]$ be the associated family defined by* (22.3)*. Then $\mathbf{z}[t]$ is a minimal isothermal patch for each t, and all of the patches in the family have the same first fundamental form.*

We say that $\mathbf{z}[t]$ is an ***isometric deformation*** from $\mathbf{x}$ to $\mathbf{y}$.

Proof. That $\mathbf{z}[t]$ is harmonic is a consequence of the third equation of (22.4). Let $E(t), F(t), G(t)$ denote the coefficients of the first fundamental form of $\mathbf{z}[t]$. We use the first two equations of (22.4). Firstly,

$$E(t) = \mathbf{z}[t]_u \cdot \mathbf{z}[t]_u = (\mathbf{x}_u \cos t - \mathbf{x}_v \sin t) \cdot (\mathbf{x}_u \cos t - \mathbf{x}_v \sin t).$$

Since $\mathbf{x}_u \cdot \mathbf{x}_u = \mathbf{x}_v \cdot \mathbf{x}_v = E$ and $\mathbf{x}_u \cdot \mathbf{x}_v = 0$, we obtain $E(t) = \mathbf{x}_u \cdot \mathbf{x}_u = E$. Similarly, $F(t) = 0$ and $G(t) = \mathbf{x}_v \cdot \mathbf{x}_v = G$ for all t. It follows that each $\mathbf{z}[t]$ is isothermal, and has the same first fundamental form as $\mathbf{x}$. ∎

Next, we observe that, up to parallel translation, all the members of an associated family have the same unit normal and tangent space.

Lemma 22.7. *Let $\mathbf{x}, \mathbf{y} : \mathcal{U} \to \mathbb{R}^3$ be conjugate harmonic isothermal patches.*

(i) *The unit normal $\mathbf{U}(t)$ of the patch $\mathbf{z}[t]$ at $\mathbf{z}[t](u, v)$ is parallel to the unit normal $\mathbf{U}$ of the patch $\mathbf{x}$ at $\mathbf{x}(u, v)$.*

(ii) *The tangent space to the patch $\mathbf{z}[t]$ at $\mathbf{z}[t](u, v)$ is parallel to the tangent space to the patch $\mathbf{x}$ at $\mathbf{x}(u, v)$.*

Proof. Part (i) follows from the formula

$$\mathbf{z}[t]_u \times \mathbf{z}[t]_v = \mathbf{x}_u \times \mathbf{x}_v, \qquad t \in \mathbb{R},$$

itself an easy consequence of (22.4). Obviously, (i) implies (ii). ∎

In the light of (ii), we can identify the unit normal of $\mathbf{z}[t]$ with the unit normal $\mathbf{U}$ of $\mathbf{x}$. Thus

Corollary 22.8. *All members of an associated family have the same Gauss map.*

In spite of the above results, the shape operator of $\mathbf{z}[t]$ will in general depend on t. Let $e(t), f(t), g(t)$ denote the coefficients of the second fundamental form of $\mathbf{z}[t]$, and let $S(t)$ denote the corresponding shape operator.

Lemma 22.9. *The second fundamental form of* $\mathbf{z}[t]\colon \mathcal{U} \to \mathbb{R}^3$ *is related to that of* $\mathbf{x}$ *by the formulas*

$$\left\{ \begin{array}{rcl} e(t) \;=\; -g(t) &=& e\cos t - f\sin t, \\ f(t) &=& e\sin t + f\cos t. \end{array} \right. \tag{22.5}$$

Furthermore,

$$\left\{ \begin{array}{rcl} S(t)\,\mathbf{z}[t]_u &=& S\,\mathbf{x}_u, \\ S(t)\,\mathbf{z}[t]_v &=& S\,\mathbf{x}_v. \end{array} \right. \tag{22.6}$$

Proof. We use the third equation of (22.4) to compute

$$e(t) = \mathbf{z}[t]_{uu} \cdot \mathbf{U}(t) = \big(\mathbf{x}_{uu}\cos t - \mathbf{x}_{uv}\sin t\big) \cdot \mathbf{U} = e\cos t - f\sin t,$$

and the rest of (22.5) also follows from (22.4).

Combining the Weingarten equations (16.32), page 522, for the isothermal patch $\mathbf{z}[t]$ and Theorem 22.6 with (22.4) and (22.5), we get

$$\begin{aligned} \lambda^2 S(t)\,\mathbf{z}[t]_u &= e(t)\mathbf{z}[t]_u + f(t)\mathbf{z}[t]_v \\ &= \big(e\cos t - f\sin t\big)\big(\mathbf{x}_u\cos t - \mathbf{x}_v\sin t\big) \\ &\qquad + \big(e\sin t + f\cos t\big)\big(\mathbf{x}_u\sin t + \mathbf{x}_v\cos t\big) \\ &= e\,\mathbf{x}_u + f\,\mathbf{x}_v = S\mathbf{x}_u. \end{aligned} \tag{22.7}$$

Similarly, $S(t)\,\mathbf{z}[t]_v = S\,\mathbf{x}_v$. ∎

We are now in a position to state a simple formula for the shape operator $S(t)$ in terms of shape operator S and complex structure J of the minimal isothermal patch $\mathbf{x}$.

Lemma 22.10. *The shape operator* $S(t)$ *of* $\mathbf{z}[t]$ *is related to the shape operator* S *of* $\mathbf{x}$ *by the formula*

$$S(t) = (\cos t)S + (\sin t)\,S\circ J. \tag{22.8}$$

In particular, the shape operator $S(\frac{\pi}{2})$ *of the conjugate patch of* $\mathbf{x}$ *equals the composition* $S\circ J$.

Proof. From (22.7) and (22.4), we get

$$\left\{ \begin{array}{rcl} S\,\mathbf{x}_u &=& S(t)\mathbf{x}_u\cos t - S(t)\mathbf{x}_v\sin t, \\ S\,\mathbf{x}_v &=& S(t)\mathbf{x}_u\sin t + S(t)\mathbf{x}_v\cos t. \end{array} \right.$$

Solving for $S(t)\mathbf{x}_u$ and $S(t)\mathbf{x}_v$, and setting $\mathbf{x}_v = J\mathbf{x}_u$, we obtain

$$\left\{ \begin{array}{rcl} S(t)\mathbf{x}_u &=& S\mathbf{x}_u\cos t + SJ\mathbf{x}_u\sin t, \\ S(t)\mathbf{x}_v &=& SJ\mathbf{x}_v\sin t + S\mathbf{x}_v\cos t. \end{array} \right. \tag{22.9}$$

Clearly, (22.9) is equivalent to (22.8). ∎

Corollary 22.11. *Let* $\mathbf{x}, \mathbf{y}: \mathcal{U} \to \mathbb{R}^3$ *be conjugate harmonic isothermal patches, and let* $\mathbf{v_p}$ *be a tangent vector to* $\mathbf{x}$ *(which by Lemma* 22.7 *can also be considered to be a tangent vector to* $\mathbf{y}$*). Then*

(i) $\mathbf{v_p}$ *is an asymptotic vector of* $\mathbf{x}$ *if and only if it is a principal vector of* $\mathbf{y}$;

(ii) $\mathbf{v_p}$ *is a principal vector of* $\mathbf{x}$ *if and only if it is an asymptotic vector of* $\mathbf{y}$.

For proofs, see Exercise 1.

22.2 Complex Derivatives

Corollary 22.2 is the reason why harmonic functions are useful for the study of minimal surfaces. Since analytic function theory is so important for the study of harmonic functions of two variables, it is natural to introduce this subject into the study of minimal surfaces in $\mathbb{R}^n$.

It is often advantageous to switch from the standard coordinates u, v in $\mathbb{R}^2$, which we have been using, to complex coordinates z and $\overline{z}$. The algebraic formulas relating the two kinds of coordinates are simple:

$$\begin{cases} z &= u + iv, \\ \overline{z} &= u - iv, \end{cases} \tag{22.10}$$

$$\begin{cases} u &= \dfrac{z + \overline{z}}{2}, \\ v &= \dfrac{z - \overline{z}}{2\,i}. \end{cases} \tag{22.11}$$

It is a little disconcerting at first to work with z and $\overline{z}$, because one tends to think that z determines $\overline{z}$. This is certainly the case when we start with u and v and define z and $\overline{z}$ by (22.10). But we can regard z and $\overline{z}$ as abstract coordinates for $\mathbb{R}^2 = \mathbb{C}$ (even though they are complex-valued), and then define u and v by (22.11). This point of view frequently leads to simpler formulas. Thus, we shall see that $\{z, \overline{z}\}$ forms a coordinate system for $\mathbb{R}^2$ that we can use in place of the standard coordinate system $\{u, v\}$.

It is convenient to introduce the differential operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial u} - i\frac{\partial}{\partial v}\right) \qquad \text{and} \qquad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial u} + i\frac{\partial}{\partial v}\right). \tag{22.12}$$

The motivation is that these operators satisfy the equations

$$\frac{\partial}{\partial z}(z) = 1, \qquad \frac{\partial}{\partial z}(\overline{z}) = 0$$
$$\frac{\partial}{\partial \overline{z}}(z) = 0, \qquad \frac{\partial}{\partial \overline{z}}(\overline{z}) = 1.$$

The Cauchy–Riemann equations (22.1) can now be expressed in the form

$$\frac{\partial}{\partial \overline{z}} f(z, \overline{z}) = 0.$$

From (22.10), we also have

$$dz = du + i dv, \qquad d\overline{z} = du - i dv \tag{22.13}$$

and we can write

$$|dz|^2 = dz\, d\overline{z} = du^2 + dv^2 \qquad \text{and} \qquad \boldsymbol{\Delta} = \frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2} = 4\frac{\partial^2}{\partial \overline{z} \partial z}.$$

Even though a patch $\mathbf{x}$ in $\mathbb{R}^n$ consists of an n-tuple of real functions, it is very useful in minimal surface theory to apply $\partial/\partial z$ to $\mathbf{x}$.

Definition 22.12. *The* ***complex derivative*** *of a patch* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ *is given by*

$$\frac{\partial \mathbf{x}}{\partial z} = \frac{1}{2}(\mathbf{x}_u - i\,\mathbf{x}_v),$$

where $z = u + iv$.

Given $\mathbf{x}$, we shall write

$$\big(\phi_1[\mathbf{x}], \ldots, \phi_n[\mathbf{x}]\big) = \frac{\partial \mathbf{x}}{\partial z} = \frac{1}{2}\left(\frac{\partial x_1}{\partial u} - i\frac{\partial x_1}{\partial v}, \ \ldots\ , \frac{\partial x_n}{\partial u} - i\frac{\partial x_n}{\partial v}\right). \tag{22.14}$$

With this notation, it is an easy matter to check the identities

$$4\sum_{k=1}^{n} \phi_k[\mathbf{x}]^2 = \mathbf{x}_u \cdot \mathbf{x}_u - \mathbf{x}_v \cdot \mathbf{x}_v - 2i\,\mathbf{x}_u \cdot \mathbf{x}_v = E - G - 2\,i\,F, \tag{22.15}$$

and

$$4\sum_{k=1}^{n} \big|\phi_k[\mathbf{x}]\big|^2 = \mathbf{x}_u \cdot \mathbf{x}_u + \mathbf{x}_v \cdot \mathbf{x}_v = E + G. \tag{22.16}$$

The next theorem shows that an isothermal minimal patch gives rise to an n-tuple of complex analytic functions such that the sum of the squares of the components equals zero. As we shall see below, this alternative description of an isothermal patch yields much useful information, because it allows the use of powerful theorems from complex analysis.

Theorem 22.13. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ *be a patch. Then*

(i) $\mathbf{x}$ *is harmonic if and only if* (22.14) *is complex analytic*;

(ii) $\mathbf{x}$ *is isothermal if and only if* $\displaystyle\sum_{k=1}^{n} \phi_k[\mathbf{x}]^2 = 0$;

(iii) *if* $\mathbf{x}$ *is isothermal, then* $\mathbf{x}$ *is regular if and only if* $\sum_{k=1}^{n} \left|\phi_k[\mathbf{x}]\right|^2 \neq 0$.

Conversely, suppose that $\mathcal{U}$ *is simply-connected, and let* $\phi_1, \ldots, \phi_n \colon \mathcal{U} \to \mathbb{C}^n$ *be complex analytic functions satisfying*

$$\sum_{k=1}^{n} \phi_k^2 = 0 \qquad \textit{and} \qquad \sum_{k=1}^{n} |\phi_k|^2 \neq 0. \tag{22.17}$$

Then there exists a regular minimal isothermal patch $\mathbf{x} \colon \mathcal{U} \to \mathbb{R}^n$ *satisfying* (22.14).

Proof. Part (i) follows from the fact that the Cauchy–Riemann equations for $\partial\mathbf{x}/\partial z$ are $\mathbf{x}_{uu} + \mathbf{x}_{vv} = 0$ and (the identity) $\mathbf{x}_{uv} - \mathbf{x}_{vu} = 0$. Then (ii) is a consequence of (22.15), and (iii) follows from (22.16).

To prove the converse, suppose that complex analytic functions ϕ_j are given satisfying (22.17). Put

$$\mathbf{x} = \left(\mathfrak{Re} \int \phi_1(z)\, dz,\ \ldots\ , \mathfrak{Re} \int \phi_n(z)\, dz \right).$$

The "simply-connected" assumption (see page 788) guarantees that the contour integrals are well defined and complex analytic, and their real parts harmonic. Moreover, (22.15) implies that $\mathbf{x}$ is isothermal, and (22.16) implies that $\mathbf{x}$ is regular. Thus $\mathbf{x}$ is a regular minimal isothermal patch. ∎

We know that a pair of conjugate harmonic functions determines a complex analytic function. Indeed,

Lemma 22.14. *Let* $\mathbf{x}, \mathbf{y} \colon \mathcal{U} \to \mathbb{R}^n$ *be conjugate harmonic patches. Then the map* $\mathbf{x} + i\mathbf{y} \colon \mathcal{U} \to \mathbb{C}^n$ *is complex analytic, and*

$$2\frac{\partial \mathbf{x}}{\partial z} = \frac{\partial}{\partial z}(\mathbf{x} + i\mathbf{y}).$$

Proof. The first statement follows from the fact that $\mathbf{x} + i\mathbf{y}$ is an n-tuple of complex functions of one variable, each of which is complex analytic. For the second statement we compute

$$\begin{aligned} \frac{\partial}{\partial z}(\mathbf{x} + i\mathbf{y}) &= \frac{1}{2}\left(\frac{\partial}{\partial u} - i\frac{\partial}{\partial v} \right)(\mathbf{x} + i\mathbf{y}) \\ &= \frac{1}{2}\big(\mathbf{x}_u - i\,\mathbf{x}_v + \mathbf{y}_v + i\mathbf{y}_u\big) = \mathbf{x}_u - i\,\mathbf{x}_v = 2\frac{\partial \mathbf{x}}{\partial z}. \end{aligned} \tag{22.18}$$

∎

This will be relevant in the next section, given the definition of the associated family on page 721.

22.3 Minimal Curves

In this section we look at functions that have some of the properties of the complex derivative of a minimal isothermal patch. First, we reformulate (22.17) as a definition, making use of the complex dot product that was defined on page 194.

Definition 22.15. *Let $\mathcal{U}$ be an open subset of $\mathbb{C}$. A* ***minimal curve*** *is a complex analytic function $\Psi\colon \mathcal{U} \to \mathbb{C}^n$ such that*

$$\Psi'(z) \cdot \Psi'(z) = 0 \tag{22.19}$$

for $z \in \mathcal{U}$. If in addition $\Psi'(z) \cdot \overline{\Psi'(z)}$ is never zero for $z \in \mathcal{U}$, we say that Ψ is a ***regular minimal curve****.*

Minimal curves were first studied by Lie[3] in [Lie]. A minimal curve can be thought of as a generalization of a minimal isothermal patch. In a different sense, a minimal curve is a generalization of a real curve in $\mathbb{R}^n$, and a complex-valued parametrized curve satisfying (22.19) is sometimes called ***isotropic***.

It is clear from the preceding section that, given a minimal curve $\Psi\colon \mathcal{U} \to \mathbb{C}^n$, the patches $\mathbf{x}, \mathbf{y}\colon \mathcal{U} \to \mathbb{R}^n$ defined by

$$\mathbf{x}(u,v) = \mathfrak{Re}\ \Psi(u+iv) \quad \text{and} \quad \mathbf{y}(u,v) = \mathfrak{Im}\ \Psi(u+iv) \tag{22.20}$$

are conjugate harmonic isothermal patches. We therefore call the pair $\mathbf{x}, \mathbf{y}$ the ***conjugate harmonic isothermal patches associated to*** Ψ. They also determine an associated family $\mathbf{z}[t]\colon \mathcal{U} \to \mathbb{R}^n$ by setting

$$\mathbf{z}[t](u,v) = \mathfrak{Re}\left(e^{-it}\Psi(u+iv)\right), \tag{22.21}$$

exactly as in (22.3). We emphasize that $\mathbf{z}[t](u,v)$ is a *real* object; it is not equal to $\mathbf{x} + i\mathbf{y}$, which is instead denoted by Ψ. Theorem 22.6 tells us that this is an isometric deformation, and that $\mathbf{z}[t]$ is a minimal isothermal patch for each t.

Let us rewrite Lemma 22.5 in terms of minimal curves.

Corollary 22.16. *Let $\Psi\colon \mathcal{U} \to \mathbb{C}^n$ be a minimal curve. Then the associated family $t \mapsto \mathbf{z}[t]$ of Ψ satisfies*

$$\mathbf{z}[t]_u - i\,\mathbf{z}[t]_v = e^{-it}\Psi'. \tag{22.22}$$

Proof. This is where we use (22.18). ∎



[3] Marius Sophus Lie (1842–1899). Norwegian mathematician. Although Lie is best known for Lie groups and Lie algebras, he made many important contributions to differential equations, surface theory, and the theory of contact transformations.

It will be convenient to write

$$\frac{1}{2}\Psi' = \frac{\partial \mathbf{x}}{\partial z} = \frac{1}{2}(\mathbf{x}_u - i\,\mathbf{x}_v) = \big(\phi_1, \ldots, \phi_n\big)$$

for the complex derivative of a patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$. Then

$$\frac{1}{2}\overline{\Psi'} = \frac{\partial \mathbf{x}}{\partial \overline{z}} = \frac{1}{2}(\mathbf{x}_u + i\,\mathbf{x}_v) = \big(\overline{\phi}_1, \ldots, \overline{\phi}_n\big).$$

Using this notation, we now study the Gauss map of a minimal curve in $\mathbb{C}^3$.

Definition 22.17. *The* ***Gauss map of a minimal curve*** Ψ *is the Gauss map of any member of an associated family of* Ψ*. As usual, we identify the Gauss map with the unit normal* $\mathbf{U}$*.*

The next lemma follows from (7.9) on page 195.

Lemma 22.18. *Let* $\Psi\colon \mathcal{U} \to \mathbb{C}^3$ *be a minimal curve. Then*

$$-i\,\Psi' \times \overline{\Psi'} = 2\left(\mathfrak{Im}(\phi_2\overline{\phi}_3),\ \mathfrak{Im}(\phi_3\overline{\phi}_1),\ \mathfrak{Im}(\phi_1\overline{\phi}_2)\right) = \|\Psi'\|^2\,\mathbf{U}.$$

Proof. It follows from (7.8) that

$$\|\Psi'\|^2 = \|\Psi' \times \overline{\Psi'}\| = 2\|\mathbf{x}_u \times \mathbf{x}_v\|.$$

Therefore,

$$\begin{aligned}\mathbf{U} = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|} &= \frac{(\Psi' \times \overline{\Psi'})/(2i)}{\|\Psi'\|^2/2} \\ &= -i\frac{\Psi' \times \overline{\Psi'}}{\|\Psi'\|^2} = \frac{2\left(\mathfrak{Im}(\phi_2\overline{\phi}_3), \mathfrak{Im}(\phi_3\overline{\phi}_1), \mathfrak{Im}(\phi_1\overline{\phi}_2)\right)}{\|\Psi'\|^2}.\end{aligned}$$

∎

At this juncture, we need the important map, which we shall denote by stereo, defined by the stereographic projection shown in Figure 22.2. It can be used to transfer information between the sphere $S^2(1)$ of unit radius in $\mathbb{R}^3$ centered at the origin and the complex plane. We already discussed the inverse of stereo in Section 12.2, but now we use complex variables.

Let $\mathbf{p} \in S^2(1)$ be any point other than the north pole $\mathbf{n} = (0, 0, 1)$. The line ℓ through $\mathbf{p}$ and $\mathbf{n}$ meets the complex plane $\mathbb{C}$ at a point $\mathbf{q}$. Let us identify $\mathbb{C}$ with the equatorial plane $\mathbb{R}^2$. Write $\mathbf{p} = (p_1, p_2, p_3)$ and $\mathbf{q} = (q_1, q_2, 0)$. We can parametrize ℓ as

$$t \mapsto t\,\mathbf{q} + (1 - t)\mathbf{n} = (tq_1,\ tq_2,\ 1 - t).$$

Thus there is a t_0 such that

$$\mathbf{p} = t_0\mathbf{q} + (1 - t_0)\mathbf{n} = (t_0q_1,\ t_0q_2,\ 1 - t_0).$$

It follows that $t_0 = 1 - p_3$, so that

$$q_1 = \frac{p_1}{t_0} = \frac{p_1}{1-p_3} \qquad \text{and} \qquad q_2 = \frac{p_2}{t_0} = \frac{p_2}{1-p_3}.$$

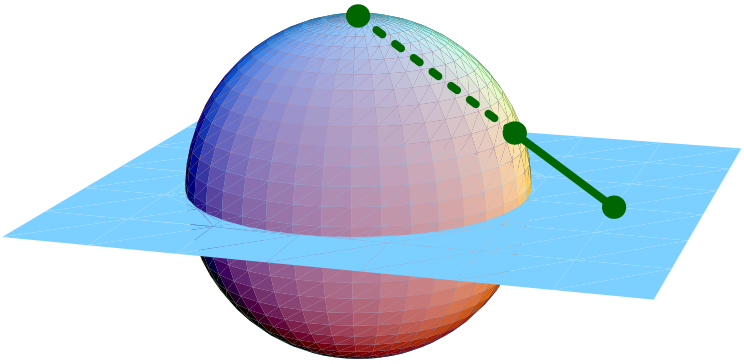


Figure 22.2: Stereographic projection from the north pole

The construction is summarized by

Definition 22.19. ***Stereographic projection*** *is the mapping* $\mathsf{stereo}\colon S^2(1)\setminus \mathbf{n} \to \mathbb{C}$ *given by*

$$\mathsf{stereo}(p_1,p_2,p_3) = \frac{p_1 + ip_2}{1-p_3}.$$

There is a simple formula for the composition of the stereographic projection and the Gauss map of a minimal curve in $\mathbb{C}^3$.

Lemma 22.20. *Let* $\Psi : \mathcal{U} \to \mathbb{C}^3$ *be a minimal curve, and write* $\Psi' = (\phi_1, \phi_2, \phi_3)$. *Let* $\mathbf{U}$ *be the unit normal vector field of the patch given by* (22.20). *Then*

$$\mathsf{stereo} \circ \mathbf{U} = \frac{\phi_3}{\phi_1 - i\,\phi_2}. \tag{22.23}$$

Proof. Starting with Lemma 22.18, we see that $\mathsf{stereo} \circ \mathbf{U}$ equals

$$\begin{aligned}
\frac{\dfrac{2\big(\mathfrak{Im}(\phi_2\overline{\phi}_3) + i\,\mathfrak{Im}(\phi_3\overline{\phi}_1)\big)}{\|\Psi'\|^2}}{1 - \dfrac{2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)}{\|\Psi'\|^2}} &= \frac{2(\mathfrak{Im}\,(\phi_2\overline{\phi}_3) + i\,\mathfrak{Im}(\phi_3\overline{\phi}_1))}{\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)} \\
&= \frac{\phi_2\overline{\phi}_3 - \overline{\phi}_2\phi_3 + i\,(\phi_3\overline{\phi}_1 - \overline{\phi}_3\phi_1)}{i\big(\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)\big)} \\
&= \frac{\phi_3(\overline{\phi}_1 + i\overline{\phi}_2) - \overline{\phi}_3(\phi_1 + i\phi_2)}{\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)}.
\end{aligned}$$

Using the fact that $(\phi_1 - i\,\phi_2)(\phi_1 + i\,\phi_2) = -\phi_3^2$, we rewrite the right-hand side as

$$\begin{aligned}\frac{\phi_3(\overline{\phi}_1 + i\,\overline{\phi}_2) + \overline{\phi}_3\left(\dfrac{\phi_3^2}{\phi_1 - i\,\phi_2}\right)}{\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)} &= \frac{\phi_3\big((\overline{\phi}_1 + i\,\overline{\phi}_2)(\phi_1 - i\,\phi_2) + |\phi_3|^2\big)}{(\phi_1 - i\,\phi_2)\big(\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)\big)} \\ &= \frac{\phi_3\big(\|\Psi'\|^2 + i\,\overline{\phi}_2\phi_1 - i\,\phi_2\overline{\phi}_1\big)}{(\phi_1 - i\,\phi_2)\big(\|\Psi'\|^2 - 2\,\mathfrak{Im}(\phi_1\overline{\phi}_2)\big)} \\ &= \frac{\phi_3}{\phi_1 - i\,\phi_2}. \ \blacksquare\end{aligned}$$

The version of Lemma 22.9 for a minimal curve is

Lemma 22.21. *Let $\Psi\colon \mathcal{U} \to \mathbb{C}^3$ be a minimal curve with unit normal vector field $\mathbf{U}$ and associated family $t \mapsto \mathbf{z}[t]$. Then the coefficients of the second fundamental form of $\mathbf{z}[t]$ are given by*

$$\left\{\begin{aligned} e(t) \;=\; -g(t) \;&=\; \mathfrak{Re}\left(\left(e^{-it}\Psi''\right)\cdot\mathbf{U}\right), \\ f(t) \;&=\; -\,\mathfrak{Im}\left(\left(e^{-it}\Psi''\right)\cdot\mathbf{U}\right). \end{aligned}\right. \tag{22.24}$$

Proof. From (22.21), we get

$$\left(e^{-it}\Psi''\right)\cdot\mathbf{U} = \left(\mathbf{z}[t]_{uu} - i\,\mathbf{z}[t]_{uv}\right)\cdot\mathbf{U} = e(t) - i\,f(t)$$

and so (22.24) follows. ∎

Theorem 22.6 implies that all of the patches in the associated family of a minimal curve are isometric. Hence by Theorem 17.5 they all have the same Gaussian curvature, and we may speak of the ***Gaussian curvature of a minimal curve***. We conclude this section by finding a formula for this quantity.

Lemma 22.22. *The Gaussian curvature of a minimal curve $\Psi\colon \mathcal{U} \mapsto \mathbb{C}^n$ is given by*

$$K = \frac{-4\big(\|\Psi'\|^2\|\Psi''\|^2 - |\Psi''\cdot\overline{\Psi'}|^2\big)}{\|\Psi'\|^6}.$$

Proof. In the notation of Definition 16.14 on page 518, $\|\Psi'\|^2 = 2\lambda^2$, and so

$$\frac{1}{4}\boldsymbol{\Delta}\log(\lambda^2) = \frac{\partial^2}{\partial\overline{z}\partial z}\log\left(\|\Psi'\|^2\right) = \frac{\partial^2}{\partial\overline{z}\partial z}\log\left(\Psi'\cdot\overline{\Psi'}\right). \tag{22.25}$$

The right-hand side equals

$$\begin{aligned}\frac{\partial}{\partial\overline{z}}\left(\frac{\Psi''\cdot\overline{\Psi'}}{\|\Psi'\|^2}\right) &= \frac{\|\Psi'\|^2\|\Psi''\|^2 - (\Psi''\cdot\overline{\Psi'})(\Psi'\cdot\overline{\Psi''})}{\|\Psi'\|^4} \\ &= \frac{\|\Psi'\|^2\|\Psi''\|^2 - |\Psi''\cdot\overline{\Psi'}|^2}{\|\Psi'\|^4}.\end{aligned}$$

Then (17.11) on page 540 and (22.25) imply that

$$K = \frac{-\boldsymbol{\Delta}\log\lambda}{\lambda^2} = -\frac{2}{\|\Psi'\|^2}\left(\frac{2\big(\|\Psi'\|^2\|\Psi''\|^2 - |\Psi''\cdot\overline{\Psi'}|^2\big)}{\|\Psi'\|^4}\right)$$
$$= \frac{-4\big(\|\Psi'\|^2\|\Psi''\|^2 - |\Psi''\cdot\overline{\Psi'}|^2\big)}{\|\Psi'\|^6}. \quad \blacksquare$$

We now turn attention to the construction of minimal curves. This will be a two-step process. In the next section, we explain how a minimal isothermal patch gives rise to a minimal curve, and in the following one show how to write down minimal curves from scratch.

22.4 Finding Conjugate Minimal Surfaces

We know that a minimal isothermal patch $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ is the real part of an n-tuple $(\psi_1, \ldots, \psi_n)$ of complex analytic functions. How do we explicitly determine $(\psi_1, \ldots, \psi_n)$ from $\mathbf{x}$?

First, let us ask how to determine a complex analytic function $f\colon \mathcal{U} \to \mathbb{C}$ such that the real part of f is a given harmonic function $h\colon \mathcal{U} \to \mathbb{R}$. Of course, f is unique only up to an imaginary constant. The standard way (found in most text books on complex variables) is to use the Cauchy–Riemann equations; this method requires integration. There is, however, a more efficient method (see [Ahlf, pages 27–28], [Boas, pages 158–159], and [Lait]) that is completely algebraic, and hence involves no integration whatever. The idea is to substitute complex variables formally for the real variables u and v in the expression $h(u, v)$.

Lemma 22.23. *Let $h\colon \mathcal{U} \to \mathbb{R}$ be a harmonic function where $\mathcal{U}$ is an open subset of $\mathbb{R}^2 = \mathbb{C}$, and let $z_0 = u_0 + iv_0 \in \mathcal{U}$.*

(i) *Let $f\colon \mathcal{U} \to \mathbb{C}$ be a complex analytic function such that $\mathfrak{Re}\, f(u+iv) = h(u,v)$ and $\mathfrak{Im}\, f(z_0) = 0$. Then*

$$f(z) = 2h\left(\frac{z+\overline{z}_0}{2}, \frac{z-\overline{z}_0}{2i}\right) - h(u_0, v_0). \tag{22.26}$$

(ii) *Conversely, if f is defined by* (22.26)*, then $f\colon \mathcal{U} \to \mathbb{C}$ is a complex analytic function such that $f(z_0) = h(u_0, v_0)$.*

Proof. For (i) we do the case $z_0 = 0$. Define g by $g(z) = \overline{f(\overline{z})}$. Then g is complex analytic, and we can express h in terms of f and g:

$$h(u,v) = \tfrac{1}{2}\big(f(u+iv) + \overline{f(u+iv)}\big) = \tfrac{1}{2}\big(f(u+iv) + g(u-iv)\big).$$

Since h is harmonic, it can be expanded in a double power series about any point $\mathbf{p} \in \mathcal{U}$. Thus, even though z is a complex variable, we may calculate

$$h\Big(\frac{z}{2}, \frac{z}{2i}\Big) = \tfrac{1}{2}\big(f(z) + g(0)\big) = \tfrac{1}{2}\big(f(z) + h(0,0)\big).$$

For (ii) we compute

$$f_u = h_u - i\,h_v \qquad \text{and} \qquad f_v = i\,h_u + h_v.$$

Therefore,

$$\frac{\partial f}{\partial \overline{z}} = 0,$$

and f satisfies the Cauchy–Riemann equations (22.1). ∎

Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a minimal isothermal patch, and write

$$\mathbf{x}(u,v) = \big(h_1(u,v),\ \ldots\ ,\ h_n(u,v)\big).$$

Having fixed $(u_0, v_0) \in \mathcal{U}$, we can apply Lemma 22.23 to each component h_j, so as to obtain a complex analytic function ψ_j for which

$$h_j(u,v) = \mathfrak{Re}\ \psi_j(u+iv) \qquad \text{and} \qquad 0 = \mathfrak{Im}\ \psi_j(z_0). \tag{22.27}$$

This allows us to record

Definition 22.24. *The mapping $\Psi = (\psi_1, \ldots, \psi_n)$ whose components satisfy* (22.27) *is called a* ***complexification*** *of* $\mathbf{x}$.

The key point is that each component of $\mathbf{x}$ is the real part of the corresponding component of Ψ. The ambiguity arising from this condition is resolved by demanding that Ψ coincides with $\mathbf{x}$ at z_0.

In the light of Lemma 22.23, the state of affairs can be summarized by

Corollary 22.25. *Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a minimal isothermal patch, where $\mathcal{U}$ is an open subset of $\mathbb{R}^2 = \mathbb{C}$ containing $(0,0)$. Let $\Psi = (\psi_1, \ldots, \psi_n)$ be the complexification of $\mathbf{x}$ such that $\mathfrak{Im}\,\Psi(0) = \mathbf{0}$. Then*

$$\Psi(z) = 2\mathbf{x}\Big(\frac{z}{2}, \frac{z}{2i}\Big) - \mathbf{x}(0,0),$$

and the conjugate harmonic isothermal patch $\mathbf{y}$ of $\mathbf{x}$ with $\mathbf{y}(0,0) = \mathbf{0}$ is given by

$$\mathbf{y}(u,v) = \mathfrak{Im}\left(2\mathbf{x}\left(\frac{u+iv}{2}, \frac{u+iv}{2\,i}\right) - \mathbf{x}(0,0)\right).$$

Enneper's Surface of Degree n

We can apply Corollary 22.25 with $\mathbf{x} = \mathsf{enneper}$, as defined on page 509. Replacing u and v by $z/2$ and $-iz/2$ respectively gives

$$\Psi(z) = \left(z - \frac{z^3}{3},\ iz + i\frac{z^3}{3},\ z^2\right). \tag{22.28}$$

This is then a minimal curve, and the conjugate to Enneper's surface is

$$\begin{aligned}\mathbf{y}(u,v) &= \mathfrak{Im}\ \Psi(u+iv)\\ &= \left(v + \frac{v^3}{3} - u^2 v,\ u + \frac{u^3}{3} - uv^2,\ 2uv\right).\end{aligned}$$

The associated family of $\mathsf{enneper}$ yields no essentially new surfaces because of the next result, which is verified in Notebook 22.

Lemma 22.26. *Let $t \mapsto \mathsf{enneper}[t]$ denote the associated family of Enneper's surface. Then $\mathsf{enneper}[t]$ coincides with a reparametrization of $\mathsf{enneper}$ rotated around the z-axis. Hence there is a Euclidean motion of $\mathbb{R}^3$ that maps $\mathsf{enneper}$ onto $\mathsf{enneper}[t]$.*

Formula (22.28) can be generalized by setting

$$\mathsf{ennepermincurve}[n](z) = \left(z - \frac{z^{2n+1}}{2n+1},\ iz + \frac{i\,z^{2n+1}}{2n+1},\ \frac{2z^{n+1}}{n+1}\right). \tag{22.29}$$

If we write

$$\mathsf{ennepermincurve}[n]' = (\phi_1, \phi_2, \phi_3),$$

it is easy to check that

$$\phi_1^2 + \phi_2^2 + \phi_3^2 = 0,$$

and that $|\phi_1|^2 + |\phi_2|^2 + |\phi_3|^2$ is never zero. Theorem 22.13 now asserts that the real part of (22.29) is a minimal surface.

Like Enneper's surface (the case $n = 1$), the patch

$$(u,v) \mapsto \mathfrak{Re}\,(\mathsf{ennepermincurve}[n](u+iv))$$

is hard to visualize because of its many self-intersections. As an aid to understanding this surface in Notebook 22, we determine the polar parametrization of each surface in its associated family. Figure 22.3 displays the case of $n = 2$ (undeformed, so $t = 0$) with a maximum radius $r_{\max}$ of 1.0, 1.6, 2.2 respectively. The curves on the right are the projections of the "polar boundary" corresponding to $r_{\max}$, and make the intersections self-evident.

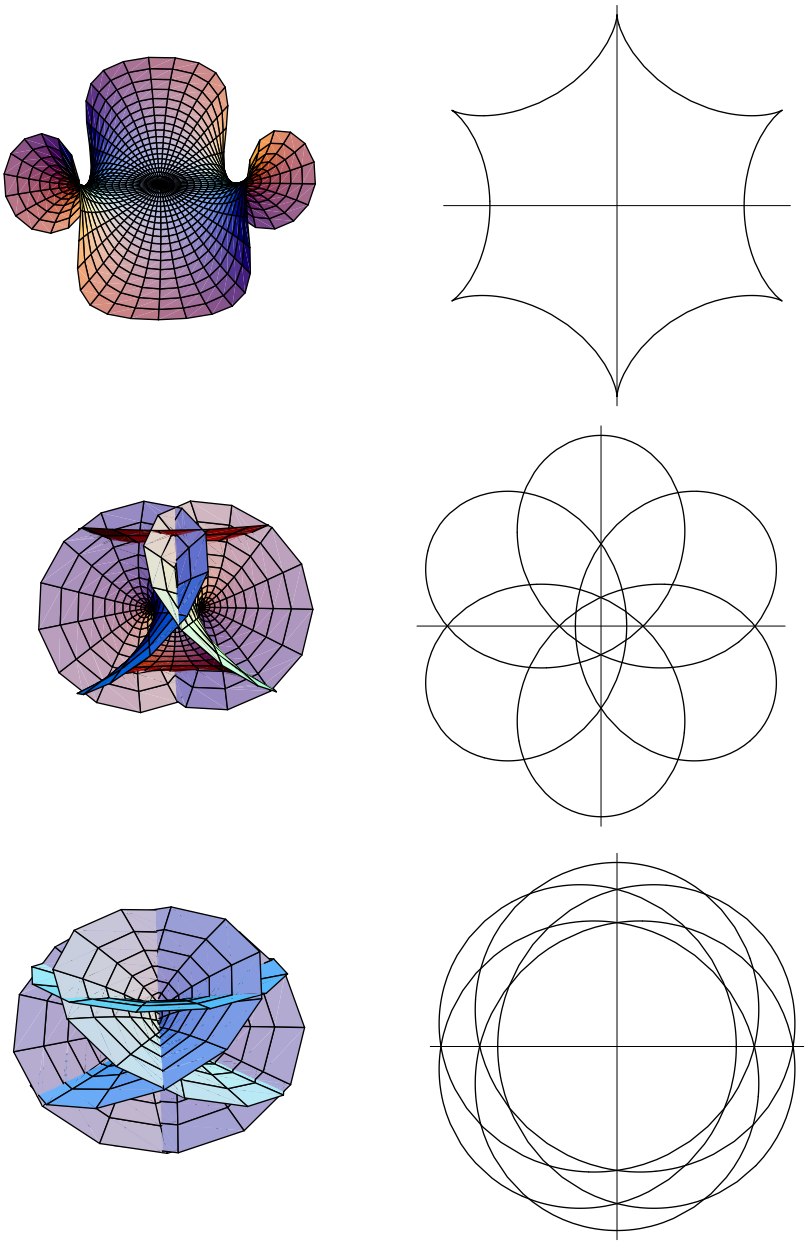


Figure 22.3: Polar plots of the surface enneper[2] with increasing radii

22.5 The Weierstrass Representation

The proof of the following result is almost obvious, yet it is of fundamental importance.

Theorem 22.27. *Let f and g be complex analytic functions defined in a region $\mathcal{U}$ of the complex plane $\mathbb{C}$. Fix $z_0 \in \mathcal{U}$, and define $\Psi = (\psi_1, \psi_2, \psi_3) \colon \mathcal{U} \to \mathbb{C}^3$ by*

$$\left\{ \begin{array}{rcl} \psi_1(z) &=& \displaystyle\int_{z_0}^{z} \frac{f(w)}{2}\big(1 - g(w)^2\big)\, dw, \\ \psi_2(z) &=& \displaystyle\int_{z_0}^{z} \frac{i f(w)}{2}\big(1 + g(w)^2\big)\, dw, \\ \psi_3(z) &=& \displaystyle\int_{z_0}^{z} f(w) g(w)\, dw. \end{array} \right. \tag{22.30}$$

Then Ψ is a minimal curve.

Proof. The complex derivative of Ψ equals

$$\big(\tfrac{1}{2} f(z)(1 - g(z)^2),\ \tfrac{1}{2} i f(z)(1 + g(z)^2),\ f(z)g(z)\big)\,, \tag{22.31}$$

and this is set up precisely to ensure that (22.19) is satisfied. ∎

We call Ψ the ***Weierstrass minimal curve*** starting at z_0 determined by $f(z)$ and $g(z)$. Put

$$\left\{ \begin{array}{rcl} \mathbf{x}(u,v) &=& \mathfrak{Re}\,\big(\psi_1(u,v),\ \psi_2(u,v),\ \psi_3(u,v)\big), \\ \mathbf{y}(u,v) &=& \mathfrak{Im}\,\big(\psi_1(u,v),\ \psi_2(u,v),\ \psi_3(u,v)\big), \end{array} \right. \tag{22.32}$$

so that Ψ is a complexification of $\mathbf{x}$. Since each ψ_j is complex analytic, we know that

$$\mathbf{x}_{uu} + \mathbf{x}_{vv} = \mathbf{y}_{uu} + \mathbf{y}_{vv} = 0,$$

and $\mathbf{x}, \mathbf{y}$ are both minimal isothermal patches.

Definition 22.28. *The map $\mathbf{x}$ is called the **Weierstrass patch** and $\mathbf{y}$ the **conjugate Weierstrass patch**, starting at z_0, determined by $f(z)$ and $g(z)$.*

Proposition 22.29. *The metric of both the Weierstrass patch and the conjugate Weierstrass patch is given by*

$$ds^2 = \frac{1}{4}|f(z)|^2\big(1 + |g(z)|^2\big)^2 |dz|^2.$$

In particular, $\mathbf{x}$ and $\mathbf{y}$ are isometric. They are regular except where f is zero or singular.

Proof. From (22.16), page 726, we have

$$\begin{aligned} E \;=\; G \;&=\; 2\sum_{k=1}^{3} |\phi_k[\mathbf{x}]|^2 \\ &=\; \frac{1}{8}\Big(|f|^2|1-g^2|^2 + |f|^2|1+g^2|^2 + 4|fg|^2\Big) = \tfrac{1}{4}|f|^2(1+|g|^2)^2. \end{aligned}$$

The result follows since $F = 0$ and $du^2 + dv^2 = |dz|^2$. ∎

As a first example, consider

$$f(z) = -e^{-z} \qquad \text{and} \qquad g(z) = -e^{z}.$$

Equation (22.31) yields

$$\Psi'(z) = \big(i\sin z,\; -i\cos z,\; 1\big).$$

Taking $z_0 = 0$ gives

$$\Psi(z) = \big(i(1-\cos z),\; -i\sin z,\; z\big).$$

With the aid of the formulas

$$\text{(22.33)} \qquad \begin{cases} \sin z \;=\; \sin u\cosh v + i\cos u\sinh v, \\ \cos z \;=\; \cos u\cosh v - i\sin u\sinh v, \end{cases}$$

it can be verified that the associated family coincides with the deformation from a helicoid to a catenoid; the resulting formulas are close to (16.8) on page 505.

Although we required the functions f, g above to be complex analytic, this is too restrictive a hypothesis in practice. Indeed, many of the well-known minimal surfaces can only be constructed using Weierstrass patches in which f and g are allowed to have isolated singularities. More precisely, we shall take f, g to be meromorphic functions, a concept we next recall.

Fix $z_0 \in \mathbb{C}$, $a > 0$, and let $\mathcal{U}$ be the "punctured disk" $\{z \mid 0 < |z - z_0| < a\}$. A complex analytic function $f\colon \mathcal{U} \to \mathbb{C}$ is said to have a ***pole or order*** k at z_0 (where $k \geqslant 1$) if

$$\lim_{z\to z_0} (z - z_0)^k f(z)$$

exists, but $|(z - z_0)^{k-1} f(z)| \to \infty$ as $z \to z_0$. Extending the value of a function to z_0 to make it continuous whenever possible, we may say that f has a pole of order k at z_0 if and only if $1/f$ has a ***zero or order*** k, which is perhaps easier to imagine. A function $f\colon \mathbb{C} \to \mathbb{C}$ is called ***meromorphic*** provided it is defined at all points apart from a (possibly infinite) number of poles.

A good example of the necessity to allow poles is provided by Henneberg's minimal surface, page 511, which is given by

Lemma 22.30. *The Weierstrass patch determined by the functions*

$$f(z) = 1 - \frac{1}{z^4} \qquad \text{and} \qquad g(z) = z$$

is a reparametrization of the surface $(u, v) \mapsto \frac{1}{2}\mathsf{henneberg}(u, -v)$.

The function f has a pole of order 4 at $z_0 = 0$.

A converse to Theorem 22.27 also holds:

Lemma 22.31. *Let* $\Psi : \mathcal{U} \to \mathbb{C}^3$ *be a minimal curve and write* $\Psi' = (\phi_1, \phi_2, \phi_3)$. *Suppose that* $\phi_1 - i\phi_2$ *is not identically zero. Define*

$$f = \phi_1 - i\phi_2, \qquad \text{and} \qquad g = \frac{\phi_3}{\phi_1 - i\phi_2}. \tag{22.34}$$

Then f *and* g *give rise to the Weierstrass representation of* Ψ; *that is,*

$$\Psi' = \big(\tfrac{1}{2}f(1 - g^2),\ \tfrac{1}{2}i\,f(1 + g^2),\ f\,g\big).$$

Proof. It is easy to check that if f and g are defined by (22.34), then

$$\phi_1 = \tfrac{1}{2}f(1 - g^2), \qquad \phi_2 = \tfrac{1}{2}if(1 + g^2) \qquad \text{and} \qquad \phi_3 = fg. \ \blacksquare$$

The associated family of the Weierstrass patch determined by meromorphic functions f, g is given by substituting (22.32) into (22.3) or (22.21). In so doing, we immediately discover

Lemma 22.32. *The patch* $\mathbf{z}[t]$ *is the Weierstrass patch determined by the meromorphic functions* $e^{-it}f$ *and* g.

Let us specialize the result of Lemma 22.22 to a Weierstrass patch.

Theorem 22.33. *The Gaussian curvature of a Weierstrass patch determined by meromorphic functions* f *and* g *is given by*

$$K = \frac{-16|g'|^2}{|f|^2(1 + |g|^2)^4}. \tag{22.35}$$

The same formula holds for the Gaussian curvature of each member of the associated family, including the conjugate Weierstrass patch determined by f, g.

Proof. We shall follow the proof (rather than the statement) of Lemma 22.22. We have

$$\boldsymbol{\Delta} \log|f(z)| = 2\frac{\partial^2}{\partial\bar{z}\partial z}\big(\log|f(z)|^2\big) = 2\frac{\partial^2}{\partial\bar{z}\partial z}\big(\log f(z) + \log\overline{f(z)}\big) = 0,$$

and

$$\begin{aligned}\boldsymbol{\Delta}\log\bigl(1+|g(z)|^2\bigr) &= 4\frac{\partial^2}{\partial\bar z\partial z}\log\bigl(1+g(z)\overline{g(z)}\bigr)\\ &= 4\frac{\partial}{\partial\bar z}\left(\frac{g'(z)\overline{g(z)}}{1+|g(z)|^2}\right)=\frac{4|g'(z)|^2}{(1+|g(z)|^2)^2}.\end{aligned}$$

We know from Proposition 22.29 that the scaling factor λ equals $|f|(1+|g|^2)$. It therefore follows that

$$\begin{aligned}\frac{1}{4}K=\frac{-\boldsymbol{\Delta}\log\bigl(|f|(1+|g|^2)\bigr)}{|f|^2(1+|g|^2)^2} &= \frac{-\boldsymbol{\Delta}\log(1+|g|^2)}{|f|^2(1+|g|^2)^2}\\ &= \frac{-\dfrac{4|g'|^2}{(1+|g|^2)^2}}{|f|^2(1+|g|^2)^2}=\frac{-4|g'|^2}{|f|^2(1+|g|^2)^4}.\ \blacksquare\end{aligned}$$

It follows that the Gaussian curvature of a Weierstrass patch determined by meromorphic functions f, g vanishes precisely at the zeros of g'. Hence if g' is not identically zero, the zeros of the Gaussian curvature are isolated. Note that the zeros of the Gaussian curvature of any minimal surface are both planar points and umbilic points.

There is a very simple formula for the Gauss map, or what amounts to the same thing, the unit normal, of a Weierstrass patch.

Theorem 22.34. *Let* $\mathbf{U}$ *be the unit normal of a Weierstrass patch determined by meromorphic functions* f, g*. Then* $\mathsf{stereo}\circ\mathbf{U}$*, considered as a function of a complex variable, is precisely the function* g*. Furthermore,*

$$\mathbf{U}(z)=\left(\frac{2\,\mathfrak{Re}\,g(z)}{|g(z)|^2+1},\ \frac{2\,\mathfrak{Im}\,g(z)}{|g(z)|^2+1},\ \frac{|g(z)|^2-1}{|g(z)|^2+1}\right). \tag{22.36}$$

Proof. Lemma 22.20 on page 730 implies that

$$\mathsf{stereo}\circ\mathbf{U}=\frac{\phi_3}{\phi_1-i\,\phi_2}=\frac{f\,g}{f(1-g^2)/2+f(1+g^2)/2}=g.$$

We also know (from page 371) that the inverse of stereographic projection is the function $\mathbb{C}\to S^2(1)$ given by

$$\mathsf{stereo}^{-1}(w)=\boldsymbol{\Upsilon}(w)=\left(\frac{2\,\mathfrak{Re}\,w}{|w|^2+1},\ \frac{2\,\mathfrak{Im}\,w}{|w|^2+1},\ \frac{|w|^2-1}{|w|^2+1}\right).$$

Formula (22.36) now follows. $\blacksquare$

We have already noted in Theorem 22.6 that the Gauss map is the same for all members of an associated family. So (22.36) is the formula for the unit normal of any member of the associated family of a Weierstrass patch.

Richmond's Minimal Surface

An interesting sequence of minimal surfaces more complicated than either the catenoid or Enneper's surface is generated by the Weierstrass representation when

$$f(z) = \frac{1}{z^2} \qquad \text{and} \qquad g(z) = z^{n+1}.$$

These surfaces have "one planar end," in the sense that, far away from a center, their Gaussian curvature K tends to zero, and in this region they resemble the plane. Figure 22.4 illustrates the case $n = 1$, which is named after Richmond.[4]

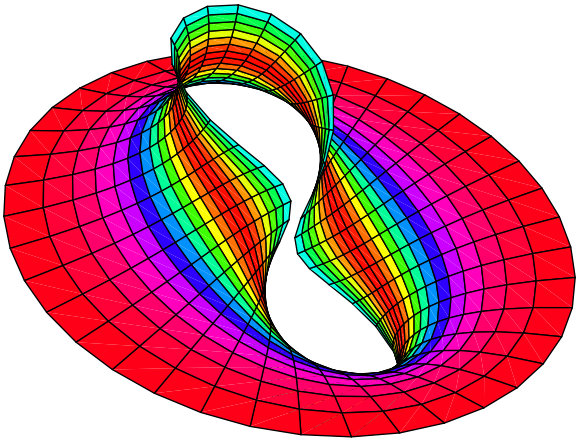


Figure 22.4: Richmond's minimal surface

Integrating, we can define the minimal curve by

$$\Psi[n](z) = \left(-\frac{1}{2z} - \frac{z^{2n+1}}{4n+2},\ -\frac{i}{2z} + \frac{i\,z^{2n+1}}{4n+2},\ \frac{z^n}{n} \right).$$

Like (22.29), this can be studied effectively using polar coordinates, and we can define an associated family

$$\begin{aligned} \boldsymbol{\rho}[n][t](r,\theta) = \Big(& \frac{-\cos(t+\theta)}{2r} - \frac{r^{2n+1}\cos\big(t-(2n+1)\theta\big)}{4n+2}, \\ & -\frac{\sin(t+\theta)}{2r} + \frac{r^{2n+1}\sin\big(t-(2n+1)\theta\big)}{4n+2},\ \frac{r^n\cos(t-n\theta)}{n} \Big). \end{aligned}$$

Fortuitously, the Gaussian curvature K is a function of r alone, and approaches its supremum 0 as r tends to zero or infinity.

---

[4] Herbert William Richmond (1863–1948). English mathematician.

We return to Richmond's surface by setting $n = 1$, and seek the minimum value of the Gaussian curvature of $\boldsymbol{\rho}[1]$. Calculations in Notebook 22 show that $r = (3/5)^{1/4}$ is a critical point of K, and it can be checked (for example, by plotting K as a function of r) that this is an absolute minimum.

22.6 Minimal Surfaces via Björling's Formula

In 1844, Björling[5] asked if it is possible to find a minimal surface containing a given analytic strip [Bj]. This question is frequently called Björling's problem. H. A. Schwarz[6] (see [Schwa, pages 179–189]) gave an elegant solution to it. For refinements of Schwarz's method see [Nits, pages 139–145] and [DHKW, pages 120–135]. We first make this notion precise.

Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ be a curve. By a ***holomorphic extension*** of $\boldsymbol{\alpha}$, we mean a complex analytic mapping $\widetilde{\boldsymbol{\alpha}}\colon \mathcal{R} \to \mathbb{C}^3$ defined on a rectangle

$$\mathcal{R} = \{u + iv : u \in (a,b),\ v \in (c,d)\} \quad \text{with} \quad c < 0 < d,$$

and $\widetilde{\boldsymbol{\alpha}}(u) = \boldsymbol{\alpha}(u)$ for all $u \in (a,b)$. Such an extension is unique if it exists (in which case $\boldsymbol{\alpha}$ is itself *real analytic*), and so we can safely omit the tilde and use the same symbol $\boldsymbol{\alpha}$ for the extension. We shall also denote $\mathcal{R}$ simply by $(a,b) \times (c,d)$.

Definition 22.35. *Let $\boldsymbol{\alpha}, \boldsymbol{\gamma}\colon (a,b) \to \mathbb{R}^3$ be curves such that*

$$\|\boldsymbol{\gamma}\| = 1 \qquad \textit{and} \qquad \boldsymbol{\alpha}' \cdot \boldsymbol{\gamma} = 0. \tag{22.37}$$

Suppose that there exist holomorphic extensions $\boldsymbol{\alpha}, \boldsymbol{\gamma}\colon (a,b) \times (c,d) \to \mathbb{C}^3$, such that (22.37) *holds also for $z \in (a,b) \times (c,d)$. Fix $z_0 \in (a,b) \times (c,d)$. Then the* ***Björling curve*** *corresponding to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ is given by*

$$\mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](z) = \boldsymbol{\alpha}(z) - i \int_{z_0}^{z} \boldsymbol{\gamma}(z) \times \boldsymbol{\alpha}'(z)\, dz.$$

We now prove that a Björling curve is in fact a minimal curve, and we show how it is related to $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

Theorem 22.36. *With the same hypotheses as Definition* 22.35*, we have*

---

[5] Emanuel Gabriel Björling (1808–1872). Swedish mathematician. Professor at the University of Uppsala.

[6] 

Herman Amandus Schwarz (1843–1921). German mathematician, who made fundamental contributions to the theories of minimal surfaces and conformal mappings. His work on eigenvalues lead to what we now call the Cauchy-Schwarz inequality.

(i) *the mapping* $z \mapsto \mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](z)$ *is a minimal curve*;

(ii) *for real* u, *the map* $u \mapsto \mathfrak{Re}\big(\mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](u)\big) - \boldsymbol{\alpha}(u)$ *is constant*;

(iii) *for real* u, $\mathbf{U}(u, 0) = \boldsymbol{\gamma}(u)$, *where* $\mathbf{U}$ *denotes the unit normal to the patch*

$$(u, v) \mapsto \mathfrak{Re}\big(\mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](u + iv)\big);$$

Proof. Write $\Psi = \mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}]$; then $\Psi' = \boldsymbol{\alpha}' - i\boldsymbol{\gamma} \times \boldsymbol{\alpha}'$. Using (7.7), page 195, and (22.37), we find that

$$\begin{aligned} \Psi' \cdot \Psi' &= (\boldsymbol{\alpha}' - i\boldsymbol{\gamma} \times \boldsymbol{\alpha}') \cdot (\boldsymbol{\alpha}' - i\boldsymbol{\gamma} \times \boldsymbol{\alpha}') \\ &= \boldsymbol{\alpha}' \cdot \boldsymbol{\alpha}' - (\boldsymbol{\gamma} \times \boldsymbol{\alpha}') \cdot (\boldsymbol{\gamma} \times \boldsymbol{\alpha}') - 2i\,\boldsymbol{\alpha}' \cdot (\boldsymbol{\gamma} \times \boldsymbol{\alpha}') \\ &= \boldsymbol{\alpha}' \cdot \boldsymbol{\alpha}' - (\boldsymbol{\gamma} \cdot \boldsymbol{\gamma})(\boldsymbol{\alpha}' \cdot \boldsymbol{\alpha}') \;=\; 0. \end{aligned}$$

Hence Ψ is a minimal curve.

Now, (ii) holds because $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\gamma}(u)$ are real whenever u is itself real. To prove (iii), we observe that

$$\Psi'(u) = \boldsymbol{\alpha}'(u) - i\boldsymbol{\gamma}(u) \times \boldsymbol{\alpha}'(u) \quad \text{and} \quad \overline{\Psi'(u)} = \boldsymbol{\alpha}'(u) + i\boldsymbol{\gamma}(u) \times \boldsymbol{\alpha}'(u)$$

for $u \in \mathbb{R}$. Hence by (22.37),

$$\Psi'(u) \cdot \overline{\Psi'(u)} = 2\boldsymbol{\alpha}'(u) \cdot \boldsymbol{\alpha}'(u),$$

and by (7.3) on page 193,

$$\Psi'(u) \times \overline{\Psi'(u)} = 2i\,\boldsymbol{\alpha}'(u) \times \big(\boldsymbol{\gamma}(u) \times \boldsymbol{\alpha}'(u)\big) = 2i\big(\boldsymbol{\alpha}'(u) \cdot \boldsymbol{\alpha}'(u)\big)\boldsymbol{\gamma}(u)$$

Using Lemma 22.18, we find that

$$\mathbf{U}(u, 0) = \frac{\Psi'(u) \times \overline{\Psi'(u)}}{i\|\Psi'(u)\|^2} = \boldsymbol{\gamma}(u). \;\blacksquare$$

We can use Theorem 22.36 to construct many interesting minimal surfaces. For example, let $\boldsymbol{\alpha}\colon (a, b) \to \mathbb{R}^2$ be a regular plane curve which has a regular holomorphic extension $\boldsymbol{\alpha}\colon (a, b) \times (c, d) \to \mathbb{C}^2$. We take $\boldsymbol{\gamma} = J\boldsymbol{\alpha}'/\|\boldsymbol{\alpha}'\|$. Then

$$\boldsymbol{\gamma} \times \boldsymbol{\alpha}' = -\|\boldsymbol{\alpha}'\|\,\mathbf{E}_3,$$

where $\mathbf{E}_3$ is the unit vector field perpendicular to $\mathbb{R}^2$ in $\mathbb{R}^3$. Thus

$$\mathsf{bjoerling}[\boldsymbol{\alpha}, \boldsymbol{\gamma}](z) = \boldsymbol{\alpha}(z) + i\left(\int_{z_0}^{z} \sqrt{\boldsymbol{\alpha}'(z) \cdot \boldsymbol{\alpha}'(z)}\, dz\right)\mathbf{E}_3.$$

This leads to the following definition.

Definition 22.37. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a plane curve, and suppose that $\boldsymbol{\alpha}$ has a holomorphic extension $\boldsymbol{\alpha}\colon (a,b)\times(c,d) \to \mathbb{C}^2$. Write $\boldsymbol{\alpha} = (a_1, a_2)$. We define*

$$\mathsf{bjmincurve}[\boldsymbol{\alpha}](z) = \left(a_1(z),\ a_2(z),\ i\int_{z_0}^{z} \sqrt{a_1'(z)^2 + a_2'(z)^2}\, dz\right).$$

We need a result about the intersection of a plane and a surface in $\mathbb{R}^3$.

Theorem 22.38. *Let $\boldsymbol{\beta}$ be a unit-speed curve which lies on the intersection of a regular surface $\mathcal{M} \subset \mathbb{R}^3$ and a plane $\mathit{\Pi}$. Suppose that $\mathcal{M}$ meets $\mathit{\Pi}$ perpendicularly along $\boldsymbol{\beta}$. Then $\boldsymbol{\beta}$ is a geodesic in $\mathcal{M}$.*

Proof. Let $\mathbf{U}$ denote the unit normal vector field to $\mathcal{M}$. Then $\boldsymbol{\beta}'$, $\boldsymbol{\beta}''$ and $\mathbf{U}\circ\boldsymbol{\beta}$ all lie in $\mathit{\Pi}$, and $\boldsymbol{\beta}'$ is perpendicular to $\mathbf{U}\circ\boldsymbol{\beta}$. Since $\boldsymbol{\beta}$ is a unit-speed curve, $\boldsymbol{\beta}''$ is perpendicular to $\boldsymbol{\beta}'$. Hence $\boldsymbol{\beta}''$ is a multiple of $\mathbf{U}\circ\boldsymbol{\beta}$, and so $\boldsymbol{\beta}$ is a geodesic in $\mathcal{M}$. ∎

Theorem 22.39. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^2$ be a unit-speed plane curve that has a holomorphic extension $\boldsymbol{\beta}\colon (a,b)\times(c,d) \to \mathbb{C}^2$, where $c < 0 < d$. Then*

$$\mathfrak{Re}\big(\mathsf{bjmincurve}\,[\boldsymbol{\beta}]\big)\colon (a,b)\times(c,d) \to \mathbb{R}^3$$

is a minimal surface which contains $\boldsymbol{\beta}$ as a geodesic.

Proof. Writing $\Psi = \mathsf{bjmincurve}[\boldsymbol{\beta}]$, it is immediate that

$$\Psi'[\boldsymbol{\beta}] \cdot \Psi'[\boldsymbol{\beta}] = 0,$$

so that $\Psi[\boldsymbol{\beta}]$ really is a minimal curve. Furthermore, if we put $\boldsymbol{\beta} = (b_1, b_2, b_3)$, then Lemma 22.18 expresses the unit normal of $\Psi[\boldsymbol{\beta}]$ as

$$\mathbf{U} = \frac{2\left(\mathfrak{Im}\left(-i\overline{b_2'}\sqrt{b_1'^2 + b_2'^2}\right),\ \mathfrak{Im}\left(i\overline{b_1'}\sqrt{b_1'^2 + b_2'^2}\right),\ \mathfrak{Im}\left(b_1'\overline{b_2'}\right)\right)}{|b_1'|^2 + |b_2'|^2 + |\sqrt{b_1'^2 + b_2'^2}\,|^2}.$$

Since $b_1'(z), b_2'(z)$ are real for $z \in \mathbb{R}$, we get

$$\mathbf{U}(u,0) \;=\; \frac{2\left(-b_2'\sqrt{b_1'^2 + b_2'^2},\ b_1'\sqrt{b_1'^2 + b_2'^2},\ 0\right)}{2(b_1'^2 + b_2'^2)}\bigg|_{z=(u,0)} = \frac{(-b_2',\ b_1',\ 0)}{\sqrt{b_1'^2 + b_2'^2}}\bigg|_{z=(u,0)}.$$

Therefore, $\mathbf{U}$ is perpendicular to $\mathbf{E}_3$, the unit normal of the xy-plane $\mathit{\Pi}$. Now Theorem 22.38 implies that $\boldsymbol{\beta}$ is a geodesic in $\mathfrak{Re}\,\Psi[\boldsymbol{\beta}]$. ∎

Corollary 22.40. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^2$ be a regular plane curve that has a holomorphic extension $\boldsymbol{\alpha}\colon (a,b)\times(c,d) \to \mathbb{C}^2$, where $c < 0 < d$. Then*

$$\mathfrak{Re}\big(\mathsf{bjmincurve}[\boldsymbol{\alpha}]\big)\colon (a,b)\times(c,d) \longrightarrow \mathbb{R}^3$$

is a minimal curve that contains $\boldsymbol{\alpha}$ as a pregeodesic.

Proof. Write $\Psi = \mathsf{bjmincurve}[\boldsymbol{\beta}]$ again. Then there exists an increasing real analytic function $h\colon (e,f) \to (c,d)$ such that $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$ is a unit-speed plane curve with a holomorphic extension. Since $\Psi[\boldsymbol{\beta}](z) = \Psi[\boldsymbol{\alpha}](h(z))$ and

$$\Psi[\boldsymbol{\beta}]'(z) = \Psi[\boldsymbol{\alpha}]\big(h(z)\big)h'(z),$$

it follows that $\Psi[\boldsymbol{\beta}]$ is a minimal curve with the same range as $\boldsymbol{\alpha}$. Since $\boldsymbol{\beta}$ is a geodesic in $\mathfrak{Re}\,\Psi[\boldsymbol{\beta}] = \mathfrak{Re}\,\Psi[\boldsymbol{\alpha}]$, we conclude that $\boldsymbol{\alpha}$ is a pregeodesic in $\mathfrak{Re}\,\Psi[\boldsymbol{\alpha}]$. ∎

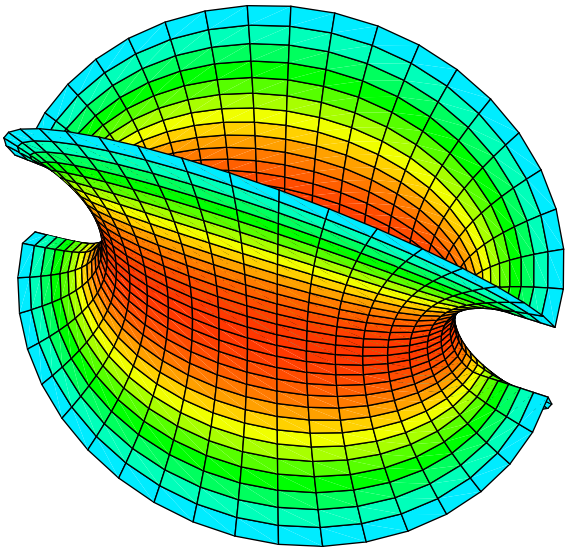


Figure 22.5: A minimal surface with an ellipse as geodesic

As a first example, let us find a minimal surface that contains the cycloid

$$t \mapsto \big(t - \sin t,\ 1 - \cos t\big)$$

as a geodesic. The minimal curve obtained from Definition 22.37 is

$$\Big(z - \sin z,\ 1 - \cos z,\ -4i\cos\frac{z}{2}\Big).$$

To take its real part we use (22.33), and deduce that the parametrization of the corresponding minimal surface is

$$\Big(u - \cosh v \sin u,\ 1 - \cos u \cosh v,\ -4\sin\frac{u}{2}\sinh\frac{v}{2}\Big).$$

Not surprisingly, the resulting minimal surface is precisely Catalan's, given its description on page 510.

It is clear that a catenoid contains a circle as a geodesic, but what is the minimal surface that contains an ellipse as a geodesic? (See [Schwa, page 182].) We can solve this problem using the above methods, by first using the formula

$$s(t) = b\,\mathrm{E}\big(t \,\big|\, 1 - \frac{a^2}{b^2}\big)$$

for the arc length of the ellipse $x^2/a^2 + y^2/b^2 = 1$. This allows us to define the corresponding minimal curve

$$\Phi(z) = \Big(a\cos z,\ b\sin z,\ ib\,\mathrm{E}\big(z \,\big|\, 1 - \frac{a^2}{b^2}\big)\Big).$$

A computation from Notebook 22 then produces the associated family, and Figure 22.5 plots the minimal surface

$$\mathbf{x}(u,v) = \Big(2\cos u\cosh v,\ \cosh v\sin u,\ -\mathfrak{Im}\ \mathrm{E}\big(u+iv \,\big|\, -3\big)\Big),$$

corresponding to $a = 2$, $b = 1$ and $t = 0$.

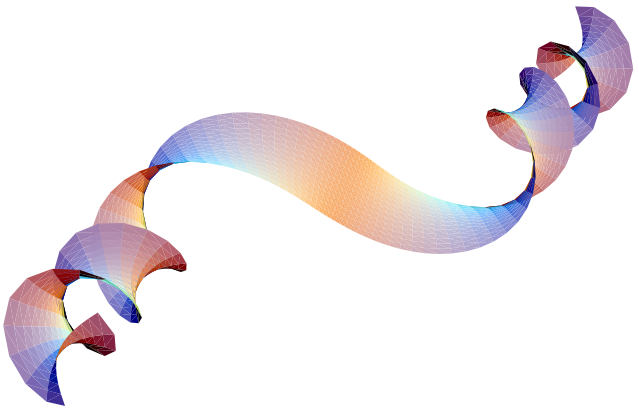


Figure 22.6: A minimal surface generated by a clothoid

Many more examples are investigated in Notebook 22; see Exercise 13 and Figure 22.11 on page 754. We conclude this section by illustrating the minimal surface generated by one of the most elegant of all plane curves, namely the clothoid. This gives rise to a minimal curve

$$\Phi(z) = \left(\sqrt{\pi}\,\mathrm{FresnelS}\Big(\frac{x}{\sqrt{\pi}}\Big),\ \sqrt{\pi}\,\mathrm{FresnelC}\Big(\frac{z}{\sqrt{\pi}}\Big),\ iz\right),$$

and the surface in Figure 22.6 is (a rotated version of) the member of the associated family with $t = \pi/4$. Since $t \neq 0$, this surface no longer contains the initial clothoid; as t tends to $\pi/2$ the corresponding surface resembles more and more a helicoid.

22.7 Costa's Minimal Surface

One of the most interesting minimal surfaces was defined by Costa in 1984 (see [Costa1]). Hoffman and Meeks (see [HoMe]) showed that Costa's minimal surface is a complete embedded surface. The term "complete" will be properly defined on page 889, while "embedded" means that the surface has no self-intersections. We shall adapt the treatment of Costa's minimal surface given in [BaCo] to find a parametrization that can be investigated in Notebook 22. For this, we need the Weierstrass zeta function ζ. First, we explain the relevant theory of this function and also the Weierstrass function $\wp$.

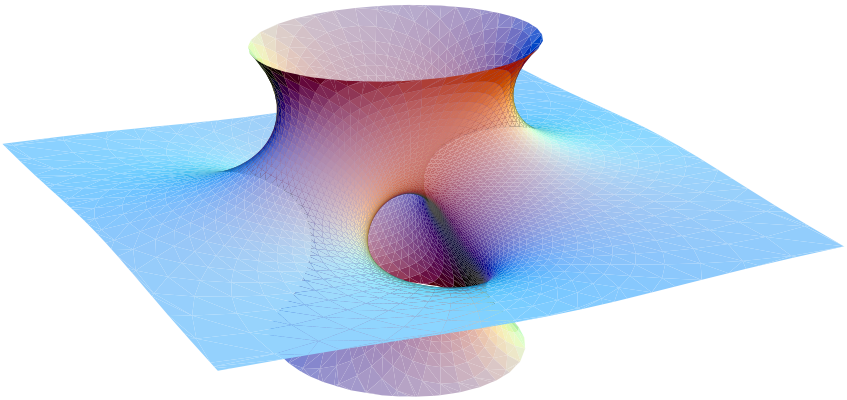


Figure 22.7: Costa's minimal surface

The Weierstrass $\wp$ function is a meromorphic function on the complex plane that is used to relate the 2-dimensional torus (as defined by Figure 11.4 on page 337) to complex cubic curves. The resulting introductory theory of elliptic curves is clearly presented in many texts, for example [Kir]. The function $\wp$ is usually defined for an arbitrary ***lattice***, meaning a set

$$\{m\omega_1 + n\omega_2 \mid m, n \in \mathbb{Z}\},$$

where ω_1, ω_2 are complex numbers with $\mathfrak{Im}(\omega_2/\omega_1) \neq 0$. There is little loss in generality in assuming that $\omega_1 = 1$, but we shall mainly consider the special case in which $\omega_1 = 1$ and $\omega_2 = i$, so that our lattice becomes the set

$$\mathscr{L} = \{m + in \mid m, n \in \mathbb{Z}\} \tag{22.38}$$

of Gaussian integers. It is also convenient to denote by $\mathscr{L}^*$ the set $\mathscr{L} \setminus \{0\}$ of nonzero lattice points.

Having defined $\mathcal{L}^*$, we may define the *Weierstrass* $\wp$ *function* by

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in \mathscr{L}^*} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right). \tag{22.39}$$

Then $\wp$ is defined for all $z \in \mathbb{C} \setminus \mathscr{L}$, but has a pole of order 2 at each lattice point. Moreover, $\wp$ is known to be ***doubly periodic***, in the sense that

$$\wp(z + m + in) = \wp(z)$$

for all $m, n \in \mathbb{Z}$ and $z \in \mathbb{C} \setminus \mathscr{L}$. Thus, $\wp$ is completely determined by its values on the *fundamental domain*

$$\mathscr{F} = \{z \in \mathbb{C} \mid z = \alpha + \beta i \text{ with } 0 \leqslant \alpha, \beta < 1\}.$$

The Weierstrass $\wp$ function satisfies an addition formula:

$$\wp(z_1 + z_2) = \frac{1}{4} \left(\frac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)} \right)^2 - \wp(z_1) - \wp(z_2). \tag{22.40}$$

The next result establishes the link with cubic equations.

Lemma 22.41. The function $\wp$ is a solution of the differential equation

$$\wp'(z)^2 = (4\wp(z)^2 - c)\wp(z), \tag{22.41}$$

where c is a nonzero constant.

Proof. We shall deduce (22.41) from a more general, but well-known, result. Namely,

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3,$$

where

$$g_2 = 60 \sum_{\omega \in \mathscr{L}^*} \frac{1}{\omega^4}, \quad \text{and} \quad g_3 = 140 \sum_{\omega \in \mathscr{L}^*} \frac{1}{\omega^6}. \tag{22.42}$$

This applies for a general lattice $\mathscr{L}$ (see, for example, [Chand, page 29]), though we claim that in our case (22.38), $g_3 = 0$. To verify this, observe that

$$(m - in)^6 + (m + in)^6 + (n - im)^6 + (n + im)^6 = 0;$$

it follows that the summands representing g_3 in (22.42) with $m^2 + n^2$ fixed cancel out. To obtain (22.41), we set $c = g_2$, and use Notebook 22 to obtain the approximation $c = 189.073$. ∎

Since $\wp'$ is an odd function of z with period 1, we see that

$$\wp'(\tfrac{1}{2}) = -\wp'(-\tfrac{1}{2}) = -\wp'(\tfrac{1}{2})$$

is zero. Similarly,

$$\wp'\big(\frac{i}{2}\big) = \wp'\big(\frac{1+i}{2}\big) = 0.$$

Moreover, $\wp'$ has no other zeros in $\mathscr{F}$ besides $1/2$, $i/2$, $(1+i)/2$. It follows from (22.41) that these three numbers are the roots of the cubic polynomial $(4\wp(z)^2 - c)\wp(z) = 0$, and it is known that $\wp((1+i)/2) = 0$. It follows that

$$e = \wp(\frac{1}{2}) = -\wp(\frac{i}{2})$$

satisfies $c = 4e^2$, and (22.41) becomes

$$\wp'(z)^2 = 4\wp(z)\big(\wp(z)^2 - e^2\big). \tag{22.43}$$

We shall also need the *Weierstrass ζ function*, which is defined by

$$\zeta(z) = \frac{1}{z} + \sum_{\omega \in \mathscr{L}^*} \Big(\frac{1}{z-\omega} + \frac{1}{\omega} + \frac{z}{\omega^2}\Big). \tag{22.44}$$

It is easy to see from (22.39) that

$$\zeta'(z) = -\wp(z). \tag{22.45}$$

The function ζ is not periodic, but it is known that

$$\zeta(z + m + in) = \zeta(z) + 2m\,\zeta(\tfrac{1}{2}) + 2n\,\zeta(\tfrac{i}{2}), \tag{22.46}$$

for integers m, n.

Lemma 22.42. *For all z, we have*

$$\wp\big(z - \tfrac{1}{2}\big) - \wp\big(z - \frac{i}{2}\big) - 2e \;=\; \frac{16e^3\wp(z)}{\wp'(z)^2}, \tag{22.47}$$

$$i\,\zeta(iz) \;=\; \zeta(z), \tag{22.48}$$

$$\zeta(\tfrac{1}{2}) = i\,\zeta(\tfrac{i}{2}) \;=\; \frac{\pi}{2}, \tag{22.49}$$

$$\zeta\big(\tfrac{1+i}{2}\big) \;=\; \frac{(1-i)\pi}{2}. \tag{22.50}$$

Proof. To verify (22.47), one uses (22.40) and (22.43) to prove

$$\wp\big(z - \tfrac{1}{2}\big) = e + \frac{2e^2}{\wp(z) - e},$$

and an analogous formula for $\wp(z - i/2)$. Equation (22.48) follows from the definition (22.44), and the second equality in (22.49) from Legendre's relation in the form

$$\zeta(\tfrac{1}{2})i - \zeta(\tfrac{i}{2}) = i\pi. \tag{22.51}$$

Finally, (22.50) follows from the formula

$$\zeta(\tfrac{1+i}{2}) = \zeta(\tfrac{1}{2}) + \zeta(\tfrac{i}{2}),$$

itself a consequence of (22.46). For more details, see [Chand, page 50]. ▌

A description of Costa's minimal surface using the Weierstrass function $\wp$ has been given by Barbosa and Colares (see [BaCo, pages 84–90], and also [HoMe]). We adapt this description, but we use instead the Weierstrass zeta function ζ to avoid one numerical integration. The use of ζ instead of $\wp$ speeds up the plotting of Costa's minimal surface by a factor of at least 50.

Costa's minimal surface can be defined as a Weierstrass patch using the functions

$$f(z) = \wp(z) \qquad \text{and} \qquad g(z) = \frac{A}{\wp'(z)}. \tag{22.52}$$

In order that Costa's minimal surface has no self-intersections, we need to take

$$A = \sqrt{2\pi c} = 2\sqrt{2\pi}\, e \approx 34.46707 \tag{22.53}$$

(see [BaCo, page 89]). The ***Costa minimal curve*** is then the minimal curve $\Psi\colon \mathbb{C} \to \mathbb{C}^3$ defined as follows. We put

$$2\,\Psi'(z) = \Big(\tfrac{1}{2} f(z)(1 - g(z)^2),\ \tfrac{1}{2} i f(z)(1 + g(z)^2),\ f(z)g(z)\Big), \tag{22.54}$$

and take Ψ to be the antiderivative of Ψ' with the normalization $\Psi((1+i)/2) = \mathbf{0}$. Next, we show how to use the Weierstrass function ζ to express Ψ without integrals.

Theorem 22.43. *The Costa minimal curve is given by* $\Psi = (\psi_1, \psi_2, \psi_3)$ *where*

$$\begin{cases} \psi_1 \;=\; \dfrac{1}{2}\Big\{ -\zeta(z) + \pi z - i\pi + \dfrac{\pi^2(1+i)}{4e} \\ \qquad\qquad + \dfrac{\pi}{2e}\Big[\zeta\big(z - \tfrac{1}{2}\big) - \zeta\big(z - \tfrac{i}{2}\big)\Big]\Big\}, \\[2ex] \psi_2 \;=\; \dfrac{i}{2}\Big\{ -\zeta(z) - \pi z + \pi - \dfrac{\pi^2(1+i)}{4e} \\ \qquad\qquad - \dfrac{\pi}{2e}\Big[\zeta\big(z - \tfrac{1}{2}\big) - \zeta\big(z - \tfrac{i}{2}\big)\Big]\Big\}, \\[2ex] \psi_3 \;=\; \dfrac{\sqrt{2\pi}}{4}\Big\{ \log\Big(\dfrac{\wp(z) - e}{\wp(z) + e}\Big) - \pi i\Big\}. \end{cases}$$

Proof. We shall only prove the expression for ψ_1. Using (22.47) and (22.53), we obtain

$$\begin{aligned} f(w)\big(1-g(w)^2\big) &= \wp(w) - \frac{A^2\wp(w)}{\wp'(w)^2} \\ &= \wp(w) - \frac{A^2}{16e^3}\Big(\wp\big(w-\tfrac{1}{2}\big) - \wp\big(w-\tfrac{i}{2}\big) - 2e\Big) \\ &= \wp(w) - \frac{\pi}{2e}\Big(\wp\big(w-\tfrac{1}{2}\big) - \wp\big(w-\tfrac{i}{2}\big) - 2e\Big) \\ &= \wp(w) + \pi - \frac{\pi}{2e}\Big(\wp\big(w-\tfrac{1}{2}\big) - \wp\big(w-\tfrac{i}{2}\big)\Big). \end{aligned}$$

Integrating both sides and using (22.49) and (22.50), we get

$$\begin{aligned} &\int_{(1+i)/2}^{z} f(w)\big(1-g(w)^2\big)\,dw \\ &\quad= \Big\{-\zeta(w) + \pi w + \frac{\pi}{2e}\Big(\zeta\big(w-\tfrac{1}{2}\big) - \zeta\big(w-\tfrac{i}{2}\big)\Big)\Big\}\Big|_{(1+i)/2}^{z} \\ &\quad= -\zeta(z) + \pi z + \frac{\pi}{2e}\Big(\zeta\big(z-\tfrac{1}{2}\big) - \zeta\big(z-\tfrac{i}{2}\big)\Big) \\ &\qquad\qquad + \zeta\big(\tfrac{1+i}{2}\big) - \frac{\pi(1+i)}{2} - \frac{\pi}{2e}\Big(\zeta\big(\tfrac{i}{2}\big) - \zeta\big(\tfrac{1}{2}\big)\Big) \\ &\quad= -\zeta(z) + \pi z + \frac{\pi}{2e}\Big(\zeta\big(z-\tfrac{1}{2}\big) - \zeta\big(z-\tfrac{i}{2}\big)\Big) - i\pi + \frac{\pi^2(1+i)}{4e}. \end{aligned}$$

The formulas for ψ_2 and ψ_3 are similar. ∎

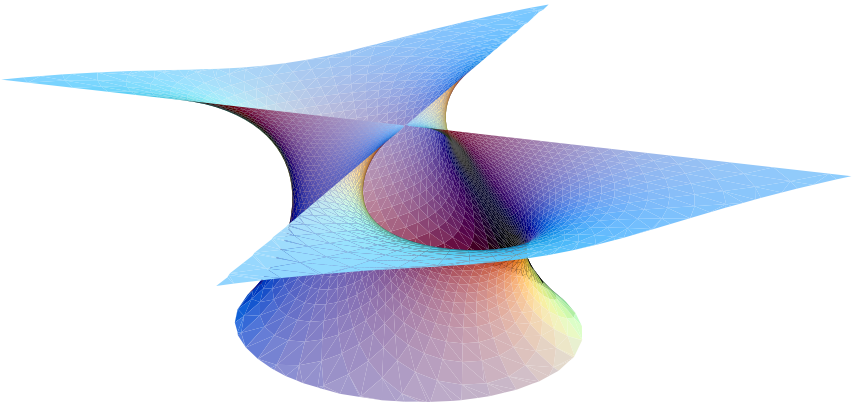


Figure 22.8: The lower part of Costa's minimal surface

Corollary 22.44. Costa's minimal surface *is given by* (x_1, x_2, x_3) *where*

$$\begin{cases} x_1(u,v) &= \frac{1}{2}\,\mathfrak{Re}\Big\{-\boldsymbol{\zeta}(u+iv)+\pi u+\frac{\pi^2}{4e} \\ &\qquad\qquad +\frac{\pi}{2e}\Big[\boldsymbol{\zeta}(u+iv-\tfrac{1}{2})-\boldsymbol{\zeta}(u+iv-\tfrac{1}{2}i)\Big]\Big\}, \\ x_2(u,v) &= \frac{1}{2}\,\mathfrak{Re}\Big\{-i\boldsymbol{\zeta}(u+iv)+\pi v+\frac{\pi^2}{4e} \\ &\qquad\qquad -\frac{\pi i}{2e}\Big[\boldsymbol{\zeta}(u+iv-\tfrac{1}{2})-\boldsymbol{\zeta}(u+iv-\tfrac{1}{2}i)\Big]\Big\}, \\ x_3(u,v) &= \frac{\sqrt{2\pi}}{4}\log\left|\frac{\boldsymbol{\wp}(u+iv)-e}{\boldsymbol{\wp}(u+i\,v)+e}\right|. \end{cases}$$

Further technical details regarding the above parametrization can be found in [FeGrMa]. It was used by Ferguson[7] to craft sculptures of Costa's surface in various materials, including snow. He produced a limited edition of fifty miniature bronze sculptures of the surface for the international conference in memory of the author in Bilbao in 2000; the inscription of each includes the expression "D_4" indicating the symmetry group that Costa's surface possesses.

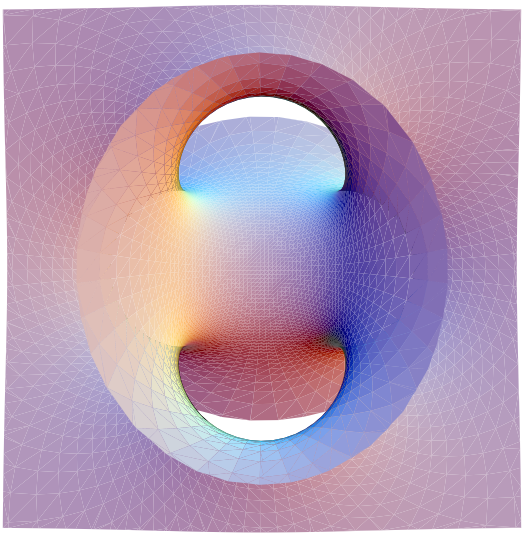
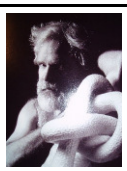


Figure 22.9: Costa's minimal surface seen from above



7 Helaman Ferguson trained as a stone mason and artist before embarking on a university career in mathematics. He now combines both disciplines to produce world-renowned sculptures, many of which represent sophisticated surfaces in differential geometry [Ferg]. He also designs algorithms for operating machinery and for scientific visualization.

22.8 Exercises

1. Give details for the proofs of Lemma 22.5, Corollary 22.11 and Lemma 7.3.

2. Show that if $\Psi\colon \mathcal{U} \to \mathbb{C}^n$ is a minimal curve, then

$$\Psi' \cdot \Psi'' = 0 \qquad \text{and} \qquad \Psi' \cdot \Psi''' + \Psi'' \cdot \Psi'' = 0.$$

3. Let $\Psi\colon \mathcal{U} \to \mathbb{C}^3$ be a minimal curve such that Ψ' and $\Psi'' \cdot \Psi''$ never vanish. Show that Ψ', Ψ'', Ψ''' are linearly independent at all points of $\mathcal{U}$.

4. Let $\Psi\colon \mathcal{U} \to \mathbb{C}^3$ be a minimal curve such that Ψ' and $\Psi'' \cdot \Psi''$ never vanish. Show that

$$\Psi' \times \Psi'' = -\left(\frac{\Psi' \times \Psi'' \cdot \Psi'''}{\Psi'' \cdot \Psi''}\right)\Psi'.$$

M **5.** Let $m \geqslant 2$ be an integer. Define ***Bour's minimal curve***[8] by

$$\Phi(z) = \left(\frac{z^{m-1}}{m-1} - \frac{z^{m+1}}{m+1},\ i\left(\frac{z^{m-1}}{m-1} + \frac{z^{m+1}}{m+1}\right),\ \frac{2z^m}{m}\right)$$

Show that Φ coincides with Enneper's minimal surface when $m = 2$. Find an explicit expression of the polar parametrization

$$(r, \theta) \mapsto \mathbf{z}[t](r\cos\theta,\ r\sin\theta)$$

of each member of the associated family, and compute its Gaussian curvature. Plot the surface $\mathbf{z}[0]$ for $m = 3$.

6. Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^3$ be any patch. Show that

$$\frac{\partial}{\partial \overline{z}}\left(\frac{e-g}{2} - if\right) = \lambda^2 \frac{\partial H}{\partial z}.$$

Conclude that for a surface of constant mean curvature the function

$$u + iv \mapsto \frac{e(u,v) - g(u,v)}{2} - i\, f(u,v)$$

is complex analytic.

M **7.** Plot the Weierstrass patch for which $f(z) = 2$ and $g(z) = z^{-1} + z^3$. The bottom of the resulting surface should resemble a catenoid and the top should resemble Enneper's minimal surface.

---

[8] Edmond Bour (1832–1866). French mathematician, who made significant contributions to analysis, algebra, geometry, and applied mechanics despite his early death.

M **8.** The ***trinoid*** is the Weierstrass patch determined by $z \mapsto 1/(z^3-1)^2$ and $z \mapsto z^2$. Plot the trinoid using polar coordinates and compute its Gaussian curvature.

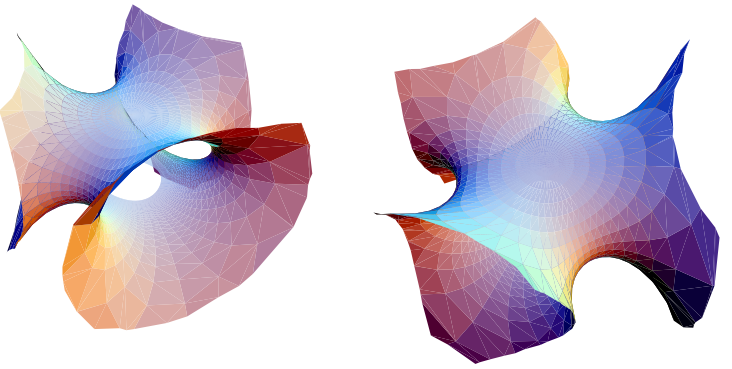


Figure 22.10: Two trinoide views

9. Show that for a Weierstrass minimal curve Ψ whose derivative is given by

$$\Psi' = \left(\tfrac{1}{2}f(1-g^2),\ \tfrac{1}{2}if(1+g^2),\ fg\right),$$

we have the formula

$$\Psi' \cdot \mathbf{U} = -fg', \tag{22.55}$$

where $\mathbf{U}$ denotes the unit normal to Ψ.

10. Prove the following facts about the Gauss map of minimal surfaces:

(a) The Gauss map of the catenoid is one-to-one; its image covers the sphere except for two points.

(b) The image of the Gauss map of the helicoid omits exactly two points; otherwise, the Gauss map sends infinitely many points onto each point in its image.

(c) The image of the Gauss map of Scherk's surface omits exactly four points; otherwise, the Gauss map maps infinitely many points onto each point in its image.

11. Let $\mathbf{x}\colon \mathcal{U} \to \mathbb{R}^n$ be a patch. The ***mean curvature vector field*** $\mathbf{H}$ of $\mathbf{x}$ is defined to be the component of

$$\frac{G\mathbf{x}_{uu} - 2F\mathbf{x}_{uv} + E\mathbf{x}_{vv}}{2(EG-F^2)}$$

that is perpendicular to $\mathbf{x}_u$ and $\mathbf{x}_v$. Show that this vanishes if and only if $\mathbf{x}$ is both isothermal and harmonic.

12. Show that the conjugate of Henneberg's surface is given by

$$\mathbf{y}(u,v) = \Big(2\cosh u \sin v - \tfrac{2}{3}\cosh 3u \sin 3v, \\ 2\cosh u \cos v + \tfrac{2}{3}\cosh 3u \cos 3v,\ 2\sinh 2u \sin 2v\Big),$$

as stated on page 512.

13. Construct a minimal surface that contains a hyperbola as a geodesic.

M **14.** Investigate the minimal surface shown in Figure 22.11, arising from the minimal curve

$$\Psi(z) = \Big(\frac{a\cos z}{1+\sin^2 z},\ \frac{a\cos z \sin z}{1+\sin^2 z},\ ia\,\mathrm{F}(z,-1)\Big),$$

where the elliptic function F is the inverse of Jacobi function am, so that $u = \mathrm{F}(\phi|m)$ if and only if $\phi = \mathrm{am}(u|m)$. The surface contains a lemnsicate as a geodesic.

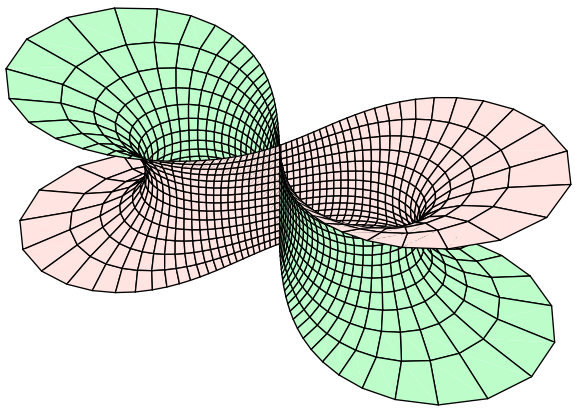


Figure 22.11: A minimal surface with a lemniscate as geodesic

M **15.** Draw a minimal surface that contains a cardioid as a geodesic.

Notebook 22

■ N22.0

This notebook implements the techniques described in the text of Chapter 22. The isometric deformations in N22.1 generalize the deformation of the helicoid to the catenoid presented in N16.2. For example, we plot a deformation of Catalan's minimal surface in N22.1.1, and we describe a sequence of surfaces that generalize Enneper's minimal surface in N22.2. For this purpose, we recall some definitions from Notebook 16:

```
catalan[a_][u_,v_]:=
  {a(u-Sin[u]Cosh[v]), a(1-Cos[u]Cosh[v]), -4a Sin[u/2]Sinh[v/2]}
enneper[u_,v_]:= {u-u^3/3+u v^2, -v+v^3/3-v u^2, u^2-v^2}
```

```
catenoid[c_][u_,v_]:= {c Cos[u]Cosh[v/c], c Cosh[v/c]Sin[u], v}
helicoid[a_,b_,c_][u_,v_]:= {a v Cos[u], b v Sin[u], c u}
```

In N22.6, we shall need some plane curves, and the arc length command from N1.3.

```
ellipse[a_,b_][t_]:= {a Cos[t], b Sin[t]}
astroid[a_][t_]:= a{Cos[t]^3, Sin[t]^3}
cycloid[a_,b_][t_]:= {a t - b Sin[t], a - b Cos[t]}
lemniscate[a_][t_]:= a{Cos[t], Sin[t]Cos[t]}/(1+Sin[t]^2)
```

```
arclength[α_][t_]:= Integrate[L[α'[tt]], tt]/.tt->t
```

The first of the following packages was introduced in N10.0:

```
<<Graphics`ImplicitPlot3D`
<<Graphics`Graphics3D`
<<Graphics`Colors`
<<Graphics`Shapes`
<<Geometry`Rotations`
```

An initial electronic subsection loads the patch henneberg from N16.4, familiar plotting options, and other programs including selectgraphics3d from N19.1.

■ N22.0.1

■ N22.1

We first introduce a command that rotates a complex number z by an angle -t and then takes the real part, since this subroutine occurs over and over again in Chapter 22.

```
rexp[t_][z_]:= ComplexExpand[Re[Exp[-I t] z]]
```

It can be used to define the family of charts, or isometric deformation, associated to conjugate minimal isothermal patches x,y:

```
deform[t_][x_,y_][u_,v_]:= ret[t][x[u,v] + I y[u,v]]
```

The Cauchy-Riemann equations correspond to the substitution

```
cr:= {D[y[u,v],u]->-D[x[u,v],v], D[y[u,v],v]->D[x[u,v],u]}
```

We can now verify the first two equations Lemma 22.5.

```
D[deform[t][x,y][u,v], u]/.cr
D[deform[t][x,y][u,v], v]/.cr
```

The second order partial derivatives of x,y are related by

```
D[cr,u]
D[cr,v]
```

These substitutions can in turn be used to compute the second derivatives of the associated family zz[t]:

```
D[deform[t][x,y][u,v], u, u]/.D[cr,u]
D[deform[t][x,y][u,v], v, v]/.D[cr,v]
D[deform[t][x,y][u,v], u, v]/.D[cr,u]
```

N22.1.1

N22.2

Complex derivatives can be defined by

```
diff[f_][u_,v_]:= (D[f,u] - I D[f,v])/2
diffc[f_][u_,v_]:= (D[f,u] + I D[f,v])/2
```

In this context, the *Mathematica* command Conjugate is cumbersome, and it is quicker to define complex conjugates by hand.

```
z:= u + I v
zc:= u - I v
diff[z][u,v]
diff[zc][u,v]
diffc[z][u,v]
diffc[zc][u,v]
```

Here, for example, is an expression for the Laplacian:

```
diffc[diff[f[u,v]][u,v]][u,v]//Simplify
```

N22.3

The definition of the family of charts associated to a minimal curve Ψ is

```
analytic2minimal[Ψ_][t_][u_,v_]:= rexp[t][Ψ[u+I v]]
```

We can experiment with the curve

```
ennepermincurve[n_][z_]:=
  {z - z^(2n+1)/(2n+1), I(z+z^(2n+1)/(2n+1)), 2z^(n+1)/(n+1)}
```

The isotropic nature of this curve is verified by computing the "norm" of its derivative:

```
Clear[n]
mc:= ennepermincurve[n]
mc'[z]
L[mc'[z]]
```

Given that this is zero, we may compute the resulting surface, for any n, and verify that it is minimal. Enneper's original surface is the case n=1.

```
n:= 1
f:=analytic2minimal[mc][π/2]
f[u,v]
mcurv[f][u,v]
```

Here is an animation based on the case n=3. Replace Show by S to run it.

```
n:= 3
Do[Show @ ParametricPlot3D[f[u,v], {u,-r,r}, {v,-r,r},
   PlotRange->Table[{-r,r},{i,3}], Axes->False, Boxed->False],
{r,.7,1.2,.025}]
```

The definition of stereographic projection using complex numbers is

```
stereo[x_,y_,z_]:= (x + I y)/(1 - z)
```

The modulus of a complex number is computed by combining the commands Abs and ComplexExpand. If z>0 then stereo[x,y,z] has modulus greater than one, a fact that is clear from Figure 22.2, plotted in N22.3.1.

```
Abs[stereo[x,y,z]]//ComplexExpand
%^2
%/. x^2+y^2->1-z^2//Simplify
```

■ N22.3.1

■ N22.4

The following definition captures Lemma 22.24.

```
harmonic2analytic[x_][z_]:= Simplify[2x[z/2, z/(2 I)] - x[0,0]]
```

Applying it to Enneper's minimal surface, we get a minimal curve already defined:

```
mc:= harmonic2analytic[enneper]
mc[z]
ennepermincurve[1][z]
% - %% //Simplify
```

We recover Enneper's surface as output of the last line in the cell

```
g[t_]:= analytic2minimal[mc][t]
g[t][u,v]
%//Together//Apart
g[π/2][u,v]
```

Other examples of minimal surfaces constructed via minimal curves are provided by

```
mc:= harmonic2analytic[helicoid[a,a,a]]
h[t_]:= analytic2minimal[mc][t]
h[t][u,v]//FullSimplify
h[π/2][u,v]
```

```
mc:= harmonic2analytic[catalan[1]]
f[t_]:= analytic2minimal[mc][t]
f[t][u,v]//FullSimplify
f[π/2][u,v]
```

Here we show that an associated family is produced by rotation.

```
z1= Rotate3D[enneper[r Cos[θ-t/2], r Sin[θ-t/2]], -t/2, 0, 0]
%//Simplify
```

```
z2= analytic2minimal[ennepermincurve[1]][t][r Cos[θ], r Sin[θ]]
%//Simplify
```

```
Simplify[z1 - z2]
```

The following isometric deformation was obtained using the methods above.

```
catalandef[a_][t_][u_,v_]:=
  {a Cos[t](u-Cosh[v]Sin[u])+a Sin[t](v-Cos[u]Sinh[v]),
   a Cos[t](1-Cos[u]Cosh[v])+a Sin[t]Sin[u]Sinh[v],
  4a(1-Cos[u/2]Cosh[v/2])Sin[t] - 4a Cos[t]Sin[u/2]Sinh[v/2]}
```

The next definition, given in Section 16.4, is almost equivalent to the output from the subsequent cell (cf. Exercise 12).

```
hennebergconj[u_,v_]:=
  {2Cosh[u]Sin[v]-(2/3)Cosh[3u]Sin[3v],
   2Cosh[u]Cos[v]+(2/3)Cosh[3u]Cos[3v], 2Sinh[2u]Sin[2v]}
```

```
mc:= harmonic2analytic[henneberg]
-analytic2minimal[mc][π/2][u,-v]//FullSimplify
```

```
ParametricPlot3D[hennebergconj[u,v]//Evaluate,
  {u,-1.5,1.5}, {v,-π/2,π/2}, PlotPoints->{80,80}, PlotRange->All]
```

The following is a reparametrization of the helicoid that arises from the construction of Section 22.4:

```
helicoid1[a_,b_,c_][u_,v_]:= {a Sinh[v]Cos[u], b Sinh[v]Sin[u], c u}
```

```
mcurv[helicoid1[a,a,c]][u,v]
```

```
mcurv[helicoid1[a,b,0]][u,v]
```

■ 22.4.1

■ N22.5

Next, we generate a Weierstrass patch and its associated family from meromorphic functions f, g. For simplicity we take z_o=0. First, we define the minimal curve:

```
weier[f_,g_][z_]:=
Module[{w,A},
  F[w_]= {(1 - g[w]^2)f[w], I(1 + g[w]^2)f[w], 2f[w]g[w]}/2;
  A= Integrate[F[w], w];
  Simplify[A/.w->z]
]
```

This definition is a little more complicated than usual in order to avoid technical problems when we use weier with other commands. To get a parametrization of the minimal isothermal patch and its associated family, we must combine it with analytic2minimal. Let us use this method to generate several standard minimal surfaces.

When making use of weier, it is convenient to make use of the shorthand described to represent functions. For example, the catenoid to helicoid deformation is obtained by

```
mc:= weier[(-E^(-#))&, (-E^#)&]
mc[z]
f[t_]:= analytic2minimal[mc][t]
```

We may now verify that the starting surface is a catenoid, and that the conjugate Weierstrass patch essentially a helicoid.

```
f[0][u,v] - catenoid[1][v,u]//Simplify
h[u_,v_]:= helicoid[1,1,1][v + π/2,-Sinh[u]] - {0,0,π/2}
f[π/2][u,v] - h[u,v] //Simplify
```

Henneberg's surface and its conjugate can be defined in the same sort of way:

```
mc:= Composition[weier[(1 - 1/#^4)&, #&], Exp]
f[t_]:= analytic2minimal[mc][t]
2f[0][u,-v]//FullSimplify
% - henneberg[u,v]
2f[π/2][u,v]//FullSimplify
```

Richmond's minimal surface is generated by

```
weier[1/#^2 &, #^(n+1) &][z]
```

The Gaussian curvature of a Weierstrass patch is given by

```
weiercurv[f_,g_][z_]:=
Module[{w},
  A= -16Abs[D[g[w],w]]^2;
  B= Abs[f[w]]^2(1 + Abs[g[w]]^2)^4;
  A/B /.w->z
]
```

N22.5.1

N22.6

The formulas for Björling's curve and minimal curve are straightforward:

```
bjoerling[α_,γ_][z_]:=
Module[{w, A},
  A= I Integrate[Cross[γ[w],α'[w]], w];
  Simplify[α[w] + A/. w->z]
]
```

```
bjmincurve[α_][z_]:=
Module[{A},
  A= arclength[α][z];
  Append[α[z], I Simplify[PowerExpand[A]]]
]
```

The command bjmincurve can be combined with analytic2minimal in order to produce the parametrization of the minimal surface containing a given curve as a geodesic. For example,

```
mc:= bjmincurve[cycloid[1,1]]
mc[z]
```

```
analytic2minimal[mc][0][u,v];
%//Simplify//PowerExpand
```

We see that the minimal surface which contains the cycloid as a pregeodesic is precisely Catalan's minimal surface.

We use an expression for the arc length of an ellipse to define a minimal curve and associated family of minimal surfaces.

```
arclength[ellipse[a,b]][t]//PowerExpand
%//InputForm
```

```
ellipsemincurve[a_,b_][z_]:= {a Cos[z], b Sin[z],
  I b EllipticE[z, 1 - a^2/b^2]}
```

```
ellipsemin[a_,b_][t_][u_,v_]:=
Module[{mc,f},
  mc= ellipsemincurve[a,b];
  f= analytic2minimal[mc][t];
  f[u,v]//FullSimplify
]
```

```
ellipsemin[2,1][0][u,v]//TeXForm
```

An example of an elliptical catenoid is

```
S @ Module[{x,y,h,xx},
  x= ellipsemin[2,1][0];
  y= catenoid[1];
  h[t_]:= 1 - (1 - Sech[t]^5)^.3;
  xx[u_,v_]:= Append[x[u,v], Hue[h[gcurv[y][u,v]]]];
  ParametricPlot3D[xx[u,v], {u,0,2π}, {v,-1,1},
    PlotPoints->{40,30}, Lighting->False]
]
```

Figure 22.5

The conjugate minimal surface of the elliptical catenoid is an analog of the helicoid:

```
ParametricPlot3D[ellipsemin[2,1][π/2][u,v], {u,-π,π}, {v,-π/2,π/2},
  PlotPoints->{48,24}];
```

Next, we find a minimal surface containing an astroid as geodesic, while N22.6.1. contains the case of the clothoid and instructions for Figure 22.6.

```
astroidcat[u_,v_]:=
Module[{mc,f},
  mc= bjmincurve[astroid[8/3]];
  f= analytic2minimal[mc][0];
  f[u,v]//FullSimplify
]
```

```
ParametricPlot3D[astroidcat[u,v], {u,-π,π}, {v,-.4,.4},
  PlotPoints->{60,12}];
```

■ N22.6.1

■ N22.7

First, we obtain numerical approximations to the constants c, e, A. The command

```
winv= WeierstrassInvariants[{1/2,I/2}]//N
c= winv[[1]]
e= Sqrt[c/4]
a= Sqrt[8π] e
```

This enables us to give abbreviations for the Weierstrass functions, which we use to record the definitions of the components of Costa's minimal surface.

```
WZ[z_]:= WeierstrassZeta[z, {c,0}]
WP[z_]:= WeierstrassP[z, {c,0}]
```

```
costa1[u_,v_]:=
  (1/2)Re[-WZ[u+I v] + π u + π^2/(4e) +
    (π/(2e))(WZ[u + I v - 1/2] - WZ[u + I v - I/2])]
```

```
costa2[u_,v_]:= (1/2)Re[-I WZ[u + I v] + π v +π^2/(4e1) -
  (π/(2e1))(I WZ[u + I v - 1/2] - I WZ[u + I v - I/2])]
```

```
costa3[u_,v_]:= (Sqrt[2π]/4)Log[Abs[(WP[u+I v] - e)/
  (WP[u+I v] + e)]]
```

In practice, we shall plot the surface using the following self-contained Module:

```
costa[u_,v_]:=
Module[{c,x},
  c= 189.07272;
  WZ[z_]= WeierstrassZeta[z, {c,0}];
  WP[z_]= WeierstrassP[z, {c,0}];
  x[1]= (1/2)Re[-WZ[u + I v] + 3.14159 u + 0.35889 +
    0.228473(WZ[u + I v - 1/2] - WZ[u + I v - I/2])];
  x[2]= (1/2)Re[-I WZ[u+I v] + 3.14159 v + 0.35889 -
    0.228473(I WZ[u + I v- 1/2] - I WZ[u + I v - I/2])];
  x[3]= 0.626657 Log[Abs[
    (WP[u + I v] - 6.87519)/(WP[u + I v] + 6.87519)]];
  Array[x,3]
]
```

```
costa[1,1]
costa1[1,1]//N
```

Our first attempt to plot Costa's minimal surface is

```
costaplot:= ParametricPlot3D[costa[u,v], {u,.001,1.001},
  {v,.001,1.001}, PlotPoints->80];
```

This plot is incorrect, so we do not display it. The difficulty is that *Mathematica* has included some polygons that have a vertex very far from the origin. To exclude them we use the command selectgraphics3d. We also need the command dip to eliminate the edges in the polygonal approximation of the surface:

```
dip[ins_][g_]:= $DisplayFunction[Insert[g,ins,{1,1}]]
```

```
S @ selectgraphics3d[costaplot[[1]], 8,
   ViewPoint->{2.9,-1.4,1.2}, Boxed->False,
   PlotRange->{{-4,4},{-4,4},{-2,2}},
   DisplayFunction->dip[EdgeForm[]]]
```

Figure 22.7

N22.7.1

N22.8

Here are some more minimal curves and associated surfaces:

```
bourmincurve[m_][z_]:=
  {z^(m-1)/(m-1)-z^(m+1)/(m+1),
  I(z^(m-1)/(m-1)+z^(m+1)/(m+1)), 2z^m/m}
```

```
bourpolar[m_][t_][r_,θ_]:=
  {r^(m-1)Cos[t-(m-1)θ]/(m-1)-r^(m+1)Cos[t-(m+1)θ]/(m+1),
   r^(m-1)Sin[t-(m-1)θ]/(m-1)+r^(m+1)Sin[t-(m+1)θ]/(m+1),
  2r^m Cos[t-m θ]/m}
```

```
bp:= bourpolar[3]
zz[v_]:= ParametricPlot3D[bp[0][r,θ], {r,0,1}, {θ,0,2π},
  PlotPoints->{24,60}, ViewPoint->v, PlotRange->All]
ar:= {zz[{0,1,2}], zz[{2,-1,1}]}
GraphicsArray[ar, GraphicsSpacing->-.1]
```

The final electronic subsection includes other Weierstrass patches, such as the trinoid that is the subject of Figure 22.9.

N22.8.1

Chapter 23

Rotation and Animation Using Quaternions

The previous chapter used complex analysis to further the study of minimal surfaces. Many applications of complex numbers to geometry can be generalized to the *quaternions*, an extended system in which the "imaginary part" of any number is a vector in $\mathbb{R}^3$. Although beyond the scope of this book, there is an extensive theory of surfaces defined by "quaternionic-holomorphic" curves [BFLPU]. Other uses include the visualization of fractals by means of iterated maps of quaternions extending the usual complex theory [DaKaSa, Wm2].

In this chapter, we describe the fundamental role that quaternions play in describing rotations of 3-dimensional space. It is a topic familiar to pure mathematicians (through the topology of Lie groups), but one that has recently assumed great popularity because of its use in computer graphics and video games. As a consequence, there is little in the chapter's text that cannot be found on the internet and elsewhere, though Notebook 23 contains a number of original animations based on the theory, hence the chapter's title.[1]

Applications are not restricted to merely viewing rotations; indeed many graphics interfaces already permit one to rotate a graphics object at the touch of a mouse. The purpose of this chapter is instead to help understand and develop the theory. For example, new surfaces can be constructed by rotating lines and curves in a nonstandard way, as in Figures 23.1 and 23.9.

During the course of the chapter, we give several descriptions of the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. All are based on representing a rotation by some sort of vector, and the fact that a rotation is uniquely specified by three parameters. In the first such description, in Section 23.1, an element $\mathbf{x}$ of $\mathbb{R}^3$ is converted (via the

[1] Added in proof: For a more extensive treatment of some of the topics in this chapter, the editors recommend the book by A.J. Hanson "Visualizing Quaternions," Elsevier, 2006.

vector cross product) into a skew-symmetric matrix A, and then exponentiated. This gives rise to a neat expression for a rotation of a given angle about a given axis, namely Theorem 23.4, whose proof is completed using quaternions.

After describing the basic operations on quaternions in Section 23.2, we define the celebrated mapping from the group $\mathscr{U}$ of unit quaternions to the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. This has a built-in ambiguity whereby plus and minus a quaternion describe the same rotation, and some of the more interesting graphics generated by Notebook 23 show tell-tale signs of this phenomenon, also present in complex analysis. The construction converts circles into figure eight curves and generates self-intersecting surfaces like that in Figure 23.1, which occur in the theory of Riemann surfaces [AhSa]. It all relates to the topological picture of $\mathbf{SO}(3)$ as a closed "ball" in $\mathbb{R}^3$ for which antipodal points of the boundary are identified, as explained in Section 23.4.

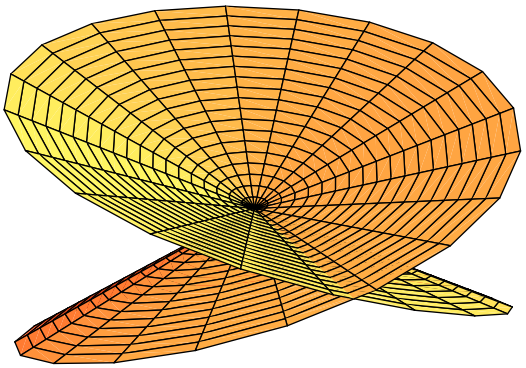


Figure 23.1: A ruled surface generated by a rotation curve

Rotation curves, the subject of Section 23.5, are simply curves whose trace lies in $\mathbf{SO}(3)$, rather than in $\mathbb{R}^3$. In practical terms, one can think of such a curve as a "static fairground ride" in which the rider is strapped inside a sphere which rotates about its center, with varying axis, and the parameter t is time. One can also associate a rotation curve to a more dynamic rollercoaster-type ride using the Frenet frame construction of Chapter 7. Section 23.7 briefly discusses this and other topics. In the intervening Section 23.6, we explain how Euler angles are used to decompose any rotation in space into the product of rotations about coordinate axes.

There is a sense in which this chapter is a bridge between past and future topics. The linear algebra reaches back to the start of the book, while Section 23.5 relates to homotopy and the theory of space curves in Chapters 6 and 7. On the other hand, the study of rotations and quaternions provides simple examples of differentiable manifolds and related techniques that we shall meet in the next chapter.

23.1 Orthogonal Matrices

As explained at the outset in Section 1.1, one can represent a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ by a matrix of size $m \times n$. In this chapter, we shall be mainly concerned with the case $m = n = 3$, but to begin with we impose only the restriction $m = n$. This being the case, one can compose two linear mappings A, B to make a third, denoted $A \circ B$, defined by

$$(A \circ B)(\mathbf{v}) = A(B(\mathbf{v})), \qquad \mathbf{v} \in \mathbb{R}^n. \tag{23.1}$$

If one regards A, B as matrices rather than mappings, then it is obvious from (23.1) that their composition $A \circ B$ corresponds to the matrix product AB. One only has to remember that it is actually B (or, more generally, the *second* factor) that is applied *first* to a column vector in $\mathbb{R}^n$.

By its very definition, composition of mappings $A, B, C\colon \mathbb{R}^n \to \mathbb{R}^n$ satisfies the associative law

$$A \circ (B \circ C) = (A \circ B) \circ C, \tag{23.2}$$

which forms part of the definition of a group (see page 131). The same is of course true of matrix multiplication, and in Section 5.1 we defined various groups by imposing restrictions on the linear transformations or matrices under consideration. In this chapter, the emphasis will be on matrices, and we begin with

Definition 23.1. *An $n \times n$ matrix A is said to be* ***orthogonal*** *if*

$$A^T A = I_n, \tag{23.3}$$

where I_n is the identity matrix of order n.

Recall that A^T denotes the transpose of A, obtained by swapping its rows and columns. Equation (23.3) therefore asserts that the *columns* of A form an orthonormal basis of $\mathbb{R}^n$. For this reason, it would be more logical to speak of an "orthonormal" matrix, though one never does.

Let $\mathbf{O}(n)$ denote the set of all orthogonal $n \times n$ matrices. Provided we represent a linear transformation with respect to a fixed orthonormal basis, $\mathbf{O}(n)$ coincides with the group $\mathsf{orthogonal}(\mathbb{R}^n)$, via equation (5.2). The fact that $\mathbf{O}(n)$ is a group can be seen directly by observing that if $A, B \in \mathbf{O}(n)$, then

$$(AB)^T(AB) = B^T A^T A B = B^T I_n B = I_n.$$

Moreover, the inverse of an orthogonal martrix is given by

$$A^{-1} = A^T,$$

from which it follows that $AA^T = I_n$, or that the *rows* of A also form an orthonormal basis.

Lemma 5.2 tells us that the determinant of any orthogonal matrix is either 1 or -1. Since $\det(AB) = (\det A)(\det B)$, it follows that the set

$$\mathbf{SO}(n) = \{A \in \mathbf{O}(n) \mid \det A = 1\}$$

is a *subgroup* of $\mathbf{O}(n)$; that is, a group in its own right with the same operation of matrix multiplication. The letter **S** stands for "special," and $\mathbf{SO}(n)$ corresponds to the group of *orientation-preserving* orthogonal transformations, or *rotations*, as they were called on page 128.

We now specialize to the case $n = 3$ for which there is already a clear notion of "rotation," meaning *a transformation leaving fixed the points on some straight line axis, and rotating any other point in a plane perpendicular to the axis.* An example of such a rotation is the transformation represented by the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{23.4}$$

where $0 \leqslant \theta < 2\pi$. This mapping leaves fixed the z-axis, but acts as a rotation by an angle θ counterclockwise in the xy-plane.

The following well-known result implies that any element of $\mathbf{SO}(3)$ is equivalent to some R_θ after a change of orthonormal basis.

Lemma 23.2. *Given a 3×3 matrix $A \in \mathbf{SO}(3)$, there exists $P \in \mathbf{SO}(3)$ and $\theta \in [0, 2\pi)$ such that*

$$P^{-1}AP = R_\theta.$$

Proof. The first step is to show that 1 is an eigenvalue of A. This is true because

$$\begin{aligned} \det(A - I_3) \;=\; \det(A - AA^T) \;&=\; \det\left(A(I_3 - A)^T\right) \\ &=\; (\det A)\det\left((I_3 - A)^T\right) \\ &=\; -\det(A - I_3) \end{aligned}$$

is zero. Since $A - I_3$ is singular, there exists a nonzero column vector $\mathbf{w}$ such that $A\mathbf{w} = \mathbf{w}$, and it is this eigenvector that determines the axis of rotation. Extend the unit vector $\mathbf{w}_3 = \mathbf{w}/\|\mathbf{w}\|$ to an orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ such that $\mathbf{w}_1 \times \mathbf{w}_2 = \mathbf{w}_3$. Being a positively oriented *orthogonal transformation*, A must restrict to a rotation in the plane generated by $\mathbf{w}_1, \mathbf{w}_2$, and there exists θ such that

$$\begin{cases} A\mathbf{w}_1 \;=\; \mathbf{w}_1\cos\theta + \mathbf{w}_2\sin\theta \\ A\mathbf{w}_2 \;=\; -\mathbf{w}_1\sin\theta + \mathbf{w}_2\cos\theta. \end{cases} \tag{23.5}$$

Now let P denote the invertible matrix whose columns are the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. By definition of matrix multiplication, the columns of AP are $A\mathbf{w}_1$, $A\mathbf{w}_2$, $A\mathbf{w}_3$, and (23.5) now implies that $AP = PR_\theta$. The result follows. ∎

Lemma 23.2 can be expressed by saying that A is ***conjugate*** to R_θ in the group $\mathbf{SO}(n)$. The two rotations are equivalent, if we are willing to change our frame of reference by pointing the z-axis in the direction of the eigenvector $\mathbf{w}$. Combining (23.6) with equation (13.18) on page 398 gives

$$\operatorname{tr} A = \operatorname{tr}(PR_\theta P^{-1}) = \operatorname{tr}(R_\theta) = 2\cos\theta + 1. \tag{23.6}$$

This means that the unsigned angle $\pm\theta$ can conveniently be recovered from the matrix A. To see how to recover the axis of rotation, see Exercise 4.

Later in this chapter, we shall be considering *curves* in the group $\mathbf{SO}(3)$. Such a curve is a mapping $\gamma\colon (a,b) \to \mathbb{R}^{3,3}$ where $(a,b) \subseteq \mathbb{R}$ and $\mathbb{R}^{3,3}$ is the space of 3×3 matrices, and the problem is to find conditions that ensure that the image of γ lies in $\mathbf{SO}(3)$. One method is to use a power series

$$\gamma(t) = I + A_1 t + A_2 t^2 + \cdots \tag{23.7}$$

where $I = I_3$, and seek appropriate conditions on the coefficients. For example, to determine A_1, we need only consider the terms up to order one, and set

$$\begin{aligned} I = \gamma(t)\gamma(t)^T &= (I + A_1 t + \cdots)(I + A_1 t + \cdots)^T \\ &= I + (A_1 + A_1^T)\, t + \cdots \end{aligned}$$

It follows that A_1 must be a *skew-symmetric* matrix, meaning ${A_1}^T = -A_1$.

Having made the crucial step, it is in fact possible to solve (23.7) using the concept of exponentiation. The ***exponential*** of a square matrix A, written e^A or $\exp A$, is defined by the usual infinite sum

$$\exp A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k, \tag{23.8}$$

with the convention that $A^0 = I$. This operation is well known to share some of the usual properties of exponentiation of real numbers. In particular, (23.8) converges, and if A, B are two square matrixes for which $AB = BA$ then

$$\exp(A + B) = (\exp A)(\exp B).$$

If $\mathbf{0}$ denotes the zero matrix then obviously $\exp(\mathbf{0}) = I$. Thus,

$$\exp(-A) = (\exp A)^{-1},$$

and the matrix (23.8) is always invertible.

Any property valid for the powers of a matrix will extend to the operation of exponentiation. For example, $(A^n)^T = (A^T)^n$ for all $n \geqslant 0$, and so $\exp(A^T) = (\exp A)^T$. A more interesting formula is

$$\det(\exp A) = \exp(\operatorname{tr} A). \tag{23.9}$$

This is immediate if A is *diagonal* with entries $\lambda_1, \ldots, \lambda_n$, for then e^A is diagonal with entries $e^{\lambda_1}, \ldots e^{\lambda_n}$, and (23.9) becomes the identity

$$\prod_{i=1}^{n} e^{\lambda_i} = e^{\lambda_1 + \cdots + \lambda_n}.$$

Having seen this, the validity of (23.9) can easily be extended to the case in which A is *diagonalizable* (over the complex numbers). For then there exists a complex invertible matrix P and a complex diagonal matrix D for which $A = PDP^{-1}$, and

$$\begin{aligned} \exp A \;=\; \exp(PDP^{-1}) \;&=\; \sum_{k=0}^{\infty} \frac{1}{k!}(PDP^{-1})^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} PD^k P^{-1} \;=\; P(\exp D)P^{-1} \end{aligned}$$

The result (23.9) now follows from the properties (13.18) already used for (23.6).

It is a general principle that, to prove a matrix equation like (23.9) (in which both sides are continuous matrix functions), it suffices to assume that A is diagonalizable. This is because an arbitrary square matrix can be approximated by a sequence of diagonalizable matrices (for example, ones that have distinct complex eigenvalues). In this way (23.9) follows in general; see [Warner] for a fuller discussion of the exponentiation of matrices. The same principle can be used to prove the *Cayley–Hamilton Theorem* (mentioned on page 402), namely that any $n \times n$ matrix A satisfies its characteristic polynomial

$$c_A(x) = \det(xI_n - A).$$

This theorem implies that the nth power A^n is always some linear combination of $I_n, A, A^2, \ldots, A^{n-1}$. The same is therefore true of $\exp A$, and a vivid illustration of this fact is provided by Theorem 23.4 below.

Lemma 23.3. *If A is a $n \times n$ skew-symmetric matrix, then* $\exp A$ *is an orthogonal matrix with determinant 1.*

Proof. We have

$$\exp(A^T A) = \exp(A^T)\exp A = \exp(-A)\exp A = (\exp A)^{-1}\exp A = I_n.$$

Given that $\operatorname{tr} A = 0$, the statement about $\det A = 1$ follows from (13.18). ▌

The importance of this lemma can be assessed by the fact that it is related to at least two topics studied in later chapters.[2]

Any skew-symmetric 3×3 matrix is uniquely determined by a column vector

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

by means of the formula

$$A[\mathbf{x}] = \begin{pmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{pmatrix}. \tag{23.10}$$

We need not have used coordinates, since (23.10) is equivalent to stating that

$$A[\mathbf{x}]\mathbf{w} = \mathbf{x} \times \mathbf{w} \tag{23.11}$$

for all $\mathbf{w} \in \mathbb{R}^3$. To understand this, consider

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{23.12}$$

and observe that the ith row of $A[\mathbf{x}]$ is (the transpose of) the vector cross product $\mathbf{e}_i \times \mathbf{x}$. Thus, the ith entry of the column vector $A[\mathbf{x}]\mathbf{w}$ equals

$$(\mathbf{e}_i \times \mathbf{x}) \cdot \mathbf{w} = \mathbf{e}_i \cdot (\mathbf{x} \times \mathbf{w}),$$

explaining the appearance of the cross product of $\mathbf{x}$ with $\mathbf{w}$.

The next result is a basic example for the mathematical theory of Lie groups [DuKo], and is also well known in the context of computer vision [Faug].

Theorem 23.4. (i) *Let* $\mathbf{x}$ *be a nonzero vector in* $\mathbb{R}^3$ *and set* $\ell = \|\mathbf{x}\|$. *Then*

$$\exp A[\mathbf{x}] = I + \frac{\sin \ell}{\ell} A[\mathbf{x}] + \frac{1 - \cos \ell}{\ell^2} A[\mathbf{x}]^2. \tag{23.13}$$

(ii) *any nonidentity* $P \in \mathbf{SO}(3)$ *can be written in this form; indeed* (23.13) *represents a rotation with axis parallel to* $\mathbf{x}$ *and angle* ℓ.

We choose to postpone the proof of (ii), which demonstrates the simplicity with which rotations can be represented via linear algebra. It is accomplished by Corollary 23.14.

---

[2]The set of skew-symmetric $n \times n$ matrices forms what is called a *Lie algebra* when one defines $[A, B] = AB - BA$ (see Exercise 5 and page 832). There exists a means of exponentiating any finite-dimensional Lie algebra to obtain a *Lie group*, thus generalizing Lemma 23.3 [Warner]. There is a more general notion of exponentiation in Riemannian geometry, whereby the tangent space of a surface acts to model the surface itself (see page 885).

Proof of (i). Let $\ell = \|\mathbf{x}\|$, and set $\mathbf{x}_1 = \mathbf{x}/\ell$. Choose a unit vector $\mathbf{x}_2$ orthogonal to $\mathbf{x}$ and let $\mathbf{x}_3 = \mathbf{x}_1 \times \mathbf{x}_2$. With respect to the basis $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, the linear transformation (23.11) associated to $A[\mathbf{x}]$ has matrix

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\ell \\ 0 & \ell & 0 \end{pmatrix}. \tag{23.14}$$

There is a similarity with (23.4), explained by the fact that (23.14) equals $A[\ell\mathbf{e}_1]$, and $A[\mathbf{x}]$ acts as ℓ times the operator J in the plane generated by $\mathbf{x}_2$ and $\mathbf{x}_3$. Both $A[\ell\mathbf{e}_1]$ and $A[\mathbf{x}]$ have eigenvalues $0, i\ell, -i\ell$. Indeed, $A[\ell\mathbf{e}_1] = PDP^{-1}$ where

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 1 \\ 0 & 1 & i \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & i\ell & 0 \\ 0 & 0 & -i\ell \end{pmatrix}.$$

The matrix D satisfies the equation

$$e^x = 1 + \frac{\sin\ell}{\ell}x + \frac{1-\cos\ell}{\ell^2}x^2$$

(with D in place of x and I_3 in place of the first 1) because each of its diagonal entries does. The same equation is satisfied by $A[\mathbf{x}]$ because

$$\begin{aligned} \exp(PDP^{-1}) &= P(\exp D)P^{-1} \\ &= I_3 + \frac{\sin\ell}{\ell}PDP^{-1} + \frac{1-\cos\ell}{\ell^2}(PDP^{-1})^2. \end{aligned}$$ ∎

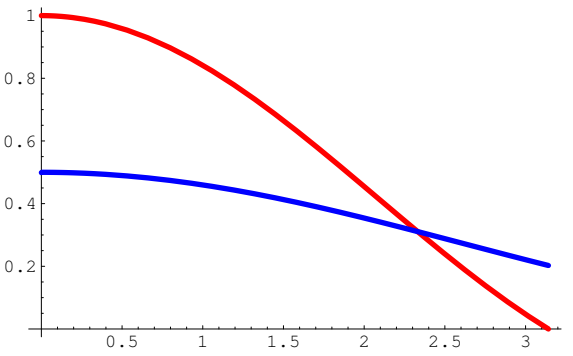


Figure 23.2: The functions $\dfrac{\sin\ell}{\ell}$ and $\dfrac{1-\cos\ell}{\ell^2}$

Figure 23.2 plots the functions occurring in (23.13) over the range $0 \leqslant \ell \leqslant \pi$ that will interest us in Section 23.4. Both tend to a limit as $\ell \to 0$, so we may allow $\mathbf{x} = \mathbf{0}$ in (23.13), as this is consistent with the formula $\exp\mathbf{0} = I$.

We shall return to the theory underlying Theorem 23.4 after the introduction of quaternions.

23.2 Quaternion Algebra

At one level, a quaternion is merely a vector with four components — that is an element of 4-space — and we explore this aspect first. By analogy to complex numbers, one denotes an individual quaternion by

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \qquad a, b, c, d \in \mathbb{R}, \tag{23.15}$$

and the set of quaternions by $\mathbb{H}$. Since $\mathbf{q}$ is determined by the vector (a, b, c, d), the set $\mathbb{H}$ can be identified with $\mathbb{R}^4$ as a vector space. But the real beauty of $\mathbb{H}$ derives from the special way in which its elements can be multiplied.

The multiplication of quaternions is based on the "fundamental formula"

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1,$$

that Hamilton[3] carved into Brougham bridge in 1843. It satisfies the associative rule (analogous to (23.2) without the $\circ$), which asserts that there is no ambiguity in computing a product of three or more quaternions written in a fixed order (this is Corollary 23.6 overleaf). For example,

$$-\mathbf{i} = \mathbf{j}^2\mathbf{i} = \mathbf{jji} = \mathbf{j}(\mathbf{ji}) = \mathbf{j}(-\mathbf{k}) = -\mathbf{jk}.$$

Indeed, it is frequently convenient to regard $\mathbf{i}, \mathbf{j}, \mathbf{k}$ as equal to the respective vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in (23.12). Their multiplication is then derived from the vector cross product in $\mathbb{R}^3$.

Now that we have defined quaternion multiplication, one can appreciate that the four coordinates a, b, c, d on the right-hand side of (23.15) are not equivalent. Indeed, $\mathbb{R}^4$ is split up into $\mathbb{R}$ and $\mathbb{R}^3$, and we can distinguish the ***real part***

$$\mathfrak{Re}\,\mathbf{q} = a,$$

and the ***imaginary part***

$$\mathfrak{Im}\,\mathbf{q} = b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

of any quaternion. Unlike for complex numbers (in which the imaginary part of $z = x + iy$ is the real number y), the imaginary part of $\mathbf{q}$ is effectively a *vector* in $\mathbb{R}^3$. One says that $\mathbf{q}$ is an ***imaginary quaternion*** if $\mathfrak{Re}\,\mathbf{q} = 0$.

We shall often write

$$\mathbf{q} = a + \mathbf{v}, \tag{23.16}$$

with the understanding that $\mathbf{v} = \mathfrak{Im}\,\mathbf{q} \in \mathbb{R}^3$.

---

[3] See page 191.

Lemma 23.5. *If* $\mathbf{q}_1 = a_1 + \mathbf{v}_1$ *and* $\mathbf{q}_2 = a_2 + \mathbf{v}_2$, *then*

$$\begin{aligned} \mathfrak{Re}(\mathbf{q}_1\mathbf{q}_2) &= a_1a_2 - \mathbf{v}_1 \cdot \mathbf{v}_2, \\ \mathfrak{Im}(\mathbf{q}_1\mathbf{q}_2) &= a_1\mathbf{v}_2 + a_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2. \end{aligned} \tag{23.17}$$

In particular, if $\mathbf{q}_1 = \mathbf{v}_1$ *and* $\mathbf{q}_2 = \mathbf{v}_2$ *are imaginary quaternions, then*

$$\mathbf{q}_1\mathbf{q}_2 = -\mathbf{v}_1 \cdot \mathbf{v}_2 + \mathbf{v}_1 \times \mathbf{v}_2. \tag{23.18}$$

Proof. This is just a matter of expanding brackets:

$$\begin{aligned} \mathbf{q}_1\mathbf{q}_2 &= (a_1 + \mathbf{v}_1)(a_2 + \mathbf{v}_2) \\ &= (a_1a_2 - \mathbf{v}_1 \cdot \mathbf{v}_2) + (a_1\mathbf{v}_2 + a_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2). \end{aligned}$$

The last statement follows from setting $a_1 = 0 = a_2$. ∎

The next result is verified in Notebook 23.

Corollary 23.6. *Multiplication of quaternions is associative:*

$$(\mathbf{q}_1\mathbf{q}_2)\mathbf{q}_3 = \mathbf{q}_1(\mathbf{q}_2\mathbf{q}_3),$$

for all $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{H}$.

Of course, quaternion multiplication is not commutative, by its very definition (but see Exercise 2).

The following will help us emphasize similarities between $\mathbb{H}$ and $\mathbb{C}$.

Definition 23.7. *Let* $\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = a + \mathbf{v} \in \mathbb{H}$.

(i) *The* ***modulus*** *or* ***norm*** *of* $\mathbf{q}$, *written* $|\mathbf{q}|$, *is given by*

$$|\mathbf{q}|^2 = a^2 + b^2 + c^2 + d^2 = a^2 + \|\mathbf{v}\|^2.$$

Moreover, $\mathbf{q}$ *is called a* ***unit quaternion*** *if* $|\mathbf{q}| = 1$.

(ii) *The* ***conjugate*** *of* $\mathbf{q}$, *written* $\overline{\mathbf{q}}$, *is given by*

$$\overline{\mathbf{q}} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k} = a - \mathbf{v}.$$

Thus, the norm of $\mathbf{q}$ is simply its norm as a vector in $\mathbb{R}^4$, as defined on page 3. Because of the analogy with complex numbers (and anticipating Lemma 23.8), we indicate the norm with a single vertical line, rather than two. For consistency, we shall also indicate the norm of a vector $\mathbf{v} \in \mathbb{R}^3$ by $|\mathbf{v}|$, rather than $\|\mathbf{v}\|$, because in this chapter we shall usually be thinking of such a vector as the imaginary quaternion $\mathbf{q} = 0 + \mathbf{v}$.

The conjugate of a quaternion $\mathbf{q}$ is formed by reversing the sign of its imaginary component. We shall use the expression $\mathfrak{Im}\,\mathbb{H}$ to indicate the space

$$\{\mathbf{q} \in \mathbb{H} \mid \overline{\mathbf{q}} = -\mathbf{q}\}$$

of imaginary quaternions. It is effectively $\mathbb{R}^3$, but using the definitions above, we can multiply two elements of $\mathfrak{Im}\,\mathbb{H}$ according to the rule (23.18). In particular, if $\mathbf{q} = \mathbf{v}$ is an imaginary quaternion, then

$$\mathbf{q}^2 = \mathbf{v}^2 = -\mathbf{v} \cdot \mathbf{v} = -|\mathbf{v}|^2 \tag{23.19}$$

is a *real* number.

The key properties underlying Definition 23.7 are encapsulated in

Lemma 23.8. *Let* $\mathbf{q}, \mathbf{q}_1, \mathbf{q}_2 \in \mathbb{H}$. *Then*

(i) $\mathbf{q}\overline{\mathbf{q}} = |\mathbf{q}|^2$;

(ii) $\overline{\mathbf{q}_1\mathbf{q}_2} = \overline{\mathbf{q}}_2\,\overline{\mathbf{q}}_1$;

(iii) $|\mathbf{q}_1\mathbf{q}_2| = |\mathbf{q}_1||\mathbf{q}_2|$.

Proof. Part (i) follows from (23.19), given that $(a + \mathbf{v})(a - \mathbf{v}) = a^2 - \mathbf{v}^2$. Part (ii) is an immediate consequence of equation (23.17), and the fact that $-\mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{v}_2 \times \mathbf{v}_1$. Given (i),

$$\begin{aligned} |\mathbf{q}_1\mathbf{q}_2|^2 &= \mathbf{q}_1\mathbf{q}_2\overline{\mathbf{q}_1\mathbf{q}_2} \\ &= \mathbf{q}_1\mathbf{q}_2\overline{\mathbf{q}}_2\,\overline{\mathbf{q}}_1 = \mathbf{q}_1|\mathbf{q}_2|^2\overline{\mathbf{q}}_1 = |\mathbf{q}_1|^2|\mathbf{q}_2|^2. \end{aligned}$$ ∎

Part (i) of this lemma has an important consequence. It exhibits explicitly the inverse of a nonzero quaternion, namely

$$\mathbf{q}^{-1} = \frac{1}{|\mathbf{q}|}\overline{\mathbf{q}}, \qquad \mathbf{q} \neq 0.$$

This means that $\mathbb{H}$ is a ***division ring*** or ***skew-field***, an algebraic structure with all the properties of a field except commutativity [Her, CoSm].

At this point, the multiplication rules allow one to regard a complex number as a quaternion for which $c = d = 0$. However, this is to miss the point, since we could have equally well chosen $\mathbf{j}$ or $\mathbf{k}$ to play the role of the symbol $i = \sqrt{-1}$ used to define complex numbers. Indeed, if $\mathbf{v} \in \mathfrak{Im}\,\mathbb{H}$ is any fixed ***unit imaginary quaternion***, we can identify the real 2-dimensional subspace

$$\langle\, 1,\, \mathbf{v}\,\rangle = \{a + b\mathbf{v} \mid a, b \in \mathbb{R}\} \tag{23.20}$$

with $\mathbb{C}$ by means of the mapping

$$a + ib \mapsto a + b\mathbf{v}.$$

The multiplications (of $\mathbb{C}$ on the one hand and $\mathfrak{Im}\,\mathbb{H}$ on the other) are consistent, because

$$\begin{aligned}(a_1 + b_1\mathbf{v})(a_2 + b_2\mathbf{v}) &= a_1a_2 + (a_1b_2 + b_1a_2)\mathbf{v} + b_1b_2\mathbf{v}^2 \\ &= (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)\mathbf{v}.\end{aligned}$$

The set of unit imaginary quaternions is parametrized by the unit sphere $S^2(1)$; in conclusion, there is a whole sphere's worth of complex planes incorporated into $\mathbb{H}$! This multiple way in which $\mathbb{H}$ extends $\mathbb{C}$ is what is understood by the term "hypercomplex geometry" when applied in relation to the quaternions.

Writing out Lemma 23.8(iii) (squared) in full gives the identity

$$(a_1{}^2 + b_1{}^2 + c_1{}^2 + d_1{}^2)(a_2{}^2 + b_2{}^2 + c_2{}^2 + d_2{}^2) = a_3{}^2 + b_3{}^2 + c_3{}^2 + d_3{}^2, \tag{23.21}$$

where (copying from Notebook 23),

$$\begin{aligned}a_3 &= a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2, \\ b_3 &= a_2b_1 + a_1b_2 - c_2d_1 + c_1d_2, \\ c_3 &= a_2c_1 + a_1c_2 + b_2d_1 - b_1d_2, \\ d_3 &= -b_2c_1 + b_1c_2 + a_2d_1 + a_1d_2.\end{aligned}$$

Now consider what happens when the components of $\mathbf{q}_1$ and $\mathbf{q}_2$ are all integers. We see that the product of two *sums of four squared integers* is itself a *sum of four squared integers*. This fact is the crucial first step in proving the following result of Lagrange.

Theorem 23.9. Any *positive integer n can be written as the sum*

$$n = a^2 + b^2 + c^2 + d^2, \qquad a, b, c, d \in \mathbb{N} \tag{23.22}$$

of four squares of nonnegative integers.

In the light of (23.21), it is sufficient to prove the result when n is a prime number, and we refer the reader to any standard text on elementary number theory, such as [Baker].

The representation (23.22) is in general far from unique. For example, it is verified in Notebook 23 that if $n = 100$ then the solutions (a, b, c, d) with $a \geqslant b \geqslant c \geqslant d$ are

$$(5,5,5,5),\ (7,5,5,1),\ (7,7,1,1),\ (8,4,4,2),\ (8,6,0,0),\ (9,3,3,1),\ (10,0,0,0).$$

Three of them are implicit in the quaternionic factorization

$$(1 + i + j + k)(3 - 4j) = 7 + 7i - j - k.$$

23.3 Unit Quaternions and Rotations

We have seen (Corollary 23.6) that quaternions obey the associative rule. The set $\mathbb{H}$ is not itself a group because the quaternion 0 has no multiplicative inverse, though the set $\mathbb{H} \setminus \{0\}$ *is* a group under multiplication. So is the set

$$\mathscr{U} = \{\mathbf{q} \in \mathbb{H} \mid |\mathbf{q}| = 1\}$$

of unit quaternions. This follows from Lemma 23.8 since if $\mathbf{q}_1, \mathbf{q}_2 \in \mathscr{U}$ then

$$|\mathbf{q}_1\mathbf{q}_2| = |\mathbf{q}_1||\mathbf{q}_2| = 1,$$

and $\mathscr{U}$ is closed under products. Also, if $\mathbf{q} \in \mathscr{U}$ then

$$\mathbf{q}^{-1} = \overline{\mathbf{q}} \in \mathscr{U},$$

showing that $\mathscr{U}$ is closed under taking inverses, and the rules on page 131 are satisfied. Readers who are not familiar with $\mathscr{U}$ may nonetheless be acquainted with its finite subgroup Q_8 of eight elements (see Exercise 3).

Since a unit quaternion has the form

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \quad \text{with} \quad a^2 + b^2 + c^2 + d^2 = 1,$$

$\mathscr{U}$ can be identified with the set

$$\big\{(a, b, c, d) \in \mathbb{R}^4 : a^2 + b^2 + c^2 + d^2 = 1\big\}. \tag{23.23}$$

This is none other than the hypersphere $S^3(1)$, consisting of all the 4-vectors a unit distance from the origin (see similar definitions on pages 236 and 662).

The aim of this section is to show that $\mathscr{U}$ is intimately related to the group $\mathbf{SO}(3)$ of rotations in $\mathbb{R}^3$. The starting point is

Lemma 23.10. *Let $\mathbf{q} \in \mathscr{U}$. If $\mathbf{w} \in \mathfrak{Im}\,\mathbb{H}$ then $\mathbf{q}\mathbf{w}\overline{\mathbf{q}}$ is an imaginary quaternion with the same norm as $\mathbf{w}$. The resulting mapping*

$$\mathbf{w} \mapsto \mathbf{q}\mathbf{w}\overline{\mathbf{q}} \tag{23.24}$$

is an orthogonal transformation of $\mathbb{R}^3$.

Proof. By hypothesis, $\overline{\mathbf{w}} = -\mathbf{w}$ and $\mathbf{w}^2 = -|\mathbf{w}|^2$. Let $\mathbf{p} = \mathbf{q}\mathbf{w}\overline{\mathbf{q}}$. The property of conjugation implies that

$$\overline{\mathbf{p}} = \overline{\overline{\mathbf{q}}\,\overline{\mathbf{w}}\,\overline{\mathbf{q}}} = \mathbf{q}(-\mathbf{w})\overline{\mathbf{q}} = -\overline{\mathbf{p}}.$$

It follows that $\mathbf{p}$ is also an imaginary quaternion; we compute

$$\begin{aligned} |\mathbf{p}|^2 \;=\; \mathbf{p}\overline{\mathbf{p}} \;&=\; (\mathbf{q}\mathbf{w}\overline{\mathbf{q}})(-\mathbf{q}\mathbf{w}\overline{\mathbf{q}}) \\ &= \mathbf{q}(-\mathbf{w}^2)\overline{\mathbf{q}} \\ &= \mathbf{q}|\mathbf{w}|^2\overline{\mathbf{q}} \;=\; |\mathbf{w}|^2. \end{aligned} \tag{23.25}$$

The mapping $\mathbf{w} \mapsto \mathbf{p}$ is certainly linear, and the last statement now follows from the proof of Theorem 5.6 on page 132. ∎

We denote the transformation (23.24) by $R[\mathbf{q}]$. Thus, if $\mathbf{q} = a + \mathbf{v} \in \mathscr{U}$, then

$$\begin{aligned} R[\mathbf{q}]\mathbf{w} &= (a+\mathbf{v})\mathbf{w}(a-\mathbf{v}) \\ &= a^2\mathbf{w} + a(\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v}) - \mathbf{v}\mathbf{w}\mathbf{v}, \qquad \mathbf{w} \in \mathfrak{Im}\,\mathbb{H}. \end{aligned} \tag{23.26}$$

The consequences of this calculation are pursued in the following key result in the subject.

Theorem 23.11. *Let* $\mathbf{q}$ *be a unit quaternion. Then* $R[\mathbf{q}]$ *represents a counterclockwise rotation by the angle* $2\arccos(\mathfrak{Re}\,\mathbf{q})$*, measured between* 0 *and* 2π*, with axis pointing in the direction of* $\mathfrak{Im}\,\mathbf{q}$*.*

The qualification *counterclockwise* is to be interpreted when one looks downwards from a "mast" pointing in the direction of $\mathfrak{Im}\,\mathbf{q}$ onto the perpendicular plane of rotation. This is equivalent to the "right-hand corkscrew rule."

Proof. Write $\mathbf{v} = \mathfrak{Im}\,\mathbf{q}$ and $\mathbf{q} = a + \mathbf{v} \in \mathscr{U}$. There is a unique number θ, with $0 \leqslant \theta \leqslant 2\pi$ such that $a = \cos(\theta/2)$, and so $|\mathbf{v}| = \sin(\theta/2) \geqslant 0$.

If $\mathbf{v} = 0$ then $\mathbf{q} = \pm 1$ and θ is 0 or 2π, which gives rise to the identity rotation, and accords with the statement of the theorem. Hence, we may assume that $|\mathbf{v}| \neq 0$. Observe too that equation (23.26) implies that

$$R[\mathbf{q}]\mathbf{v} = a^2\mathbf{v} + |\mathbf{v}|^2\mathbf{v} = \mathbf{v}. \tag{23.27}$$

Next, choose a unit imaginary quaternion $\mathbf{w}_1$ perpendicular to $\mathbf{v}$ in $\mathfrak{Im}\,\mathbb{H} = \mathbb{R}^3$. Then $\mathbf{w}_1 \cdot \mathbf{v} = 0$ and $\mathbf{v}\mathbf{w}_1 = -\mathbf{w}_1\mathbf{v} = \mathbf{v} \times \mathbf{w}_1$. If we set

$$\mathbf{w}_3 = \frac{1}{|\mathbf{v}|}\mathbf{v}, \qquad \mathbf{w}_2 = \mathbf{w}_3\mathbf{w}_1,$$

then $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthonormal basis of $\mathbb{R}^3$ with $\mathbf{w}_1 \times \mathbf{w}_2 = \mathbf{w}_3$. Moreover, using (23.26),

$$\begin{aligned} R[\mathbf{q}]\mathbf{w}_1 &= \mathbf{w}_1 \cos^2\frac{\theta}{2} + 2\mathbf{v}\mathbf{w}_1 \cos\frac{\theta}{2} - \mathbf{w}_1 \sin^2\frac{\theta}{2} \\ &= \mathbf{w}_1 \cos\theta + \mathbf{w}_2 \sin\theta. \end{aligned} \tag{23.28}$$

Similarly,

$$\begin{aligned} R[\mathbf{q}]\mathbf{w}_2 &= \mathbf{w}_2 \cos^2\frac{\theta}{2} + 2\mathbf{v}\mathbf{w}_2 \cos\frac{\theta}{2} - \mathbf{w}_2 \sin^2\frac{\theta}{2} \\ &= \mathbf{w}_2 \cos\theta - \mathbf{w}_1 \sin\theta, \end{aligned} \tag{23.29}$$

exactly as in (23.5). Equations (23.27), (23.28), (23.29) assert that the linear transformation $R[\mathbf{q}]$ has matrix (23.4) relative to the orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, and is the rotation stated in the theorem. ∎

Note that both $\mathbf{q}$ and $-\mathbf{q}$ determine the same transformation; this is related to the appearance of half-angles $\theta/2$. Computationally, it is now an easy matter to use quaternions to model rotations, such as the one that Figure 23.3 attempts to represent:

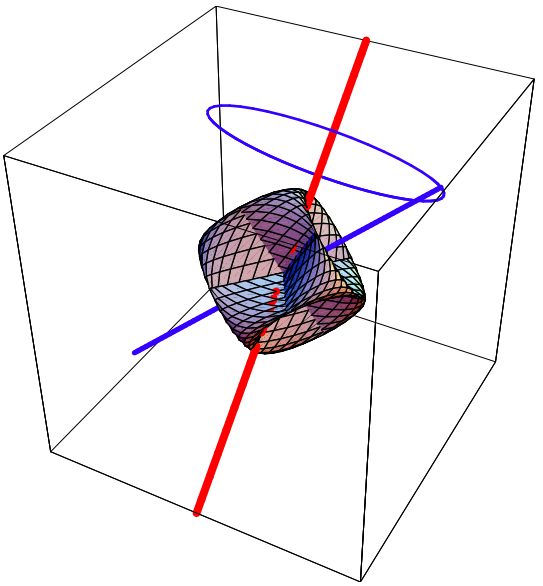


Figure 23.3: A rotation of the sine surface by $R[\frac{1}{\sqrt{2}} + \frac{1}{2}\mathbf{j} + \frac{1}{2}\mathbf{k}]$

The mapping

$$\begin{array}{ccc} \mathscr{U} & \longrightarrow & \mathbf{SO}(3) \\ \mathbf{q} & \mapsto & R[\mathbf{q}] \end{array} \tag{23.30}$$

satisfies the important property

$$R[\mathbf{q}_1\mathbf{q}_2] = R[\mathbf{q}_1] \circ R[\mathbf{q}_2]. \tag{23.31}$$

This follows from its definition (23.24), since $R[\mathbf{q}_1\mathbf{q}_2]$ is the transformation

$$\mathbf{w} \mapsto (\mathbf{q}_1\mathbf{q}_2)\mathbf{w}\overline{(\mathbf{q}_1\mathbf{q}_2)} = \mathbf{q}_1(\mathbf{q}_2\mathbf{w}\overline{\mathbf{q}}_2)\overline{\mathbf{q}}_1.$$

Equation (23.31) characterizes what is called a ***group homomorphism***, a mapping from one group to another that preserves the respective multiplication rules. (For this topic, we recommend [Her, chapter 2].)

At this stage, we choose to write $R[\mathbf{q}]$ explicitly as a 3×3 orthogonal matrix. Referring to Lemma 23.2, this matrix is obtained by conjugating (23.4) by the orthogonal matrix whose columns are the orthonormal triple $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. The result is well known and is computed in Notebook 23:

Proposition 23.12. *Let* $\mathbf{q} = (a, b, c, d)$ *be a unit quaternion. Then the rotation* $R[\mathbf{q}]$ *has matrix*

$$\begin{pmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd \\ 2bc+2ad & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 \end{pmatrix}. \tag{23.32}$$

In Notebook 23, we also verify directly that this matrix has determinant 1, even though this fact is a consequence of the proof of Theorem 23.11. It also follows from (23.26) and a continuity argument, as $\det R[\mathbf{q}]$ cannot jump from 1 to -1.

It is tempting to say that (23.32) is completely determined by the imaginary part $\mathfrak{Im}\,\mathbf{q} = (b, c, d)$ by defining

$$a = \sqrt{1-b^2-c^2-d^2}.$$

This is not quite true, because $\mathfrak{Re}\,\mathbf{q}$ may be negative, and changing the overall sign of $\mathbf{q}$ to adjust for this has the effect of reversing the sign of $\mathbf{v}$ and the angle of rotation. Nevertheless, there is a very natural way of converting a 3-vector into a rotation, and we describe this in the next section.

As an application of Theorem 23.11, consider the composition of two rotations $R[\mathbf{q}_1], R[\mathbf{q}_2]$ about axes $\mathbf{v}_1, \mathbf{v}_2$ with respective angles θ_1, θ_2. The combined rotation (23.31) is easiest to identify in the case in which the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal:

Corollary 23.13. *If* $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, *the composition* $R[\mathbf{q}_1\mathbf{q}_2]$ *is a rotation through an angle* θ *with*

$$\cos\frac{\theta}{2} = \cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2},$$

and parallel to the axis

$$\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}\,\mathbf{i} + \cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}\,\mathbf{j} + \sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}\,\mathbf{k}.$$

Proof. Without loss of generality, we may assume that $\mathbf{v}_1, \mathbf{v}_2$ are parallel to the coordinate axes x, y. Then

$$\mathbf{q}_1 = a_1 + b_1\mathbf{i}, \qquad \mathbf{q}_2 = a_2 + b_2\mathbf{j},$$

where $a_i = \cos(\theta_i/2)$ and $b_i = \sin(\theta_i/2)$. Thus,

$$\mathbf{q}_1\mathbf{q}_2 = a_1a_2 \;+\; b_1a_2\,\mathbf{i} + a_1b_2\,\mathbf{j} + b_1b_2\,\mathbf{k},$$

and we can obtain the required formulas, bearing in mind that $\mathbf{q}_1\mathbf{q}_2$ is again a *unit* quaternion. ∎

23.4 Imaginary Quaternions and Rotations

The secret of using vectors to achieve rotations relies on the following fact, which (only) at first sight appears to have nothing to do with quaternions. Any rotation can be represented by a vector $\mathbf{x}$ in $\mathbb{R}^3$ with $|\mathbf{x}| \leqslant \pi$, by means of the following rule:

> *the axis of rotation is in the direction given by* $\mathbf{x}$*, and the angle of rotation is* $|\mathbf{x}|$ *radians counterclockwise.*

The representation above is almost unique, but there is an inevitable ambiguity, this time only if $|\mathbf{x}| = \pi$. (Recall that, in this chapter, $|\mathbf{x}|$ denotes the usual Euclidean norm $\|\mathbf{x}\|$.) For in this case, both $\mathbf{x}$ and $-\mathbf{x}$ represent the same rotation by π radians or 180°. This is because a rotation of θ counterclockwise about $\mathbf{x}$ is the same thing as a rotation of $-\theta$ or $2\pi - \theta$ counterclockwise about $-\mathbf{x}$, and the two coincide if $\theta = \pi$.

The description above incorporates a significant fact. To explain it, we first define the *closed ball*

$$B^3(\pi) = \{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| \leqslant \pi\}, \tag{23.33}$$

consisting of the sphere $S^2(\pi)$ of radius π, together with all its interior points. We are therefore saying that the set $\mathbf{SO}(3)$ of rotations can be thought of as the set $B^3(\pi)$ in which antipodal points on the boundary sphere $S^2(\pi)$ are identified. For if $|\mathbf{x}| = \pi$, then $\mathbf{x}$ and $-\mathbf{x}$ are *different* points in $B^3(\pi)$ that define the *same* rotation. The idea is that a bookworm living inside the ball that tunnels to the boundary does not get outside the sphere, but re-emerges at the point inside diametrically opposite the point it left.

It is a consequence of the description above that the set $\mathbf{SO}(3)$ of rotations forms a topological space *without* boundary. Mathematically, the sort of operation that has been performed on $B^3(\pi)$ to obtain $\mathbf{SO}(3)$ is a higher-dimensional analogue of the type of identifications carried out in Section 11.2. The argument after Figure 11.16 on page 346 shows that the disk $B^2(\pi)$ in $\mathbb{R}^2$ with points of its boundary circle $S^1(1)$ identified gives rise to the cross cap surface. The latter is a realization of the real projective plane $\mathbb{RP}^2$, which parametrizes straight lines passing through the origin in $\mathbb{R}^3$. The same argument can be repeated to see that the set $\mathbf{SO}(3)$ is topologically the real projective 3-space $\mathbb{RP}^3$ parametrizing lines through the origin in $\mathbb{R}^4$; see Exercise 12.

To understand the link with quaternions, we need to represent analytically the rotation defined by the vector $\mathbf{x}$. We can do this quickly using the machinery of the previous section. Let $\ell = |\mathbf{x}|$ denote the required angle of rotation, this time with $0 < \ell \leqslant \pi$. Set

$$a = \cos\frac{\ell}{2} \geqslant 0, \qquad \mathbf{v} = \Big(\sin\frac{\ell}{2}\Big)\frac{\mathbf{x}}{\ell}. \tag{23.34}$$

Then $\mathbf{q} = a + \mathbf{v}$ is a unit quaternion, whose imaginary part $\mathbf{v}$ has norm $\sin(\ell/2)$. Theorem 23.11 implies that $R[\mathbf{q}]$ is the required rotation. Recalling the notation (23.10), we can use this to establish

Corollary 23.14. *Let $\mathbf{x} \in B^3(\pi)$. The matrix of the rotation $R[\mathbf{q}]$ (with axis $\mathbf{x}$ and counterclockwise angle $|\mathbf{x}|$) equals $\exp A[\mathbf{x}]$.*

This result is effectively Theorem 23.4(ii), whose proof we now give.

Proof. Having fixed $\mathbf{x}$, we use (23.34), and temporarily abbreviate

$$A = A[\mathbf{v}] = \frac{\sin(\ell/2)}{\ell} A[\mathbf{x}].$$

Bearing in mind the form of equation (23.13), we shall express A^2 using quaternion multiplication. Assume first that $\mathbf{w}$ is perpendicular to $\mathbf{v}$. Then

$$\begin{aligned} 2A^2\mathbf{w} = 2A(\mathbf{v} \times \mathbf{w}) &= \mathbf{v} \times (2\mathbf{v} \times \mathbf{w}) \\ &= \mathbf{v} \times (\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v}) \\ &= \frac{1}{2}\Big(\mathbf{v}(\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v}) - (\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v})\mathbf{v}\Big) \\ &= -\mathbf{v}\mathbf{w}\mathbf{v} - |\mathbf{v}|^2\mathbf{w}. \end{aligned} \tag{23.35}$$

Since $\mathbf{v}\mathbf{w} - \mathbf{w}\mathbf{v} = 2\mathbf{v}\mathbf{w}$ and $a^2 + |\mathbf{v}|^2 = 1$, we may re-write (23.26) as

$$R[\mathbf{q}]\mathbf{w} = \mathbf{w} + 2aA\mathbf{w} + 2A^2\mathbf{w}.$$

This relation also holds when $\mathbf{w}$ is parallel to $\mathbf{v}$ because of (23.27). In conclusion,

$$\begin{aligned} R[\mathbf{q}] &= I_3 + 2\cos\frac{\ell}{2}A + 2A^2 \\ &= I_3 + \frac{\sin\ell}{\ell}A[\mathbf{x}] + \frac{1-\cos\ell}{\ell^2}A[\mathbf{x}]^2, \end{aligned}$$

as expected. ∎

It follows that

$$\mathbf{SO}(3) = \{\exp A[\mathbf{x}] \mid |\mathbf{x}| \leqslant \pi\},$$

so that the composition $\exp \circ A$ maps the closed ball $B^3(\pi)$ *onto* $\mathbf{SO}(3)$.

The descriptions we have seen make it easy to model rotations, a careful choice of which can be used to view the details of a surface. Figure 23.4 is a single frame of an animation from Notebook 23 of a truncated view of the sine surface, designed to reveal its inners (the axis of rotation is visible). Lines of self-intersections meet in a more complicated singularity at the origin.

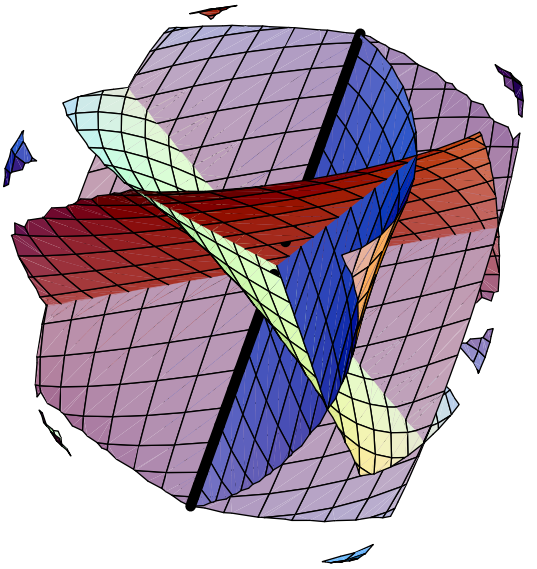


Figure 23.4: Viewing singularities of the sine surface

23.5 Rotation Curves

The theory of the preceding section taught us to associate to any vector $\mathbf{x}$ in the ball $B^3(\pi)$ a rotation in space, and to represent this rotation by the orthogonal matrix $\exp A[\mathbf{x}]$. In practice, this is best calculated using the formula of Theorem 23.4, which we now know arises from the unit quaternion

$$\mathbf{q} = \cos(\ell/2) + \frac{\sin(\ell/2)}{\ell}\mathbf{x}, \qquad \ell = |\mathbf{x}| \leqslant \pi, \tag{23.36}$$

defined by (23.34). The only ambiguity occurs when $|\mathbf{x}| = \pi$ for then $\pm\mathbf{x}$ give rise to the same unit quaternion $\mathbf{q}$.

Let $\mathbf{c}_j$ denote the j^{th} column of the matrix (23.32) representing (23.36). It is the image of the j^{th} basis element (from the list (23.12)) under the rotation. Since $a^2 + b^2 + c^2 + d^2 = 1$, we have

$$\mathbf{c}_1 = \begin{pmatrix} -2c^2 - 2d^2 + 1 \\ 2bc + 2ad \\ 2bd - 2ac \end{pmatrix} = R[\mathbf{q}] \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The fact that $R[\mathbf{q}]$ is a rotation translates into the identity

$$\mathbf{c}_1 \times \mathbf{c}_2 = \mathbf{c}_3. \tag{23.37}$$

In particular, it means that $R[\mathbf{q}]$ is completely determined by the first two columns $\mathbf{c}_1, \mathbf{c}_2$.

We can now work backwards, and attempt to construct (23.32), $R[\mathbf{q}]$ and ultimately $\mathbf{x}$ itself from the columns of (23.32). If we start from the first column, the only restriction is that it be a *unit* vector, so

$$\mathbf{c}_1 \in S^2(1).$$

However, having fixed $\mathbf{c}_1$, the next column $\mathbf{c}_2$ must be chosen perpendicular to $\mathbf{c}_1$. This having been done, $\mathbf{c}_3$ is completely determined by (23.37).

Let $\mathcal{M} = S^2(1)$. If $\mathbf{u} \in \mathcal{M}$, then the orthogonal complement

$$\mathbf{u}^\perp = \{\mathbf{w} \in \mathbb{R}^3 \mid \mathbf{u} \cdot \mathbf{w} = 0\}$$

can be identified with the tangent space $\mathcal{M}_\mathbf{u}$. In general, the set

$$\{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u} \in \mathcal{M},\ \mathbf{w} \in \mathcal{M}_\mathbf{u}\}$$

is called the ***tangent bundle*** of the surface $\mathcal{M}$, and is discussed in a more general setting in the next chapter (within Section 24.5). But for the moment, it leads to yet another description of the set $\mathbf{SO}(3)$.

Proposition 23.15. *The set* $\mathbf{SO}(3)$ *is in bijective correspondence with the* ***unit tangent bundle***

$$\{(\mathbf{u}, \mathbf{w}) \mid \mathbf{u}, \mathbf{w} \in S^2(1),\ \mathbf{u} \cdot \mathbf{w} = 0\}$$

of the unit sphere.

Note the symmetry between $\mathbf{u}$ and $\mathbf{w}$ in this definition.

Proposition 23.15 is relevant to the next topic. At the start of this book, on page 5, we defined a curve in $\mathbb{R}^n$, and in Chapters 7 we specialized to the case $n = 3$. We are now in a position to adapt this to the case in which $\mathbb{R}^3$ is replaced by $\mathbf{SO}(3)$.

Definition 23.16. *A* ***rotation curve*** *is a mapping* $\boldsymbol{\gamma}\colon [a, b] \to \mathbf{SO}(3)$ *of the form*

$$\boldsymbol{\gamma}(t) = \exp A[\boldsymbol{\alpha}(t)],$$

where $\boldsymbol{\alpha}\colon [a, b] \to \mathbb{R}^3$ *is a parametrized curve whose image lies in* $B^3(\pi)$. *The rotation curve is* ***closed*** *if* $\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b)$.

We begin with two basic examples of such curves.

(i) Fix a unit vector $\mathbf{u}$, and set

$$\boldsymbol{\alpha}(t) = t\,\mathbf{u}, \qquad -\pi \leqslant t \leqslant \pi.$$

Then $\boldsymbol{\gamma}(t) = \exp A[t\mathbf{u}]$ is a rotation, by an angle t, about a fixed axis parallel to $\mathbf{u}$. The curve $\boldsymbol{\alpha}$ is a diameter of $B^3(\pi)$, but $\boldsymbol{\gamma}$ is closed since $\exp A[-\pi\mathbf{u}] = \exp A[\pi\mathbf{u}]$.

(ii) Fix an orthonormal pair $\mathbf{u}_1, \mathbf{u}_2$ of vectors, and set

$$\boldsymbol{\beta}(t) = \Big(\mathbf{u}_1 \cos\frac{t}{2} + \mathbf{u}_2 \sin\frac{t}{2}\Big)\pi, \qquad -\pi \leqslant t \leqslant \pi.$$

The trace of $\boldsymbol{\beta}$ is half of a great circle lying on the boundary $S^2(\pi)$ of $B^3(\pi)$, and *every* element $\boldsymbol{\delta}(t) = \exp A[\boldsymbol{\beta}(t)]$ is a rotation of 180°. The rotation curve $\boldsymbol{\delta}$ is again closed since $\boldsymbol{\beta}(\pi) = \pi\mathbf{u}_2 = -\boldsymbol{\beta}(-\pi)$.

These two examples are almost identical, though this fact is partially hidden by our preference for imaginary quaternions; matters become clearer when we revert to unit quaternions. In case (i), the associated unit quaternion is

$$\mathbf{q}(t) = \cos\frac{t}{2} + \mathbf{u}\sin\frac{t}{2}, \tag{23.38}$$

whereas for (ii), $\boldsymbol{\beta}(t)$ is itself a unit imaginary quaternion. Indeed, $\boldsymbol{\beta}(t)$ has the same form as (23.38) with $\mathbf{u}_1, \mathbf{u}_2$ in place of $1, \mathbf{u}$; if we choose $\mathbf{u} = \overline{\mathbf{u}}_1\mathbf{u}_2$, then

$$\boldsymbol{\beta}(t) = \mathbf{u}_1\mathbf{q}(\theta).$$

This means that

$$\boldsymbol{\delta}(t) = R[\mathbf{u}_1]\boldsymbol{\gamma}(t),$$

and the two rotation curves differ by the fixed 180° rotation $R[\mathbf{u}_1]$. In technical language, *left translation* within the group $\mathscr{U}$ converts a diameter of $B^3(\pi)$ into half a great circle.

Surprisingly complicated examples can be obtained by starting from a curve whose trace lies in the intersection of $B^3(\pi)$ with a plane through the origin. Using standard coordinates in $\mathbb{R}^3$, consider the ellipse

$$\boldsymbol{\alpha}[a,b](t) = \big(0,\ a\cos t,\ b\sin t\big)\pi, \qquad -\pi \leqslant t \leqslant \pi, \tag{23.39}$$

where a, b are fixed positive numbers such that $a^2\cos^2 t + b^2\sin^2 t \leqslant 1$ for all t. Let $\boldsymbol{\gamma}(t) = \exp A[\boldsymbol{\alpha}[a,b](t)]$; both $\boldsymbol{\alpha}[a,b]$ and $\boldsymbol{\gamma}$ are closed curves. To describe $\boldsymbol{\gamma}$, it suffices to specify the first two columns $\boldsymbol{\gamma}(t)\mathbf{e}_1$ and $\boldsymbol{\gamma}(t)\mathbf{e}_2$ in accordance with the discussion leading to Proposition 23.15. As t varies, these generate two curves lying on $S^2(\pi)$.

This example is animated in Notebook 23 and illustrated in Figure 23.5. One can imagine a rigid body with two perpendicular arms that are constrained to follow the respective spherical curves. In the frame shown, the rotation is about to pass through a "vertex" in which one column achieves an extreme value, and the second passes through the chicane in the figure eight.

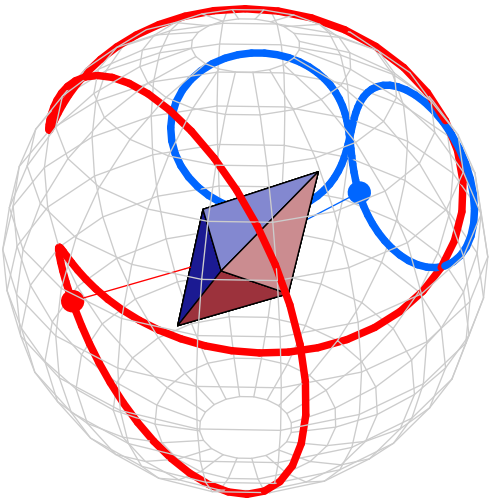


Figure 23.5: A rotation curve specified by its columns

Figure 23.6 concerns the case of (23.39) in which $a = b$ and

$$\boldsymbol{\gamma}[a](t) = \exp(A[\boldsymbol{\alpha}[a, a]]), \qquad 0 \leqslant t \leqslant 2\pi.$$

From example (ii) above, $\boldsymbol{\gamma}[1]$ represents a rotation curve corresponding to revolving two turns (4π radians or 720^{o}) around a fixed axis. At the other extreme, $\boldsymbol{\gamma}[0]$ is the "identity curve," whose trace consists of the single point $I_3 \in \mathbf{SO}(3)$. Figure 23.6 displays the intermediate curves $\boldsymbol{\gamma}[1-a]$ with $a = n/17$ and $1 \leqslant n \leqslant 16$. In the first frame, the straightforward 720^{o} rotation has been deformed slightly to $\boldsymbol{\gamma}[16/17]$ so that the axis begins to "wobble." By the time we get down to $\boldsymbol{\gamma}[4/17]$, represented by the first frame in the last row, a rigid body subject to this family of rotations will merely rock slowly about an axis lullaby-fashion (imagine Figure 23.5 with the more tranquil background).

The family

$$a \mapsto \boldsymbol{\gamma}[1 - a], \qquad 0 \leqslant a \leqslant 1$$

is a continuous deformation from the "two turns curve" $\boldsymbol{\gamma}[1]$ to the identity $\boldsymbol{\gamma}[0]$. In the language of Section 6.3, it is a *homotopy* between $\boldsymbol{\gamma}[1]$ and $\boldsymbol{\gamma}[0]$. There are other ways of demonstrating the existence of such a homotopy, such as the so-called Dirac scissors trick (see [Pen, §11.3]). There exists no such homotopy if we begin with the closed rotation curve $\boldsymbol{\gamma}[1]$ defined for $0 \leqslant t \leqslant \pi$, which corresponds to a single revolution. This fact is expressed by saying that the topological space $\mathbf{SO}(3)$, which we have mentioned is equivalent to $\mathbb{RP}^3$, is not ***simply-connected*** — there exist closed curves that cannot be deformed

to a point. It is however the case that the "double" of any closed curve can always be deformed to a point. These facts are explained in introductory texts to algebraic topology, such as [Arm].

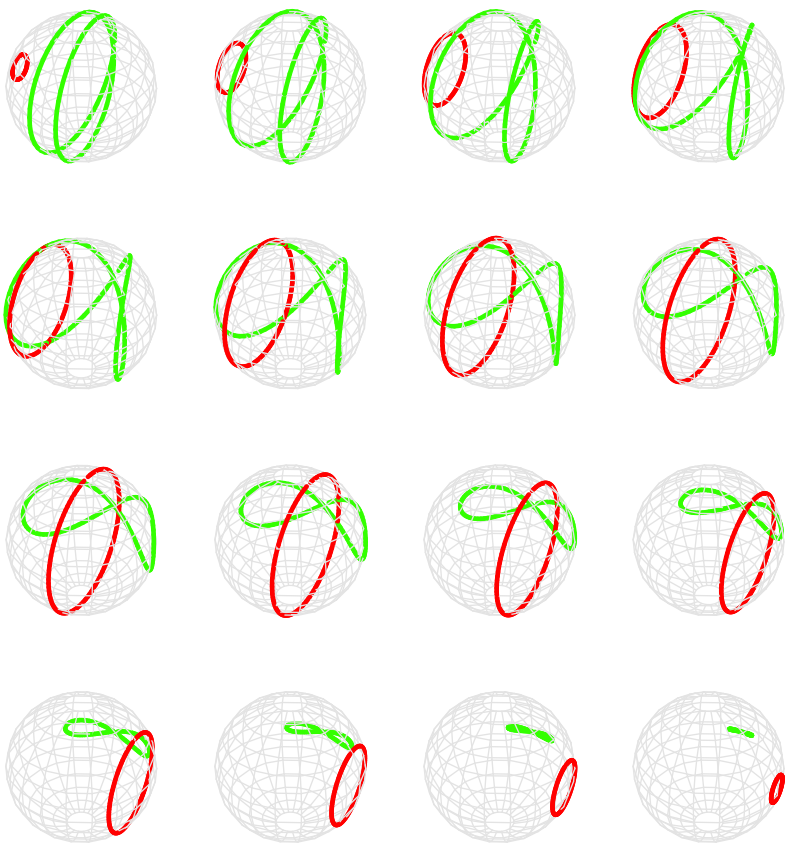


Figure 23.6: A homotopy of rotation curves $\boldsymbol{\gamma}[\frac{n}{17}]$

23.6 Euler Angles

The traditional way to specify a rotation is to break it up into the composition of three rotations about the three Cartesian axes. In practice, this is the method used in fairground machinery, or in a gyroscope to align a navigation system.

In order to provide a vivid explanation of Euler angles, we define them in relation to an aircraft taking off from runway 27L at Heathrow airport in London, on its way to Paris. Imagine first a Cartesian frame of reference $\mathbf{F}_0$ fixed in the center of the runway, with the x-axis due east, y-axis due north

(which just happens to be perpendicular to the runway), and z (for "zenith") vertically upwards. As the aircraft passes by on its take-off run, its own frame of reference $\mathbf{F}_1$ (with x pointing to the nose and y to the port or left wing) is obtained from $\mathbf{F}_0$ by means of a rotation of 180^o about the z-axis, to give the correct heading of 270^o. After takeoff, the aircraft climbs at a constant angle or elevation of 20^o; mathematically this is achieved by rotating $\mathbf{F}_1$ by -20^o about its y-axis, and produces a new frame of reference $\mathbf{F}_2$. At a height of 2000 feet, the aircraft banks left, by rolling -30^o about its nose-tail or x-axis, temporarily achieving the frame of reference $\mathbf{F}_3$.

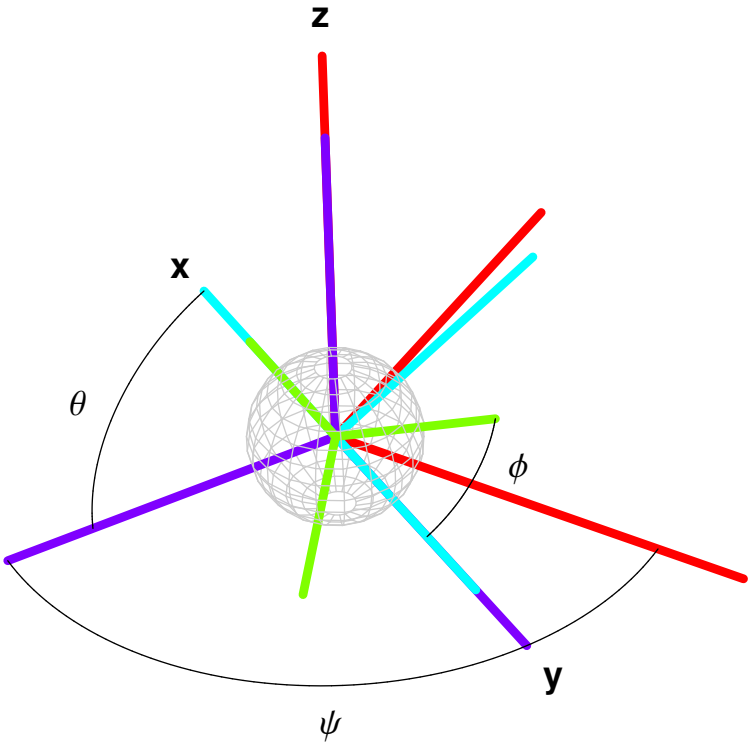


Figure 23.7: A sequence of rotations about axes $z \rightsquigarrow y \rightsquigarrow x$

To sum up, $\mathbf{F}_3$ has been obtained from $\mathbf{F}_0$ by means of a sequence of rotations through

$$\psi = 180^o, \quad \theta = -20^o, \quad \phi = -30^o \tag{23.40}$$

about the z-, y-, x-axes respectively (our notation is similar to that of [Kuipers, 4.4]). The triple $(\psi, \theta\, \phi)$ comprises the ***Euler angles*** of the single rotation P needed to move $\mathbf{F}_0$ to $\mathbf{F}_3$, and can be regarded as yet another way of representing a rotation by a 3-vector (albeit one whose three components are angles).

To be more precise, we can interpret each frame $\mathbf{F}_i$ as the orthonormal triple of vectors defining the frame's axes, and translate all the frames to a common origin. In this way, $\mathbf{F}_3$ is obtained from $\mathbf{F}_0$ by the composition

$$R[a_3 + b_3\mathbf{i}] \circ R[a_2 + b_2\mathbf{j}] \circ R[a_1 + b_1\mathbf{k}]$$

of rotations, where

$$a_3 = \cos\frac{\phi}{2}, \quad a_2 = \cos\frac{\theta}{2}, \quad a_1 = \cos\frac{\psi}{2}.$$

One can easily compute P using Corollary 23.13.

The definition of the Euler angles depends on the sequence of axes chosen to perform the respective three rotations. The above choice $z \rightsquigarrow y \rightsquigarrow x$ (meaning first z then y then x) is illustrated in Figure 23.7 but with a different choice of angles ($\psi = -2\pi/3$, $\theta = -\pi/3$, $\phi = +\pi/4$ in radians). At each step the new axes are shortened to help identify them, but the figure is best seen on screen in color. One can attempt to use any other sequence to obtain a given rotation, such as $x \rightsquigarrow y \rightsquigarrow x$, though it is useless if the same axis appears next to itself in the sequence. It follows that there is a total of $3 \cdot 2 \cdot 2 = 12$ choices for the triple, but essentially only two types: those in which the three axes are different and those in which the first and third coincide. As it happens, the standard rotation package within *Mathematica* uses the sequence $z \rightsquigarrow x \rightsquigarrow z$ (see Notebook 23).

Euler's theorem states that any element in $\mathbf{SO}(3)$ can be decomposed as the composition of three rotations about mutually orthogonal axes, in either of the above ways we care to choose. Using the theory from Section 23.3, it amounts to the following factorization result for unit quaternions. First, recall the notation (23.20), and consider the circle

$$\mathscr{U}_{\mathbf{v}} = \{a + b\mathbf{v} \mid a^2 + b^2 = 1\} \tag{23.41}$$

formed by intersecting the subspace $\langle\, 1,\, \mathbf{v}\,\rangle$ with $\mathscr{U}$.

Theorem 23.17. *Let $\mathbf{q} \in \mathscr{U}$. Then there exist*

(i) $\mathbf{q}_1 \in \mathscr{U}_{\mathbf{i}}$, $\mathbf{q}_2 \in \mathscr{U}_{\mathbf{j}}$, $\mathbf{q}_3 \in \mathscr{U}_{\mathbf{k}}$ *such that* $\mathbf{q} = \mathbf{q}_3\mathbf{q}_2\mathbf{q}_1$;

(ii) $\mathbf{p}_1 \in \mathscr{U}_{\mathbf{i}}$, $\mathbf{p}_2 \in \mathscr{U}_{\mathbf{j}}$, $\mathbf{p}_3 \in \mathscr{U}_{\mathbf{i}}$ *such that* $\mathbf{q} = \mathbf{p}_3\mathbf{p}_2\mathbf{p}_1$.

We omit the proof, referring the reader to [Kuipers, chapter 8]. A first step consists in the characterization of those quaternions that can be represented in the form $\mathbf{q}_1\mathbf{q}_2$ where $\mathbf{q}_1 \in \mathscr{U}_{\mathbf{i}}$ and $\mathbf{q}_2 \in \mathscr{U}_{\mathbf{j}}$ (see Exercise 8). To explain the nature of the problem, we prove instead the following simpler result.

Lemma 23.18. *Given $\mathbf{v} \in \mathfrak{Im}\,\mathbb{H}$, there exist $a, b, c \in \mathbb{R}$ such that*

$$\mathfrak{Im}[(a + \mathbf{i})(b + \mathbf{j})(c + \mathbf{k})] = \mathbf{v}.$$

Proof. Writing $\mathbf{v} = r\mathbf{i} + s\mathbf{j} + t\mathbf{k}$ and expanding, we need to solve the system

$$\begin{cases} a + bc &= r, \\ -b + ac &= s, \\ c + ab &= t, \end{cases} \tag{23.42}$$

that does not quite have cyclic symmetry. Eliminating b, we obtain

$$\begin{cases} a(1 + c^2) &= cs + r, \\ c(1 + a^2) &= as + t. \end{cases}$$

Eliminating c gives the quintic equation

$$a^5 - r\,a^4 + 2a^3 + (st - 2r)a^2 + (t^2 - s^2 + 1)a - r - st = 0,$$

which has at least one real root a. The last two equations in (23.42) are then linear and can be solved for b and c. ∎

There are well-documented disadvantages of the use of Euler angles to describe rotations. Not only do they depend very much on the choice of axis sequence, but are subject to an analog of ***gimbal lock*** of an inertial guidance system. Fear of gimbal lock was the bane of mission control during the Apollo moon landing and other space flights. One manifestation of the difficulty is that when an Euler angle approaches $\pi/2$, its cosine approaches 0, and it is impossible for a computer to perform division by $\cos\theta$. Mechanically, one can partly understand the problem by supposing that our aircraft is able (quickly after takeoff) to climb vertically, so that θ has become -90° in (23.40). In these circumstances, the aircraft's heading has been "lost," the third rotation could have been achieved by changing the direction of the runway, and ψ, ϕ are no longer independent parameters.

The advantage of our earlier description of rotations based on imaginary quaternions is that the mapping $B^3(\pi) \to \mathbf{SO}(3)$ defined by

$$\boldsymbol{\rho}(\mathbf{x}) = \exp A[\mathbf{x}] = R[\mathbf{q}]$$

is *regular* in the sense that the partial derivatives $\boldsymbol{\rho}_x, \boldsymbol{\rho}_y, \boldsymbol{\rho}_z$ are always linearly independent (as 3×3 matrices, or elements of $\mathbb{R}^9$), in the spirit of Definition 19.16 on page 606. This is *not* true of the mapping $\boldsymbol{\sigma}\colon \mathcal{R} \to \mathbf{SO}(3)$ defined by

$$\boldsymbol{\sigma}(\psi, \theta, \phi) = \mathbf{F}_3,$$

where $\mathcal{R} = [0, 2\pi] \times [-\pi/2, \pi/2] \times [-\pi, \pi]$ is the usual domain of definition for the Euler angles. Indeed, we verify in Notebook 23 that the set

$$\left\{ \boldsymbol{\sigma}\left(\psi, \frac{\pi}{2}, \phi\right) \;\middle|\; \psi, \phi \in \mathbb{R} \right\}$$

is a curve and not a surface in $\mathbf{SO}(3)$.

23.7 Further Topics

We conclude this chapter with some more applications. The first two link to the themes of Chapters 7 and 15 respectively.

Frenet Frames Revisited

We can associate to each point of a space curve $\boldsymbol{\alpha}\colon (a,b) \to \mathbb{R}^3$ a rotation matrix $\mathbf{F}(t)$ as follows. The parameter t need not be arc length. The columns of $\mathbf{F}(t)$ are the unit tangent, normal and binormal vectors $\mathbf{T}(t), \mathbf{N}(t), \mathbf{B}(t)$ at the point $\boldsymbol{\alpha}(t)$. These are mutually orthogonal and by definition,

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

It follows that $\mathbf{F}(t)$ lies in $\mathbf{SO}(3)$, and represents the rotation necessary to transform the standard frame of reference $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ to $\mathbf{F}(t)$, in analogy to the aeronautical example of the previous section.

The assignment of the unit tangent vector $\mathbf{T}(t)$, thought of as a point on $S^2(1)$, to $\boldsymbol{\alpha}(t)$ is a type of Gauss map for the curve, although it could be argued that using $\mathbf{N}(t)$ in place of $\mathbf{T}(t)$ is closer to the Gauss definition for surfaces. The fact is that we may as well associate the *whole Frenet frame* $\mathbf{F}(t)$ to a point $\boldsymbol{\alpha}(t)$ where $\boldsymbol{\kappa}(t) \neq 0$. The resulting assignment

$$t \mapsto \boldsymbol{\alpha}(t) \mapsto \mathbf{F}(t), \tag{23.43}$$

ultimately mapping the trace of $\boldsymbol{\alpha}$ to $\mathbf{SO}(3)$, is an example of a "higher order" Gauss map, of the sort that is much studied in current research.

The composition (23.43) describes a rotation curve. The latter may be closed even if $\boldsymbol{\alpha}$ is not, and the best illustration of this phenomenon is the helix

$$\mathsf{helix}[a,b,c](t) = \big(a\cos t,\ b\sin t,\ ct\big)$$

(see Section 7.3), which is circular if $a = b$. Since

$$\mathsf{helix}[a,b,c](t+2\pi) = \mathsf{helix}[a,b,c](t) + (0,\ 0,\ 2\pi c),$$

it follows that $\mathbf{F}(t+2\pi) = \mathbf{F}(0)$, and

$$\mathbf{F}\colon [0,2\pi] \longrightarrow \mathbf{SO}(3)$$

is a closed rotation curve.

A more subtle example is the cubic curve

$$\mathsf{twicubic}(t) = (t,\ t^2,\ t^3) \tag{23.44}$$

introduced on page 202. When $|t|$ is very large, we may ignore t and t^2 in comparison to t^3, and the curve approximates the vertical straight line $t \mapsto (0,0,t^3)$

for which $\mathbf{T}(t) = (0, 0, 1)$ independently of t (be it positive or negative). On the other hand, the curvature $\boldsymbol{\kappa}$[twicubic] computed on page 205 never vanishes. Thus, it is always possible to define $\mathbf{N}$ and $\mathbf{B}$, and one verifies that

$$\begin{aligned} \lim_{t\to\pm\infty} \mathbf{N}(t) &= (0,\ -1,\ 0), \\ \lim_{t\to\pm\infty} \mathbf{B}(t) &= (1,\ 0,\ 0) \end{aligned} \tag{23.45}$$

(see Exercise 10). As a consequence, adjoining the single point

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

makes $\mathbf{F}(t)$ into a smooth closed curve in $\mathbf{SO}(3)$. What is really going on here is that (23.44) can be extended to a mapping $\mathbb{RP}^1 \to \mathbb{RP}^3$ in the context of projective geometry (see Exercise 12), and $\mathbf{F}$ too has values in $\mathbb{RP}^3$.

Figure 23.8 shows the situation for Viviani's curve, which is somewhat of a special case, as the original curve already lies on the sphere. It displays the traces of $\mathbf{T}$ and $\mathbf{B}$ (the latter with cusps), in the style of previous figures. Despite first appearances, the unit tangent vector

$$\mathbf{T}(t) = \frac{\sqrt{2}}{\sqrt{3+\cos t}}\Big(\sin t,\ \cos t,\ \cos\frac{t}{2}\Big) \tag{23.46}$$

is not itself the intersection of the sphere with a circular cylinder (see Exercise 9).

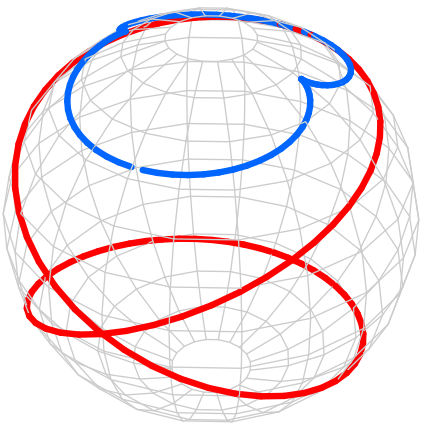


Figure 23.8: The moving Frenet frame on Viviani's curve

It is tempting to think that *any* curve $\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3)$ arises as the Frenet frame $\mathbf{F}(t)$ of a suitable space curve $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$. The Frenet formulas, in the version of Theorem 7.13 on page 203, show that this is not however the case, since $\mathbf{T}'(t)$ must be parallel to the second column $\mathbf{N}(t)$ of $\boldsymbol{\gamma}(t)$. Thus, the rotation curve

$$\mathbf{F} : (a,b) \to \mathbf{SO}(3)$$

defined by the Frenet frame of $\boldsymbol{\beta}$ is completely determined by the unit tangent mapping $\mathbf{T}\colon (a,b) \to S^2(1)$. This provides a significant constraint on the trace of $\mathbf{F}$, which is best understood using the description of $\mathbf{SO}(3)$ given by Proposition 23.15.

Generalized Surfaces of Revolution

On page 471 we mentioned Darboux's point of view, whereby a surface of revolution is viewed as the surface swept out by the one-parameter group of motions corresponding to rotation about a fixed axis. This motivated the definition of generalized helicoid, for which a similar interpretation holds. Now that we have developed machinery to describe a general rotation curve

$$\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3),$$

we could equally well apply this to a given space curve $\boldsymbol{\beta}\colon (c,d) \to \mathbb{R}^3$, so as to obtain the chart

$$\mathbf{x}(u,v) = \boldsymbol{\gamma}(u)\boldsymbol{\beta}(v).$$

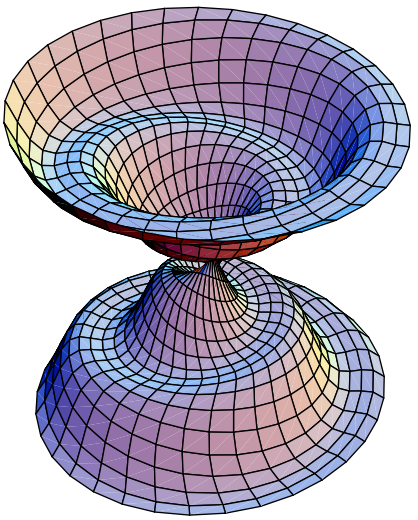


Figure 23.9: A surface generated by applying a rotation curve to a helix

A natural question is what happens when $\boldsymbol{\gamma}$ is one of the ellipses (23.39). Figure 23.1 was obtained by taking

$$\boldsymbol{\gamma}(u) = \exp A[(0,\ \tfrac{9}{10}\cos u,\ \tfrac{27}{100}\sin u)]$$

in the notation of (23.10), and $\boldsymbol{\beta}$ to be the straight line

$$\boldsymbol{\beta}(v) = (v,\ 0,\ 0).$$

Figure 23.9 applies a similar rotation curve to the space curve

$$\boldsymbol{\beta}(v) = \big(\cos 5v,\ \sin 5v,\ 10v\big).$$

Rotations in $\mathbb{R}^4$

The astute reader may question the extent to which quaternions are essential for an analysis of the rotation group. While it is true that Theorem 23.4 provided the main computational power of the chapter, one should not underestimate the mathematical significance of Theorem 23.11, particularly in extending the theory to $\mathbb{R}^n$. To finish, we merely state a generalization of Lemma 23.10.

Lemma 23.19. *Let $\mathbf{q}_1, \mathbf{q}_2 \in \mathscr{U}$. If $\mathbf{p} \in \mathbb{H}$ then $\mathbf{q}_1\mathbf{p}\overline{\mathbf{q}}_2$ is a quaternion with the same norm as $\mathbf{p}$. The resulting mapping*

$$\mathbf{p} \mapsto \mathbf{q}_1\mathbf{p}\overline{\mathbf{q}}_2 \tag{23.47}$$

is an orthogonal transformation of $\mathbb{R}^4$ and determines a matrix in $\mathbf{SO}(4)$.

Proof. It suffices to modify the step (23.25) for the previous proof to work. ∎

This leads to a description of $\mathbf{SO}(4)$ based on $\mathscr{U} \times \mathscr{U}$. Further to (23.30), there is an associated homomorphism of groups

$$\mathscr{U} \times \mathscr{U} \longrightarrow \mathbf{SO}(4).$$

Exactly two elements of $\mathscr{U} \times \mathscr{U}$ (namely, $(\mathbf{q}_1, \mathbf{q}_2)$ and $(-\mathbf{q}_1, -\mathbf{q}_2)$) map to the same rotation of $\mathbb{R}^4$. Elements of the type $(\mathbf{q}, \mathbf{q})$ really do map to $\mathbf{SO}(3)$, in the sense that they leave fixed the direction generated by $1 \in \mathbb{H}$, in the same way that (23.4) is effectively a rotation of $\mathbb{R}^2$ rather than $\mathbb{R}^3$.

The ability to describe a rotation by two separate elements $\mathbf{q}_1, \mathbf{q}_2$ is a special feature of the group of rotations in *four* dimensions. This led in the 1970s to a fuller understanding by mathematicians of gauge theories in theoretical physics, a resurgence of interest in quaternions [Atiyah], and unforeseen developments in pure mathematics.

23.8 Exercises

1. Prove that the matrix $A = A[\mathbf{x}]$ satisfies

(a) $A^3 = -\|\mathbf{x}\|^2 A$.

(b) $A^2 = \mathbf{x}\mathbf{x}^T - \|\mathbf{x}\|^2 I_3$.

2. Let $\mathbf{p}, \mathbf{q} \in \mathbb{H}$. Under what conditions on $\mathbf{p}$ and $\mathbf{q}$ is it true that $\mathbf{pq} = \mathbf{qp}$?

3. Show that the eight elements

$$1, \quad -1, \quad i, \quad j, \quad k, \quad -i, \quad -j, \quad -k$$

form a group under quaternion multiplication.

4. Let $P \in \mathbf{SO}(3)$, $P^2 \neq I_3$. Show that P represents a rotation about an axis parallel to the vector $\mathbf{x}$, where $A[\mathbf{x}] = P - P^T$.

5. Prove that if X, Y are skew-symmetric $n \times n$ matrices then so is $XY - YX$. Now suppose that $n = 3$ and that $X = A[\mathbf{x}]$, $Y = A[\mathbf{y}]$. Use (23.11) to verify that

$$XY - YX = A[\mathbf{x} \times \mathbf{y}],$$

and relate this fact to the Jacobi identity (24.46) on page 832.

6. Let $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a complex 2×2 matrix satisfying $A^T \overline{A} = I_2$, where $\overline{A}$ is the matrix obtained from A by complex-conjugating its four elements and I_2 is the identity. Show that the complex number $\det A = \alpha\delta - \beta\gamma$ has modulus one, and that if $\det A = 1$ then $\gamma = -\overline{\beta}$ and $\delta = \overline{\alpha}$.

7. Prove that the set of matrices

$$\mathbf{SU}(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix} \;\middle|\; |\alpha|^2 + |\beta|^2 = 1 \right\}$$

arising from the previous question is a group under matrix multiplication. By writing each complex entry in real and imaginary parts, show that its elements are in bijective correspondence with those of the group $\mathscr{U}$ of unit quaternions.

8. Show that a unit quaternion

$$\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

can be written as a product $\mathbf{q} = \mathbf{q}_1\mathbf{q}_2$ with $\mathbf{q}_1 \in \mathscr{U}_\mathbf{i}$ and $\mathbf{q}_2 \in \mathscr{U}_\mathbf{j}$ if and only if $ad - bc = 0$. See (23.41) for the notation.

M **9.** Plot the projection of the curve (23.46) on the xy-plane, and verify that it is not a circle. Find analogous expressions for the normal and binormal vectors to Viviani's curve.

M **10.** Verify the limits (23.45), and plot the rotation curve $\mathbf{F}(t)$ for a twisted cubic.

M **11.** Let $\boldsymbol{\beta}\colon (a,b) \to \mathbb{R}^3$ be a curve with speed $v = v(t)$ whose Frenet frame defines $\boldsymbol{\gamma}\colon (a,b) \to \mathbf{SO}(3)$. Use Theorem 7.13 to show that

$$\boldsymbol{\gamma}'(t) = \boldsymbol{\gamma}(t)\, A[\mathbf{x}(t)],$$

where $\mathbf{x}(t) = (-\boldsymbol{\tau}(t),\, 0,\, \boldsymbol{\kappa}(t))v$.

12. Define an equivalence relation $\sim$ on $S^n(1) \subset \mathbb{R}^{n+1}$ by writing $\mathbf{a} \sim \mathbf{b}$ if and only if $\mathbf{a} = \pm\mathbf{b}$. The resulting quotient space is called ***real projective space*** of dimension n and denoted by $\mathbb{RP}^n$. The case $n = 2$ was the subject of Section 11.5. Define $p\colon S^n(1) \mapsto \mathbb{RP}^n$ by $p(\mathbf{a}) = \{\mathbf{a}, -\mathbf{a}\}$. Deduce from Theorem 23.11 that there is a bijective correspondence between $\mathbf{SO}(3)$ and $\mathbb{RP}^3$.

Notebook 23

■ N23.0

This notebook uses orthogonal matrices and quaternion algebra to rotate and animate surfaces. We shall use the sine surface, introduced in Notebook 10, as a prime example for illustrations. Viewed from afar it is approximately cubical.

```
sinsurface[u_,v_]:= {Sin[u], Sin[v], Sin[u+v]}
```

```
ParametricPlot3D[sinsurface[u,v], {u,0,2π}, {v,0,2π},
  PlotRange->Table[{-3,3},{i,3}]]
```

The sphere will serve as a background to various space curves. Viviani's curve was first defined in N7.5, and is used in N23.6.

```
sphere[a_][u_,v_]:= a{Cos[v]Cos[u], Cos[v]Sin[u], Sin[v]}
viviani[a_][t_]:= a{1+Cos[t], Sin[t], 2Sin[t/2]}
```

The first of the following packages can be used to multiply quaternions numerically, though more basic techniques are required to handle the algebra.

```
<<Algebra`Quaternions`
<<Graphics`Shapes`
```

We reload the definitions of L, Frenet frame commands, and the curvature $\kappa 2$ of a plane curve in N23.0.1. This subsection also contains the code for Figure 23.1.

■ N23.0.1

■ N23.1

The 3 by 3 identity matrix (already used in N7.1) appears inside the definition of the skew-symmetric matrix A[x] associated to a 3-vector x:

```
id= IdentityMatrix[3]
```

```
Unprotect[A]
  A[x_]:= Table[Cross[id[[i]], x], {i,3}]
Protect[A]
```

```
x:= {b,c,d}
A[x]//MatrixForm
```

In *Mathematica*, powers of matrices must be computed with MatrixPower to avoid merely raising each entry to the given power. The same is true for exponentiation.

```
M:= A[x]
M^3//MatrixForm
MatrixPower[M,3]//MatrixForm
```

Adding a decimal point after any entry of the vector v has the same effect as using the command N.

```
Exp[A[{1,2,3}]]//MatrixForm
MatrixExp[A[{1,2,3.}]]//MatrixForm
```

We next compute exp[A[x]] in the simpler way suggested by Theorem 23.4, though its evaluation when x=0 causes a problem owing to division by zero.

```
EA[x_]:=
Module[{l},
  l= L[x];
  id + Sin[l] A[x]/l + (1-Cos[l]) A[x].A[x]/l^2
]
```

```
xx:= {1,2,3.}
EA[xx]//MatrixForm
Sum[MatrixPower[A[xx], k]/k!, {k,0,12}] - EA[xx]
```

■ N23.1.1

■ N23.2

The Quaternions package represents a quaternion q as a linear combination of I,J,K.

```
q= Quaternion[1,4,3,2]
FromQuaternion[q]
```

At another level, q is merely a vector with four components, or an element of 4-space. We customize the following command to represent q in this way:

```
from[q_]:= Table[q[[i]], {i,4}]
```

```
from[q]
```

Complex numbers can be regarded as quaternions whose imaginary part is a vector parallel to the x-axis:

```
p= ToQuaternion[1 + 7I]
FromQuaternion[p]
```

Like complex numbers, quaternions can be multiplied in a consistent way, and the product is indicated ** in *Mathematica*. This multiplication is not commutative (order matters), but it *is* associative (brackets are not needed).

```
Clear[x]
p= Quaternion[0,a,b,c]
q= Quaternion[0,s,t,u]
r= Quaternion[0,x,y,z]
from[p**q - q**p]
```

```
from[(p**q)**r - p**(q**r)]//Simplify
```

The product of two *imaginary* quaternions equals minus their scalar product plus their cross product. But to form the product of two generic quaternions p,q, we need to add terms originating from their real parts:

```
p= Quaternion[f,a,b,c]
q= Quaternion[w,x,y,z]
from[p**q]//ColumnForm
%/.{f->0, w->0}
```

If z is a complex number, the product of z with its conjugate equals its modulus (denoted Abs[z] in *Mathematica)* squared, and is therefore a *real* number:

```
z= 1 + 5I
z'= Conjugate[z]
z z'
Abs[z]
Clear[z]
```

A similar thing happens for the conjugate q' of a quaternion q (the correct notation is q overlined). The product of qq' is effectively the norm-squared of q. (Warning: use Abs and not Norm!). A *unit quaternion* is a quaternion u whose norm is 1.

```
q= Quaternion[1,2,3,4]
q'= Conjugate[q]
q**q'
Abs[q]
Norm[q]
q/Abs[q]//Abs
```

Conjugation of quaternions obeys the rule (pq)' = q'p' in which the order of factors is reversed:

```
p:= Quaternion[1,2,3,4]
q:= Quaternion[0,5,6,0]
p'= Conjugate[p]
q'= Conjugate[q]
Conjugate[p**q]
p'**q'
q'**p'
```

Here we find the components of a product, in notation used in the text:

```
q[i_]:= Quaternion[a[i], b[i], c[i], d[i]]
from[q[1]**q[2]]//ColumnForm
```

Output from the next cell lists sums of four squares totalling 100.

```
Do[n:= a^2+b^2+c^2+d^2; If[n==100, Print[{a,b,c,d}]],
  {a,0,10}, {b,0,a}, {c,0,b}, {d,0,c}]
```

■ N23.3

The following cell converts a unit quaternion q, input as a 4-vector like {a,b,c,d}, into the rotation matrix R[q]. It uses very little in the way of special commands; e[1],e[2],e[3] stand for the unit imaginary quaternions i,j,k.

```
Unprotect[R];
R[q_]:=
Module[{e,p,u},
  e[1]= Quaternion[0,1,0,0];
  e[2]= Quaternion[0,0,1,0];
  e[3]= Quaternion[0,0,0,1];
  u= Quaternion[q[[1]], q[[2]], q[[3]], q[[4]]];
  Table[(u**e[j]**Conjugate[u])[[i]], {i,2,4}, {j,3}]
]
Protect[R]
```

In theory, one could normalize the 4-vector q inside the program; in practice this procedure clashes with the Quaternions package and it is best to assume already that Abs[q]=1. Here is an example:

```
p:= {1,2,3,4}
q:= p/L[p]
M= R[q]
M//MatrixForm
```

Now we check that the matrix has determinant 1 and is orthogonal:

```
Det[M]
MatrixForm[M.Transpose[M]]//Simplify
```

Here is the general case:

```
q:= {a,b,c,d}
R[q]//TeXForm
```

The code for Figure 23.3 in the hidden subsection makes use of the customized plotting program para contained in an initialization cell in N23.4.

N23.3.1

N23.4

We now compute the unit quaternion and rotation matrix associated to a 3-vector x. The matrix coincides with EA[x], which we can therefore use in computations.

```
x:= {1,2,.3}
l:= L[x]
q= Prepend[Sin[l/2]x/l, Cos[l/2]]
R[q]//MatrixForm
EA[x]//MatrixForm
```

It is the *columns* c[v,j] of EA[x] that are the images of the basis elements i,j,k under the rotation. This makes them a little awkward to extract, but we do so by realizing that they are the *rows* of the transposed matrix:

```
c[x_,j_]:= Transpose[EA[x]][[j]]
```

These columns are also obtained by premultiplying the standard basis of column vectors, the first of which is represented below by {1,0,0} (*Mathematica* understands that this is to be interpreted as a column to make sense of the matrix multplication). The fact that R[v] is a rotation is equivalent to asserting that c[v,j] is an orthonormal basis with the third column equal to the cross product of the first two:

```
x:= {.1,2,3}/30
EA[x].{1,0,0}
c[x,1]
Cross[c[x,1], c[x,2]] - c[x,3]
```

The next command is an abbreviation for ParametricPlot3D with some useful options. Setting a plot range is an important technique for fixing the frame of reference

when performing an animation, so the constant a is inserted as a first variable. The symbol A stands for a vector-valued function f[u,v] to plot, plus the ranges of u,v, plus any additional options required.

```
para[a_,A__]:= ParametricPlot3D[A, Compiled->False, Axes->False,
  PlotRange->{{-a,a},{-a,a},{-a,a}}]
```

We rotate the sine surface about a fixed direction determined by the unit vector

```
W:= {1,2,3}/Sqrt[14]
```

The first rotation of each "run" is one of π radians, and an end-value of t=19/20 makes sure that a frame is not repeated in the movie. An increment of 39/400 rather than 1/20 ensures that the problematic EA[0] is never realized. The red axis indicates the direction of W, while thin black axis is attached to the surface. Replace Show by S to activate the animation.

```
Module[{ax,z,y},
  ss= sinsurface;
  ax= Graphics3D[{Hue[0],Thickness[.015], Line[{-3W,3W}]}];
  z[t_]:= para[3, EA[π t W].ss[u,v], {u,0,2π}, {v,0,2π}];
  y[t_]:= para[3, EA[π t W].{s,0,0}, {s,-3,3}];
  Do[Show[y[t], z[t], ax], {t,-1, 19/20, 39/400}]
]
```

N23.4.1

N23.5

We shall now illustrate the effect of rotating a diamond-shaped box (actually an ellipsoid plotted with few PlotPoints), which one can imagine being strapped inside of! First, we define a couple of space curves, incorporating a parameter c that must be chosen to be sufficiently small that the curves lie in the ball of radius π.

```
X[c_,t_]:= c{0,Cos[t],.3Sin[t]}π
Y[c_,t_]:= .999{Cos[t], 2Sin[t], c}π/Sqrt[Cos[t]^2+4Sin[t]^2+c^2]
```

We then define a background sphere, parametrized by its GrayLevel:

```
sfera[gl_]:= WireFrame[{GrayLevel[gl], Sphere[1,20,15]}]
```

Each column of EA[X[c,t]] corresponds to a curve on the sphere. The rotation matrix is completely determined by any two of its columns that must always be perpendicular. Because of this, it is sufficient to attach two perpendicular arms of unit length to the box and force them to trace out the two figure eights that the circle in imaginary quaternion space surprisingly generates. This is accomplished by the next program.

```
ride[c_,s_]:=
Module[{o,ax,g,r},
  o= {0,0,0};
  f[i_,t_]:= EA[X[c,t]][[i]];
  curv[i_]:=para[1, Append[f[i,t], {Hue[.6(i-1)], Thickness[.01]}],
    {t,0,2π}, PlotPoints->100];
  ax[t_]:= Graphics3D[{{Hue[0],Line[{o, f[1,t]}]}, {Hue[.6],
    Line[{o, f[2,t]}]}, {PointSize[.03],Hue[ 0],Point[f[1,t]]},
    {PointSize[.03],Hue[.6],Point[f[2,t]]}}];
  g[u_,v_]:= {Cos[u]Cos[v], Sin[u]Cos[v], 2Sin[v]}/4;
  dia[t_]:= para[1, EA[X[c,t]].g[u,v], {u,0,2π}, {v,0,2π},
    PlotPoints->5];
  Show[sfera[4/5], ax[s], dia[s], curv[1], curv[2], Boxed->False]
]
```

```
S @ ride[.9,2.7]
```

Figure 23.5

Try varying the two parameters in the definition of ride, to change the static picture. The second curve defined above can be used by changing Y to X (in this way the second definition of X will overwrite the first). The program can be animated by replacing Show by S in the definition of ride, and evaluating the next cell.

```
Do[ride[.9, s], {s, 0, 39π/20, π/20}]
```

The animation can be viewed at (say) 5 frames per second by setting

Graphics Options->Animation->AnimationDisplayTime->0.2

in the Preferences menu. Another animation is included after instructions for Figure 23.6 in the following subsection.

■ N23.5.1

N23.6

The more traditional way to specify a rotation is to break it up into the composition of three rotations about the three Cartesian axes with respective Euler angles θ, ϕ, ψ. This can be accomplished using the add-on

```
<<Geometry`Rotations`
```

The package's convention is to use the axis sequence z,x,z to describe the rotation. Here is the resulting matrix:

```
RotationMatrix3D[θ, ϕ, ψ]
```

An analog of gimbal lock obviously occurs when the second angle is 0 or π.

```
RotationMatrix3D[α, π, ψ]
```

We can evaluate the Euler sequence z,y,x using the command EA, and verify the loss of freedom produced by setting the middle angle to be $\pi/2$.

```
Rx:= EA[id[[1]] α]//PowerExpand
Ry:= EA[id[[2]] π/2]
Rz:= EA[id[[3]] γ]//PowerExpand
Rx.Ry.Rz//MatrixForm
```

N23.6.1

N23.7

Frenet frame calculations enable us to plot Figure 23.8, and its animation in N23.7.1:

```
υ:= viviani[1]
tnb[t_]:= {tangent[υ][t], normal[υ][t], binormal[υ][t]}
cc[i_,t_]:= Append[tnb[t][[i]], {Hue[.3(i-1)], Thickness[.01]}]
cc[3,1]//N
```

```
zz[i_]:= para[1, cc[i,t], {t,.1,4π}, PlotPoints->100]
S[sfera[4/5], zz[1], zz[3], Boxed->False]
```

Figure 23.8

The methods can be used to generate new surfaces:

```
Module[{f},
  f[u_]= {Cos[4u], Sin[4u], 10u};
  S @ ParametricPlot3D[EA[Z[.4,v]].f[u], {u,-3,3}, {v,0,2π},
    Boxed->False, Axes->False, PlotPoints->35,
    BoxRatios->{1,1,1}]
]
```

Figure 23.9

```
para[4, EA[Z[.9,v]].{0,3+(Sin[v]+2)^2 Cos[u]/9,
  (Sin[v]+2)^2/9 Sin[u]}, {u,0,2π}, {v,0,2π}]
```

N23.7.1

N23.8

The formula in Section 23.7 for the equatorial projection of the unit tangent vector to Viviani's curve is obtained from

```
tt= tangent[viviani[1]][t]
```

The final two cells solve Exercises 9 and 10 respectively.

```
z1:= ParametricPlot[{tt[[1]],tt[[2]]}, {t,0,4π},
  AspectRatio->Automatic]
z2:= {Hue[0], Circle[{0,-.1},.8]}//Graphics
{z1,z2}
```

```
γ[t_]:= {t, t^2, t^3}
κ[γ][t]
normal[γ][100.]
```

Chapter 24

Differentiable Manifolds

We have reached a stage for which it is beneficial to take on board the theory of manifolds. This more abstract theory is a vehicle in which the definitions of Chapter 9 for Euclidean spaces extend in a natural way. The resulting concepts will provide us with a framework with which to pursue the intrinsic study of surfaces begun in Chapter 17.

The modern concept of a differentiable manifold, due to Weyl[1] (see [Weyl]), did not appear until the early twentieth century. A *manifold*, generally speaking, is a topological space which resembles Euclidean space locally. A *differentiable manifold* is a manifold $\mathcal{M}$ for which this resemblance is sharp enough to allow partial differentiation and consequently all the features of differential calculus on $\mathcal{M}$. The study of differentiable manifolds involves topology, since differentiability implies continuity, but metric properties of Euclidean space are not *a priori* included. In this chapter, we define the notion of differentiable manifold and some of the standard apparatus associated with it. Metrics and distances are discussed in the following two chapters.

The set $\mathbb{R}^n$ of n-tuples of real numbers is not only a vector space, but also a topological space, and the vector operations are continuous with respect to the topology. In addition, there is the notion of differentiability of real-valued functions on $\mathbb{R}^n$: we say that $f\colon \mathbb{R}^n \to \mathbb{R}$ is *differentiable* provided that the partial derivatives

$$\frac{\partial^{i_1+\cdots+i_r} f}{\partial u_1^{i_1} \dots \partial u_r^{i_r}}$$



1 Hermann Klaus Hugo Weyl (1885–1955). German mathematician, who together with É. Cartan created the basis for the modern theory of Lie groups and their representations. With his application of group theory to quantum mechanics, he set up the modern theory of the subject. He was professor in Zürich and Göttingen. From 1933 until he retired in 1952, he worked at the Institute of Advanced Study at Princeton.

of all orders exist. Such functions are called C^∞ *functions*. They contrast with, on the one hand, the C^k *functions* (the functions which have continuous partial derivatives whenever $i_1+\cdots+i_r \leqslant k$ for some fixed positive integer k) and, on the other hand, the *real analytic functions* (functions which have convergent power series). If we denote these respective classes by $C^\infty(\mathbb{R}^n)$, $C^k(\mathbb{R}^n)$, $C^\omega(\mathbb{R}^n)$, then we have the strict inclusions

$$C^\omega(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n) \subset C^k(\mathbb{R}^n).$$

We choose to base our definitions on the class of C^∞ functions; use of C^k requires too complicated an approach, and use of C^ω is too restrictive.

In Section 24.1, we give the definition of differentiable manifold and describe some simple examples. The n-dimensional sphere is made into a differentiable manifold by introducing maps that generalize the stereographic map studied in Sections 8.6 and 22.3. The algebra $\mathfrak{F}(\mathcal{M})$ of C^∞ functions on a differentiable manifold is described in Section 24.2. The ensuing definitions generalize those already given in Chapter 12 for functions on and between surfaces. The tangent space to a differentiable manifold is defined in Section 24.3, and derives from the analogous Section 9.3. Maps between differentiable manifolds are studied in Section 24.4, and vector fields and tensor fields in Sections 24.5 and 24.6.

We shall assume that the reader is familiar with all the basic information about calculus of several variables, including the Implicit Function Theorem and existence theorems for ordinary differential equations, already used in previous chapters. This material can be found in, for example, [Spiv].

24.1 The Definition of a Differentiable Manifold

Recall the definition of the ***natural coordinate functions*** of $\mathbb{R}^n$ given on page 273; these are the mappings $u_i\colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$u_i(p_1, \ldots, p_n) = p_i$$

for $i = 1, \ldots, n$, and we shall resolutely distinguish the functions u_i and the numbers p_i. A function $\Psi\colon \mathbb{R}^m \to \mathbb{R}^n$ is ***differentiable***, ***continuous*** or ***linear*** if and only if each $u_i \circ \Psi$ is differentiable, continuous or linear, respectively. This is consistent with Definition 9.7 on page 268.

We now make precise the "local resemblance" referred to on the previous page.

Definition 24.1. *A* ***patch*** *on a topological space $\mathcal{M}$ is a pair $(\mathbf{x},\mathcal{U})$, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$ and*

$$\mathbf{x}\colon \mathcal{U} \longrightarrow \mathbf{x}(\mathcal{U}) \subset \mathcal{M}$$

is a homeomorphism of $\mathcal{U}$ onto an open set $\mathbf{x}(\mathcal{U})$ of $\mathcal{M}$.

Here $\mathbf{x}$ is called the ***local homeomorphism*** of the patch, and $\mathbf{x}(\mathcal{U})$ the ***coordinate neighborhood***. Frequently, we refer to "the patch $\mathbf{x}$" when the domain $\mathcal{U}$ is understood. Let

$$x_i = u_i \circ \mathbf{x}^{-1} \colon \mathbf{x}(\mathcal{U}) \longrightarrow \mathbb{R} \tag{24.1}$$

for $i = 1, \ldots, n$. Then x_i is called the ***ith coordinate function*** and $(x_1, \ldots, x_n)$ is called a ***system of local coordinates*** for $\mathcal{M}$. The coordinate functions $x_1, \ldots, x_n$ contain the same information as the local homeomorphism $\mathbf{x}$. Often, we write $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$, with the meaning that

$$\mathbf{x}^{-1}(\mathbf{p}) = \big(x_1(\mathbf{p}), \ldots, x_n(\mathbf{p})\big)$$

for all $\mathbf{p} \in \mathbf{x}(\mathcal{U})$.

We are now ready to define the notion of differentiable manifold. Intuitively, the idea is this: in studying the geography of the earth, it is a great convenience to use geographical maps or *patches* instead of examining the earth directly. A collection of geographical maps that covers the earth is called an ***atlas***, and it gives a complete picture of the earth. Roughly, we shall follow the same procedure with differentiable manifolds.

Definition 24.2. *An* ***atlas*** $\mathfrak{A}$ *on a topological space* $\mathcal{M}$ *is a collection of patches on* $\mathcal{M}$ *such that all the patches map from open subsets of the same Euclidean space* $\mathbb{R}^n$ *into* $\mathcal{M}$*, and* $\mathcal{M}$ *is the union of all the* $\mathbf{x}(\mathcal{U})$*'s such that* $(\mathbf{x},\mathcal{U}) \in \mathfrak{A}$*. A topological space* $\mathcal{M}$ *equipped with an atlas is called a* ***topological manifold***.

Let $\mathfrak{A}$ be an atlas on a topological space $\mathcal{M}$. Notice that if $(\mathbf{x},\mathcal{U})$ and $(\mathbf{y},\mathcal{V})$ are two patches in $\mathfrak{A}$ such that $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V}) = \mathcal{W}$ is a nonempty subset of $\mathcal{M}$, then the map

$$\mathbf{x}^{-1} \circ \mathbf{y} \colon \mathbf{y}^{-1}(\mathcal{W}) \longrightarrow \mathbf{x}^{-1}(\mathcal{W}) \tag{24.2}$$

is a homeomorphism between open subsets of $\mathbb{R}^n$. We call $\mathbf{x}^{-1} \circ \mathbf{y}$ a ***change of coordinates***. We are now in a position to state

Definition 24.3. *A* ***differentiable manifold*** *is a paracompact topological space* $\mathcal{M}$ *equipped with an atlas* $\mathfrak{A}$ *such that for any two patches* $(\mathbf{x},\mathcal{U}), (\mathbf{y},\mathcal{V})$ *in* $\mathfrak{A}$ *with* $\mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V}) = \mathcal{W}$ *nonempty, the change of coordinates* (24.2) *is differentiable (that is, of class* C^∞*) in the ordinary Euclidean sense. The* ***dimension*** *of the manifold* $\mathcal{M}$ *(denoted by* $\dim \mathcal{M}$*) is the number* n *in Definition* 24.2.

Remarks

(1) Topological terms applied to a manifold refer to its underlying topological space; for example, we shall assume that a given differentiable manifold is connected, unless stated otherwise. Manifolds share many of the local properties of Euclidean space $\mathbb{R}^n$, for example, local connectedness and

local compactness, but the paracompact assumption needs to be added. Strictly speaking, the hypothesis of Hausdorff should also be added in the definition to exclude certain pathological examples. The exact meaning of these words is not important for us, but their definitions can be found in any book on general topology, such as [Kelley].

(2) A function of n complex variables is called ***holomorphic*** if it is so in each variable separately (see page 721), and in this case it is expandable in a complex power series in a neighborhood of each point where it is defined. The definition of ***complex manifold*** is the same as that of a differentiable manifold, except that the local homeomorphisms are required to map from open subsets of $\mathbb{C}^n$, and the changes of coordinates (24.2) are required to be holomorphic rather than C^∞. Any such complex manifold is, in particular, a real manifold of dimension $2n$.

Implicitly associated with $\mathfrak{A}$ is a possibly larger atlas, which is a theoretical nicety that is needed to declare when two manifolds are really the same.

Definition 24.4. *The* ***completion*** $\widetilde{\mathfrak{A}}$ *of an atlas* $\mathfrak{A}$ *is the collection*

$$\widetilde{\mathfrak{A}} = \big\{\, (\mathbf{x},\mathcal{U}) \mid \mathbf{x}^{-1}\circ\mathbf{y} \text{ and } \mathbf{y}^{-1}\circ\mathbf{x} \text{ are differentiable for all } (\mathbf{y},\mathcal{V}) \in \mathfrak{A} \,\big\}.$$

We say that an atlas $\mathfrak{A}$ *is* ***complete*** *if it coincides with its completion.*

We do not distinguish between differentiable manifolds $(\mathcal{M},\mathfrak{A}_1)$ and $(\mathcal{M},\mathfrak{A}_2)$ when the atlases $\mathfrak{A}_1$ and $\mathfrak{A}_2$ have the same completion. When we speak of a "differentiable manifold $\mathcal{M}$," we shall mean a differentiable manifold $\mathcal{M}$ equipped with a specific complete atlas.

There are two simple but important ways to construct new manifolds from old.

Definition 24.5. *Let* $\mathcal{M}$ *be a differentiable manifold defined by a complete atlas* $\mathfrak{A}$*, and let* $\mathcal{V}$ *be an open subset of* $\mathcal{M}$*. Define*

$$\mathfrak{A}\,|\,\mathcal{V} = \big\{\, (\mathbf{x},\mathcal{U}) \in \mathfrak{A} \mid \mathbf{x}(\mathcal{U}) \subseteq \mathcal{V} \,\big\}.$$

Evidently, $\mathfrak{A}\,|\,\mathcal{V}$ *is an atlas of* $\mathcal{V}$*. We call* $\mathcal{V}$ *equipped with the atlas* $\mathfrak{A}(V)$ *an* ***open submanifold*** *of* $\mathcal{M}$*.*

Definition 24.6. *Let* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *be differentiable manifolds of dimensions* n_1 *and* n_2 *defined by atlases* $\mathfrak{A}_1$ *and* $\mathfrak{A}_2$*. If* $(\mathbf{x},\mathcal{U})$ *is in* $\mathfrak{A}_1$ *and* $(\mathbf{y},\mathcal{V})$ *is in* $\mathfrak{A}_2$*, we define* $\mathbf{x}\times\mathbf{y}\colon \mathcal{U}\times\mathcal{V} \to \mathcal{M}_1\times\mathcal{M}_2$ *by*

$$(\mathbf{x}\times\mathbf{y})(\mathbf{p},\mathbf{q}) = \big(\mathbf{x}(\mathbf{p}),\mathbf{y}(\mathbf{q})\big).$$

Then $\mathcal{M}_1 \times \mathcal{M}_2$ equipped with the completion of the atlas

$$\left\{ (\mathbf{x} \times \mathbf{y},\, \mathcal{U} \times \mathcal{V}) \mid (\mathbf{x},\mathcal{U}) \in \mathfrak{A}_1,\ (\mathbf{y},\mathcal{V}) \in \mathfrak{A}_2 \right\}$$

is a differentiable manifold of dimension $n_1 + n_2$, called the ***product*** *of $\mathcal{M}_1$ and $\mathcal{M}_2$.*

For the proof that $\mathcal{M}_1 \times \mathcal{M}_2$ is actually a differentiable manifold, see Exercise 1.

Examples of Differentiable Manifolds

The Euclidean space $\mathbb{R}^n$ is made into a differentiable manifold as follows. Let $\mathbf{1}\colon \mathbb{R}^n \to \mathbb{R}^n$ denote the identity map

$$(u_1, , \ldots, u_n)\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n.$$

Then $(\mathbf{1}, \mathbb{R}^n)$ constitutes an atlas for $\mathbb{R}^n$ all by itself. This formalizes the fact that the notion of manifold is a generalization of Euclidean space.

The n-dimensional sphere or hypersphere represents an example where the concept of atlas is essential. Let $a > 0$ and put

$$S^n(a) = \left\{ (t_0, \ldots, t_n) \in \mathbb{R}^{n+1} \;\middle|\; \sum_{j=0}^{n} t_j^2 = a^2 \right\},$$

as on page 236. We shall make $S^n(a)$ into a differentiable manifold by defining an atlas

$$\mathfrak{A} = \left\{ (\mathsf{north}, \mathbb{R}^n),\ (\mathsf{south}, \mathbb{R}^n) \right\},$$

where north and south are the "stereographic injections." Analytically these are patches

$$\mathsf{north}\colon \mathbb{R}^n \longrightarrow S^n(a) \setminus \{\mathbf{n}\} \qquad \text{and} \qquad \mathsf{south}\colon \mathbb{R}^n \longrightarrow S^n(a) \setminus \{\mathbf{s}\},$$

where $\mathbf{n} = (a, 0, \ldots, 0)$ is the "north pole," and $\mathbf{s} = (-a, 0, , \ldots, 0)$ is the "south pole." They are defined by setting

$$\mathsf{north} = (\Phi_0, \ldots, \Phi_n), \qquad \mathsf{south} = (\Psi_0, \ldots, \Psi_n),$$

where

$$\Phi_0 = -\Psi_0 = a \left(\frac{-a^2 + \sum_{j=1}^{n} u_j^2}{a^2 + \sum_{j=1}^{n} u_j^2} \right), \qquad \text{and} \qquad \Phi_k = \Psi_k = \frac{2a^2 u_k}{a^2 + \sum_{j=1}^{n} u_j^2}$$

for $k = 1, \ldots, n$. In the special case $n = 2$ and $a = 1$, north coincides with $\boldsymbol{\Upsilon}$ on page 250. It is also the inverse of the stereographic projection described by

Definition 22.19 and Figure 22.2 on page 730 with (u_1, u_2) a point of the equatorial plane $\mathbb{R}^2$; south is obtained using straight lines passing through $(0, 0, -a)$ instead of $(a, 0, 0)$. It is easily verified that north and south are injective, and the compositions $\mathsf{north} \circ \mathsf{south}^{-1}$ and $\mathsf{south} \circ \mathsf{north}^{-1}$ are differentiable.

To conclude this subsection, we insert a result that provides a whole host of examples.

Lemma 24.7. *A regular surface $\mathcal{M}$ in $\mathbb{R}^n$ is a differentiable manifold.*

Proof. A regular surface in $\mathbb{R}^n$ was defined on page 297. For the atlas of $\mathcal{M}$ we choose the regular injective patches on $\mathcal{M}$. Corollary 10.30, page 300, implies that each change of coordinates is differentiable. Hence $\mathcal{M}$ is a differentiable manifold. ▮

24.2 Differentiable Functions on Manifolds

Differentiable manifolds are locally like Euclidean space. For their study it will be important to transfer to manifolds as much of the differential calculus of Euclidean space as we can. First on the agenda is the notion of differentiability for real-valued functions on a differentiable manifold. The definition of this notion is almost the same as the corresponding definition that we gave on page 301.

Definition 24.8. *Let $f\colon \mathcal{W} \to \mathbb{R}$ be a function defined on an open subset $\mathcal{W}$ of a differentiable manifold $\mathcal{M}$. We say that f is **differentiable** at $\mathbf{p} \in \mathcal{W}$, provided that for some patch $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ with $\mathcal{U} \subset \mathbb{R}^n$ and $\mathbf{p} \in \mathbf{x}(\mathcal{U}) \subset \mathcal{W}$, the composition $f \circ \mathbf{x}\colon \mathcal{U} \to \mathbb{R}$ is differentiable (in the ordinary Euclidean sense) at $\mathbf{x}^{-1}(\mathbf{p})$. If f is differentiable at all points of $\mathcal{W}$, we say that f is **differentiable** on $\mathcal{W}$.*

This definition illustrates the force behind the definition of differentiable manifold. One might think that it should be necessary to require that $f \circ \mathbf{x}$ be differentiable for *every* patch $(\mathbf{x}, \mathcal{U})$ in the atlas of $\mathcal{M}$. However, this fact is a consequence of the definition:

Lemma 24.9. *The definition of differentiability of a real-valued function on a differentiable manifold does not depend on the choice of patch.*

Proof. If $(\mathbf{x}, \mathcal{U})$ and $(\mathbf{y}, \mathcal{V})$ are patches on a manifold $\mathcal{M}$, then the change of coordinates $\mathbf{x}^{-1} \circ \mathbf{y}$ is differentiable by the definition of differentiable manifold. We can write $f \circ \mathbf{y}$ as

$$f \circ \mathbf{y} = (f \circ \mathbf{x}) \circ (\mathbf{x}^{-1} \circ \mathbf{y}).$$

Since the composition of the Euclidean-differentiable functions is differentiable, the differentiability of $f \circ \mathbf{x}$ implies the differentiability of $f \circ \mathbf{y}$, and conversely. ▮

Next, we consider the totality of differentiable functions on a differentiable manifold and define some algebraic structure on it.

Definition 24.10. *Let $\mathcal{M}$ be a differentiable manifold. We put*

$$\mathfrak{F}(\mathcal{M}) = \{\, f\colon \mathcal{M} \to \mathbb{R} \mid f \text{ is differentiable} \,\}.$$

We call $\mathfrak{F}(\mathcal{M})$ the ***algebra of real-valued differentiable functions*** *$\mathcal{M}$.*

For $a, b \in \mathbb{R}$ and $f, g \in \mathfrak{F}(\mathcal{M})$ the functions $af + bg$ and fg are defined by

$$(af + bg)(\mathbf{x}) = af(\mathbf{x}) + bg(\mathbf{x}) \qquad \text{and} \qquad (fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$$

for $\mathbf{x} \in \mathcal{M}$. Also, we identify any $a \in \mathbb{R}$ with the constant function a given by $a(\mathbf{x}) = a$ for $\mathbf{x} \in \mathcal{M}$.

Let us note some of the algebraic properties of $\mathfrak{F}(\mathcal{M})$.

Lemma 24.11. *Let $\mathcal{M}$ be a differentiable manifold. Then $\mathfrak{F}(\mathcal{M})$ is a commutative ring with identity and an algebra over the real numbers $\mathbb{R}$.*

Proof. Let $f, g \in \mathfrak{F}(\mathcal{M})$ and $a, b \in \mathbb{R}$. If $(\mathbf{x},\mathcal{U})$ is a patch on $\mathcal{M}$, then $f \circ \mathbf{x}$ and $g \circ \mathbf{x}$ are differentiable in the ordinary Euclidean sense; hence $a(f \circ \mathbf{x}) + b(g \circ \mathbf{x})$ and $(f \circ \mathbf{x})(g \circ \mathbf{x})$ are Euclidean differentiable. It follows easily that both $af + bg$ and fg are differentiable. Also, constant functions are differentiable, and the identity of the ring $\mathfrak{F}(\mathcal{M})$ is $1 \in \mathbb{R}$. Associativity, commutativity and distributivity are easy to prove. ∎

Note also that if $f \in \mathfrak{F}(\mathcal{M})$ is never zero, then $1/f \in \mathfrak{F}(\mathcal{M})$.

It will be important to know when we can extend a real-valued function on an open set of $\mathcal{M}$ to a function that is differentiable on *all* of $\mathcal{M}$.

Lemma 24.12. *Let $\mathcal{M}$ be a differentiable manifold, and let $\mathbf{p} \in \mathcal{M}$. If $\mathcal{W}$ is an open neighborhood of $\mathbf{p}$, then there exist a function $k \in \mathfrak{F}(\mathcal{M})$ and open sets $\mathcal{P}$ and $\mathcal{Q}$ such that $\mathbf{p} \in \mathcal{P} \subset \mathcal{M} \setminus \mathcal{Q} \subset \mathcal{W}$ and*

(i) $0 \leqslant k(\mathbf{x}) \leqslant 1$ *for* $\mathbf{x} \in \mathcal{M}$;

(ii) $k(\mathbf{x}) = 1$ *for* $\mathbf{x} \in \mathcal{P}$;

(iii) $k(\mathbf{x}) = 0$ *for* $\mathbf{x} \in \mathcal{Q}$.

Proof. For $c > 0$ let

$$\mathcal{B}_c = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n \;\middle|\; \sum_{j=1}^{n} p_j^2 < c \right\}.$$

There exist real numbers a, b with $0 < a < b$ and patches $(\mathbf{x}, \mathcal{B}_a)$, $(\mathbf{x}, \mathcal{B}_b)$ in the (completion of the) atlas of $\mathcal{M}$ such that

$$\mathbf{p} \in \mathbf{x}(\mathcal{B}_a) \subset \mathbf{x}(\mathcal{B}_b) \subset \overline{\mathbf{x}(\mathcal{B}_b)} \subset \mathcal{W},$$

where $\overline{\mathbf{x}(\mathcal{B}_b)}$ is the closure of $\mathbf{x}(\mathcal{B}_b)$. Set $\mathcal{P} = \mathbf{x}(\mathcal{B}_a)$ and $\mathcal{Q} = \mathcal{M} \setminus \overline{\mathbf{x}(\mathcal{B}_b)}$.

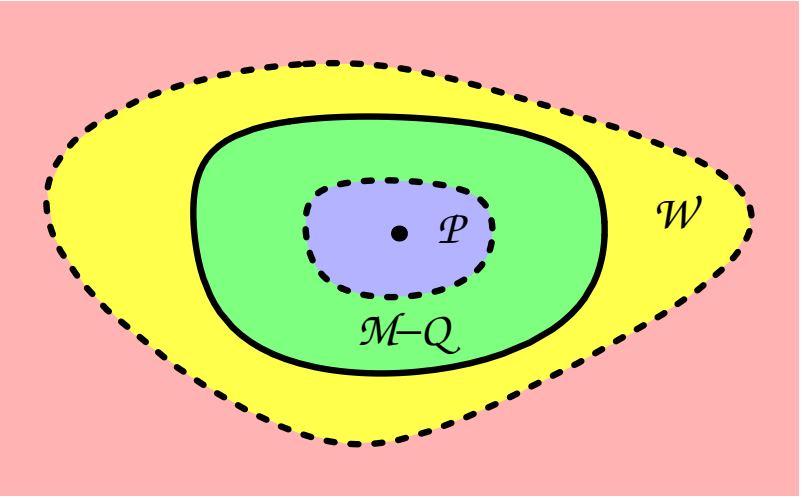


Figure 24.1: Neighborhoods of $\mathbf{p}$ and $\mathcal{M} \setminus \mathcal{W}$

Define $g\colon \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} \exp\left(\dfrac{1}{x-b} - \dfrac{1}{x-a}\right) & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Then g is a C^∞ function (but not a C^ω one). This fact is obvious at all points except a and b. It is clear at least that g is continuous at a and b. Moreover, it follows from L'Hôpital's rule that

$$\begin{aligned} \lim_{x \downarrow a} \frac{g(x) - g(a)}{x - a} &= \lim_{x \downarrow a} \frac{\exp\left(\dfrac{1}{x-b} - \dfrac{1}{x-a}\right)}{x - a} \\ &= \exp\left(\frac{1}{a-b}\right) \lim_{x \downarrow a} \frac{\exp\left(-\dfrac{1}{x-a}\right)}{x - a} = \exp\left(\frac{1}{a-b}\right) \lim_{y \downarrow 0} \frac{y}{e^{\frac{1}{y}}} \\ &= \exp\left(\frac{1}{a-b}\right) \lim_{y \downarrow 0} \frac{\dfrac{-1}{y^2}}{\dfrac{-1}{y^2} e^{\frac{1}{y}}} = \exp\left(\frac{1}{a-b}\right) \lim_{y \downarrow 0} e^{-\frac{1}{y}} = 0. \end{aligned}$$

In a similar fashion, it can be shown that all other derivatives of g exist and vanish at both a and b. The functions $G\colon \mathbb{R} \to \mathbb{R}$ and $\psi\colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$G(x) = \frac{\displaystyle\int_x^b g(t)\,dt}{\displaystyle\int_a^b g(t)\,dt}, \qquad \psi(p_1, \ldots, p_n) = G\big(p_1^2 + \cdots + p_n^2\big)$$

are also differentiable. We define $k\colon \mathcal{M} \to \mathbb{R}$ by

$$k(\mathbf{q}) = \begin{cases} \psi\big(\mathbf{x}^{-1}(\mathbf{q})\big), & \text{for } \mathbf{q} \in \mathbf{x}(\mathcal{B}_b), \\ 0, & \text{otherwise.} \end{cases}$$

Then k has all the required properties. ▌

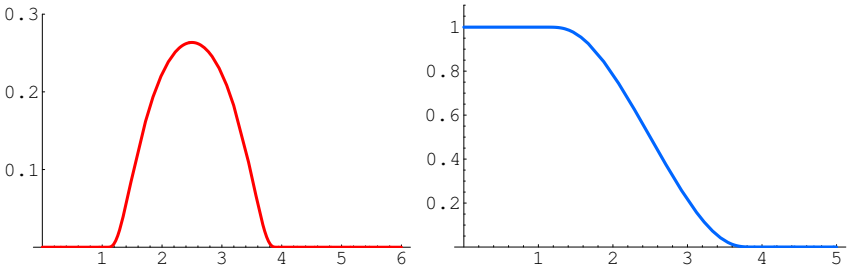


Figure 24.2: The functions $g(x)$ and $G(x)$ with $a = 1$, $b = 4$

If f is a real-valued differentiable function on an open subset $\mathcal{W}$ of a differentiable manifold $\mathcal{M}$, it might not be possible to extend f to a differentiable function defined on all of $\mathcal{M}$, as there might be some bizarre behavior of f at the boundary of $\mathcal{W}$. However, we now show that each point of $\mathcal{W}$ has a neighborhood $\mathcal{P}$ such that we can extend the *restriction* $f|\mathcal{P}$ to all of $\mathcal{M}$.

Lemma 24.13. *Let $\mathcal{W} \subset \mathcal{M}$ be an open neighborhood of $\mathbf{p} \in \mathcal{M}$, and suppose that $f \in \mathfrak{F}(\mathcal{W})$. Then there exist $\widetilde{f} \in \mathfrak{F}(\mathcal{M})$ and an open set $\mathcal{P}$ with $\mathbf{p} \in \mathcal{P} \subseteq \mathcal{W}$ such that $\widetilde{f}|\mathcal{P} = f|\mathcal{P}$. We call $\widetilde{f}$ a* ***globalization*** *of f.*

Proof. By Lemma 24.12 there are neighborhoods $\mathcal{P}$ of $\mathbf{p}$ and $\mathcal{Q}$ of $\mathcal{M} \setminus \mathcal{W}$ with the properties:

(i) $\mathcal{P} \subseteq \mathcal{W}$;

(ii) there is a function $k \in \mathfrak{F}(\mathcal{M})$ such that $k(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathcal{P}$ and $k(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathcal{Q}$.

Define $\widetilde{f}\colon \mathcal{M} \to \mathbb{R}$ by

$$\widetilde{f}(\mathbf{q}) = \begin{cases} (kf)(\mathbf{q}) & \text{for } \mathbf{q} \in \mathcal{W}, \\ 0 & \text{for } \mathbf{q} \in \mathcal{M} \setminus \mathcal{W}. \end{cases}$$

Then $\widetilde{f} \in \mathfrak{F}(\mathcal{M})$, and on $\mathcal{P}$ we have $\widetilde{f} = kf = f$. ▌

The coordinate functions (24.1) are elements of $\mathfrak{F}(\mathbf{x}(\mathcal{U}))$. By applying Lemma 24.13 to the monomials $(x_i)^m$, we see that $\mathfrak{F}(\mathcal{M})$ is *infinite-dimensional*. By contrast, the analogs of Lemmas 24.12 and 24.13 are false for C^ω functions.

Having defined the notion of real-valued differentiable function on a differentiable manifold, we are ready to define what it means for a map between manifolds to be differentiable.

Definition 24.14. *Let $\mathcal{M},\mathcal{N}$ be differentiable manifolds, and let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a map. We say that Ψ is **differentiable** provided $\mathbf{y}^{-1} \circ \Psi \circ \mathbf{x}$ is differentiable for every patch $(\mathbf{x},\mathcal{U})$ in the atlas of $\mathcal{M}$ and every patch $(\mathbf{y},\mathcal{V})$ in the atlas of $\mathcal{N}$, where the compositions are defined. A **diffeomorphism** between manifolds $\mathcal{M}$ and $\mathcal{N}$ is a differentiable map $\Phi\colon \mathcal{M} \to \mathcal{N}$ that has a differentiable inverse $\Phi^{-1}\colon \mathcal{N} \to \mathcal{M}$. If such a map Φ exists, $\mathcal{M}$ and $\mathcal{N}$ are said to be **diffeomorphic**. A map $\Psi\colon \mathcal{M} \to \mathcal{N}$ is called a **local diffeomorphism** provided each $\mathbf{p} \in \mathcal{M}$ has a neighborhood $\mathcal{W}$ such that $\Psi|\mathcal{W}\colon \mathcal{W} \to \Psi(\mathcal{W})$ is a diffeomorphism.*

The following lemma is an easy consequence of the definitions and the fact that the corresponding lemma for $\mathbb{R}^n$ is known.

Lemma 24.15. *Suppose $\mathcal{M} \overset{\Phi}{\to} \mathcal{N} \overset{\Psi}{\to} \mathcal{P}$ are differentiable maps between differentiable manifolds. Then the composition $\Psi \circ \Phi\colon \mathcal{M} \to \mathcal{P}$ is differentiable. If Φ and Ψ are diffeomorphisms, then so is $\Psi \circ \Phi$ and*

$$(\Psi \circ \Phi)^{-1} = \Phi^{-1} \circ \Psi^{-1}.$$

The coordinates (24.1) are examples of differentiable functions $\mathbf{x}(\mathcal{U}) \to \mathbb{R}$. Moreover,

Lemma 24.16. *Let $\mathcal{M}$ be an n-dimensional differentiable manifold, and let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch. Write $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$. Then*

(i) *$\mathbf{x}$ is a differentiable mapping between the manifolds $\mathcal{U}$ and $\mathcal{M}$;*

(ii) *$\mathbf{x}^{-1}\colon \mathbf{x}(\mathcal{U}) \to \mathbb{R}^n$ is differentiable.*

Proof. Let $\mathbf{1}\colon \mathbb{R}^n \to \mathbb{R}^n$ be the identity map. For each patch $(\mathbf{y},\mathcal{V})$ in $\mathfrak{A}$, the maps $\mathbf{y}^{-1} \circ \mathbf{x} \circ \mathbf{1}$ and $\mathbf{1} \circ \mathbf{x}^{-1} \circ \mathbf{y}$ are $\mathbb{R}^n$-differentiable. By definition $\mathbf{x}$ and $\mathbf{x}^{-1}$, considered as maps between the manifolds $\mathcal{U}$ and $\mathcal{M}$, are differentiable. ▌

A differentiable map between manifolds induces a correspondence between the algebras of differentiable functions on each manifold.

Lemma 24.17. *Let $\Phi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable mapping between manifolds. Then $f \in \mathfrak{F}(\mathcal{N})$ implies $f \circ \Phi \in \mathfrak{F}(\mathcal{M})$.*

Proof. Let $(\mathbf{x},\mathcal{U})$ be a patch on $\mathcal{M}$ and $(\mathbf{y},\mathcal{V})$ a patch on $\mathcal{N}$. By hypothesis, $f \circ \mathbf{y}\colon \mathcal{V} \to \mathbb{R}$ and $\mathbf{y}^{-1} \circ \Phi \circ \mathbf{x}$ is differentiable. Hence the composition

$$f \circ \Phi \circ \mathbf{x} = (f \circ \mathbf{y}) \circ (\mathbf{y}^{-1} \circ \Phi \circ \mathbf{x})$$

is differentiable. Since this is true for every patch $(\mathbf{x},\mathcal{U})$ in the atlas of $\mathcal{M}$, it follows that $f \circ \Phi$ is differentiable. ∎

24.3 Tangent Vectors on Manifolds

By their very nature differentiable manifolds are "curved" spaces, and they can be very complicated objects to study. In comparison, a vector space is much simpler. But a vector space has a great deal of structure that facilitates the study of differentiable manifolds. In this section, we define the notion of tangent space to a differentiable manifold $\mathcal{M}$ at a point $\mathbf{p} \in \mathcal{M}$. It can be thought of as the best linear approximation to $\mathcal{M}$ at $\mathbf{p}$.

The notion of a vector tangent to a curve or surface in $\mathbb{R}^n$ is intuitively clear, as we have seen in Chapters 1, 7, and 10; we want to define a similar concept for an arbitrary differentiable manifold. However, if we try to generalize directly the notion of tangent vector, we face a difficulty: the elementary definition of tangent vector that we gave in Section 10.5 makes a tangent vector to a surface a tangent vector to $\mathbb{R}^n$. But a surface, or more generally an arbitrary manifold, is not *a priori* contained in any Euclidean space, so we need a definition of tangent vector that does not depend on any such assumption.

Working backwards, let us suppose we have a suitable definition of a vector $\mathbf{v_p}$ tangent to a manifold $\mathcal{M}$ at a point $\mathbf{p} \in \mathcal{M}$. Then, just as in elementary calculus, we can speak of the derivative $\mathbf{v_p}[f]$ of $f \in \mathfrak{F}(\mathcal{M})$ in the direction $\mathbf{v_p}$. Roughly speaking, the number $\mathbf{v_p}[f]$ is the ordinary derivative of f at $\mathbf{p}$ along a curve leaving $\mathbf{p}$ in the direction $\mathbf{v_p}$. Now for a fixed $\mathbf{v_p}$, the function $\mathfrak{F}(\mathcal{M}) \to \mathbb{R}$ that maps f into $\mathbf{v_p}[f]$ has the essential properties always possessed by differentiation: linearity and the Leibniz[2] product rule. Thus, in standard mathematical fashion, we shall define a tangent vector to be a function that has



[2] Baron Gottfried Wilhelm Leibniz (1646–1716). German mathematician. A cofounder of calculus. Although Leibniz discovered calculus a few years later than Newton, it is the notation of Leibniz (such as dt and $\int$) that has gained the widest acceptance.

these properties. Although it is not clear from the outset that this definition will yield an object that has all the intuitive properties of a tangent vector, in fact, it does.

Definition 24.18. *Let* $\mathbf{p}$ *be a point of a manifold* $\mathcal{M}$*. A* ***tangent vector*** $\mathbf{v_p}$ *to* $\mathcal{M}$ *at* $\mathbf{p}$ *is a real-valued function* $\mathbf{v_p}\colon \mathfrak{F}(\mathcal{M}) \to \mathbb{R}$ *such that*

$$\begin{aligned} \mathbf{v_p}[a\,f + b\,g] &= a\,\mathbf{v_p}[f] + b\,\mathbf{v_p}[g] && \text{(the } \textbf{\textit{linearity}} \text{ property)},\\ \mathbf{v_p}[fg] &= f(\mathbf{p})\mathbf{v_p}[g] + g(\mathbf{p})\mathbf{v_p}[f] && \text{(the } \textbf{\textit{Leibnizian}} \text{ property)}, \end{aligned}$$

for all $a, b \in \mathbb{R}$ *and* $f, g \in \mathfrak{F}(\mathcal{M})$*.*

Clearly, a tangent vector to a surface in $\mathbb{R}^3$ as defined in Section 10.5 gives rise to a tangent vector in the sense of this definition.

We can easily give some nontrivial examples of tangent vectors. First, we need some new notation.

Definition 24.19. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *be a patch on a differentiable manifold* $\mathcal{M}$ *and write* $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$*. For* $f \in \mathfrak{F}(\mathcal{M})$ *and* $\mathbf{p} \in \mathbf{x}(\mathcal{U})$*, write* $\mathbf{p} = \mathbf{x}(\mathbf{q})$*, and define*

$$\frac{\partial f}{\partial x_i}(\mathbf{p}) = \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{p}} = \left.\frac{\partial (f \circ \mathbf{x})}{\partial u_i}\right|_{\mathbf{q}} \tag{24.3}$$

for $i = 1, \ldots, n$*. Here, as usual, the* u_i*'s are the natural coordinate functions of* $\mathbb{R}^n$*, and the ordinary Euclidean partial derivative appears on the right-hand side of* (24.3)*.*

Note that for $\mathcal{M} = \mathbb{R}^n$ and $\mathbf{x}$ the identity map, the right-hand side of (24.3) reduces to the ordinary partial derivative. In general, we can write

$$\frac{\partial f}{\partial x_i} = \frac{\partial (f \circ \mathbf{x})}{\partial u_i} \circ \mathbf{x}^{-1}.$$

Lemma 24.20. *Let* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *be a patch on a differentiable manifold* $\mathcal{M}$ *and write* $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$*. For* $\mathbf{p} \in \mathbf{x}(\mathcal{U})$*, and for each* $i = 1, \ldots, n$*, the function*

$$\left.\frac{\partial}{\partial x_i}\right|_{\mathbf{p}} : \mathfrak{F}(\mathcal{M}) \longrightarrow \mathbb{R}$$

defined by

$$\left.\frac{\partial}{\partial x_i}\right|_{\mathbf{p}} [f] = \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{p}}$$

is a tangent vector to $\mathcal{M}$ *at* $\mathbf{p}$*.*

Proof. Put $\mathbf{p} = \mathbf{x}(\mathbf{q})$, where $\mathbf{q} \in \mathcal{U}$. Let $a, b \in \mathbb{R}$ and $f, g \in \mathfrak{F}(\mathcal{M})$. We have

$$\begin{aligned}\frac{\partial(af+bg)}{\partial x_i}\bigg|_{\mathbf{p}} &= \frac{\partial}{\partial u_i}\big((af+bg)\circ\mathbf{x}\big)\bigg|_{\mathbf{q}} = \frac{\partial}{\partial u_i}\big(a(f\circ\mathbf{x})+b(g\circ\mathbf{x})\big)\bigg|_{\mathbf{q}} \\ &= a\frac{\partial(f\circ\mathbf{x})}{\partial u_i}\bigg|_{\mathbf{q}} + b\frac{\partial(g\circ\mathbf{x})}{\partial u_i}\bigg|_{\mathbf{q}} = a\frac{\partial f}{\partial x_i}\bigg|_{\mathbf{p}} + b\frac{\partial g}{\partial x_i}\bigg|_{\mathbf{p}}.\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial(fg)}{\partial x_i}\bigg|_{\mathbf{p}} &= \frac{\partial}{\partial u_i}\big((fg)\circ\mathbf{x}\big)\bigg|_{\mathbf{q}} = \frac{\partial}{\partial u_i}\big((f\circ\mathbf{x})(g\circ\mathbf{x})\big)\bigg|_{\mathbf{q}} \\ &= (f\circ\mathbf{x})(\mathbf{q})\frac{\partial(g\circ\mathbf{x})}{\partial u_i}\bigg|_{\mathbf{q}} + (g\circ\mathbf{x})(\mathbf{q})\frac{\partial(f\circ\mathbf{x})}{\partial u_i}\bigg|_{\mathbf{q}} \\ &= f(\mathbf{p})\frac{\partial g}{\partial x_i}\bigg|_{\mathbf{p}} + g(\mathbf{p})\frac{\partial f}{\partial x_i}\bigg|_{\mathbf{p}}. \blacksquare\end{aligned}$$

Now that we know that there exist some tangent vectors at every point $\mathbf{p}$, let us consider the set of *all* tangent vectors at $\mathbf{p}$.

Definition 24.21. *Let $\mathcal{M}$ be a differentiable manifold, and let $\mathbf{p}$ be a point in $\mathcal{M}$. The* ***tangent space*** *to $\mathcal{M}$ at $\mathbf{p}$ is the set of all tangent vectors to $\mathcal{M}$ at $\mathbf{p}$, and is denoted by $\mathcal{M}_{\mathbf{p}}$. Thus, $\mathcal{M}_{\mathbf{p}}$ equals the set of $\mathbf{v}_{\mathbf{p}}\colon \mathfrak{F}(\mathcal{M}) \to \mathbb{R}$ satisfying*

$$\mathbf{v}_{\mathbf{p}}[af+bg] = a\,\mathbf{v}_{\mathbf{p}}[f] + b\,\mathbf{v}_{\mathbf{p}}[g] \quad \textit{and} \quad \mathbf{v}_{\mathbf{p}}[fg] = f(\mathbf{p})\mathbf{v}_{\mathbf{p}}[g] + g(\mathbf{p})\mathbf{v}_{\mathbf{p}}[f],$$

for $a, b \in \mathbb{R}$ and $f, g \in \mathfrak{F}(\mathcal{M})$.

Lemma 24.22. *If $\mathcal{M}$ is a differentiable manifold and $\mathbf{p}$ is a point in $\mathcal{M}$, then the tangent space $\mathcal{M}_{\mathbf{p}}$ is naturally a vector space.*

Proof. We make $\mathcal{M}_{\mathbf{p}}$ into a vector space over $\mathbb{R}$ as follows. Let $\mathbf{v}_{\mathbf{p}}, \mathbf{w}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$, $a \in \mathbb{R}$ and $f \in \mathfrak{F}(\mathcal{M})$. Then we define $\mathbf{v}_{\mathbf{p}} + \mathbf{w}_{\mathbf{p}}$ and $a\,\mathbf{v}_{\mathbf{p}}$ by

$$\begin{cases} (\mathbf{v}_{\mathbf{p}} + \mathbf{w}_{\mathbf{p}})[f] = \mathbf{v}_{\mathbf{p}}[f] + \mathbf{w}_{\mathbf{p}}[f], \\ (a\,\mathbf{v}_{\mathbf{p}})[f] = a\,\mathbf{v}_{\mathbf{p}}[f]. \end{cases}$$

It is easily verified by the same sort of proof as that of Lemma 24.20 that both $\mathbf{v}_{\mathbf{p}} + \mathbf{w}_{\mathbf{p}}$ and $a\,\mathbf{v}_{\mathbf{p}}$ belong to $\mathcal{M}_{\mathbf{p}}$. Thus, addition and scalar multiplication makes $\mathcal{M}_{\mathbf{p}}$ into a vector space. $\blacksquare$

We would now like to prove that $\mathcal{M}_{\mathbf{p}}$ is a finite-dimensional vector space whose dimension is the same as that of $\mathcal{M}$. Furthermore, we have

$$\left\{ \frac{\partial}{\partial x_1}\bigg|_{\mathbf{p}}, \cdots, \frac{\partial}{\partial x_n}\bigg|_{\mathbf{p}} \right\}$$

as a natural candidate for a basis of $\mathcal{M}_{\mathbf{p}}$. However, there are some technical difficulties to overcome. First, we need

Lemma 24.23. *Let $\mathcal{M}$ be a differentiable manifold and $\mathbf{p} \in \mathcal{M}$.*

(i) *If $f \in \mathfrak{F}(\mathcal{M})$ can be expressed as the product of two functions $g, h \in \mathfrak{F}(\mathcal{M})$ both of which vanish at $\mathbf{p}$, then $\mathbf{v}_{\mathbf{p}}[f] = 0$ for all $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$.*

(ii) *If c is a constant function on $\mathcal{M}$, then $\mathbf{v}_{\mathbf{p}}[c] = 0$ for all $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$.*

(iii) *Tangent vectors are **local**; that is, if $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$, then the number $\mathbf{v}_{\mathbf{p}}[f]$ for $f \in \mathfrak{F}(\mathcal{M})$ depends only on the values of f in a neighborhood of $\mathbf{p}$.*

(iv) *Let h be a differentiable function defined only on a neighborhood $\mathcal{P}$ of $\mathbf{p}$. If $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$, then for any globalization $\widetilde{h}$ of h the value $\mathbf{v}_{\mathbf{p}}[\widetilde{h}]$ is the same; we call this value $\mathbf{v}_{\mathbf{p}}[h]$.*

Proof. For (i) suppose that $f = g\,h$ where $g, h \in \mathfrak{F}(\mathcal{M})$ and $g(\mathbf{p}) = h(\mathbf{p}) = 0$. Then for $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$ we have by the Leibnizian property that

$$\mathbf{v}_{\mathbf{p}}[f] = \mathbf{v}_{\mathbf{p}}[g\,h] = g(\mathbf{p})\mathbf{v}_{\mathbf{p}}[h] + h(\mathbf{p})\mathbf{v}_{\mathbf{p}}[g] = 0.$$

Next, for (ii) we first observe that

$$\mathbf{v}_{\mathbf{p}}[1] = \mathbf{v}_{\mathbf{p}}[1^2] = 1 \cdot \mathbf{v}_{\mathbf{p}}[1] + 1 \cdot \mathbf{v}_{\mathbf{p}}[1] = 2\,\mathbf{v}_{\mathbf{p}}[1],$$

so that $\mathbf{v}_{\mathbf{p}}[1] = 0$. For an arbitrary constant c we then have by the linearity property that

$$\mathbf{v}_{\mathbf{p}}[c] = \mathbf{v}_{\mathbf{p}}[c \cdot 1] = c\,\mathbf{v}_{\mathbf{p}}[1] = 0.$$

Since parts (iii) and (iv) are rephrasings of the same statement, we prove (iv). Let $\widetilde{h}$ and $\widehat{h}$ be globalizations of h which agree with h on a neighborhood $\mathcal{W}$ of $\mathbf{p}$, and let $g = \widetilde{h} - \widehat{h}$. Then g vanishes on $\mathcal{W}$. By Lemma 24.12, there exist a neighborhood $\mathcal{P} \subset \mathcal{W}$ of $\mathbf{p}$ and a function $k \in \mathfrak{F}(\mathcal{M})$ such that $k(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathcal{P}$ and $k(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathcal{M} \setminus \mathcal{W}$ (we have replaced k on page 815 by $1 - k$). Then $g = g\,k$ and $g(\mathbf{p}) = k(\mathbf{p}) = 0$. Hence by part (i) we have $0 = \mathbf{v}_{\mathbf{p}}[g] = \mathbf{v}_{\mathbf{p}}[\widetilde{h}] - \mathbf{v}_{\mathbf{p}}[\widehat{h}]$, so that $\mathbf{v}_{\mathbf{p}}[\widetilde{h}] = \mathbf{v}_{\mathbf{p}}[\widehat{h}]$. ∎

Next, we need a fact about $\mathfrak{F}(\mathbb{R}^n)$.

Lemma 24.24. *Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and suppose that $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Then there exist differentiable functions $g_i\colon \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, n$ such that*

$$g = \sum_{i=1}^{n} (u_i - a_i) g_i + g(\mathbf{a}). \tag{24.4}$$

Proof. Fix $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n$, and consider the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(s) = g\big(st_1 + (1-s)a_1,\ \ldots\ , st_n + (1-s)a_n\big) = g\big(s\mathbf{t} + (1-s)\mathbf{a}\big).$$

Since g is differentiable, so is f, and

$$f'(s) = \sum_{i=1}^{n} (t_i - a_i) \frac{\partial g}{\partial u_i}\big(s\mathbf{t} + (1-s)\mathbf{a}\big).$$

Now

$$\int_0^1 f'(s)\,ds = f(1) - f(0) = g(\mathbf{t}) - g(\mathbf{a}),$$

and so the fundamental theorem of calculus implies that

$$\begin{aligned} g(\mathbf{t}) - g(\mathbf{a}) \ &= \ \int_0^1 f'(s)\,ds \\ &= \ \sum_{i=1}^{n} (t_i - a_i) \int_0^1 \frac{\partial g}{\partial u_i}\big(s\mathbf{t} + (1-s)\mathbf{a}\big)\,ds \\ &= \ \sum_{i=1}^{n} \big(u_i(\mathbf{t}) - a_i\big) g_i(\mathbf{t}), \end{aligned}$$

where

$$g_i(\mathbf{t}) = \int_0^1 \frac{\partial g}{\partial u_i}\big(s\mathbf{t} + (1-s)\mathbf{a}\big)\,ds.$$

Clearly, the g_i's are differentiable. Thus we get (24.4). ∎

Note that the proof of this lemma fails for C^k functions. If g is a C^k function, we can only say that each g_i is a C^{k-1} function.

We are now able to prove:

Theorem 24.25. (Basis Theorem) *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on a differentiable manifold $\mathcal{M}$ with $\mathbf{p} \in \mathbf{x}(\mathcal{U})$ and write $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$. If $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$, then*

$$\mathbf{v_p} = \sum_{i=1}^{n} \mathbf{v_p}[x_i] \left.\frac{\partial}{\partial x_i}\right|_\mathbf{p}.$$

Proof. It suffices to show that

$$\mathbf{v_p}[f] = \sum_{i=1}^{n} \mathbf{v_p}[x_i] \left.\frac{\partial f}{\partial x_i}\right|_\mathbf{p} \tag{24.5}$$

for all $f \in \mathfrak{F}(\mathcal{M})$. To this end, let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a globalization of $f \circ \mathbf{x}$ near $\mathbf{x}^{-1}(\mathbf{p})$. By Lemma 24.24, there exist differentiable functions $g_i\colon \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, n$ such that

$$g = \sum_{i=1}^{n} \big(u_i - u_i(\mathbf{x}^{-1}(\mathbf{p}))\big) g_i + g\big(\mathbf{x}^{-1}(\mathbf{p})\big).$$

Then near $\mathbf{p}$ we have

$$f = \sum_{i=1}^{n} \big(x_i - x_i(\mathbf{p})\big) f_i + f(\mathbf{p}), \tag{24.6}$$

where $f_i = g_i \circ \mathbf{x}^{-1}$. Lemma 24.23 (iii) allows us to use (24.6) to compute $\mathbf{v}_\mathbf{p}[f]$. It follows that

$$\begin{aligned} \mathbf{v}_\mathbf{p}[f] &= \sum_{i=1}^{n} \Big\{ f_i(\mathbf{p}) \mathbf{v}_\mathbf{p}[x_i - x_i(\mathbf{p})] + \big(x_i(\mathbf{p}) - x_i(\mathbf{p})\big) \mathbf{v}_\mathbf{p}[f_i] \Big\} + \mathbf{v}_\mathbf{p}[f(\mathbf{p})] \\ &= \sum_{i=1}^{n} f_i(\mathbf{p}) \mathbf{v}_\mathbf{p}[x_i], \end{aligned}$$

because $x_i(\mathbf{p})$ and $f(\mathbf{p})$ are constants. For the same reasons, from (24.6) it follows that

$$\left.\frac{\partial f}{\partial x_j}\right|_\mathbf{p} = \sum_{i=1}^{n} f_i(\mathbf{p}) \frac{\partial x_i}{\partial x_j}(\mathbf{p}) = f_j(\mathbf{p}),$$

because[3]

$$\frac{\partial x_i}{\partial x_j}(\mathbf{p}) = \frac{\partial (x_i \circ \mathbf{x})}{\partial u_j}\big(\mathbf{x}^{-1}(\mathbf{p})\big) = \frac{\partial u_i}{\partial u_j}\big(\mathbf{x}^{-1}(\mathbf{p})\big) = \delta_{ij}.$$

Hence

$$\left(\sum_{j=1}^{n} \mathbf{v}_\mathbf{p}[x_j] \frac{\partial}{\partial x_j} \right)\Bigg|_\mathbf{p} [f] = \sum_{j=1}^{n} \mathbf{v}_\mathbf{p}[x_j] \left.\frac{\partial f}{\partial x_j}\right|_\mathbf{p} = \sum_{j=1}^{n} \mathbf{v}_\mathbf{p}[x_j] f_j(\mathbf{p}) = \mathbf{v}_\mathbf{p}[f].$$

Thus we get (24.5). ▌

We have seen in Lemmas 24.12 and 24.13 that to show the existence of nonconstant differentiable functions on a differentiable manifold $\mathcal{M}$, it was necessary to use C^∞ functions and not C^ω functions. Furthermore, in the proofs of Lemma 24.24 and Theorem 24.25 it is necessary to use C^∞ functions and not C^k functions.

---

[3]Here δ_{ij} denotes the Kronecker delta function, defined by $\delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$

Corollary 24.26. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on a differentiable manifold $\mathcal{M}$. Let $\mathbf{p} \in \mathbf{x}(\mathcal{U})$ and write $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$. Then the vectors*

$$\left.\frac{\partial}{\partial x_1}\right|_{\mathbf{p}}, \ldots, \left.\frac{\partial}{\partial x_n}\right|_{\mathbf{p}}$$

form a basis for the tangent space $\mathcal{M}_{\mathbf{p}}$. Hence the dimension of each tangent space $\mathcal{M}_{\mathbf{p}}$ as a vector space is the same as the dimension of $\mathcal{M}$ as a manifold.

Proof. The given vectors span the tangent space $\mathcal{M}_{\mathbf{p}}$ by Theorem 24.25. To prove they are linearly independent, suppose that $a_1, \ldots, a_n \in \mathbb{R}$ are such that

$$\sum_{i=1}^{n} a_i \left.\frac{\partial}{\partial x_i}\right|_{\mathbf{p}} = 0.$$

Then

$$0 = \sum_{i=1}^{n} a_i \left.\frac{\partial x_j}{\partial x_i}\right|_{\mathbf{p}} = \sum_{i=1}^{n} a_i \frac{\partial (x_j \circ \mathbf{x})}{\partial u_i}\big(\mathbf{x}^{-1}(\mathbf{p})\big) = \sum_{i=1}^{n} a_i \delta_{ij} = a_j$$

for $j = 1, \ldots, n$. ∎

In order to differentiate functions on manifolds (that is, apply tangent vectors to them) as easily as we would differentiate functions on $\mathbb{R}^n$, we shall need the following result that uses the same notation as Lemma 9.6 on page 267.

Lemma 24.27. (The Chain Rule) *Let $\mathcal{M}$ be a differentiable manifold. Suppose $g_1, \ldots, g_k \in \mathfrak{F}(\mathcal{M})$ and $h \in \mathfrak{F}(\mathbb{R}^k)$. Let $f = h \circ (g_1, \ldots, g_k)$. Then for $\mathbf{v}_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}$ we have*

$$\mathbf{v}_{\mathbf{p}}[f] = \sum_{i=1}^{k} \frac{\partial h}{\partial u_i}\big(g_1(\mathbf{p}), \ldots, g_k(\mathbf{p})\big)\mathbf{v}_{\mathbf{p}}[g_i]. \tag{24.7}$$

Proof. Let $g = (g_1, \ldots, g_k)$. Define $\mathbf{w}\colon \mathfrak{F}(\mathbb{R}^k) \to \mathbb{R}$ by

$$\mathbf{w}[h] = \mathbf{v}_{\mathbf{p}}[h \circ g]$$

for $h \in \mathfrak{F}(\mathbb{R}^k)$. Then it is easy to check from the definition that $\mathbf{w}$ is an element of the tangent space $\mathbb{R}^k_{g(\mathbf{p})}$, because it is linear and Leibnizian. Using Theorem 24.25 with the patch $(u_1, \ldots, u_k)$ at $g(\mathbf{p})$ we get

$$\mathbf{w}[h] = \sum_{i=1}^{k} \mathbf{w}[u_i] \frac{\partial h}{\partial u_i}\big(g(\mathbf{p})\big). \tag{24.8}$$

Now $\mathbf{w}[h] = \mathbf{v}_{\mathbf{p}}[h \circ g] = \mathbf{v}_{\mathbf{p}}[f]$ and $\mathbf{w}[u_i] = \mathbf{v}_{\mathbf{p}}[u_i \circ g] = \mathbf{v}_{\mathbf{p}}[g_i]$, so that (24.8) becomes

$$\mathbf{v}_{\mathbf{p}}[f] = \sum_{i=1}^{k} \frac{\partial h}{\partial u_i}\big(g(\mathbf{p})\big)\mathbf{v}_{\mathbf{p}}[g_i]. \tag{24.9}$$

But (24.9) is another way of writing (24.7). ∎

Note that (24.7) allows us to transfer standard differentiation formulas from $\mathbb{R}^n$ to manifolds. For example, if $\mathbf{v_p}$ is a tangent vector to a manifold $\mathcal{M}$ at $\mathbf{p}$ and $f \in \mathfrak{F}(\mathcal{M})$, then $\mathbf{v_p}[\sin f] = \cos(f(\mathbf{p}))\mathbf{v_p}[f]$.

For abstract surface theory, it will be useful to have alternative notation for tangent vectors.

Definition 24.28. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on a differentiable manifold $\mathcal{M}$, and let $\mathbf{q} \in \mathcal{U}$. Then*

$$\mathbf{x}_{u_i}(\mathbf{q})\colon \mathfrak{F}(\mathcal{M}) \longrightarrow \mathbb{R}$$

is the operator given by

$$\mathbf{x}_{u_i}(\mathbf{q})[f] = \left.\frac{\partial(f \circ \mathbf{x})}{\partial u_i}\right|_{\mathbf{q}} \tag{24.10}$$

for $\mathbf{q} \in \mathcal{U}$ and $f \in \mathfrak{F}(\mathcal{M})$.

The point is that (24.10) succeeds in generalizing the derivatives $\mathbf{x}_u, \mathbf{x}_v$ defined by (10.3) on page 289 to the more abstract setting of this chapter.

24.4 Induced Maps

In the previous section, we showed that to each point $\mathbf{p}$ of a differentiable manifold $\mathcal{M}$ there is associated a vector space called the tangent space $\mathcal{M}_\mathbf{p}$. In the present section we shall show how a differentiable map $\Psi\colon \mathcal{M} \to \mathcal{N}$ between differentiable manifolds $\mathcal{M}$ and $\mathcal{N}$ gives rise to a linear map between tangent spaces, in complete analogy to Definition 9.9 on page 268. Just as the tangent space is the best linear approximation of a differentiable manifold, the tangent map is the best linear approximation to a differentiable map between manifolds.

Definition 24.29. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map between differentiable manifolds $\mathcal{M}$ and $\mathcal{N}$, and let $\mathbf{p} \in \mathcal{M}$. The* ***tangent map*** *of Ψ at $\mathbf{p}$ is the map*

$$\Psi_{*\mathbf{p}}\colon \mathcal{M}_\mathbf{p} \longrightarrow \mathcal{N}_{\Psi(\mathbf{p})}$$

given by

$$\Psi_{*\mathbf{p}}(\mathbf{v_p})[f] = \mathbf{v_p}[f \circ \Psi] \tag{24.11}$$

for each $f \in \mathfrak{F}(\mathcal{N})$ and $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$.

In order for this definition to make sense, we must be sure that the image of $\Psi_{*\mathbf{p}}$ is actually contained in the tangent space $\mathcal{N}_{\Psi(\mathbf{p})}$.

Lemma 24.30. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map, and let $\mathbf{p} \in \mathcal{M}$, $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$. Define*

$$\Psi_{*\mathbf{p}}(\mathbf{v_p})\colon \mathfrak{F}(\mathcal{N}) \longrightarrow \mathbb{R}$$

by (24.11). *Then $\Psi_{*\mathbf{p}}(\mathbf{v_p}) \in \mathcal{N}_{\Psi(\mathbf{p})}$.*

Proof. For example, we show that $\Psi_{*\mathbf{p}}(\mathbf{v_p})$ is Leibnizian. Using the fact that $\mathbf{v_p}$ is Leibnizian, we have

$$\begin{aligned}\Psi_{*\mathbf{p}}(\mathbf{v_p})[fg] &= \mathbf{v_p}[(fg)\circ\Psi] = \mathbf{v_p}\big[(f\circ\Psi)(g\circ\Psi)\big]\\ &= (f\circ\Psi)(\mathbf{p})\mathbf{v_p}[g\circ\Psi] + (g\circ\Psi)(\mathbf{p})\mathbf{v_p}[f\circ\Psi]\\ &= f\big(\Psi(\mathbf{p})\big)\Psi_{*\mathbf{p}}(\mathbf{v_p})[g] + g\big(\Psi(\mathbf{p})\big)\Psi_{*\mathbf{p}}(\mathbf{v_p})[f]. \ \blacksquare\end{aligned}$$

The next results establish properties of tangent maps.

Lemma 24.31. *Suppose that $\mathcal{M}$ is an n-dimensional differentiable manifold and that $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. Write $\mathbf{x}^{-1} = (x_1, \dots, x_n)$. Then*

$$\mathbf{x}_{*\mathbf{q}}\left(\left.\frac{\partial}{\partial u_i}\right|_{\mathbf{q}}\right) = \mathbf{x}_{u_i}(\mathbf{q}) = \left.\frac{\partial}{\partial x_i}\right|_{\mathbf{x}(\mathbf{q})} \tag{24.12}$$

for $\mathbf{q} \in \mathcal{U}$ and $i = 1, \dots, n$.

Proof. For $f \in \mathfrak{F}(\mathcal{M})$ we have by (24.3) that

$$\mathbf{x}_{*\mathbf{q}}\left(\left.\frac{\partial}{\partial u_i}\right|_{\mathbf{q}}\right)[f] = \left.\frac{\partial}{\partial u_i}\right|_{\mathbf{q}}[f\circ\mathbf{x}] = \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}(\mathbf{q})} = \left.\frac{\partial}{\partial x_i}\right|_{\mathbf{x}(\mathbf{q})}[f]. \ \blacksquare$$

The tangent map $\Psi_{*\mathbf{p}}$ is a kind of dual of the homomorphism $\mathfrak{F}(\mathcal{N}) \to \mathfrak{F}(\mathcal{M})$ defined by mapping a function f to $f \circ \Psi$.

Lemma 24.32. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map and $\mathbf{p} \in \mathcal{M}$.*

(i) *The map $\Psi_{*\mathbf{p}}\colon \mathcal{M}_{\mathbf{p}} \to \mathcal{N}_{\Psi(\mathbf{p})}$ is a linear map between vector spaces.*

(ii) *If $\mathcal{M} \xrightarrow{\Psi} \mathcal{N} \xrightarrow{\Phi} P$ are differentiable maps, then*

$$(\Phi\circ\Psi)_{*\mathbf{p}} = \Phi_{*\Psi(\mathbf{p})}\circ\Psi_{*\mathbf{p}}.$$

(iii) *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on $\mathcal{M}$ at $\mathbf{p}$ and $\mathbf{y}\colon \mathcal{V} \to \mathcal{N}$ a patch on $\mathcal{N}$ at $\Psi(\mathbf{p})$. Write $\mathbf{x}^{-1} = (x_1, \dots, x_m)$ and $\mathbf{y}^{-1} = (y_1, \dots, y_n)$, where $m = \dim\mathcal{M}$ and $n = \dim\mathcal{N}$. Then*

$$\Psi_{*\mathbf{p}}\left(\left.\frac{\partial}{\partial x_j}\right|_{\mathbf{p}}\right) = \sum_{i=1}^{n}\frac{\partial(y_i\circ\Psi)}{\partial x_j}(\mathbf{p})\left.\frac{\partial}{\partial y_i}\right|_{\Psi(\mathbf{p})} \tag{24.13}$$

for $j = 1, \dots, m$.

Proof. The proofs of (i) and (ii) are straightforward. To prove (iii), we use the basis theorem (Theorem 24.25). Since

$$\Psi_{*\mathbf{p}}\left(\left.\frac{\partial}{\partial x_j}\right|_{\mathbf{p}}\right) \in \mathcal{N}_{\Psi(\mathbf{p})},$$

and

$$\left\{\left.\frac{\partial}{\partial y_1}\right|_{\Psi(\mathbf{p})}, \ldots, \left.\frac{\partial}{\partial y_n}\right|_{\Psi(\mathbf{p})}\right\}$$

is a basis for $\mathcal{N}_{\Psi(\mathbf{p})}$, we can write

$$\Psi_{*\mathbf{p}}\left(\left.\frac{\partial}{\partial x_j}\right|_{\mathbf{p}}\right) = \sum_{i=1}^{n} a_{ij} \left.\frac{\partial}{\partial y_i}\right|_{\Psi(\mathbf{p})}. \tag{24.14}$$

To compute the a_{ij}'s, we apply both sides of (24.14) to y_k. We have

$$\left(\sum_{i=1}^{n} a_{ij} \left.\frac{\partial}{\partial y_i}\right|_{\Psi(\mathbf{p})}\right)[y_k] = a_{kj}$$

while

$$\Psi_{*\mathbf{p}}\left(\left.\frac{\partial}{\partial x_j}\right|_{\mathbf{p}}\right)[y_k] = \left(\left.\frac{\partial}{\partial x_j}\right|_{\mathbf{p}}\right)[y_k \circ \Psi] = \frac{\partial(y_k \circ \Psi)}{\partial x_j}(\mathbf{p}).$$

Hence

$$a_{ij} = \frac{\partial(y_i \circ \Psi)}{\partial x_j}(\mathbf{p}),$$

and we get (24.13). ∎

Definition 24.33. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map, where $\mathcal{M}, \mathcal{N}$ are differentiable manifolds of dimensions m, n respectively. Let $\mathbf{p} \in \mathcal{M}$, let $(\mathbf{x}, \mathcal{U})$ be a patch on $\mathcal{M}$ at $\mathbf{p}$ and $(\mathbf{y}, \mathcal{V})$ a patch on $\mathcal{N}$ at $\Psi(\mathbf{p})$. The* ***Jacobian matrix*** *$\mathscr{J}(\Psi)(\mathbf{p})$ of Ψ at $\mathbf{p}$ relative to $\mathbf{x}$ and $\mathbf{y}$ is the matrix of $\Psi_{*\mathbf{p}}$ relative to the bases*

$$\left\{\left.\frac{\partial}{\partial x_1}\right|_{\mathbf{p}}, \ldots, \left.\frac{\partial}{\partial x_m}\right|_{\mathbf{p}}\right\} \quad \textit{and} \quad \left\{\left.\frac{\partial}{\partial y_1}\right|_{\Psi(\mathbf{p})}, \ldots, \left.\frac{\partial}{\partial y_n}\right|_{\Psi(\mathbf{p})}\right\}.$$

Explicitly, $\mathscr{J}(\Psi)(\mathbf{p})$ is the matrix $\left(\dfrac{\partial(y_i \circ \Psi)}{\partial x_j}(\mathbf{p})\right)$.

We can now get the manifold formulation of the Inverse Function Theorem from multivariable calculus:

Theorem 24.34. (Inverse Function Theorem) *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map, and let $\mathbf{p} \in \mathcal{M}$. Suppose* $\dim \mathcal{M} = \dim \mathcal{N} = n$. *Then the following conditions are equivalent*:

(i) $\Psi_{*\mathbf{p}}$ *is a linear isomorphism*;

(ii) *the Jacobian matrix* $\mathscr{J}(\Psi)(\mathbf{p})$ *is invertible*;

(iii) *there exist neighborhoods* $\mathcal{P}$ *of* $\mathbf{p}$ *and* $\mathcal{Q}$ *of* $\Psi(\mathbf{p})$ *such that the restriction* $\Psi\,|\,\mathcal{P}\colon \mathcal{P} \to \mathcal{Q}$ *has a differentiable inverse* $(\Psi\,|\,\mathcal{P})^{-1}\colon \mathcal{Q} \to \mathcal{P}$.

Proof. That (i) implies (ii) is a standard fact of linear algebra. Also, it is easy to see from Lemma 24.32(ii) that (iii) implies (i).

To show that (ii) implies (iii), consider $\mathbf{y}^{-1} \circ \Psi \circ \mathbf{x}$, where $(\mathbf{x},\mathcal{U})$ is a patch on $\mathcal{M}$ with $\mathbf{p} \in \mathbf{x}(\mathcal{U})$ and $(\mathbf{y},\mathcal{V})$ is a patch on $\mathcal{N}$ with $\Psi(\mathbf{p}) \in \mathbf{y}(\mathcal{V})$. Without loss of generality, we can assume that $\mathbf{p} = \mathbf{x}(\mathbf{0})$ and $\Psi(\mathbf{p}) = \mathbf{y}(\mathbf{0})$ so that

$$(\mathbf{y}^{-1} \circ \Psi \circ \mathbf{x})(\mathbf{0}) = \mathbf{0}.$$

Write $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$ and $\mathbf{y}^{-1} = (y_1, \ldots, y_n)$. By assumption

$$\det\left(\frac{\partial(u_i \circ \mathbf{y}^{-1} \circ \Psi \circ \mathbf{x})}{\partial u_j}(\mathbf{0})\right) = \det\left(\frac{\partial(y_i \circ \Psi)}{\partial x_j}(\mathbf{p})\right) = \det \mathscr{J}(\Psi)(\mathbf{p}) \neq 0.$$

The Inverse Function Theorem for $\mathbb{R}^n$ implies that $\mathbf{y}^{-1} \circ \Psi \circ \mathbf{x}$ has a local inverse. Hence so does Ψ. ∎

Corollary 24.35. *Let* $(x_1, \ldots, x_n)$ *be a coordinate system for* $\mathcal{M}$ *at* $\mathbf{p}$*, where each* x_j *is defined on a neighborhood* $\mathcal{W}$ *of* $\mathbf{p}$*. Then functions* $f_1, \ldots, f_n \in \mathfrak{F}(\mathcal{W})$ *form a coordinate system for* $\mathcal{M}$ *at* $\mathbf{p}$ *if and only if*

$$\det\left(\frac{\partial f_i}{\partial x_j}\right) \neq 0$$

on $\mathcal{W}$.

Definition 24.36. *A differentiable map* $\Psi\colon \mathcal{M} \to \mathcal{N}$ *is called* ***regular****, provided* $\Psi_{*\mathbf{p}}$ *is a nonsingular linear transformation for each* $\mathbf{p} \in \mathcal{M}$.

Corollary 24.37. *A map* $\Psi\colon \mathcal{M} \to \mathcal{N}$ *is a local diffeomorphism if and only if each* $\Psi_{*\mathbf{p}}$ *is a linear isomorphism.*

Next, we discuss the notion of curve in a differentiable manifold using the concepts we have developed. In what follows I denotes an open interval of the real line $\mathbb{R}$ and d/du the natural coordinate vector field on it. For each $t \in I$ we have a canonical element of the tangent space $\mathbb{R}_t$ to $\mathbb{R}$ at the point t, namely

$$\left.\frac{d}{du}\right|_t.$$

Definition 24.38. *A **curve** on a differentiable manifold $\mathcal{M}$ is a differentiable function $\boldsymbol{\alpha}\colon I \to \mathcal{M}$. If $t \in I$ the **velocity vector** of $\boldsymbol{\alpha}$ at t is the tangent vector*

$$\boldsymbol{\alpha}'(t) = \boldsymbol{\alpha}_{*t}\left(\left.\frac{d}{du}\right|_t\right) \in \mathcal{M}_{\boldsymbol{\alpha}(t)}.$$

We do not exclude the possibility that the interval I is half-infinite or all of $\mathbb{R}$. One can usually arrange to have zero in the domain I of a curve $\boldsymbol{\alpha}$; it is a convenient reference point. A curve on a surface in $\mathbb{R}^3$ is also a curve in the sense of this definition and the two notions of velocity vector are equivalent. We shall say that a curve $\boldsymbol{\alpha}$ ***starts*** at $\mathbf{p} \in \mathcal{M}$ provided that $\boldsymbol{\alpha}(0) = \mathbf{p}$, and that $\boldsymbol{\alpha}$ has ***initial velocity*** $\mathbf{v_p} \in \mathcal{M}_\mathbf{p}$, provided $\boldsymbol{\alpha}'(0) = \mathbf{v_p}$. We leave the reader to reconcile the notation $\boldsymbol{\alpha}'(t)$ in Definition 24.38 with its usual meaning when $\boldsymbol{\alpha}$ is merely a differentiable mapping $\mathbb{R} \to \mathbb{R}$.

Lemma 24.39. *Let $\boldsymbol{\alpha}\colon I \to \mathcal{M}$ be a curve. Then the value of the velocity vector $\boldsymbol{\alpha}'(t)$ on a function $f \in \mathfrak{F}(\mathcal{M})$ is given by*

$$\boldsymbol{\alpha}'(t)[f] = \frac{d(f \circ \boldsymbol{\alpha})}{du}(t).$$

Proof. We have

$$\boldsymbol{\alpha}'(t)[f] = \boldsymbol{\alpha}_{*t}\left(\left.\frac{d}{du}\right|_t\right)[f] = \frac{d(f \circ \boldsymbol{\alpha})}{du}(t). \blacksquare$$

Lemma 24.39 makes rigorous the earlier suggestion that for a tangent vector $\mathbf{v_p}$ the number $\mathbf{v_p}[f]$ is the derivative of f in the direction $\mathbf{v_p}$. This is because for any curve $\boldsymbol{\alpha}$ with initial velocity $\mathbf{v_p}$ we have

$$\mathbf{v_p}[f] = \frac{d(f \circ \boldsymbol{\alpha})}{du}(0).$$

Lemma 24.40. *Let $\boldsymbol{\alpha}$ be a curve in $\mathcal{M}$, and let $(x_1, \ldots, x_n)$ be a coordinate system at $\boldsymbol{\alpha}(t) \in \mathcal{M}$. Then*

$$\boldsymbol{\alpha}'(t) = \sum_{i=1}^{n} \frac{d(x_i \circ \boldsymbol{\alpha})}{du}(t) \left.\frac{\partial}{\partial x_i}\right|_{\boldsymbol{\alpha}(t)}.$$

Proof. This is an immediate consequence of Theorem 24.25. $\blacksquare$

A definition of the acceleration vector of a curve in an abstract manifold requires a *connection* on the manifold, and is given in Section 25.1.

Next, we prove the generalization to differentiable manifolds of Lemma 1.9 on page 7. Although the following lemma can be proved by showing that it is a special case of Lemma 24.27, it is easier to prove it directly.

Lemma 24.41. (The Chain Rule for curves on a manifold) *Let* $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ *and* $h\colon (c,d) \to (a,b)$ *be differentiable. Put* $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$. *Then*

$$\boldsymbol{\beta}' = (\boldsymbol{\alpha}' \circ h)\, h'.$$

Proof. We have

$$\begin{aligned} \boldsymbol{\beta}'(t) \;&=\; \boldsymbol{\beta}_{*t}\left(\frac{d}{du}\Big|_t\right) \;=\; \boldsymbol{\alpha}_{*h(t)} \circ h_{*t}\left(\frac{d}{du}\Big|_t\right) \\ &=\; \boldsymbol{\alpha}_{*h(t)}\left(h'(t)\frac{d}{du}\Big|_{h(t)}\right) = h'(t)\,\boldsymbol{\alpha}'\big(h(t)\big). \end{aligned}$$

∎

24.5 Vector Fields on Manifolds

The notion of derivation of an algebra comes up frequently in differential geometry. Let us give the general definition and then specialize it to show that a vector field is a derivation.

Definition 24.42. *Let* A *be an algebra* (*which is not necessarily associative nor commutative*) *over a field* $\mathbb{F}$. *A* ***derivation*** *of* A *is a mapping* $D\colon A \to A$ *such that*

$$\begin{cases} D(af+bg) \;=\; aDf + bDg, \\ D(fg) \;=\; f(Dg) + (Df)g \end{cases}$$

for all $a, b \in \mathbb{F}$ *and* $f, g \in A$.

Shortly, we shall define a vector field as a derivation of a special sort. But first let us study derivations in general.

Definition 24.43. *If* D_1 *and* D_2 *are derivations of an algebra, then the* ***bracket*** $[D_1, D_2]$ *of* D_1 *and* D_2 *is defined by*

$$[D_1, D_2] = D_1D_2 - D_2D_1.$$

The proofs of the next two lemmas are easy.

Lemma 24.44. *If* D_1 *and* D_2 *are derivations of an algebra, then so are* $[D_1, D_2]$, $D_1 + D_2$ *and* aD_1 *for* $a \in \mathbb{F}$.

Lemma 24.45. *Let* D_1, D_2 *and* D_3 *be derivations of an algebra. Then*:

(i) $[D_1, D_2] = -[D_2, D_1]$.

(ii) *The* Jacobi identity *holds*:

$$\big[[D_1, D_2], D_3\big] + \big[[D_3, D_1], D_2\big] + \big[[D_2, D_3], D_1\big] = 0.$$

We sometimes abbreviate the Jacobi identity to

$$\mathop{\mathfrak{S}}_{1,2,3} \big[[D_1, D_2], D_3\big] = 0,$$

where $\mathfrak{S}$ denotes the cyclic sum. It is an important ingredient in

Definition 24.46. *Let $\mathbb{F}$ be a field. A* ***Lie algebra*** *over $\mathbb{F}$ is a vector space V over $\mathbb{F}$ with a bracket $[\ ,\ ]\colon V \times V \to V$ such that*

$$\begin{cases} [X, X] &= 0, \\ [aX + bY, Z] &= a[X, Z] + b[Y, Z], \\ [Z, aX + bY] &= a[Z, X] + b[Z, Y], \\ \mathop{\mathfrak{S}}\limits_{X,Y,Z} \big[X, [Y, Z]\big] &= 0 \end{cases}$$

for each $X, Y, Z \in V$ and $a, b \in \mathbb{F}$.

We remarked at the foot of page 773 that the set of $n \times n$ skew-symmetric matrices can be made into a Lie algebra. Here is an another class of examples:

Lemma 24.47. *Let A be a commutative algebra over a field $\mathbb{F}$ and denote by $\mathfrak{D}(A)$ the set of derivations of A. Then $\mathfrak{D}(A)$ is a module over A and a Lie algebra over $\mathbb{F}$.*

Derivations arise in many different contexts in differential geometry, and the one of immediate interest provides us with an efficient way to extend the definition of a vector field on $\mathbb{R}^n$, given in Chapter 9, to manifolds.

Definition 24.48. *A* ***vector field*** $\mathbf{X}$ *on a differentiable manifold $\mathcal{M}$ is a derivation of the algebra $\mathfrak{F}(\mathcal{M})$ of real-valued differentiable functions on $\mathcal{M}$.*

Thus a vector field $\mathbf{X}$ on $\mathcal{M}$ is a mapping $\mathbf{X}\colon \mathfrak{F}(\mathcal{M}) \to \mathfrak{F}(\mathcal{M})$ satisfying

$$\begin{aligned} \mathbf{X}[af + bg] &= a\mathbf{X}[f] + b\mathbf{X}[g] \qquad \text{(the linearity property)}, \\ \mathbf{X}[fg] &= f\mathbf{X}[g] + g\mathbf{X}[f] \qquad \text{(the Leibnizian property)}, \end{aligned}$$

for $a, b \in \mathbb{R}$ and $f, g \in \mathfrak{F}(\mathcal{M})$. Also, the ***bracket*** $[\mathbf{X}, \mathbf{Y}]$ of vector fields $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$[\mathbf{X}, \mathbf{Y}] = \mathbf{XY} - \mathbf{YX}.$$

Note that $\mathbf{XY}$ (defined by $\mathbf{XY}[f] = \mathbf{X}(\mathbf{Y}[f])$) is a well-defined operator on $\mathfrak{F}(\mathcal{M})$, but not a vector field because the Leibnizian property is not satisfied. We put

$$\mathfrak{X}(\mathcal{M}) = \{\, \mathbf{X} \mid \mathbf{X} \text{ is a vector field on } \mathcal{M} \,\}.$$

Lemma 24.47 tells us that $\mathfrak{X}(\mathcal{M})$ is a module over $\mathfrak{F}(\mathcal{M})$ and a Lie algebra over $\mathbb{R}$. In other words, we may multiply vector fields by either real numbers or real-valued functions. Moreover, we have a bracket "multiplication" between vector fields that yields another vector field in $\mathfrak{X}(\mathcal{M})$. The bracket is *not* multilinear with respect to functions — there is a more complicated rule that is straightforward to prove:

Lemma 24.49. *Let $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$ and $f, g \in \mathfrak{F}(\mathcal{M})$. Then*

$$[f\mathbf{X},\, g\mathbf{Y}] = fg[\mathbf{X}, \mathbf{Y}] + f\mathbf{X}[g]\,\mathbf{Y} - g\mathbf{Y}[f]\,\mathbf{X}. \tag{24.15}$$

The definition of vector field that we have given is very abstract. Probably a more intuitive notion of vector field is that of a function that assigns to each point $\mathbf{p}$ of a manifold $\mathcal{M}$ a tangent vector $\mathbf{X}_\mathbf{p}$. We now show that this intuitive notion is equivalent to the definition that we have given. First, we introduce

Definition 24.50. *The **tangent bundle** of a differentiable manifold $\mathcal{M}$ is the set*

$$T(\mathcal{M}) = \big\{\, (\mathbf{p}, \mathbf{v}_\mathbf{p}) \;\big|\; \mathbf{v}_\mathbf{p} \in \mathcal{M}_\mathbf{p},\ \ \mathbf{p} \in \mathcal{M} \,\big\}.$$

This object is both an important source of new manifolds, and a means of interpreting the definition of vector field. The next result is the subject of Exercise 9.

Theorem 24.51. *The tangent bundle $T(\mathcal{M})$ of a differentiable manifold $\mathcal{M}$ is naturally a differentiable manifold whose dimension is twice the dimension of $\mathcal{M}$. Furthermore, the **projection map** $\pi\colon T(\mathcal{M}) \to \mathcal{M}$ defined by $\pi(\mathbf{p}, \mathbf{v}_\mathbf{p}) = \mathbf{p}$ is a differentiable map.*

Lemma 24.52. *Associated with each vector field $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ is a **section** of $T(\mathcal{M})$, that is, a differentiable map $\mathbf{X}\colon \mathcal{M} \to T(\mathcal{M})$ such that $\pi \circ \mathbf{X} = \mathbf{1}_\mathcal{M}$, where $\mathbf{1}_\mathcal{M}$ denotes the identity map on $\mathcal{M}$. In particular, $\mathbf{X}$ is a map that associates a tangent vector $\mathbf{X}_\mathbf{p} \in \mathcal{M}_\mathbf{p}$ with each point $\mathbf{p} \in \mathcal{M}$.*

Conversely, each section of the tangent bundle gives rise in a natural way to an element of $\mathfrak{X}(\mathcal{M})$.

From now on we consider vector fields as derivations of $\mathfrak{F}(\mathcal{M})$, or as sections of $T(\mathcal{M})$, whichever seems more convenient. If $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$, we write $\mathbf{X}_\mathbf{p}$ for the tangent vector in $\mathcal{M}_\mathbf{p}$ determined by $\mathbf{X}$.

Definition 24.53. *Let $(x_1, \ldots, x_n)$ be a system of local coordinates for $\mathcal{M}$ at $\mathbf{p}$ defined on an open set $\mathcal{W} \subset \mathcal{M}$. Then*

$$\frac{\partial}{\partial x_i}\colon \mathfrak{F}(\mathcal{W}) \longrightarrow \mathfrak{F}(\mathcal{W})$$

is the operator that assigns to each function $f \in \mathfrak{F}(\mathcal{W})$ its partial derivative $\partial f/\partial x_i$.

The next lemma is an obvious consequence of Lemma 24.20.

Lemma 24.54. *For each $i = 1, \ldots, n$, and $\mathbf{p} \in \mathcal{W}$, we have*

$$\frac{\partial}{\partial x_i} \in \mathfrak{X}(\mathcal{W}) \qquad \text{and} \qquad \left.\frac{\partial}{\partial x_i}\right|_{\mathbf{p}} \in \mathcal{M}_{\mathbf{p}}.$$

We are now in a position to give a version of the Basis Theorem on page 823 for vector fields.

Corollary 24.55. *Let $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ and let $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$ be a system of local coordinates defined on an open set $\mathcal{W}$. Then $\mathbf{X}$ can be written as*

$$\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}[x_i] \frac{\partial}{\partial x_i} \tag{24.16}$$

on $\mathcal{W}$.

Proof. It follows from Theorem 24.25 that

$$\mathbf{X}_{\mathbf{p}} = \sum_{i=1}^{n} \mathbf{X}_{\mathbf{p}}[x_i] \left.\frac{\partial}{\partial x_i}\right|_{\mathbf{p}}. \tag{24.17}$$

Since (24.17) holds for all $\mathbf{p} \in \mathcal{W}$, we get (24.16). ∎

We now consider the effect of a differentiable map $\Psi\colon \mathcal{M} \to \mathcal{N}$ on vector fields. Unfortunately, it is not always possible to transfer vector fields on $\mathcal{M}$ to vector fields on $\mathcal{N}$. The problem is that if $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ and $\mathbf{p}, \mathbf{q} \in \mathcal{M}$ are points such that $\Psi(\mathbf{p}) = \Psi(\mathbf{q})$, we may have that

$$\Psi_{*\mathbf{p}}(\mathbf{X}_{\mathbf{p}}) \neq \Psi_{*\mathbf{q}}(\mathbf{X}_{\mathbf{q}}).$$

An example of the phenomenon occurs for the vector field $\mathbf{X}$ in $\mathbb{R}^3$ defined by

$$\mathbf{X}_{(x,y,z)} = (-y,\ x,\ 0),$$

illustrated in Figure 24.3. The projection $(x, y, z) \mapsto (y, z)$ to the yz-plane "muddles up" the vectors. By contrast, mapping (x, y, z) to (x, y) gives rise to an unambiguous vector field in the xy-plane, and corresponds to the favorable case formulated next.

Definition 24.56. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map, and let $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$, $\mathbf{Y} \in \mathfrak{X}(\mathcal{N})$. We say that $\mathbf{X}$ and $\mathbf{Y}$ are* ***Ψ-related*** *provided that*

$$\Psi_{*\mathbf{p}}(\mathbf{X}_{\mathbf{p}}) = \mathbf{Y}_{\Psi(\mathbf{p})}$$

for all $\mathbf{p} \in \mathcal{M}$.

We use the notation $\mathbf{X}^{\Psi} = \mathbf{Y}$ or $\Psi_*(\mathbf{X}) = \mathbf{Y}$, but beware that $\mathbf{Y}$ is not in general determined by $\mathbf{X}$ if Ψ is not surjective.

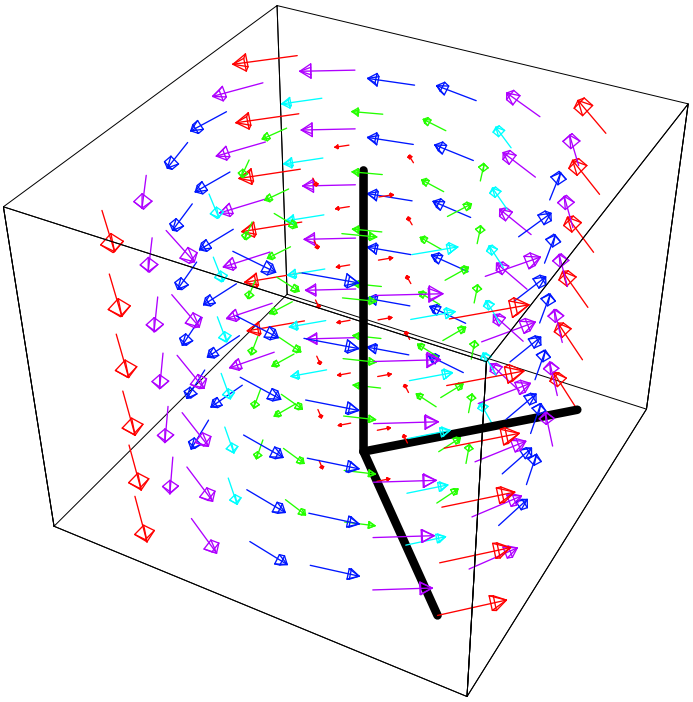


Figure 24.3: A swirling vector field in $\mathbb{R}^3$

The proof of the following results is straightforward.

Lemma 24.57. *Let $\Psi\colon \mathcal{M} \to \mathcal{N}$ be a differentiable map.*

(i) *A vector field $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ is Ψ-related to a vector field $\mathbf{Y} \in \mathfrak{X}(\mathcal{N})$ if and only if*

$$\mathbf{Y}[f] \circ \Psi = \mathbf{X}[f \circ \Psi]$$

for all $f \in \mathfrak{F}(\mathcal{N})$.

(ii) *Let $\mathbf{W}, \mathbf{X} \in \mathfrak{X}(\mathcal{M})$ and $\mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{N})$ with $\mathbf{W}^\Psi = \mathbf{Y}$ and $\mathbf{X}^\Psi = \mathbf{Z}$. Then $[\mathbf{W}, \mathbf{X}]$ is Ψ-related to $[\mathbf{Y}, \mathbf{Z}]$, so*

$$[\mathbf{W}^\Psi, \mathbf{X}^\Psi] = [\mathbf{Y}, \mathbf{Z}] = [\mathbf{W}, \mathbf{X}]^\Psi.$$

(iii) *If $\Psi\colon \mathcal{M} \to \mathcal{N}$ is a diffeomorphism, then for any $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ there exists a unique $\mathbf{Y} \in \mathfrak{X}(\mathcal{N})$ such that $\mathbf{X}^\Psi = \mathbf{Y}$.*

The vector-field version of Lemma 24.31 is

$$\mathbf{x}_*\Big(\frac{\partial}{\partial u_i}\Big) = \frac{\partial}{\partial x_i} = \mathbf{x}_{u_i} \circ \mathbf{x}^{-1}. \tag{24.18}$$

From it, we may deduce an important vanishing result:

Corollary 24.58. *Let $\mathcal{M}$ be an n-dimensional differentiable manifold and let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch, where $\mathcal{U}$ is an open subset of $\mathbb{R}^n$. Then*

$$\left[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right] = \left[\mathbf{x}_{u_i}, \mathbf{x}_{u_j}\right] = 0,$$

for $1 \leqslant i, j \leqslant n$.

Proof. We have

$$\left[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right] = \left[\mathbf{x}_*\left(\frac{\partial}{\partial u_i}\right), \mathbf{x}_*\left(\frac{\partial}{\partial u_j}\right)\right] = \mathbf{x}_*\left(\left[\frac{\partial}{\partial u_i}, \frac{\partial}{\partial u_j}\right]\right) = 0,$$

proving the second equation. ∎

24.6 Tensor Fields

It will also be necessary to consider *tensor fields*. Tensor analysis was developed by Ricci-Curbastro[4] (see [Ricci]). The modern invariant notation described below was introduced by Koszul [Kosz].

Instead of developing the most general situation, it will be sufficient for our purposes to treat two special cases.

Definition 24.59. *A **covariant tensor field** of degree r on a differentiable manifold $\mathcal{M}$ is a mapping*

$$\alpha\colon \underbrace{\mathfrak{X}(\mathcal{M}) \times \cdots \times \mathfrak{X}(\mathcal{M})}_{r \text{ times}} \longrightarrow \mathfrak{F}(\mathcal{M})$$

that satisfies

$$\begin{aligned}(24.19)\quad \alpha(\mathbf{X}_1, \ldots, f_i\mathbf{Y}_i + g_i\mathbf{Z}_i, \ldots, \mathbf{X}_r) \;=\; & f_i\,\alpha(\mathbf{X}_1, \ldots, \mathbf{Y}_i, \ldots, \mathbf{X}_r) \\ & + g_i\,\alpha(\mathbf{X}_1, \ldots, \mathbf{Z}_i, \ldots, \mathbf{X}_r)\end{aligned}$$

*for all $f_i, g_i \in \mathfrak{F}(\mathcal{M})$, $\mathbf{X}_1, \ldots, \mathbf{X}_r$, $\mathbf{Y}_i, \mathbf{Z}_i \in \mathfrak{X}(\mathcal{M})$, and for each index i in turn. A **vectorvariant tensor field** of degree r is a mapping*

$$\Phi\colon \underbrace{\mathfrak{X}(\mathcal{M}) \times \cdots \times \mathfrak{X}(\mathcal{M})}_{r \text{ times}} \longrightarrow \mathfrak{X}(\mathcal{M})$$

that satisfies (24.19).



4 Gregorio Ricci-Curbastro (1853–1925). Italian mathematician, professor at the University of Padua. He invented the absolute differential calculus between 1884 and 1894. It became the foundation of the tensor analysis that was used by Einstein in his theory of general relativity (see [HaEl]).

Remarks

(1) Condition (24.19) asserts that α and Φ are *multilinear with respect to functions*. On the other hand, the Lie bracket $[\ ,\ ]$ is not a tensor field because it is not multilinear with respect to functions; instead it satisfies the more complicated formula (24.15).

(2) Sometimes a covariant tensor field of degree r is said to be a tensor field of type $(r, 0)$, and a vectorvariant tensor field of degree r is said to be a tensor field of type $(r, 1)$. It is also possible to define tensor fields of type (r, s), but we ignore the case $s \geqslant 2$.

(3) There is actually a definition in linear algebra that is very similar to the definition of tensor field. Applying it to the tangent space at a point $\mathbf{p}$ to a differentiable manifold $\mathcal{M}$ gives

Definition 24.60. *A* ***covariant tensor*** *at* $\mathbf{p} \in \mathcal{M}$ *is a multilinear map*

$$\alpha\colon \underbrace{\mathcal{M}_\mathbf{p} \times \cdots \times \mathcal{M}_\mathbf{p}}_{r \text{ times}} \longrightarrow \mathbb{R}$$

that satisfies

$$\begin{aligned}\alpha(\mathbf{x}_1, \ldots, a_i\mathbf{y}_i + b_i\mathbf{z}_i, \ldots, \mathbf{x}_r) \;=\; & a_i\,\alpha(\mathbf{x}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{x}_r) \\ & + b_i\,\alpha(\mathbf{x}_1, \ldots, \mathbf{z}_i, \ldots, \mathbf{x}_r)\end{aligned}$$

for all $a_i, b_i \in \mathbb{R}$, *and all* $\mathbf{x}_1, \ldots, \mathbf{x}_r$, $\mathbf{y}_i, \mathbf{z}_i \in \mathcal{M}_\mathbf{p}$. *A* ***vectorvariant tensor*** *at* $\mathbf{x}$ *is defined similarly, by replacing* $\mathbb{R}$ *with* $\mathcal{M}_\mathbf{p}$.

On page 834 we noted that some differentiable mappings between manifolds $\mathcal{M}$ and $\mathcal{N}$ do not always induce a map between the algebras of vector fields of $\mathcal{M}$ and $\mathcal{N}$. The situation is much better with covariant tensor fields.

Definition 24.61. *Let* $\Psi\colon \mathcal{M} \to \mathcal{N}$ *be a differentiable mapping, and let* α *be a covariant tensor field of degree* r *on* $\mathcal{N}$. *Then the* ***pullback*** *of* α *is the covariant tensor field* $\Psi^*(\alpha)$ *on* $\mathcal{M}$ *given by*

$$\Psi^*(\alpha)(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \alpha\big(\Psi_{*p}(\mathbf{v}_1), \ldots, \Psi_{*p}(\mathbf{v}_r)\big),$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *are arbitrary tangent vectors at an arbitrary point* $\mathbf{p} \in \mathcal{M}$.

The following lemma is easy to prove.

Lemma 24.62. *Let* $\Psi\colon \mathcal{M} \to \mathcal{N}$ *be a differentiable mapping, and let* α *and* β *be covariant tensor fields of degree* r *on* $\mathcal{N}$. *Also, let* $f, g \in \mathfrak{F}(\mathcal{N})$. *Then*

$$\Psi^*(f\alpha + g\beta) = (f \circ \Psi)\,\Psi^*(\alpha) + (g \circ \Psi)\,\Psi^*(\beta).$$

Let us single out one of the simplest types of covariant tensor fields.

Definition 24.63. *A* **1-*form*** *on a manifold $\mathcal{M}$ is an $\mathfrak{F}(\mathcal{M})$-linear map*

$$\omega\colon \mathfrak{X}(\mathcal{M}) \longrightarrow \mathfrak{F}(\mathcal{M}).$$

In other words, a 1-form ω has the property that

$$\omega(f\,\mathbf{X} + g\,\mathbf{Y}) = f\,\omega(\mathbf{X}) + g\omega(\mathbf{Y})$$

for $f, g \in \mathfrak{F}(\mathcal{M})$ and $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$. We denote the collection of 1-forms on $\mathcal{M}$ by $\mathfrak{X}(\mathcal{M})^*$, and we make it into a module over $\mathfrak{F}(\mathcal{M})$ by defining

$$(f\,\omega + g\,\theta)(\mathbf{X}) = f\,\omega(\mathbf{X}) + g\,\theta(\mathbf{X})$$

for $f, g \in \mathfrak{F}(\mathcal{M})$, $\omega, \theta \in \mathfrak{X}(\mathcal{M})^*$ and $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$.

Lemma 24.64. *Suppose $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ and $\mathbf{p} \in \mathcal{M}$ are such that $\mathbf{X}_\mathbf{p} = 0$. Then $\omega(\mathbf{X})(\mathbf{p}) = 0$ for all $\omega \in \mathfrak{X}(\mathcal{M})^*$.*

Proof. Let $(\mathbf{x},\mathcal{U})$ be a patch on $\mathcal{M}$ at $\mathbf{p}$ with $\mathbf{x}^{-1} = (x_1, \ldots, x_n)$. Using Corollary 24.55, we have

$$\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}[x_i] \frac{\partial}{\partial x_i}$$

near $\mathbf{p}$. Also, $\mathbf{X}_\mathbf{p} = 0$ implies that $\mathbf{X}[x_i](\mathbf{p}) = 0$ for $i = 1, \ldots, n$. Hence,

$$\omega(\mathbf{X})(\mathbf{p}) = \omega\left(\sum_{i=1}^{n} \mathbf{X}[x_i] \frac{\partial}{\partial x_i}\right)(\mathbf{p}) = \sum_{i=1}^{n} \mathbf{X}[x_i](\mathbf{p})\, \omega\left(\frac{\partial}{\partial x_i}\right)(\mathbf{p}) = 0. \;\blacksquare$$

If V is any vector space over any field $\mathbb{F}$ (in our case, almost invariably $\mathbb{R}$), its dual space is denoted by V^*. It is the new vector space

$$V^* = \{\, \alpha\colon V \to \mathbb{F} \mid \alpha \text{ is linear} \,\}$$

consisting of linear functionals, or linear mappings from V to the field. We see that Definition 24.63 extends this concept to modules, so that $\mathfrak{X}(\mathcal{M})^*$ is the dual module of $\mathfrak{X}(\mathcal{M})$. On the other hand, we can evaluate 1-forms at a point in the same way that we evaluate vector fields at a point. If we do this, we obtain from ω an element of the ***cotangent space*** $(\mathcal{M}_\mathbf{p})^*$ to $\mathcal{M}$ at each fixed point $\mathbf{p} \in \mathcal{M}$, as follows.

Definition 24.65. *Let $\omega \in \mathfrak{X}(\mathcal{M})^*$ and $\mathbf{p} \in \mathcal{M}$. Then $\omega_\mathbf{p}\colon \mathcal{M}_\mathbf{p} \to \mathbb{R}$ is defined by*

$$\omega_\mathbf{p}(\mathbf{v}) = \omega(\mathbf{X})(\mathbf{p}),$$

where $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ is any vector field for which $\mathbf{X}_\mathbf{p} = \mathbf{v}$.

Lemma 24.66. *The definition of $\omega_{\mathbf{p}}$ does not depend on the choice of the vector field $\mathbf{X}$.*

Proof. Let $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ be vector fields such that $\mathbf{X}_{\mathbf{p}} = \widetilde{\mathbf{X}}_{\mathbf{p}} = \mathbf{v}$. Then $\mathbf{X} - \widetilde{\mathbf{X}}$ vanishes at $\mathbf{p}$ so by Lemma 24.64,

$$\omega(\mathbf{X})(\mathbf{p}) - \omega(\widetilde{\mathbf{X}})(\mathbf{p}) = \omega(\mathbf{X} - \widetilde{\mathbf{X}})(\mathbf{p}) = 0. \blacksquare$$

It is easy to see now that $\omega_{\mathbf{p}} \in (\mathcal{M}_{\mathbf{p}})^*$. The ***cotangent vector*** $\omega_{\mathbf{p}}$ is also called a "1-form," and only the context makes clear the precise meaning of this terminology. There is a particularly important kind of 1-form defined on $\mathcal{M}$:

Definition 24.67. *For $f \in \mathfrak{F}(\mathcal{M})$ the **differential** df of f is the 1-form defined by*

$$df(\mathbf{X}) = \mathbf{X}[f]$$

for $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$.

We conclude this chapter with

Lemma 24.68. *Let $(x_1, \ldots, x_n)$ be a system of coordinates defined on an open set $\mathcal{W} \subset \mathcal{M}$. Then*

(i) *For each $\mathbf{p} \in \mathcal{W}$, the 1-forms $dx_1(\mathbf{p}), \ldots, dx_n(\mathbf{p})$ form a basis of $(\mathcal{M}_{\mathbf{p}})^*$. Hence*

$$\{\, \omega_{\mathbf{p}} \mid \omega \in \mathfrak{X}(\mathcal{M})^* \,\} = (\mathcal{M}_{\mathbf{p}})^*.$$

(ii) *The set $\{dx_1, \ldots, dx_n\}$ is a basis for $\mathfrak{X}(\mathcal{W})^*$. Thus for any $\omega \in \mathfrak{X}(\mathcal{W})^*$, we can write*

$$\omega = \sum_{i=1}^{n} \omega\left(\frac{\partial}{\partial x_i}\right) dx_i.$$

24.7 Exercises

1. If $\mathcal{M}_1$ and $\mathcal{M}_2$ are differentiable manifolds, show that $\mathcal{M}_1 \times \mathcal{M}_2$ is a differentiable manifold.

2. Show that each tangent space $\mathcal{M}_{\mathbf{p}}$ to a differentiable manifold $\mathcal{M}$ is itself a differentiable manifold.

3. Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be defined by $\Phi(t) = t^3$. Show that the complete atlas $\widetilde{\mathfrak{A}}_2$ on $\mathbb{R}$ containing the patch $(\Phi, \mathbb{R})$ is different from the usual complete atlas $\widetilde{\mathfrak{A}}_1$ containing the identity patch $(\mathbf{1}, \mathbb{R})$. Show that $(\mathbb{R}, \widetilde{\mathfrak{A}}_1)$ and $(\mathbb{R}, \widetilde{\mathfrak{A}}_2)$ are nonetheless diffeomorphic.

4. Recall the definition of real projective space $\mathbb{RP}^n$ in Exercise 12 of the previous chapter, and define again $p\colon S^n(1) \mapsto \mathbb{RP}^n$ by $p(\mathbf{a}) = \{\mathbf{a}, -\mathbf{a}\}$. Let

$$\mathcal{P}_j = \{\, p(a_1, \ldots, a_n) \mid a_j \neq 0\}.$$

Show that:

(a) Each $\mathcal{P}_j$ is an open subset of $\mathbb{RP}^n$.

(b) $\mathbb{RP}^n$ is the union of all the $\mathcal{P}_j$'s.

(c) For each j, there is a homeomorphism $\psi_j\colon \mathcal{P}_j \to \mathbb{R}^n$.

(d) $\mathbb{RP}^n$ is a compact differentiable manifold of dimension n.

5. On $\mathbb{R}^3$ consider the vector fields

$$\mathbf{X} = x^2 \frac{\partial}{\partial x} + y \frac{\partial}{\partial z} \qquad \text{and} \qquad \mathbf{Y} = y^3 \frac{\partial}{\partial y},$$

and the function $f\colon \mathbb{R}^3 \to \mathbb{R}$ defined by $f(p_1, p_2, p_3) = {p_1}^2 p_2 p_3$. Compute

(a) $[\mathbf{X}, \mathbf{Y}]_{(1,0,1)}$

(b) $(f\,\mathbf{Y})_{(1,0,1)}$

(c) $(\mathbf{Y}f)(1, 0, 1)$

(d) $f_{*(1,0,1)}\big(\mathbf{Y}_{(1,0,1)}\big)$.

Give examples of vector fields $\mathbf{X}$ and $\mathbf{Y}$ on $\mathbb{R}^2$ for which

$$\mathbf{X}_{(0,0)} = \mathbf{Y}_{(0,0)} \qquad \text{but} \qquad [\mathbf{X}, \mathbf{Y}]_{(0,0)} \neq \mathbf{0}.$$

6. Let $\Phi\colon \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $\Phi(p_1, p_2, p_3) = (p_1, p_2)$, and let $\mathbf{Z}$ be the vector field on $\mathbb{R}^3$ given by

$$\mathbf{Z} = \cos z \frac{\partial}{\partial x} + \sin z \frac{\partial}{\partial y} + \frac{\partial}{\partial z},$$

where x, y, z are the natural coordinate functions of $\mathbb{R}^3$. Show that there is no vector field $\mathbf{Y}$ on $\mathbb{R}^2$ for which

$$\Phi_{*\mathbf{p}}(\mathbf{Z}) = \mathbf{Y}_{\Phi(\mathbf{p})}$$

for all $\mathbf{p} \in \mathbb{R}^3$.

7. Set $\mathcal{W} = \{(u, v) \in \mathbb{R}^2 \mid uv > 0\}$. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $g\colon \mathcal{W} \to \mathbb{R}^3$ be defined by

$$f(u, v) = \big(u^2 \sin v,\ e^{uv}\big) \quad \text{and} \quad g(u, v) = \big(u v^2,\ \log(uv),\ v \sin u\big).$$

Compute the matrices of f_* and g_* with respect to the standard bases of vector fields on $\mathbb{R}^2$ and $\mathbb{R}^3$.

8. Prove Lemma 24.62.

9. Define, for each patch $(\mathbf{x},\mathcal{U})$ on an n-dimensional differentiable manifold $\mathcal{M}$, an associated patch $\widetilde{\mathbf{x}}\colon \mathcal{U}\times\mathbb{R}^n \to T(\mathcal{M})$ on the tangent bundle of $\mathcal{M}$ by setting

$$\widetilde{\mathbf{x}}(\mathbf{p},\mathbf{q}) = \left(\mathbf{x}(\mathbf{p}),\ \ \mathbf{x}_{*\mathbf{p}}\left(\sum_{i=1}^{n} q_i \frac{\partial}{\partial u_i}\right)\right),$$

where $\mathbf{p}\in\mathcal{U}$ and $\mathbf{q}=(q_1,\ldots,q_n)$. Show that the tangent bundle $T(\mathcal{M})$ becomes a differentiable manifold with atlas

$$\left\{\ (\widetilde{\mathbf{x}},\ \mathcal{U}\times\mathbb{R}^n)\ \middle|\ (\mathbf{x},\mathcal{U}) \text{ is a patch on } \mathcal{M}\ \right\}.$$

10. Define, for fixed $a>0$, a mapping from an open set of $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$ by

$$\mathsf{stereo}(q_0,q_1,\ldots,q_n) = \frac{a}{a-q_0}(q_1,\ldots,q_n), \qquad q_0 < a. \tag{24.20}$$

Referring to page 813, show that the composition $\mathsf{stereo}\circ\mathsf{north}$ equals the identity map from $\mathbb{R}^n$ to itself. Observe that (24.20) generalizes the mapping of the same name on page 730 to the case of arbitary n and a.

Notebook 24

■ N24.0

This notebook has little underlying mathematical content. Instead, its purpose is to describe how the three figures, two in Section 24.2 and one in Section 24.5, were built up. We explain their construction, step by step, in some detail.

The third figure was plotted using the vector field package

```
<<Graphics`PlotField3D`
```

To investigate the gradient of a function in N24.5, we shall need the following program, originally introduced in N9.4.

```
grad[f_, coor_]:= Table[D[f, coor[[i]]], {i,Length[coor]}]
```

The only overriding definition is our usual command to suppress displays as the default. While hardly necessary in this notebook, it prevents duplicating plots in running the code for Figure 24.2, and we adopt it as a matter of consistent practice.

```
$DisplayFunction= Identity
Unprotect[S]
  S[X__]:= Show[X, DisplayFunction->(Display[$Display,#] &)]
Protect[S]
```

■ N24.2

Figure 24.1 is purely schematic, but is the sort of diagram that can serve in many different situations. It is built around three closed boundary curves. We label each with an index i. This runs from 1 (the outer one) to 3 (the inner one), in order that each corresponding region is a subset of the previous one.

The smooth closed curves are obtained by deforming an ellipse, parametrized by the function f. We add on smaller multiples of trigonometric functions with double and

triple the period, in the style of Fourier series, though mixing the phases of the two components f,g will make the appearance of the curves more casual.

```
f[t_]:= {3Cos[t], 2Sin[t]}
g[t_]:= {Cos[3t], Sin[2t]}
h[1,t_]:= 3.2f[t+2] + g[t]
h[2,t_]:= 2.2h[3,t]
h[3,t_]:= f[t] + g[t+1]/5
```

In the setting of Lemma 24.12, the first curve is the boundary of the assigned open set $\mathfrak{W}$, whereas the last two are boundaries of the open sets whose existence is guaranteed. To emphasize the affinity between the last two curves, we have given them the same shape by defining the second function to be a multiple of the third.

Using the three functions f[i,t], we next define the vertices of three polygons that will be used to color the interiors of the respective curves.

```
vert[i_]:= Table[h[i, 2π n/30], {n,0,30}]
```

To enhance the plotting, we shall use the following abbreviations. The colors are assigned in the same order as the curves, outer to inner. For example, co[[3]] represents the green interior of the smallest polygon with vertex set vert[3]. This ensures that it is the last of the three to be plotted and is therefore visible "on top" of the previous colors.

```
co:= {RGBColor[1,1,.3], RGBColor[.5,1,.5], RGBColor[.7,.7,1]}
th:= Thickness[1/125]
da:= Dashing[{.01,.02}]
```

The text is placed in script font using the instruction

```
txs[A_,v_]:= Text[FontForm[A,{"Script",24}], v]
```

The above commands are brought together in the following program, in which y[1], y[2], y[3] represent the three colored polygons, y[4] is the central point, and y[5] consists of the labels of the various sets. These objects are converted to z[i] using the command Graphics, and z[6] achieves a plot of the three curves simultaneously. The outmost pink coloring is added as Background.

```
Module[{ta,y,z},
  y[i_]:= {co[[i]], Polygon[vert[i]]};
  y[4]= {PointSize[.02], Point[{0,0}]};
  y[5]= {txs["W",{8,.5}], txs["P",{1.5,0}], txs["M-Q",{0,-3.5}]};
  z[i_]:= Graphics[y[i]];
  z[6]:= ParametricPlot[Table[h[i,t],{i,3}]//Evaluate, {t,0,2π},
     Axes->False, PlotStyle->{{th,da}, th, {th,da}}];
  S[Array[z,6], Background->RGBColor[1,.7,.7],
     PlotRange->{{-12,12},{-9,8}}, ImageSize->400]
]
```

Figure 24.1

The size of the text does not change when the image is manually rescaled. Thus, to fix a desirable ratio between the labels' font size and overall plot for insertion into the text of Chapter 24, we use the ImageSize command.

The commands for Figure 24.2 are relatively straightforward. We begin by defining the relevant exponential function.

```
g[a_,b_][x_]:= If[a<x<b, Exp[1/(x-b) - 1/(x-a)], 0]
```

In the Plot command below, it is essential to stipulate an appropriate PlotRange to prevent the exponential graph going off the scale. Remember that the symbols S@ must be added to start of the third line to view the plot.

```
f:= g[-3,3]
f[2]
Plot[f[x], {x,-5,5}, PlotRange->{0,1}]
```

It is not possible to integrate the function f in closed form, so *Mathematica* leaves the result as stated.

```
Integrate[f[x],{x,0,2}]
```

For plotting purposes though, a numerical result will suffice.

```
G[a_,b_][x_]:= NIntegrate[g[a,b][u], {u,x,b}]
H[a_,b_][x_]:= G[a,b][x]/G[a,b][a]
```

```
Plot[H[1,4][x], {x,0,5}]
```

The GraphicsSpacing option makes the two plots of Figure 12.2 fit snugly together, and ImageSize is used again to ensure that the axis labels end up neither too small nor too big.

The axis labels themselves can be adjusted using the Ticks command, though this can sometimes be avoided by a judicious choice of PlotRange which affects not just the graph itself but also the axis labels.

```
Module[{z},
  z[1]= Plot[g[1,4][x],{x,0,6}, PlotRange->{0,.3},
    Ticks->{Automatic, {.1,.2,.3}}, PlotStyle->{th,Hue[0]}];
  z[2]= Plot[H[1,4][x], {x,0,5}, PlotRange->{0,1.1},
    PlotStyle->{th,Hue[.6]}];
  S @ GraphicsArray[Array[z,2], GraphicsSpacing->0,
     ImageSize->600]
]
```

Figure 24.2

We have separated the PlotStyle options in order to color the curves differently, but they share a common thickness abbreviation th defined above.

■ N24.5

Figure 24.3 is plotted using the standard package PlotField3D loaded in N24.0. The vector field in question does not in fact depend on z, though in theory it could by means of a definition less simple than the one we actually give:

```
X[x_,y_,z_]:= {-y, x, 0};
```

Thick black lines represent the Cartesian axes, and are constructed with the command Line by first defining the origin o and the extremities a,b,c. The axes then constitute a single set, to which the command Graphics3D has been appended to obtain an object z[1] that can be inserted into the command Show or S.

```
Module[{o,a,b,c,z},
  o= {0,0,0};
  a= {1,-1,0};
  b= {1,1,0};
  c= {0,0,2};
  z[1]= {Thickness[1/80], Line[{o,a}], Line[{o,b}], Line[{o,c}]}
    //Graphics3D;
  z[2]= PlotVectorField3D[X[x,y,z], {x,-1,1}, {y,-1,1}, {z,0,2},
    PlotPoints->6, VectorHeads->True, ColorFunction->Hue];
  S @ Array[z,2]
]
```

Figure 24.3

The related program PlotGradientField3D can be used to plot vector fields that arise as gradients. The vector field X above is conformally related to the gradient field defined next.

```
F:= ArcTan[y/x]
grad[F, {x,y,z}]//Simplify
```

```
PlotGradientField3D[F, {x,-1,1}, {y,-1,1}, {z,-1,1},
  PlotPoints->6, VectorHeads->True, ColorFunction->GrayLevel]
```

The use of the ColorFunction option effectively eliminates arrows close to the z-axis that tend to confuse the overall plot.

Chapter 25

Riemannian Manifolds

Our ultimate goal is to study *abstract surfaces*, that is, 2-dimensional manifolds which have a notion of metric compatible with their manifold structure (see Definition 25.21 on page 858). The theory of differentiable manifolds developed in the previous chapter is insufficient on its own to measure lengths and areas. For that we need to impose additional structure so that we end up with a generalization of a metric as defined in Section 12.1. Riemann[1] in his thesis [Riem] (written in 1854 but published only in 1867) introduced such a metric ds^2 in n dimensions and described how the curvature of ds^2 can be measured.

In this chapter, we develop key concepts like that of a covariant derivative and a pseudo-Riemannian metric that "bolt on" to the definition of differentiable manifold. It is just as easy to start by defining these concepts for arbitrary dimensions, which is what we do first, though we shall concentrate on 2 dimensions in the last two chapters of this book.

Covariant derivatives are defined and their basic properties are developed in Section 25.1. They will play a particularly important role in a further discussion of geodesics in Chapter 26. The theory of covariant derivatives or "connections,"



[1] Georg Friedrich Bernhard Riemann (1826–1866). German mathematician, buried in Verbania in northern Italy. Although he published few papers, Riemann's ideas have had a profound effect on the development of mathematics and physics. In an early paper he clarified the notion of integral by defining what is now called the Riemann integral. Riemann's Habilitationsvortrag [Riem], *Über die Hypothesen welche der Geometrie zu Grunde liegen* (impressing even Gauss), became a classic, and its results were incorporated into Einstein's general relativity. The Riemann hypothesis, which concerns the zeros of the complex analytic function

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z},$$

has enormous implications for number theory but remains unsolved.

as they are called in a more abstract setting, is one of the cornerstones of contemporary research in differential geometry. The notion was introduced by Levi-Civita[2] in his book *The Absolute Differential Calculus* [LeviC] to explain parallelism in Riemannian manifolds.

Pseudo-Riemannian manifolds are defined in Section 25.2; it is a significant fact that every pseudo-Riemannian manifold has a natural covariant derivative (Definition 25.18). A Riemannian manifold is a pseudo-Riemannian manifold whose metric is positive definite. Section 25.3 is devoted to bridging the gap between the more modern techniques of Sections 25.1 and 25.2 on the one hand, and the classical treatment of metrics on the other.

We begin the final section by making precise the way in which the Christoffel symbols fail to be the components of a tensor. We then define the celebrated Riemann curvature tensor R and state a number of standard results, deferring proofs to the Exercises. When R is expressed in terms of coordinates, one recognizes formulas from the discussion of the Peterson–Mainardi–Codazzi equations in Section 19.3. This will enable us to pin down the precise relationship between the Riemann and Gaussian curvature for an abstract surface in the next chapter.

25.1 Covariant Derivatives

Since it is important to study vector fields on a differentiable manifold $\mathcal{M}$, it is natural to try to differentiate vector fields on $\mathcal{M}$. For $\mathbb{R}^n$, there is an obvious way to do this. Recall that $\mathfrak{X}(\mathbb{R}^n)$ denotes the space of differentiable vector fields defined on $\mathbb{R}^n$. Examples are the vector fields

$$\frac{\partial}{\partial u_1}, \ldots, \frac{\partial}{\partial u_n} \tag{25.1}$$

determined by the natural coordinate functions $u_1, \ldots, u_n$.

Now consider an arbitrary vector field $\mathbf{Y} \in \mathfrak{X}(\mathbb{R}^n)$; in terms of (25.1), we can write

$$\mathbf{Y} = \sum_{i=1}^{n} f_i \frac{\partial}{\partial u_i}$$

for some functions $f_1, \ldots, f_n \in \mathfrak{F}(\mathbb{R}^n)$. As a matter of fact, in calculus the vector field $\mathbf{Y}$ is frequently identified with the n-tuple $(f_1, \ldots, f_n)$. The obvious candidate for the derivative of $\mathbf{Y}$ in the direction of a vector $\mathbf{X}$ is then



2 Tullio Levi-Civita (1873–1941). Italian mathematician, professor at Padua and Rome. He also wrote on hydrodynamics and rational mechanics. Levi-Civita was an outspoken opponent of fascism and was forced to give up teaching in 1938.

$$\mathbf{D}_{\mathbf{X}}\mathbf{Y} = \big(\mathbf{X}[f_1], \ldots, \mathbf{X}[f_n]\big), \tag{25.2}$$

as defined in Lemma 9.32 on page 277. In the more abstract setting we are about to consider, the same quantity will be denoted by the slightly different notation $D_{\mathbf{X}}\mathbf{Y}$, as a reminder of the fact that differentiation "D" no longer takes place in a Euclidean space.

For a general differentiable manifold, there is no automatic way to differentiate vector fields. Instead, we must add a new element of structure:

Definition 25.1. *A* ***connection*** *or* ***covariant derivative*** *on a differentiable manifold $\mathcal{M}$ is a map*

$$D\colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}),$$

which for $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ also defines a map

$$D_{\mathbf{X}}\colon \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$$

by setting $D(\mathbf{X}, \mathbf{Y}) = D_{\mathbf{X}}\mathbf{Y}$. It is required to have the properties:

$$D_{f\mathbf{X}+g\mathbf{Y}} = fD_{\mathbf{X}} + gD_{\mathbf{Y}}, \tag{25.3}$$

$$D_{\mathbf{X}}(\mathbf{Y}+\mathbf{Z}) = D_{\mathbf{X}}\mathbf{Y} + D_{\mathbf{X}}\mathbf{Z}, \tag{25.4}$$

$$D_{\mathbf{X}}(f\mathbf{Y}) = \mathbf{X}[f]\mathbf{Y} + fD_{\mathbf{X}}\mathbf{Y}, \tag{25.5}$$

for $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{M})$ and $f, g \in \mathfrak{F}(\mathcal{M})$.

Remarks

(1) A connection D is not a tensor field in the sense of Section 24.6. Despite the linearity condition (25.3), it is not linear with respect to functions in its *second* argument.

(2) A connection is however linear with respect to real numbers in its second argument, as equations (25.4) and (25.5) imply that

$$D_{\mathbf{X}}(a\,\mathbf{Y} + b\,\mathbf{Z}) = a\,D_{\mathbf{X}}\mathbf{Y} + b\,D_{\mathbf{X}}\mathbf{Z}, \tag{25.6}$$

for $a, b \in \mathbb{R}$ and $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{M})$.

(3) Using a partition of unity, it is possible to show that every differentiable manifold has a connection, though there is in general no canonical choice. We *shall* see that every Riemannian metric has a special connection associated with it.

Example: The Connection on Euclidean Space

Define

$$D\colon \mathfrak{X}(\mathbb{R}^n) \times \mathfrak{X}(\mathbb{R}^n) \to \mathfrak{X}(\mathbb{R}^n)$$

by first putting

$$D_{\partial/\partial u_i} \frac{\partial}{\partial u_j} = 0,$$

for $1 \leqslant i, j \leqslant n$. More generally, for

$$\mathbf{Y} = \sum_{i=1}^{n} f_i \frac{\partial}{\partial u_i},$$

we are now forced by (25.5) and (25.6) to require that

$$D_{\mathbf{X}}\mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}[f_i] \frac{\partial}{\partial u_i}, \tag{25.7}$$

for $1 \leqslant i \leqslant n$. In fact, (25.7) defines a natural connection on $\mathbb{R}^n$; moreover, (25.7) is equivalent to (25.2).

There is a variant of a connection analogous to the definition of $\boldsymbol{D}_{\mathbf{v}}$ that we gave in Section 9.5. Let $\mathcal{M}$ be a differentiable manifold and let $\mathbf{p} \in \mathcal{M}$. A connection D on $\mathcal{M}$ gives rise to a map

$$D\colon \mathcal{M}_{\mathbf{p}} \times \mathfrak{X}(\mathcal{M}) \to \mathcal{M}_{\mathbf{p}},$$

defined as follows.

Definition 25.2. *Let $\mathbf{v}$ be a tangent vector to $\mathcal{M}$ at $\mathbf{p} \in \mathcal{M}$. Choose a vector field $\mathbf{V} \in \mathfrak{X}(\mathcal{M})$ such that $\mathbf{V}_{\mathbf{p}} = \mathbf{v}$. For $\mathbf{W} \in \mathfrak{X}(\mathcal{M})$ we define $D_{\mathbf{v}}\mathbf{W}$ to be the tangent vector in $\mathcal{M}_{\mathbf{p}}$ given by*

$$D_{\mathbf{v}}\mathbf{W} = \left(D_{\mathbf{V}}\mathbf{W}\right)_{\mathbf{p}}.$$

Lemma 25.3. *Fix $\mathbf{W} \in \mathfrak{X}(\mathcal{M})$. The definition of $D_{\mathbf{v}}\mathbf{W}$ is independent of the choice of $\mathbf{V} \in \mathfrak{X}(\mathcal{M})$ such that $\mathbf{V}_{\mathbf{p}} = \mathbf{v}$.*

Proof. Let $\mathbf{V}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$ be such that $\mathbf{V}_{\mathbf{p}} = \mathbf{Y}_{\mathbf{p}}$, and let $(x_1, \ldots, x_n)$ be a coordinate system defined in an open set containing $\mathbf{p}$. By Corollary 24.55, page 834, we can write

$$\mathbf{V} = \sum_{i=1}^{n} \mathbf{V}[x_i] \frac{\partial}{\partial x_i} \qquad \text{and} \qquad \mathbf{Y} = \sum_{i=1}^{n} \mathbf{Y}[x_i] \frac{\partial}{\partial x_i}.$$

Then (25.3) and (25.5) imply that

$$\begin{aligned}(D_{\mathbf{V}}\mathbf{W})_{\mathbf{p}} &= \sum_{i=1}^{n}\Big(\mathbf{V}[x_i]D_{\partial/\partial x_i}\mathbf{W}\Big)_{\mathbf{p}}\\ &= \sum_{i=1}^{n}\mathbf{V}[x_i](\mathbf{p})\Big(D_{\partial/\partial x_i}\mathbf{W}\Big)_{\mathbf{p}}\\ &= \sum_{i=1}^{n}\mathbf{Y}[x_i](\mathbf{p})\Big(D_{\partial/\partial x_i}\mathbf{W}\Big)_{\mathbf{p}}\\ &= (D_{\mathbf{Y}}\mathbf{W})_{\mathbf{p}}.\ \blacksquare\end{aligned}$$

It is important to realize that the statement of the previous lemma refers to the *first* argument of $D_{\mathbf{v}}\mathbf{w}$. In general, it is not true that $\mathbf{Y}_{\mathbf{p}} = \mathbf{Z}_{\mathbf{p}}$ implies $(D_{\mathbf{X}}\mathbf{Y})_{\mathbf{p}} = (D_{\mathbf{X}}\mathbf{Z})_{\mathbf{p}}$, as can be seen by defining $\mathbf{Z} = f\mathbf{Y}$ for some $f \in \mathfrak{F}(M)$. We can however make a useful statement using the next concept.

Definition 25.4. *Let $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$. An **integral curve** of the vector field $\mathbf{X}$ is any curve $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ for which $\boldsymbol{\alpha}'(t) = \mathbf{X}_{\boldsymbol{\alpha}(t)}$ for all $t \in (a,b)$.*

The statement which we are after deals with the case in which two vector fields agree along an integral curve:

Lemma 25.5. *Let $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be an integral curve of $\mathbf{X}$. If $\mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{M})$ satisfy $\mathbf{Y}_{\boldsymbol{\alpha}(t)} = \mathbf{Z}_{\boldsymbol{\alpha}(t)}$ whenever $a < t < b$, then*

$$(D_{\mathbf{X}}\mathbf{Y})_{\boldsymbol{\alpha}(t)} = (D_{\mathbf{X}}\mathbf{Z})_{\boldsymbol{\alpha}(t)}$$

for $a < t < b$.

Proof. We can assume without loss of generality that there is a system of coordinates $(x_1, \ldots, x_n)$ defined on an open set $\mathcal{W} \subset \mathcal{M}$ such that the trace of $\boldsymbol{\alpha}$ is contained in $\mathcal{W}$.

It suffices to show that for any $\mathbf{Y} \in \mathfrak{X}(\mathcal{M})$, if $\mathbf{Y}_{\boldsymbol{\alpha}(t)}$ vanishes identically then so does $(D_{\mathbf{X}}\mathbf{Y})_{\boldsymbol{\alpha}(t)}$. By Corollary 24.55, we can write

$$\mathbf{Y} = \sum_{i=1}^{n}\mathbf{Y}[x_i]\frac{\partial}{\partial x_i} \tag{25.8}$$

on $\mathcal{W}$. Suppose that $t \mapsto \mathbf{Y}_{\boldsymbol{\alpha}(t)}$ vanishes identically. Then (25.8) implies that

$$\mathbf{Y}[x_i]\big(\boldsymbol{\alpha}(t)\big) = 0$$

for $a < t < b$ and $i = 1, \ldots, n$. When we differentiate (25.8) we obtain

$$D_{\mathbf{X}}\mathbf{Y} = \sum_{i=1}^{n}\left(\mathbf{X}\mathbf{Y}[x_i]\frac{\partial}{\partial x_i} + \mathbf{Y}[x_i]D_{\mathbf{X}}\frac{\partial}{\partial x_i}\right),$$

which restricts to $\boldsymbol{\alpha}$ as

$$(D_{\mathbf{X}}\mathbf{Y})_{\boldsymbol{\alpha}(t)} = \sum_{i=1}^{n} \left(\mathbf{X}\mathbf{Y}[x_i] \frac{\partial}{\partial x_i} \right) \Big|_{\boldsymbol{\alpha}(t)}. \tag{25.9}$$

But by Lemma 24.39, we have

$$\mathbf{X}\mathbf{Y}[x_i]\big(\boldsymbol{\alpha}(t)\big) = \mathbf{X}_{\boldsymbol{\alpha}(t)}\big[\mathbf{Y}[x_i]\big] = \boldsymbol{\alpha}'(t)\big[\mathbf{Y}[x_i]\big] = \frac{d(\mathbf{Y}[x_i] \circ \boldsymbol{\alpha})}{dt} = 0.$$

Thus the coefficients in (25.9) all vanish, and so $(D_{\mathbf{X}}\mathbf{Y})_{\boldsymbol{\alpha}(t)} = 0$ for $a < t < b$. ∎

In Section 1.2, we considered vector fields along a curve in $\mathbb{R}^n$. This notion generalizes to manifolds.

Definition 25.6. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve in a differentiable manifold $\mathcal{M}$. A **vector field along** $\boldsymbol{\alpha}$ is a function $\mathbf{Y}$ that assigns to each $t \in (a,b)$ a tangent vector $\mathbf{Y}(t) \in \mathcal{M}_{\boldsymbol{\alpha}(t)}$.*

Lemma 25.5 allows us to define unambiguously the covariant derivative of a vector field along a curve:

Definition 25.7. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be an injective regular curve in a differentiable manifold $\mathcal{M}$, and let $\mathbf{Y}$ be a vector field along $\boldsymbol{\alpha}$. Then $\mathbf{Y}'(t)$ is the vector field along $\boldsymbol{\alpha}$ defined by*

$$\mathbf{Y}'(t) = D_{\boldsymbol{\alpha}'(t)}\widetilde{\mathbf{Y}},$$

where $\widetilde{\mathbf{Y}} \in \mathfrak{X}(\mathcal{M})$ is any vector field such that $\widetilde{\mathbf{Y}}_{\boldsymbol{\alpha}(t)} = \mathbf{Y}(t)$ for $a < t < b$.

Vector fields along a curve in a manifold $\mathcal{M}$ have most of the properties enjoyed by vector fields along a curve in $\mathbb{R}^n$. For example, the following lemma generalizes Lemma 1.11 on page 8.

Lemma 25.8. *Let $\mathbf{Y}$ and $\mathbf{Z}$ be vector fields along a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ in a differentiable manifold $\mathcal{M}$, and let $f\colon (a,b) \to \mathbb{R}$ be differentiable. Then*

$$\begin{aligned} (\mathbf{Y} + \mathbf{Z})' &= \mathbf{Y}' + \mathbf{Z}', \\ (f\mathbf{Y})' &= f'\mathbf{Y} + f\mathbf{Y}'. \end{aligned} \tag{25.10}$$

Proof. We prove the second equation (25.10). Let $\widetilde{f} \in \mathfrak{F}(\mathcal{M})$ and $\widetilde{\mathbf{Y}} \in \mathfrak{X}(\mathcal{M})$ be such that $\widetilde{f}(\boldsymbol{\alpha}(t)) = f(t)$ and $\widetilde{\mathbf{Y}}(\boldsymbol{\alpha}(t)) = \mathbf{Y}(t)$ for $a < t < b$. Then

$$\begin{aligned} (f\mathbf{Y})'(t) &= D_{\boldsymbol{\alpha}'(t)}(\widetilde{f}\widetilde{\mathbf{Y}}) \\ &= \boldsymbol{\alpha}'(t)[\widetilde{f}]\,\widetilde{\mathbf{Y}}_{\boldsymbol{\alpha}(t)} + \widetilde{f}\big(\boldsymbol{\alpha}(t)\big) D_{\boldsymbol{\alpha}'(t)}\widetilde{\mathbf{Y}} \\ &= (\widetilde{f} \circ \boldsymbol{\alpha})'(t)\,\mathbf{Y}(t) + f(t)\,\mathbf{Y}'(t) \;=\; f'(t)\,\mathbf{Y}(t) + f(t)\,\mathbf{Y}'(t). \end{aligned}$$

∎

There is also a chain rule for vector fields along a curve.

Lemma 25.9. *Suppose that* $\mathbf{Y}$ *is a vector field along a curve* $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ *in a differentiable manifold* $\mathcal{M}$*, and let* $h\colon (c,d) \to (a,b)$ *be differentiable. Then*

$$(\mathbf{Y} \circ h)'(t) = h'(t)\mathbf{Y}'\big(h(t)\big) \tag{25.11}$$

for $c < t < d$.

Proof. If $\widetilde{\mathbf{Y}} \in \mathfrak{X}(\mathcal{M})$ is such that $\widetilde{\mathbf{Y}}\big(\boldsymbol{\alpha}(h(t))\big) = \mathbf{Y}(h(t))$ for $c < t < d$, then by the chain rule for curves (Lemma 24.41 on page 831), we have

$$\begin{aligned}(\mathbf{Y} \circ h)'(t) \;&=\; D_{(\boldsymbol{\alpha}\circ h)'(t)}\widetilde{\mathbf{Y}} \;=\; D_{h'(t)(\boldsymbol{\alpha}'\circ h)(t)}\widetilde{\mathbf{Y}} \\ &=\; h'(t)\, D_{(\boldsymbol{\alpha}'\circ h)(t)}\widetilde{\mathbf{Y}} \;=\; h'(t)\mathbf{Y}'\big(h(t)\big). \;\blacksquare\end{aligned}$$

Now we are ready to define the notion of acceleration of a curve in a manifold. It depends on the choice of covariant derivative.

Definition 25.10. *Let* $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ *be a curve in a differentiable manifold* $\mathcal{M}$*, and let* D *be a connection on* $\mathcal{M}$*. The* ***acceleration*** *of* $\boldsymbol{\alpha}$ *with respect to* D *is the vector field*

$$t \mapsto D_{\boldsymbol{\alpha}'(t)}\boldsymbol{\alpha}'(t),$$

which we write more simply as $D_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$*. We say that a curve* $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ *is a* ***geodesic*** *with respect to* D *provided*

$$D_{\boldsymbol{\alpha}'(t)}\boldsymbol{\alpha}'(t) = 0$$

for $a < t < b$.

We have denoted the acceleration of a curve $\boldsymbol{\alpha}$ in a manifold $\mathcal{M}$ by $D_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ to avoid confusion with the acceleration of a curve in $\mathbb{R}^n$. For example, if $\boldsymbol{\alpha}$ is a curve in a surface $\mathcal{M} \subset \mathbb{R}^n$, the two accelerations $D_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ and $\boldsymbol{\alpha}''$ are usually different. This is explained in Section 26.3 in the next chapter, which includes a justification for the general definition of geodesic which we have just given.

Next, we prove the manifold version of equation (1.16) on page 16, which tells us the relation between the acceleration of a curve $\boldsymbol{\alpha}$ and the acceleration of a reparametrization of $\boldsymbol{\alpha}$.

Lemma 25.11. *Let* D *be a connection on a differentiable manifold* $\mathcal{M}$*, and let* $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ *be a curve on* $\mathcal{M}$*. Suppose that* $\boldsymbol{\beta} = \boldsymbol{\alpha} \circ h$ *is a reparametrization of* $\boldsymbol{\alpha}$*, where* $h\colon (c,d) \to (a,b)$ *is differentiable. Then*

$$D_{\boldsymbol{\beta}'}\boldsymbol{\beta}' = \big((D_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}') \circ h\big)\, h'^2 + (\boldsymbol{\alpha}' \circ h)\, h''. \tag{25.12}$$

Proof. From Lemma 24.41 we know that

$$\boldsymbol{\beta}' = (\boldsymbol{\alpha}' \circ h)\, h'.$$

We differentiate this equation and use the chain rule (25.11) in combination with (25.10). For $c < t < d$, we obtain

$$\begin{aligned} D_{\boldsymbol{\beta}'(t)}\boldsymbol{\beta}'(t) &= D_{\boldsymbol{\beta}'(t)}\big(h'(t)\,\boldsymbol{\alpha}'(h(t))\big) \\ &= h''(t)\,\boldsymbol{\alpha}'(h(t)) + h'(t)\, D_{\boldsymbol{\beta}'(t)}\boldsymbol{\alpha}'(h(t)) \\ &= h''(t)\,\boldsymbol{\alpha}'(h(t)) + h'(t)\, D_{h'(t)\boldsymbol{\alpha}'(h(t))}\boldsymbol{\alpha}'(h(t)) \\ &= h''(t)\,\boldsymbol{\alpha}'(h(t)) + h'(t)^2\, D_{\boldsymbol{\alpha}'(h(t))}\boldsymbol{\alpha}'(h(t)). \end{aligned}$$

Since this holds for all t, we may rewrite the result as (25.12). ∎

25.2 Pseudo-Riemannian Metrics

After having studied 1-forms on a manifold in Section 24.6, it is natural to investigate covariant tensor fields of degree 2. In this generality, not much can be said. To make progress one must make a restrictive assumption on the covariant tensor field.

Definition 25.12. *Suppose that σ is a covariant tensor field of degree 2 on a differentiable manifold $\mathcal{M}$. We say that σ is a* ***symmetric bilinear form*** *provided that*

$$\sigma(\mathbf{X}, \mathbf{Y}) = \sigma(\mathbf{Y}, \mathbf{X}) \tag{25.13}$$

for all $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$. Similarly, σ is called an ***antisymmetric bilinear form*** *if*

$$\sigma(\mathbf{X}, \mathbf{Y}) = -\sigma(\mathbf{Y}, \mathbf{X}) \tag{25.14}$$

for all $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$.

We shall concentrate on symmetric bilinear forms that satisfy a nondegeneracy condition that we now explain. Notice that any covariant tensor field σ of degree 2 on a manifold $\mathcal{M}$ (whether or not it is symmetric) gives rise to a tensor $\sigma_{\mathbf{p}}$ on each tangent space $\mathcal{M}_{\mathbf{p}}$. Explicitly, $\sigma_{\mathbf{p}}$ is given as follows. Let $\mathbf{x}, \mathbf{y} \in \mathcal{M}_{\mathbf{p}}$; then

$$\sigma_{\mathbf{p}}(\mathbf{x}, \mathbf{y}) = \sigma(\mathbf{X}, \mathbf{Y})(\mathbf{p}),$$

where $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$ are chosen so that $X_{\mathbf{p}} = \mathbf{x}$ and $Y_{\mathbf{p}} = \mathbf{y}$. Since σ is a tensor field, it is linear in both its arguments with respect to functions. This fact implies that the definition of $\sigma_{\mathbf{p}}$ does not depend on the choice of $\mathbf{X}$ and $\mathbf{Y}$ extending $X_{\mathbf{p}} = \mathbf{x}$ and $Y_{\mathbf{p}} = \mathbf{y}$.

Definition 25.13. *A symmetric bilinear form σ on a differentiable manifold $\mathcal{M}$ is called* ***nondegenerate****, provided that $\sigma_{\mathbf{p}}$ is nondegenerate on each tangent space $\mathcal{M}_{\mathbf{p}}$ for each point $\mathbf{p} \in \mathcal{M}$, meaning*

$$\sigma_{\mathbf{p}}(\mathbf{x}, \mathbf{y}) = 0 \quad \textit{for all} \quad \mathbf{y} \in \mathcal{M}_{\mathbf{p}} \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}.$$

A nondegenerate symmetric bilinear form is called a ***pseudo-Riemannian metric****.*

A theorem from linear algebra states that a symmetric bilinear form σ on an n-dimensional vector space V over $\mathbb{R}$ can be diagonalized. This means that there exists a basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ of V for which

$$\sigma(\mathbf{f}_i, \mathbf{f}_j) = 0 \quad \text{whenever } i \neq j.$$

To understand this fact in a different way, we can start with any given basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ and note that the real numbers $\sigma(\mathbf{e}_1, \mathbf{e}_j)$ are the entries of a *symmetric* matrix S. An equivalent theorem now asserts that there exists an orthogonal matrix P (recall Section 23.1) for which $P^{-1}SP = P^TSP$ is diagonal. A similar idea occurs in the proof of Lemma 23.2 on page 770, and the jth column of P expresses the new basis vector $\mathbf{f}_j$ in terms of the $\mathbf{e}_i$.

In the case that σ is nondegenerate, we can postmultiply P by a diagonal matrix of square roots so that the entries of P^TSP become each ± 1. It follows that the basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ can be chosen so that

$$\sigma(\mathbf{f}_i, \mathbf{f}_i) = \pm 1,$$

for each i with $1 \leqslant i \leqslant n$. Moreover, the number p of plus signs and the number $q = n - p$ of minus signs does not depend on the choice of the basis $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$. The pair (p, q) is called the ***signature*** of the symmetric bilinear form.

A particularly important special case occurs when the signature of the symmetric bilinear form is $(n, 0)$. In this case, one says that σ is ***positive definite***. The next definition extends that on page 362 at the start of Chapter 12.

Definition 25.14. *A symmetric bilinear form σ on a differentiable manifold $\mathcal{M}$ is called a* ***Riemannian metric*** *if $\sigma_{\mathbf{p}}$ is positive definite on $\mathcal{M}_{\mathbf{p}}$ for each point $\mathbf{p} \in \mathcal{M}$; equivalently,*

$$\sigma_{\mathbf{p}}(\mathbf{x}, \mathbf{x}) > 0 \quad \textit{for all nonzero } \mathbf{x} \in \mathcal{M}_{\mathbf{p}}.$$

We shall see that any regular surface $\mathcal{M}$ in $\mathbb{R}^n$ acquires such a form σ.

We shall use the word ***metric*** to mean either a pseudo-Riemannian metric or a Riemannian metric. Usually we denote a metric by $\langle\ ,\ \rangle$ and if it is positive definite, we write

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

for any tangent vector $\mathbf{x}$ to $\mathcal{M}$. Beware that when $\langle\ ,\ \rangle$ is not positive definite, $\langle \mathbf{x}, \mathbf{x} \rangle$ can be zero even if $\mathbf{x}$ is not.

Definition 25.15. *A **pseudo-Riemannian manifold** is a differentiable manifold $\mathcal{M}$ equipped with a metric $\langle\ ,\ \rangle$. If $\mathbf{x}, \mathbf{y}$ are tangent vectors to $\mathcal{M}$, we call $\langle \mathbf{x}, \mathbf{y} \rangle$ the **inner product** of $\mathbf{x}$ and $\mathbf{y}$. Similarly, we can speak about the inner product of vector fields. If $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$, then $\langle \mathbf{X}, \mathbf{Y} \rangle$ is the function, or rather element of $\mathfrak{F}(\mathcal{M})$, defined by*

$$\langle \mathbf{X}, \mathbf{Y} \rangle(\mathbf{p}) = \langle \mathbf{X}_\mathbf{p}, \mathbf{Y}_\mathbf{p} \rangle, \qquad \mathbf{p} \in \mathcal{M}.$$

*A **Riemannian manifold** is a differentiable manifold $\mathcal{M}$ equipped with a positive definite metric $\langle\ ,\ \rangle$.*

Example: The Metric on Euclidean Space

There is a natural positive definite metric on $\mathbb{R}^n$ given by

$$\left\langle \sum_{i=1}^{n} f_i \frac{\partial}{\partial u_i}, \sum_{j=1}^{n} g_j \frac{\partial}{\partial u_j} \right\rangle = \sum_{i=1}^{n} f_i g_i. \tag{25.15}$$

This metric coincides with the standard dot product $\cdot$ on $\mathbb{R}^n$ that we introduced in Section 1.1, and it is this object that "induces" the metric defined in Section 12.1 on surfaces.

In the light of the discussion following Definition 25.13, a more general metric on $\mathbb{R}^n$ is obviously given by

$$\left\langle \sum_{i=1}^{n} f_i \frac{\partial}{\partial u_i}, \sum_{j=1}^{n} g_j \frac{\partial}{\partial u_j} \right\rangle = \sum_{i=1}^{n} a_i f_i g_i, \tag{25.16}$$

where $a_1, \ldots, a_n$ are nonzero constants. The metric (25.16) is positive definite if and only if all the a_i's are positive. We shall make use of (25.16) with $n = 3$ and $(a_1, a_2, a_3) = (1, 1, -1)$ in the next chapter (see the explanation of Figure 26.1 on page 876). An even more general metric is given by

$$\left\langle \sum_{i=1}^{n} f_i \frac{\partial}{\partial u_i}, \sum_{j=1}^{n} g_j \frac{\partial}{\partial u_j} \right\rangle = \sum_{i,j=1}^{n} a_{ij} f_i g_j, \tag{25.17}$$

where $A = (a_{ij})$ denotes a symmetric matrix of constants whose determinant is nonzero. The metric is positive definite if and only if the eigenvalues of A (which are necessarily real) are all positive.

The notion of gradient that we defined on page 274 has a generalization to pseudo-Riemannian manifolds.

Definition 25.16. *Let θ be a 1-form on a pseudo-Riemannian manifold $\mathcal{M}$. The **dual** of θ is the vector field, denoted $\theta^\sharp$, defined by*

$$\langle \theta^\sharp, \mathbf{Y} \rangle = \theta(\mathbf{Y}) \qquad \textit{for all } \mathbf{Y} \in \mathfrak{X}(\mathcal{M}).$$

Thus, $\theta^\sharp \in \mathfrak{X}(\mathcal{M})$. In the special case when $\theta = df$ for some $f \in \mathfrak{F}(\mathcal{M})$ we call $df^\sharp$ the ***gradient*** *of f and denote it by* $\mathsf{grad}\, f$*; it is given by*

$$\langle \mathsf{grad}\, f, \mathbf{Y}\rangle = \mathbf{Y}[f].$$

That $\theta^\sharp$ is well defined is a consequence of an elementary fact:

Lemma 25.17. Let V be a vector space with an inner product $\langle\ ,\ \rangle$. Any vector $\mathbf{x} \in V$ is completely determined by the mapping $V \to \mathbb{R}$ defined by $\mathbf{y} \mapsto \langle \mathbf{x}, \mathbf{y}\rangle$.

Proof. If $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}', \mathbf{y}\rangle$ for all $\mathbf{y} \in V$, then $\langle \mathbf{x} - \mathbf{x}', \mathbf{y}\rangle = 0$ for all $\mathbf{y} \in V$. It follows from the nondegeneracy of $\langle\ ,\ \rangle$ that $\mathbf{x} = \mathbf{x}'$. ∎

It is fortunate for the study of pseudo-Riemannian manifolds that there is a canonical choice for the covariant derivative.

Definition 25.18. Let $\mathcal{M}$ be a pseudo-Riemannian manifold. The ***Riemannian connection***

$$\nabla\colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}) \tag{25.18}$$

of $\mathcal{M}$ is defined by

$$\begin{aligned} 2\langle \nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}\rangle \;=\; & \mathbf{X}\langle \mathbf{Y}, \mathbf{Z}\rangle + \mathbf{Y}\langle \mathbf{X}, \mathbf{Z}\rangle - \mathbf{Z}\langle \mathbf{X}, \mathbf{Y}\rangle \\ & - \langle \mathbf{X}, [\mathbf{Y}, \mathbf{Z}]\rangle - \langle \mathbf{Y}, [\mathbf{X}, \mathbf{Z}]\rangle + \langle \mathbf{Z}, [\mathbf{X}, \mathbf{Y}]\rangle \end{aligned} \tag{25.19}$$

for $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{M})$.

The proof of the following lemma is straightforward, though we refer the reader to [ON2, chapter 3].

Lemma 25.19. The mapping ∇ defined by (25.19) *has the following properties*:

$$\nabla_{f\mathbf{X}+g\mathbf{Y}} = f\nabla_{\mathbf{X}} + g\nabla_{\mathbf{Y}}, \tag{25.20}$$

$$\nabla_{\mathbf{X}}(\mathbf{Y} + \mathbf{Z}) = \nabla_{\mathbf{X}}\mathbf{Y} + \nabla_{\mathbf{X}}\mathbf{Z}, \tag{25.21}$$

$$\nabla_{\mathbf{X}}(f\mathbf{Y}) = \mathbf{X}[f]\mathbf{Y} + f\nabla_{\mathbf{X}}\mathbf{Y}, \tag{25.22}$$

$$\nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} = [\mathbf{X}, \mathbf{Y}], \tag{25.23}$$

$$\mathbf{X}\langle \mathbf{Y}, \mathbf{Z}\rangle = \langle \nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}\rangle + \langle \mathbf{Y}, \nabla_{\mathbf{X}}\mathbf{Z}\rangle, \tag{25.24}$$

for $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(\mathcal{M})$, $f, g \in \mathfrak{F}(\mathcal{M})$ and $a, b \in \mathbb{R}$. Conversely, if a mapping (25.18) *satisfies* (25.20)–(25.24)*, then it is given by* (25.19).

Remarks

(1) Lemma 25.17 implies that $\nabla_{\mathbf{X}}\mathbf{Y}$ is completely determined by (25.19).

(2) The first three properties coincide with those of Definition 25.1 and confirm that the mapping ∇ really *is* a covariant derivative.

(3) A general connection D on a differentiable manifold may or may not satisfy condition (25.23), which expresses the vanishing of a tensor called the ***torsion*** (see Exercise 4), without reference to the metric. A connection that does satisfy (25.23) is said to be ***torsion-free***.

The following characterization is a restatement of the converse of the previous lemma. Its proof is left as Exercise 5.

Lemma 25.20. *Let $\langle\ ,\ \rangle$ be a pseudo-Riemannian metric on a manifold $\mathcal{M}$. Then there is a unique connection ∇ on $\mathcal{M}$ that satisfies* (25.23) *and* (25.24).

It is customary to use ∇ (rather than D) to denote what we have called *the Riemannian connection*. Often, it is called the ***Levi-Civita connection*** (see the footnote on page 848).

25.3 The Classical Treatment of Metrics

In the next chapter, we shall specialize the general theory of pseudo-Riemannian manifolds to the 2-dimensional case. We know (from Lemma 24.7) that a regular surface $\mathcal{M}$ in $\mathbb{R}^n$ is a differentiable manifold. We can now move on to consider the geometry of surfaces freed from the requirement that they be embedded in $\mathbb{R}^3$ or even $\mathbb{R}^n$.

Definition 25.21. *A 2-dimensional differentiable manifold equipped with a Riemannian or pseudo-Riemannian metric $\langle\ ,\ \rangle$ is called an* ***abstract surface***.

The metric on an abstract surface is a generalization of the induced metric that we studied in Chapter 12. Lemma 24.7 implies:

Lemma 25.22. *A regular surface $\mathcal{M}$ equipped with the metric induced from $\mathbb{R}^n$ is an abstract surface.*

In the literature, one frequently sees a metric in n dimensions written as

$$ds^2 = \sum_{i,j=1}^{n} g_{ij}\,dx_i\,dx_j, \tag{25.25}$$

though in this book we do not generally adopt the convention of placing coordinate indices in an upper position. We gave an intuitive interpretation of (25.25)

in Chapter 12 in terms of infinitesimals for $n = 2$. To make sense of it in the terminology of this chapter, let $\mathcal{M}$ be a Riemannian manifold with metric $\langle\ ,\ \rangle$. We write

$$\langle\ ,\ \rangle = ds^2$$

and choose a coordinate system $(x_1, \dots, x_n)$ on $\mathcal{M}$. This coordinate system gives rise to 1-forms $dx_1, \dots, dx_n$.

The ***symmetric product*** of dx_i and dx_j is the covariant tensor field of degree 2 denoted in one of the equivalent ways

$$dx_i\, dx_j, \qquad dx_i \odot dx_j, \qquad dx_i \vee dx_j,$$

and defined (using the juxtaposition notation) by

$$\begin{aligned}(dx_i\, dx_j)(\mathbf{X}, \mathbf{Y}) &= \tfrac{1}{2}\big(dx_i(\mathbf{X})dx_j(\mathbf{Y}) + dx_i(\mathbf{Y})dx_j(\mathbf{X})\big)\\ &= \tfrac{1}{2}\big(\mathbf{X}[x_i]\mathbf{Y}[x_j] + \mathbf{Y}[x_i]\mathbf{X}[x_j]\big)\end{aligned}$$

for $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$. Insertion of the factor $1/2$ is to some extent justified by the next lemma. Finally, let

$$g_{ij} = \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle$$

for $1 \leqslant i, j \leqslant n$. The g_{ij}'s are called the ***components*** of the metric $\langle\ ,\ \rangle$.

Lemma 25.23. *A metric $ds^2 = \langle\ ,\ \rangle$ is given in terms of the g_{ij}'s by the formula*

$$ds^2 = \sum_{i,j=1}^{n} g_{ij} dx_i dx_j. \tag{25.26}$$

Proof. By definition

$$ds^2\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}\right) = \left\langle \frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l} \right\rangle = g_{kl}.$$

On the other hand, we have

$$(dx_i\, dx_j)\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}\right) = \frac{1}{2}\frac{\partial x_i}{\partial x_k}\frac{\partial x_j}{\partial x_l} + \frac{1}{2}\frac{\partial x_i}{\partial x_l}\frac{\partial x_j}{\partial x_k} = \tfrac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}),$$

so that

$$\left(\sum_{i,j=1}^{n} g_{ij} dx_i dx_j\right)\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}\right) = \tfrac{1}{2}\sum_{i,j=1}^{n} g_{ij}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) = g_{kl}. \tag{25.27}$$

It therefore follows that

$$ds^2\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}\right) = \left(\sum_{i,j=1}^{n} g_{ij} dx_i dx_j\right)\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}\right),$$

for $1 \leqslant k, l \leqslant n$. For general $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$, we write

$$\mathbf{X} = \sum_{k=1}^{n} a_k \frac{\partial}{\partial x_k} \quad \text{and} \quad \mathbf{Y} = \sum_{l=1}^{n} b_l \frac{\partial}{\partial x_l};$$

then (25.27) implies that

$$\begin{aligned} ds^2(\mathbf{X}, \mathbf{Y}) &= \sum_{kl=1}^{n} a_k b_l \left(\sum_{i,j=1}^{n} g_{ij} dx_i dx_j \right) \left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l} \right) \\ &= \left(\sum_{i,j=1}^{n} g_{ij} dx_i dx_j \right) (\mathbf{X}, \mathbf{Y}). \end{aligned}$$

Since (25.28) holds for all $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$, we get (25.26). ∎

Here is some additional classical terminology.

Definition 25.24. *Let g_{ij} $(1 \leqslant i, j \leqslant n)$ be the components of a metric with respect to a coordinate system $(x_1, \ldots, x_n)$; then (g^{ij}) is the inverse matrix of (g_{ij}). Frequently, the entries g^{ij} are called the* ***upper g ij's****, while the entries g_{ij} are called the* ***lower g ij's****. Finally, we put*

$$\mathscr{G} = \det(g_{ij}). \tag{25.28}$$

Lemma 25.25. *For any coordinate system $(x_1, \ldots, x_n)$ on a Riemannian manifold $\mathcal{M}$, each 1-form dx_i is dual to*

$$\sum_{h=1}^{n} g^{ih} \frac{\partial}{\partial x_h}$$

for $i = 1, \ldots, n$; that is,

$$dx_i{}^\sharp = \sum_{h=1}^{n} g^{ih} \frac{\partial}{\partial x_h}.$$

Proof. By definition of inverse matrix, we have

$$\delta_{ij} = \sum_{h=1}^{n} g^{ih} g_{hj} = \sum_{h=1}^{n} g^{ih} \left\langle \frac{\partial}{\partial x_h}, \frac{\partial}{\partial x_j} \right\rangle = \left\langle \sum_{h=1}^{n} g^{ih} \frac{\partial}{\partial x_h}, \frac{\partial}{\partial x_j} \right\rangle.$$

On the other hand,

$$\left\langle dx_i{}^\sharp, \frac{\partial}{\partial x_j} \right\rangle = dx^i \left(\frac{\partial}{\partial x_j} \right) = \delta_{ij}.$$

Then Lemma 25.17 implies that $dx_i{}^\sharp$ and

$$\sum_{h=1}^{n} g^{ih} \frac{\partial}{\partial x_h}$$

coincide. ∎

25.4 Christoffel Symbols in Riemannian Geometry

The Riemannian connection can be used as a means to define Christoffel symbols on a pseudo-Riemannian manifold $\mathcal{M}$. We shall see in the next chapter that, when $\mathcal{M}$ is a surface, these are essentially the same objects defined in Section 17.3. However we shall use lower case notation γ^i_{jk} for the time being, to avoid possible confusion. (It is also closer to the notation of the notebooks.)

The Christoffel symbols arise when we attempt to express the Riemannian connection in terms of local coordinates.

Definition 25.26. *Let $\langle\ ,\ \rangle$ be a metric on an n-dimensional manifold, and let ∇ be the associated Riemannian connection. If $(x_1, \ldots, x_n)$ is a system of local coordinates on $\mathcal{M}$, then the* ***Christoffel symbols*** *γ^k_{ij} of ∇ relative to $(x_1, \ldots, x_n)$ are given by*

$$\gamma^k_{ij} = dx_k\Big(\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j}\Big) \tag{25.29}$$

for $1 \leqslant i, j, k \leqslant n$.

The next lemma is an easy consequence of this definition and Theorem 24.25. It is an effective alternative definition of the Christoffel symbols for Riemannian manifolds and (as we shall soon see) surfaces.

Lemma 25.27. *Let $(x_1, \ldots, x_n)$ be a coordinate system on a manifold $\mathcal{M}$ endowed with a metric $ds^2 = \langle\ ,\ \rangle$, and let ∇ be the Riemannian connection of ds^2. Then*

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \sum_{k=1}^{n} \gamma^k_{ij} \frac{\partial}{\partial x_k}, \tag{25.30}$$

for $1 \leqslant i, j \leqslant n$.

Replacing "∇" by "D," it is of course possible to define the Christoffel symbols for a general connection on a manifold. But there is a symmetry property of the Christoffel symbols of a Riemannian connection that may not hold for a general connection.

Lemma 25.28. *The Christoffel symbols of a Riemannian connection ∇ satisfy the symmetry relation*

$$\gamma^k_{ij} = \gamma^k_{ji} \tag{25.31}$$

for $i, j, k = 1, \ldots, n$.

Proof. Equation (25.31) is an immediate consequence of two facts: (i) ∇ is torsion-free (equation (25.23)), and (ii) the brackets of coordinate vector fields vanish (Corollary 24.58 on page 836). ∎

Next, we derive the classical expression for the γ^i_{jk}'s in terms of the g_{ij} and the g^{ij}. In practice, the following result provides the most efficient way of computing the Christoffel symbols, and it is the basis for a program in Notebook 25.

Lemma 25.29. *The Christoffel symbols γ^i_{jk} of a Riemannian connection are given by*

$$\gamma^i_{jk} = \frac{1}{2}\sum_{h=1}^{n} g^{ih}\left(\frac{\partial g_{hj}}{\partial x_k} + \frac{\partial g_{hk}}{\partial x_j} - \frac{\partial g_{jk}}{\partial x_h}\right). \tag{25.32}$$

Proof. Since brackets of coordinate vector fields are zero, (25.19) reduces to

$$\begin{aligned} 2\left\langle \nabla_{\frac{\partial}{\partial x_j}}\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_h}\right\rangle &= \frac{\partial}{\partial x_j}\left\langle \frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_h}\right\rangle + \frac{\partial}{\partial x_k}\left\langle \frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_h}\right\rangle - \frac{\partial}{\partial x_h}\left\langle \frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_k}\right\rangle \\ &= \frac{\partial g_{kh}}{\partial x_j} + \frac{\partial g_{jh}}{\partial x_k} - \frac{\partial g_{jk}}{\partial x_h}. \end{aligned}$$

It follows from the definition of γ^i_{jk} and Lemma 25.25 that

$$\gamma^i_{jk} \;=\; dx_i\Big(\nabla_{\frac{\partial}{\partial x_j}}\frac{\partial}{\partial x_k}\Big) = \left\langle \nabla_{\frac{\partial}{\partial x_j}}\frac{\partial}{\partial x_k},\; dx_i{}^\sharp\right\rangle = \sum_{h=1}^{n} g^{ih}\left\langle \nabla_{\frac{\partial}{\partial x_j}}\frac{\partial}{\partial x_k},\; \frac{\partial}{\partial x_h}\right\rangle.$$

The result now follows. ∎

Finally, we generalize (25.30).

Lemma 25.30. *Let $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(\mathcal{M})$, and let $(x_1, \ldots, x_n)$ be a coordinate system on a Riemannian manifold $\mathcal{M}$. Then*

$$\nabla_{\mathbf{X}}\mathbf{Y} = \sum_{k=1}^{n}\left(\mathbf{X}\mathbf{Y}[x_k] + \sum_{i,j=1}^{n}\mathbf{X}[x_i]\mathbf{Y}[x_j]\gamma^k_{ij}\right)\frac{\partial}{\partial x_k}. \tag{25.33}$$

Proof. We have

$$\begin{aligned} \nabla_{\mathbf{X}}\mathbf{Y} \;&=\; \nabla_{\sum \mathbf{X}[x_i]\frac{\partial}{\partial x_i}}\left(\sum_{j=1}^{n}\mathbf{Y}[x_j]\frac{\partial}{\partial x_j}\right) \\ &= \sum_{i,j=1}^{n}\mathbf{X}[x_i]\left(\frac{\partial}{\partial x_i}\mathbf{Y}[x_j]\frac{\partial}{\partial x_j} + \mathbf{Y}[x_j]\nabla_{\frac{\partial}{\partial x_i}}\frac{\partial}{\partial x_j}\right) \\ &= \sum_{j=1}^{n}\mathbf{X}\mathbf{Y}[x_j]\frac{\partial}{\partial x_j} + \sum_{i,j,k=1}^{n}\left(\mathbf{X}[x_i]\mathbf{Y}[x_j]\gamma^k_{ij}\frac{\partial}{\partial x_k}\right) \\ &= \sum_{k=1}^{n}\left\{\mathbf{X}\mathbf{Y}[x_k] + \sum_{i,j=1}^{n}\mathbf{X}[x_i]\mathbf{Y}[x_j]\gamma^k_{ij}\right\}\frac{\partial}{\partial x_k}. \;\blacksquare \end{aligned}$$

The Christoffel symbols are therefore used to modify the notion of differentiation that a patch provides, so as to obtain a patch-independent derivative.

25.5 The Riemann Curvature Tensor

According to Lemma 25.27, the Christoffel symbols are the components of the mapping

$$\nabla\colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$$

defined by $(\mathbf{X}, \mathbf{Y}) \mapsto \nabla_{\mathbf{X}}\mathbf{Y}$. We know from the definition of a covariant derivative that this mapping is *not* a vectorvariant tensor because it is not linear over $\mathfrak{X}(M)$ (see page 836). For this reason, a change of chart results in a rather complicated change of the symbols γ^i_{jk}, which we describe next.

Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ and $\widetilde{\mathbf{x}} = \mathbf{y}\colon \mathcal{V} \to \mathcal{M}$ be two charts with corresponding systems of coordinates x_k and y_q. Then we may write

$$\frac{\partial}{\partial x_k} = \sum_{q=1}^{n} \frac{\partial y_q}{\partial x_k} \frac{\partial}{\partial y_q}, \qquad k = 1, \ldots, n, \tag{25.34}$$

where $\mathscr{J} = (\partial y_q / \partial x_k)$ denotes the Jacobian matrix of the change of patch. The situation is essentially that described on page 828, and (25.34) corresponds to (24.13) in which Φ is the identity mapping. In this section, we shall however blur the distinction between functions on the open set $\mathcal{W} = \mathbf{x}(\mathcal{U}) \cap \mathbf{y}(\mathcal{V})$ of $\mathcal{M}$ and functions on the open sets $\mathbf{x}^{-1}(\mathcal{W})$ and $\mathbf{y}^{-1}(\mathcal{W})$ of $\mathbb{R}^n$.

We may now repeat the proof of Lemma 25.30 by taking $\mathbf{X}, \mathbf{Y}$ to be vector fields arising from the first chart. From (25.34) and properties of ∇, we get

$$\begin{aligned} \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial x_k} &= \sum_{q=1}^{n} \left(\frac{\partial^2 y_q}{\partial x_j \partial x_k} \frac{\partial}{\partial y_q} + \frac{\partial y_q}{\partial x_k} \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial y_q} \right) \\ &= \sum_{q=1}^{n} \frac{\partial^2 y_q}{\partial x_j \partial x_k} \frac{\partial}{\partial y_q} + \sum_{p,q=1}^{n} \frac{\partial y_p}{\partial x_j} \frac{\partial y_q}{\partial x_k} \nabla_{\frac{\partial}{\partial y_p}} \frac{\partial}{\partial y_q}. \end{aligned} \tag{25.35}$$

Using Lemma 25.27, this can be expressed as

$$\sum_{i=1}^{n} \gamma^i_{jk} \frac{\partial}{\partial x_i} = \sum_{r=1}^{n} \left(\frac{\partial^2 y_r}{\partial x_j \partial x_k} + \sum_{p,q=1}^{n} \frac{\partial y_p}{\partial x_j} \frac{\partial y_q}{\partial x_k} \widetilde{\gamma}^r_{pq} \right) \frac{\partial}{\partial y_r},$$

where $\widetilde{\gamma}^r_{pq}$ are Christoffel symbols associated to the patch $\mathbf{y}$. A final application of (25.34) yields

$$\sum_{i=1}^{n} \gamma^i_{jk} \frac{\partial y_r}{\partial x_i} = \frac{\partial^2 y_r}{\partial x_j \partial x_k} + \sum_{p,q=1}^{n} \frac{\partial y_p}{\partial x_j} \frac{\partial y_q}{\partial x_k} \widetilde{\gamma}^r_{pq}, \tag{25.36}$$

for $r = 1, \ldots, n$. It is the presence of the first term on the right-hand side that violates the transformation rule for tensors.

It is a remarkable fact that one may nonetheless take derivatives of the Christoffel symbols and combine them with quadratic terms in the same symbols, so as to obtain quantities that transform more simply than (25.36). The book [McCl] contains a good exposition of the calculations that led Riemann to discover the curvature tensor, defined next.

Definition 25.31. *The* ***Riemann curvature tensor*** *is the mapping*

$$R\colon \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \longrightarrow \mathfrak{X}(\mathcal{M})$$

defined[3] *by*

$$R(\mathbf{X},\mathbf{Y},\mathbf{Z}) = \nabla_{\mathbf{X}}(\nabla_{\mathbf{Y}}\mathbf{Z}) - \nabla_{\mathbf{Y}}(\nabla_{\mathbf{X}}\mathbf{Z}) - \nabla_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z}. \tag{25.37}$$

As its name suggests, *the mapping R is a vectorvariant tensor field* (one of degree 3), in the sense of Definition 24.59 on page 836. This fact is not obvious, but can easily be verified, using properties of the Lie bracket $[\mathbf{X},\mathbf{Y}]$ of vector fields defined on page 832 (see Exercise 6).

There are a number of tensors associated to R, all of which define essentially the same object. For example, for each $\mathbf{X},\mathbf{Y}$, we may define a mapping $R_{\mathbf{XY}}\colon \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ by

$$R_{\mathbf{XY}}(\mathbf{Z}) = R(\mathbf{X},\mathbf{Y},\mathbf{Z}).$$

Abbreviating the left-hand side, we may replace $R(\mathbf{X},\mathbf{Y},\mathbf{Z})$ by $R_{\mathbf{XY}}\mathbf{Z}$, which (as we are about to see) is a more useful expression. Furthermore, we may manufacture a covariant tensor field, by combining (25.37) with the metric:

$$R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) = \langle R(\mathbf{X},\mathbf{Y},\mathbf{Z}),\mathbf{W}\rangle = \langle R_{\mathbf{XY}}\mathbf{Z},\mathbf{W}\rangle.$$

The use of the same symbol R in these slightly different contexts does not cause confusion.

Lemma 25.32. *The Riemann curvature tensor has the following symmetries:*

$$R_{\mathbf{XY}} = -R_{\mathbf{YX}}, \tag{25.38}$$

$$\langle R_{\mathbf{XY}}\mathbf{Z},\mathbf{W}\rangle = -\langle R_{\mathbf{XY}}\mathbf{W},\mathbf{Z}\rangle, \tag{25.39}$$

$$\langle R_{\mathbf{XY}}\mathbf{Z},\mathbf{W}\rangle = \langle R_{\mathbf{ZW}}\mathbf{X},\mathbf{Y}\rangle, \tag{25.40}$$

$$R_{\mathbf{XY}}\mathbf{Z} + R_{\mathbf{YZ}}\mathbf{X} + R_{\mathbf{ZX}}\mathbf{Y} = \mathbf{0}. \tag{25.41}$$

Equation (25.41) is the ***first Bianchi identity*** (see the footnote on page 483).

---

[3] Approximately 40% of books on Riemannian manifolds define the curvature to be the negative of the one given here, but this sign will pop up again in Lemma 25.34.

Algebraically, the curvature tensor R is a complicated object, and it is useful to study it by applying so-called ***contractions*** — linear mappings from one space of tensors to another (simpler) one. There are two such contractions one can apply to R; to define them it is convenient to choose a set of vector fields $\{\mathbf{E}_1, \cdots, \mathbf{E}_n\}$ on an open set of $\mathcal{M}$, which are orthonormal at each point.

Definition 25.33. *The* ***Ricci*** *and* ***scalar*** *curvatures of a Riemannian manifold are defined by*

$$\rho(\mathbf{X}, \mathbf{Y}) = \sum_{m=1}^{n} \langle R_{\mathbf{X}\,\mathbf{E}_m}\mathbf{E}_m, \mathbf{Y}\rangle, \tag{25.42}$$

and

$$s = \sum_{l=1}^{n} \rho(\mathbf{E}_l, \mathbf{E}_l) = \sum_{l,m=1}^{n} \langle R_{\mathbf{E}_l\mathbf{E}_m}\mathbf{E}_m, \mathbf{E}_l\rangle. \tag{25.43}$$

It is inherent in this definition that the quantities ρ and s are independent of the choice of orthonormal basis. The reason for this will become clearer after we have discussed their coordinate descriptions.

The scalar curvature is merely a function, and its value at a given point of $\mathcal{M}$ is a number, which is an "invariant" of the Riemannian structure. The Ricci tensor ρ is a symmetric bilinear form (see Definition 25.12 and Exercise 8); its value at each point is determined by a symmetric $n \times n$ matrix or $n(n+1)/2$ numbers, though only 1 is necessary in dimension 2. For $n \geqslant 3$, the Ricci tensor is therefore exactly the same sort of tensor as the metric itself. This fact allows one to combine the Ricci tensor and the metric in the same equation, and Einstein's equations of general relativity do just that (see [HaEl, chapter 3]).

n	R	ρ	s
2	1	1	1
3	6	6	1
4	20	10	1
5	50	15	1
6	105	21	1
7	196	28	1

The nature of the Riemann tensor is less obvious, but by analysing the symmetries (25.38)–(25.41) in dimension n, it can be shown that R is determined by a total of $n^2(n^2-1)/12$ numbers at each point (see, for example, [Spivak]). The table above displays this numerical information for $n \leqslant 7$, and shows that in

general the Ricci tensor accounts for a relatively small part of the full curvature tensor R. Nonetheless, it is of vital importance in lower dimensions. When $n = 2$, R is completely determined by s or, as we shall see, the Gaussian curvature K. When $n = 3$, the Ricci tensor ρ determines all of R; when $n = 4$, it determines exactly *half* of R.

Pick a chart $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$, with associated coordinates $x_1, \dots, x_n$, and set

$$R_{ijkl} = R\Big(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}, \frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_i}\Big).$$

The strange order of the indices is to retain consistency with the definition of the related scalar quantities $R^i{}_{jkl}$ (adopted from [HaEl, chapter 2]) by means of the formula

$$R\Big(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}, \frac{\partial}{\partial x_j}\Big) = \sum_{i=1}^{n} R^i{}_{jkl} \frac{\partial}{\partial x_i}.$$

The two are related by

$$R_{ijkl} = \sum_{m=1}^{n} g_{im} R^m{}_{jkl} \qquad \text{or} \qquad R^i{}_{jkl} = \sum_{m=1}^{n} g^{im} R_{mjkl}.$$

This sort of index shifting is common in tensor calculus, and we have already seen it in Lemma 25.25. The metric can also be used to perform contractions; with our conventions, the Ricci and scalar curvatures are given by

$$\begin{aligned} \rho_{jl} = \rho\Big(\frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_l}\Big) &= \sum_{i=1}^{n} R^i{}_{jil} = \sum_{i,k=1}^{n} g^{ik} R_{ijkl}, \\ s = \sum_{j,l=1}^{n} g^{jl} \rho_{jl} &= \sum_{i,j,k,l=1}^{n} g^{jl} g^{ik} R_{ijkl}. \end{aligned} \tag{25.44}$$

The curvature tensor itself is given in local coordinates by

Lemma 25.34. *With the notation above,*

$$-R^i{}_{jkl} = \frac{\partial \gamma^i_{jk}}{\partial x_l} - \frac{\partial \gamma^i_{jl}}{\partial x_k} + \sum_{m=1}^{n} \big(\gamma^m_{jk}\gamma^i_{lm} - \gamma^m_{jl}\gamma^i_{km}\big). \tag{25.45}$$

Proof. Because the brackets between the vector fields $\partial/\partial x_i$ vanish, we obtain

$$\begin{aligned} R\left(\frac{\partial}{\partial x_k}, \frac{\partial}{\partial x_l}, \frac{\partial}{\partial x_j}\right) &= \nabla_{\frac{\partial}{\partial x_k}}\Big(\nabla_{\frac{\partial}{\partial x_l}} \frac{\partial}{\partial x_j}\Big) - \nabla_{\frac{\partial}{\partial x_l}}\Big(\nabla_{\frac{\partial}{\partial x_k}} \frac{\partial}{\partial x_j}\Big) \\ &= \nabla_{\frac{\partial}{\partial x_k}}\Big(\sum_{m=1}^{n} \gamma^m_{lj} \frac{\partial}{\partial x_m}\Big) - \nabla_{\frac{\partial}{\partial x_l}}\Big(\sum_{m=1}^{n} \gamma^m_{kj} \frac{\partial}{\partial x_m}\Big). \end{aligned}$$

The result follows from (25.31) and steps of the type (25.35). ∎

When we expand (25.45) for $n = 2$, we obtain formulas that resemble the right-hand side of equations (19.11) on page 601. We shall be more explicit in the next chapter, once we have identified the Christoffel symbols γ^i_{jk} with those defined for a surface in Chapter 17.

25.6 Exercises

1. Complete the proof of Lemma 25.8.

2. Give examples of vector fields $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ on $\mathbb{R}$ such that $\mathbf{Y}_0 = \mathbf{Z}_0$, but $(D_{\mathbf{X}}\mathbf{Y})_0 \neq (D_{\mathbf{X}}\mathbf{Z})_0$.

3. Prove Lemmas 25.19 and 25.20.

4. The ***torsion*** of a connection D on a differentiable manifold $\mathcal{M}$ is defined by

$$T(\mathbf{X}, \mathbf{Y}) = D_{\mathbf{X}}\mathbf{Y} - D_{\mathbf{Y}}\mathbf{X} - [\mathbf{X}, \mathbf{Y}].$$

Show that T is a vectorvariant tensor field on $\mathcal{M}$ that satisfies

$$T(\mathbf{Y}, \mathbf{X}) = -T(\mathbf{X}, \mathbf{Y})$$

(refer to page 836).

5. Prove Lemma 25.20 by simplifying the right-hand side of (25.19).

6. Verify that the curvature tensor, as defined in (25.37) on page 864, is a vectorvariant tensor field.

7. Deduce from the previous exercise that if $\mathbf{x}$ and $\widetilde{\mathbf{x}} = \mathbf{y}$ are two charts with respective curvature components $R_{ijkl}, \widetilde{R}_{pqrs}$ then

$$R_{ijkl} = \sum_{p,q,r,s=1}^{n} \widetilde{R}_{pqrs} \frac{\partial y_p}{\partial x_i} \frac{\partial y_q}{\partial x_j} \frac{\partial y_r}{\partial x_k} \frac{\partial y_s}{\partial x_l}. \tag{25.46}$$

8. Show that the Ricci tensor satisfies

$$\rho(\mathbf{X}, \mathbf{Y}) = \rho(\mathbf{Y}, \mathbf{X}),$$

and that for $n = 2$,

$$\rho(\mathbf{X}, \mathbf{Y}) = \tfrac{1}{2} s \langle \mathbf{X}, \mathbf{Y} \rangle,$$

where s is the scalar curvature. See also Theorem 26.4 on page 874 of the next chapter.

Notebook 25

N25.0

This notebook anticipates some results in Chapter 26, by applying the methods of Chapter 25 to surfaces. Its aim is two-fold. First, to give a program for computing the Christoffel symbols of surfaces, based on Lemma 25.29. We shall confirm experimentally that the new symbols coincide with those defined in Chapter 17. The second aim is to carry out computations with the Riemann tensor, and to define the scalar curvature.

All the definitions of Chapter 25 are valid in n dimensions. There is no real difficulty in computing in higher dimensions, provided n is fixed in each particular problem. We shall however restrict to n=2 in this notebook, so that it is easier to compare our formulas with those of earlier notebooks. There are in any case a number of refined computer programs available for performing tensor calculus. Many of these incorporate routines for computing Christoffel symbols and the Riemann curvature tensor.

For specific calculations (that work well in any single-digit dimension), the editors have benefited from GRTensor, a program that is freely available from the website http://grtensor.phy.queensu.ca/ in versions for both *Mathematica* and *Maple*.

N25.0.1 includes definitions of some surfaces on which to test the theory, and the programs metric from N12.1, brioschi from N17.1, and christoff2 from N17.3.

N25.0.1

N25.4

As in Notebook 17, we shall use M to denote a symmetric matrix of functions defining a first fundamental form. Two variables need to be added to view the matrix:

```
M:= metric[monkeysaddle]
M[u,v]//MatrixForm
```

The following miniprogram computes the Christoffel symbols using Lemma 25.29, a well-known formula of tensor calculus. Inside the Module, the matrix entries A[[i,j]] are the "g lower ij," and B[[i,j]] the inverse components "g upper ij."

```
christoff3[M_][u_,v_]:=
Module[{si,U,B,gd},
  si= Simplify[#]/.{U[1]->u,U[2]->v} &;
  A= M[U[1],U[2]];
  B= Inverse[A];
  gd[i_,j_,k_]:= D[A[[i,j]],U[k]];
  γ[i_,j_,k_]:= Sum[B[[i,m]](gd[m,j,k]+gd[m,k,j]-gd[j,k,m])/2,
    {m,2}]//si;
  Flatten[Table[γ[i,j,k], {i,2}, {j,2}, {k,j,2}]]
]
```

Inside the Module, differentiation and summation is performed using the dummy variables U[1],U[2]. The output is a list of the six (potentially distinct) Christoffel symbols, but once the program is run, typing γ[i,j,k] will yield an individual symbol.

```
christoff3[M][u,v]
γ[1,2,1]
```

By their very definition, the Christoffel symbols are unchanged when j,k are swapped:

```
γ[1,2,1] - γ[1,1,2] //Simplify
```

```
x[u_,v_]:= {u^5, Exp[v^3], Sin[u v]}
M:= metric[x]
christoff3[M][u,v]
```

The following output confirms that the Christoffel symbols defined in Chapter 25 coincide with those of Chapter 17 (at least for the complicated patch defined above!).

```
christoff3[M][u,v] - christoff2[M][u,v] //Simplify
```

■ N25.5

The Riemann curvature tensor can be computed using Lemma 25.34. The next program first runs christoff3 to find the Christoffel symbols determined by the metric M. These are then used to obtain the components $R^i{}_{jkl}$ of the Riemann tensor with one upper index.

```
riemann[M_][i_][j_,k_,l_][u_,v_]:=
Module[{si,U},
  si= Simplify[#]/.{U[1]->u,U[2]->v} &;
  christoff3[M][U[1],U[2]];
  -D[γ[i,j,k],U[l]] + D[γ[i,j,l],U[k]] -
    Sum[γ[m,j,k]γ[i,l,m] - γ[m,j,l]γ[i,k,m], {m,2}]//si
]
```

The components R_{ijkl} with all lower indices are given next. They vanish if the first (or last) two indices are equal.

```
riemann[M_][i_,j_,k_,l_][u_,v_]:=
Module[{A},
  A[m_]:= riemann[M][m][j,k,l][u,v];
  Sum[M[u,v][[i,m]]A[m],{m,2}]//Simplify
]
```

```
M:= metric[monkeysaddle]
R[i_,j_,k_,l_]:= riemann[M][i,j,k,l][u,v]
R[1,1,1,2]
R[1,2,2,2]
R[1,2,1,2]
```

The Ricci and scalar curvatures are defined by

```
ricci[M_][j_,l_][u_,v_]:= Sum[riemann[M][i][j,i,l][u,v], {i,2}]
scalar[M_][u_,v_]:=
  Sum[ricci[M][j,l][u,v]Inverse[M[u,v]][[j,l]], {j,2}, {l,2}]
```

We display the scalar curvature of an arbitrary metric, and prove that this coincides with twice the Gaussian curvature.

```
MM[u_,v_]:= {{e[u,v],f[u,v]},{f[u,v],g[u,v]}}
MM[u,v]//MatrixForm
scalar[MM][u,v]//Simplify
% - 2brioschi[MM][u,v]//Simplify
```

Chapter 26

Abstract Surfaces and Their Geodesics

The revolution in surface theory had been initiated by Gauss in the first half of the nineteenth century. Before Gauss, surface theory meant the study of surfaces in ordinary space $\mathbb{R}^3$. Gauss's *Theorema Egregium* (Theorem 17.5, page 536), stating that the curvature K of a surface in $\mathbb{R}^3$ does not depend on how it is embedded in $\mathbb{R}^3$, led mathematicians such as Riemann to question the necessity of a surface being embedded in $\mathbb{R}^3$ in the first place. In this way, the idea of an *abstract surface* was born. Whilst regular surfaces in $\mathbb{R}^3$ *are* abstract surfaces (see page 858), there exist abstract surfaces that cannot be viewed inside $\mathbb{R}^3$.

Section 26.1 pins down the exact relation between the previous chapter's notation for the metric and Christoffel symbols of a Riemannian manifold, and that of Chapters 13 and 17. We are then at liberty to provide examples of abstract surfaces, and we spend part of Section 26.2 explaining the equivalent Poincaré metrics on the disk and upper half-plane. These examples have constant Gaussian curvature, and computations carried out in Notebook 26 with Brioschi's formula effectively verify this fact.

A discussion of *orientability* precedes the definition of *geodesic curvature* for abstract surfaces in Section 26.3. We then define acceleration in terms of the Riemannian connection, so as to characterize geodesics on a Riemannian manifold. We show that this approach is consistent with that of Chapter 17, and Section 26.4 concludes with a brief discussion of Euclid's parallel postulate.

The *exponential map*, which naturally maps a given tangent space to the surface itself, is defined in Section 26.5. A *geodesic polar patch* provides a standard form for many abstract metrics and simplifies the resulting calculations. We also prove the *Gauss Lemma* on page 887. Length minimizing properties of geodesics are discussed in the last section, where it is shown that an abstract surface is a metric space.

26.1 Christoffel Symbols on Abstract Surfaces

The concept of abstract surface was defined on page 858 of the previous chapter. The formulas for E, F, G on page 362 make sense for a patch on an abstract surface; all that needs to be done is to use the abstract surface metric in place of the induced metric. We begin by making precise the relation between the g_{ij}'s introduced in Section 25.3 and E, F, G. We denote the natural coordinate functions of $\mathbb{R}^2$ either by u, v or by u_1, u_2.

Lemma 26.1. *Let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a patch on an abstract surface $\mathcal{M}$ endowed with a metric ds^2. Then*

$$\text{(26.1)} \qquad \begin{cases} g_{11} \circ \mathbf{x} = E, & g_{12} \circ \mathbf{x} = F, \\ g_{21} \circ \mathbf{x} = F, & g_{22} \circ \mathbf{x} = G. \end{cases}$$

Hence the pullback (see page 837) of ds^2 by $\mathbf{x}$ is given by

$$\mathbf{x}^*\big(g_{11}{dx_1}^2 + 2g_{12}\,dx_1dx_2 + g_{22}{dx_2}^2\big) = E\,du^2 + 2F\,dudv + G\,dv^2.$$

Proof. It follows from Corollary 24.58 on page 836 that

$$g_{ij} \circ \mathbf{x} = \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle \circ \mathbf{x} = \left\langle \mathbf{x}_*\left(\frac{\partial}{\partial u_i}\right), \mathbf{x}_*\left(\frac{\partial}{\partial u_j}\right) \right\rangle \circ \mathbf{x} = \big\langle \mathbf{x}_{u_i}, \mathbf{x}_{u_j} \big\rangle.$$

In particular, $g_{11} \circ \mathbf{x} = \big\langle \mathbf{x}_{u_1}, \mathbf{x}_{u_1} \big\rangle = E$, and the other formulas of (26.1) follow. Also,

$$\begin{aligned} &\mathbf{x}^*\big(g_{11}{dx_1}^2 + 2g_{12}dx_1dx_2 + g_{22}{dx_2}^2\big) \\ &\quad = (g_{11} \circ \mathbf{x})(\mathbf{x}^*\,dx_1)^2 + 2(g_{12} \circ \mathbf{x})(\mathbf{x}^*\,dx_1)(\mathbf{x}^*\,dx_2) + (g_{22} \circ \mathbf{x})(\mathbf{x}^*\,dx_2)^2 \\ &\quad = (g_{11} \circ \mathbf{x})d(x_1 \circ \mathbf{x})^2 + 2(g_{12} \circ \mathbf{x})d(x_1 \circ \mathbf{x})d(x_2 \circ \mathbf{x}) + (g_{22} \circ \mathbf{x})d(x_2 \circ \mathbf{x})^2 \\ &\quad = (g_{11} \circ \mathbf{x})du^2 + 2(g_{12} \circ \mathbf{x})dudv + (g_{22} \circ \mathbf{x})dv^2 \\ &\quad = E\,du^2 + 2F\,dudv + G\,dv^2. \ \blacksquare \end{aligned}$$

The coefficients g^{ij} are determined by the inverse matrix

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} = \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

Thus,

Lemma 26.2. *The inverse coefficients g^{ij} associated to ds^2 are given by*

$$\text{(26.2)} \qquad \begin{cases} g^{11} \circ \mathbf{x} = \dfrac{G}{EG - F^2}, & g^{12} \circ \mathbf{x} = \dfrac{-F}{EG - F^2}, \\ g^{21} \circ \mathbf{x} = \dfrac{-F}{EG - F^2}, & g^{22} \circ \mathbf{x} = \dfrac{E}{EG - F^2}. \end{cases}$$

We defined the Christoffel symbols (denoted Γ^i_{jk}) for a surface $\mathcal{M}$ in $\mathbb{R}^3$ on page 538. Since Theorem 17.7 determines them purely in terms of E, F, G and their derivatives, we can just use the same formulas for a general abstract surface. The resulting definitions are consistent with those of the Christoffel symbols on a Riemannian manifold (denoted γ^i_{jk} on page 861). Indeed, the relation between the two versions is analogous to (26.1):

Lemma 26.3. *Let $\mathcal{M}$ be an abstract surface with a metric given locally by*

$$\langle\ ,\ \rangle = ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2.$$

Define Γ^i_{jk} by equation (17.7) *on page* 539*, and $\boldsymbol{\gamma}^i_{jk}$ relative to a patch $\mathbf{x}$ by* (25.32) *on page* 862*. Then*

$$\boldsymbol{\gamma}^i_{jk} \circ \mathbf{x} = \Gamma^i_{jk}, \qquad 1 \leqslant i, j, k \leqslant 2.$$

Proof. As an example, to show that $\boldsymbol{\gamma}^1_{11} \circ \mathbf{x}$ given by (25.32) coincides with Γ^1_{11} given by (17.7), we compute

$$\begin{aligned} 2\boldsymbol{\gamma}^1_{11} &= \sum_{h=1}^{2} g^{1h}\left\{2\frac{\partial g_{h1}}{\partial x_1} - \frac{\partial g_{11}}{\partial x_h}\right\} \\ &= g^{11}\frac{\partial g_{11}}{\partial x_1} + 2g^{12}\frac{\partial g_{21}}{\partial x_1} - g^{12}\frac{\partial g_{11}}{\partial x_2}. \end{aligned} \tag{26.3}$$

On the other hand, (26.1) implies that

$$\frac{\partial g_{11}}{\partial x_1} \circ \mathbf{x} = \mathbf{x}_*\left(\frac{\partial}{\partial u}\right)[g_{11}] = \frac{\partial(g_{11} \circ \mathbf{x})}{\partial u} = E_u.$$

Similarly,

$$\frac{\partial g_{21}}{\partial x_1} \circ \mathbf{x} = F_u \qquad \text{and} \qquad \frac{\partial g_{11}}{\partial x_2} \circ \mathbf{x} = E_v.$$

We substitute these expressions and (26.2) into (26.3) and get

$$2\boldsymbol{\gamma}^1_{11} \circ \mathbf{x} = \frac{GE_u - 2FF_u + FE_v}{EG - F^2} = 2\Gamma^1_{11}. \qquad \blacksquare$$

The pre-Gauss notion of Gaussian curvature K was introduced in Chapter 13, though Brioschi's formula (17.3) on page 533 shows that K depends only on the metric. We use the same formula therefore to *define* the Gaussian curvature K of an abstract surface. Definition 25.33 on page 865 describes another invariant of a Riemannian manifold, namely its *scalar curvature* s.

Theorem 26.4. *On an abstract surface, the scalar curvature equals twice the Gaussian curvature.*

Proof. In the light of Lemma 26.3, the expressions on the right-hand side of equations (19.11) on page 601 are examples of equation (25.45) on page 866. They are none other than $R^2{}_{121}$, $R^1{}_{121}$, $R^1{}_{212}$, $R^2{}_{212}$, respectively. Combining (25.44) with (19.11), we therefore have

$$\begin{cases} \rho_{11} = R^2{}_{121} = EK, \\ \rho_{12} = R^1{}_{112} = FK, \\ \rho_{21} = R^2{}_{221} = FK, \\ \rho_{22} = R^1{}_{212} = GK. \end{cases}$$

Together with Lemma 26.2, this yields

$$\begin{aligned} s = \sum_{i,j=1}^{2} g^{ij}\rho_{ij} &= \frac{1}{EG-F^2}\Big(GEK - FFK - FFK + EGK\Big) \\ &= 2K. \end{aligned}$$

This result was verified independently in Notebook 25.

Isometries and local isometries between abstract surfaces can be defined in exactly the same way as we defined isometries and local isometries between surfaces in $\mathbb{R}^n$ in Section 12.2 on page 365. Lemmas 12.6, 12.7, 12.10 and Corollary 12.8 all hold for abstract surfaces. In particular, a local isometry preserves E, F, G and the Christoffel symbols, and Gauss's *Theorema Egregium* (Theorem 17.5) is valid for isometries between abstract surfaces.

26.2 Examples of Abstract Metrics

There are many interesting metrics that do not arise from the standard dot product on $\mathbb{R}^3$. We describe a few, and state expressions for their Gaussian curvature, as computed in Notebook 26 using Brioschi's formula.

The first two examples have constant negative Gaussian curvature. We have studied many examples of surfaces in $\mathbb{R}^3$ with K constant and negative in Chapters 15 and 21, including the pseudosphere and the surfaces of Dini and Kuen. But the resulting metrics are much more complicated to describe than those immediately below. This is one very good reason to study an abstract surface without the requirement that the surface be realized in $\mathbb{R}^3$.

The Poincaré Metric on the Disk

Let u, v be standard coordinates on the disk

$$D^2(a) = \big\{\, (u,v) \in \mathbb{R}^2 \mid u^2 + v^2 < a^2 \,\big\}$$

of radius a. We define the ***Poincaré*[1] *disk metric*** by

$$ds^2 = \frac{4a^4}{(a^2 - u^2 - v^2)^2}\big(du^2 + dv^2\big), \tag{26.4}$$

for $(u, v) \in D^2(a)$. We compute the Gaussian curvature of this metric in Notebook 26; it has constant negative curvature $-1/a^2$.

It can be shown that any surface with $K \equiv -1/a^2$ is locally isometric to a portion of the Riemannian manifold $(D^2(a),\, ds^2)$. This applies in particular to the patch pseudosphere$[a]$ on page 480, defined on $\mathcal{U} = (0, 2\pi) \times (-\pi, \pi)$, but the resulting isometry realizes $\mathcal{U}$ as a *proper* subset of $D^2(a)$. It is not possible to obtain (26.4) by means of a single patch $\mathbf{y}\colon D^2(a) \to \mathbb{R}^n$ in the manner of Section 12.1 (see [dC1]).

The metric (26.4) can however be obtained by mapping $D^2(a)$ onto one half of a hyperboloid of two sheets in $\mathbb{R}^3$, when the latter is given the pseudo-Riemannian metric of signature $(2, 1)$ defined by

$$\left\langle \sum_{i=1}^{3} f_i \frac{\partial}{\partial u_i},\ \sum_{j=1}^{3} g_j \frac{\partial}{\partial u_j} \right\rangle = f_1 g_1 + f_2 g_2 - f_3 g_3. \tag{26.5}$$

This metric is a special case of (25.16) on page 856, and is often called the ***Lorentzian metric***. To emphasize that we are using this and not the Euclidean metric (25.15), we shall write $\mathbb{R}^{2,1}$ in place of $\mathbb{R}^3$.

Let $P = (x, y, z)$ be a point with $z > 0$ on the hyperboloid

$$x^2 + y^2 - z^2 = -a^2$$

where a is a positive constant (see page 314). The line joining P to the apex $(0, 0, -a)$ of the lower sheet intersects the xy-plane in

$$(u,\ v,\ 0) = \Big(\frac{ax}{z + a},\ \frac{ay}{z + a},\ 0\Big).$$

The mapping that sends this point back to P is illustrated in Figure 26.1. It determines a patch $\mathbf{y}\colon D^2(a) \to \mathbb{R}^{2,1}$, where

$$\mathbf{y}(u, v) = a\left(\frac{2au}{a^2 - u^2 - v^2},\ \frac{2av}{a^2 - u^2 - v^2},\ \frac{a^2 + u^2 + v^2}{a^2 - u^2 - v^2}\right)$$

(the calculation is in Notebook 26). The image of $\mathbf{y}$ is the whole upper sheet of the hyperboloid, and $\mathbf{y}$ is a Cartesian version of the patch hyperboloid2$[a, a, a]$ defined on page 313.

---

[1] 

Jules Henri Poincaré (1854–1912). French mathematician, widely regarded as the leading mathematician of his time. He developed the concept of automorphic functions and invented much of algebraic topology. He also worked in algebraic geometry, analytic functions of several complex variables, and number theory.

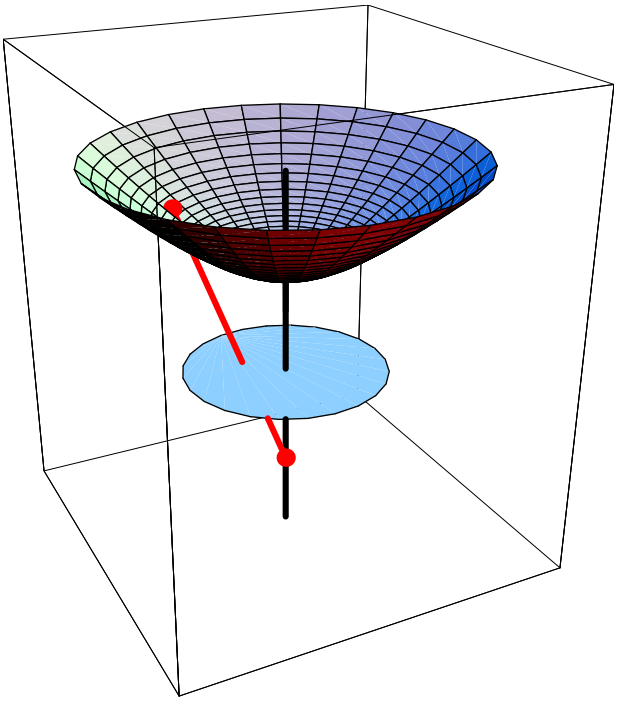


Figure 26.1: Using the hyperboloid of two sheets

The reason that $\mathbf{y}$ gives rise to the metric (26.4) is that, again using (26.5) and Notebook 26, we find

$$\langle \mathbf{y}_u, \mathbf{y}_u \rangle = \frac{4a^4}{(a^2 - u^2 - v^2)^2}, \quad \langle \mathbf{y}_u, \mathbf{y}_v \rangle = 0, \quad \langle \mathbf{y}_v, \mathbf{y}_v \rangle = \frac{4a^4}{(a^2 - u^2 - v^2)^2}.$$

These three quantities are the analogues of the functions E, F, G defined on page 362, and we can say that ds^2 in (26.4) has been "induced" by the patch $\mathbf{y}$.

The Poincaré Metric on the Upper Half-Plane

Let (x, y) be the standard coordinates on the upper half-plane

$$\mathbb{R}^2_+ = \{\, (x, y) \in \mathbb{R}^2 \mid y > 0 \,\},$$

and let a again be a positive constant. We define the ***Poincaré half-plane metric*** on this set by

$$ds^2 = \frac{a^2}{y^2}\big(dx^2 + dy^2\big). \tag{26.6}$$

Like (26.4), the metric (26.6) also has constant negative curvature $-1/a^2$. In fact, there exists a mapping of $D^2(a)$ onto $\mathbb{R}^2_+$ that is an isometry for the respective Poincaré metrics. This map is given by

$$f(u,v) = \left(\frac{-2av}{(a-u)^2+v^2},\ \frac{a^2-u^2-v^2}{(a-u)^2+v^2} \right). \tag{26.7}$$

Complex numbers can be used (see Exercise 4) to show that f maps the disk $D^2(a)$ onto $\mathbb{R}^2_+$. Figure 26.2 illustrates this fact by plotting part of the logarithmic spiral

$$\mathsf{logspiral}[1, -\tfrac{1}{100}](t) = e^{-t/100}\big(\cos t,\ \sin t\big)$$

from page 23 inside $D^2(1)$ and its image under f (with $a = 1$). Although scales have been omitted for clarity, it also illustrates the fact that $f(0,0) = (0,1)$.

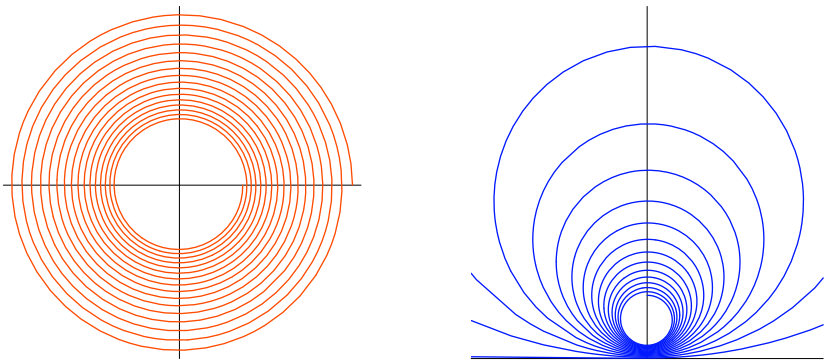


Figure 26.2: A logarithmic spiral inside $D^2(1)$ and part of its image under f

The metric (26.6) can be generalized by putting

$$ds^2 = \frac{a^m}{y^m}dx^2 + \frac{a^n}{y^n}dy^2, \tag{26.8}$$

where m, n are positive integers. We show in Notebook 26 that (26.8) has constant Gaussian curvature if and only if $n = 2$.

A Flat Metric on a Torus

The torus is a rather complicated surface in $\mathbb{R}^3$, as its nonparametric form requires a quartic polynomial in x, y, z. There is a much simpler version of the torus in $\mathbb{R}^4$, however. Let us define a patch

$$\mathsf{flattorus}[a,b]\colon (0,2\pi)\times(0,2\pi) \longrightarrow \mathbb{R}^4 \tag{26.9}$$

by

$$\mathsf{flattorus}[a,b](u,v) = \big(a\cos u,\ a\sin u,\ b\cos v,\ b\sin v\big). \tag{26.10}$$

An easy computation gives $E = a^2$, $G = b^2$ and $F = 0$. Since E and G are constant, it is immediate from Corollary 17.4, page 534, that the Gaussian curvature of $\mathsf{flattorus}[a,b]$ vanishes identically.

The form of (26.10) shows that the image of (26.9) can be identified with the Cartesian product $S^1(a) \times S^1(b)$ of two unit circles, and therefore with a torus of revolution in $\mathbb{R}^3$. However the latter has nonzero Gaussian curvature K. To emphasize the difference, $\mathsf{flattorus}[a,b]$ is called a ***flat torus***. Another consequence of (26.10) is that $\mathsf{flattorus}[a,b]$ is a submanifold of the 3-dimensional hypersphere of radius $\sqrt{a^2+b^2}$ in $\mathbb{R}^4$ centered at the origin (see page 236).

The Schwarzschild Surface

Our last example is not strictly speaking an abstract surface, since the metric we shall describe is *pseudo*-Riemannian. The ***Schwarzschild*[2] *metric*** is the 4-dimensional metric defined by

$$ds^2 = \frac{dv^2}{\left(1-\dfrac{2m}{v}\right)} + v^2 d\theta^2 + (v\sin\theta)^2 d\phi^2 - \left(1-\frac{2m}{v}\right)du^2. \tag{26.11}$$

Taking θ and ϕ to be constants in (26.11), we get a 2-dimensional metric on either the upper half-plane or the lower half-plane:

$$ds^2 = \frac{dv^2}{\left(1-\dfrac{2m}{v}\right)} - \left(1-\frac{2m}{v}\right)du^2. \tag{26.12}$$

We define the ***Schwarzschild surface*** to be the abstract surface consisting of either the upper half-plane or the lower half-plane together with the metric (26.12).

See [ON2, page 152] for a discussion of the Schwarzschild surface. Many generalizations of (26.11) occur in the theory of general relativity and (in higher dimensions) supergravity.

26.3 The Abstract Definition of Geodesic Curvature

In this section we generalize the curvature of a curve in the plane, namely $\boldsymbol{\kappa 2}$, to curves in an arbitrary surface. This is accomplished by replacing the acceleration $\boldsymbol{\alpha}''$ by $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$, but first we need to discuss orientability.

---

[2] 

Karl Schwarzschild (1873–1916). German mathematician. Professor at Göttingen and Potsdam. He died soon after returning from military service in Russia during World War I.

For surfaces in $\mathbb{R}^3$, the existence of a continuous unit normal vector field is a convenient intuitive criterion for orientability. This criterion is not in general available for abstract surfaces, but we can define orientability in terms of the existence of a complex structure. The definition for surfaces in $\mathbb{R}^3$ given on page 332 goes over almost verbatim to abstract surfaces:

Definition 26.5. *Let $\mathcal{M}$ be an abstract surface. Then $\mathcal{M}$ is* ***orientable*** *provided each tangent space $\mathcal{M}_{\mathbf{p}}$ has a complex structure $J_{\mathbf{p}}$ such that $\mathbf{p} \mapsto J_{\mathbf{p}}$ is a continuous function. An* ***oriented*** *abstract surface $(\mathcal{M}, J)$ consists of an orientable abstract surface together with a choice of the complex structure $\mathbf{p} \mapsto J_{\mathbf{p}}$.*

Note that we can regard J as an $\mathfrak{F}(\mathcal{M})$-linear map $J\colon \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ such that J^2 is equal to minus the identity. We also need an explicit formula for J on an abstract surface.

Lemma 26.6. *Suppose $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch on an oriented surface $(\mathcal{M}, J)$. Assume that $\mathcal{M}$ has a Riemannian metric, which when written in terms of $\mathbf{x}$ is $ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$. Then either*

$$J\mathbf{x}_u = \frac{-F\mathbf{x}_u + E\mathbf{x}_v}{\sqrt{EG - F^2}} \quad \text{and} \quad J\mathbf{x}_v = \frac{-G\mathbf{x}_u + F\mathbf{x}_v}{\sqrt{EG - F^2}}, \tag{26.13}$$

or the same equations hold with J replaced by $-J$.

Proof. One cannot expect to specify J uniquely, since there is an overall sign ambiguity in its definition. Since $\mathbf{x}$ is regular, $\mathbf{x}_u, \mathbf{x}_v$ are linearly independent, and we can write

$$J\mathbf{x}_u = p\mathbf{x}_u + q\mathbf{x}_v,$$

for some functions p and q, where $q \neq 0$. Then

$$E = \|\mathbf{x}_u\|^2 = \|J\mathbf{x}_u\|^2 = p^2E + 2pqF + q^2G$$

and

$$0 = \langle J\mathbf{x}_u, \mathbf{x}_u \rangle = pE + qF.$$

Hence $p = -qF/E$ and

$$E = q^2\left(\frac{F^2}{E} - 2\frac{F^2}{E} + G\right) = \frac{q^2(EG - F^2)}{E}.$$

Thus

$$q = \frac{\pm E}{\sqrt{EG - F^2}} \qquad \text{and} \qquad p = \frac{\mp F}{\sqrt{EG - F^2}},$$

and we obtain the first equation of (26.13); the second equation is proved similarly. ∎

Definition 26.7. *Let $(\mathcal{M}, J)$ be an oriented abstract surface and $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ a curve. Then for those values of t for which $\boldsymbol{\alpha}'(t) \neq 0$ the* ***geodesic curvature*** *$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]$ of $\boldsymbol{\alpha}$ is given by*

$$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t) = \frac{\big\langle \nabla_{\boldsymbol{\alpha}'(t)} \boldsymbol{\alpha}'(t),\ J\boldsymbol{\alpha}'(t) \big\rangle}{\big\|\boldsymbol{\alpha}'(t)\big\|^3}. \tag{26.14}$$

The motivation for (26.14) is equation (17.18) on page 542; in fact (26.14) reduces to (17.18) for surfaces in $\mathbb{R}^3$.

Next, we show that the geodesic curvature $\boldsymbol{\kappa}_g$ behaves in exactly the same way as $\boldsymbol{\kappa 2}$ under reparametrizations.

Theorem 26.8. *Let $\mathcal{M}$ be an oriented abstract surface, let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve and let $\boldsymbol{\gamma}\colon (c,d) \to \mathcal{M}$ be a reparametrization of $\boldsymbol{\alpha}$. Write $\boldsymbol{\gamma} = \boldsymbol{\alpha} \circ h$, where $h\colon (c,d) \to (a,b)$ is piecewise-differentiable. Then the geodesic curvatures of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are related by*

$$\boldsymbol{\kappa}_g[\boldsymbol{\gamma}](t) = (\operatorname{sign} h'(t))\, \boldsymbol{\kappa}_g[\boldsymbol{\alpha}]\big(h(t)\big) \tag{26.15}$$

wherever $h'(t)$ is nonzero.

Proof. It follows from Lemma 24.41, page 831, and Lemma 25.11, page 853, that

$$\boldsymbol{\kappa}_g[\boldsymbol{\gamma}] = \frac{\big\langle h'^2 \left(\nabla_{\boldsymbol{\alpha}'} \boldsymbol{\alpha}'\right) \circ h + h'' \boldsymbol{\alpha}' \circ h,\ h' J(\boldsymbol{\alpha}' \circ h) \big\rangle}{\big\|(\boldsymbol{\alpha}' \circ h) h'\big\|^3} = (\operatorname{sign} h')\, \boldsymbol{\kappa}_g[\boldsymbol{\alpha}] \circ h,$$

since $\langle \boldsymbol{\alpha}' \circ h,\ J(\boldsymbol{\alpha}' \circ h) \rangle = 0$. ∎

The geodesic curvature is especially simple for unit-speed curves.

Lemma 26.9. *Let $\boldsymbol{\beta}\colon (a,b) \to \mathcal{M}$ be a unit-speed curve in an oriented abstract surface $\mathcal{M}$. Then*

$$\nabla_{\boldsymbol{\beta}'} \boldsymbol{\beta}' = \boldsymbol{\kappa}_g[\boldsymbol{\beta}] J \boldsymbol{\beta}'. \tag{26.16}$$

Proof. The proof is almost word-for-word that of Lemma 1.21 on page 16. ∎

26.4 Geodesics on Abstract Surfaces

First, we duplicate part of Definition 25.10 in the Riemannian context.

Definition 26.10. *Let $\mathcal{M}$ be an abstract surface and $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ a curve. Then $\boldsymbol{\alpha}$ is a* ***geodesic*** *provided $\nabla_{\boldsymbol{\alpha}'} \boldsymbol{\alpha}' = 0$, where ∇ denotes the Riemannian connection of $\mathcal{M}$ defined on page* 857.

To motivate this definition, let us first consider the case of a surface $\mathcal{M} \subset \mathbb{R}^n$. Since a curve $\boldsymbol{\alpha}$ in $\mathcal{M}$ is also a curve in $\mathbb{R}^n$, there are two competing accelerations of $\boldsymbol{\alpha}$, namely, the second derivative $\boldsymbol{\alpha}''$ and the covariant derivative $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$. The two are related by

Lemma 26.11. *Let $\mathcal{M}$ be a surface in $\mathbb{R}^n$ and let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve. Then the tangential component of the $\mathbb{R}^n$-acceleration $\boldsymbol{\alpha}''$ of $\boldsymbol{\alpha}$ is the abstract surface acceleration $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ of $\boldsymbol{\alpha}$.*

Proof. Let $\mathbf{X}$ and $\mathbf{Y}$ be vector fields tangent to $\mathcal{M}$. A careful examination of the fundamental formula (25.19), page 857, allows us to conclude that $\nabla_{\mathbf{X}}\mathbf{Y}$ is the tangential component of $D_{\mathbf{X}}\mathbf{Y}$, where D denotes the natural connection of $\mathbb{R}^n$. In particular, $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ is the tangential component of $\boldsymbol{\alpha}'' = D_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$. ∎

More can be said in the case $n = 3$:

Lemma 26.12. *Let $\mathcal{M}$ be an orientable surface in $\mathbb{R}^3$ and choose a unit normal vector field $\mathbf{U}$ to $\mathcal{M}$. If $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ is a curve, then*

$$\boldsymbol{\alpha}'' = \nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}' + \big(S(\boldsymbol{\alpha}') \cdot \boldsymbol{\alpha}'\big)\mathbf{U}, \tag{26.17}$$

where S denotes the shape operator of $\mathcal{M}$ defined by $\mathbf{U}$.

Proof. The decomposition of $\boldsymbol{\alpha}''$ into its tangential and normal components is

$$\boldsymbol{\alpha}'' = \nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}' + (\boldsymbol{\alpha}'' \cdot \mathbf{U})\mathbf{U}.$$

But from Lemma 26.17 and Lemma 13.4, page 387, we know that $\boldsymbol{\alpha}'' \cdot \mathbf{U} = S(\boldsymbol{\alpha}') \cdot \boldsymbol{\alpha}'$, and so we get (26.17). ∎

We show next that the differential equations for a geodesic on an abstract surface are exactly the same as those for a surface in $\mathbb{R}^3$, given in Lemma 18.14 on page 566.

Lemma 26.13. *Let $\mathcal{M}$ be an abstract surface, and let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch that parametrizes $\mathcal{M}$. Then the geodesics on $\mathcal{M}$ are determined by the two second-order differential equations* (18.13)*, where $\Gamma^i_{jk} = \gamma^i_{jk} \circ \mathbf{x}$ are the Christoffel symbols of $\mathbf{x}$.*

Proof. Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a curve whose trace is contained in $\mathbf{x}(\mathcal{U})$, and let $\mathbf{X} \in \mathfrak{X}(\mathcal{M})$ be a vector field such that $\mathbf{X}_{\boldsymbol{\alpha}(t)} = \boldsymbol{\alpha}'(t)$ for $a < t < b$. Write $\mathbf{x}^{-1} = (x_1, x_2)$; then $x_1 = u \circ \mathbf{x}^{-1}$ and $x_2 = v \circ \mathbf{x}^{-1}$, where u and v are the natural coordinate functions of $\mathbb{R}^2$. Since $x_1 \circ \boldsymbol{\alpha} = u$ and $x_2 \circ \boldsymbol{\alpha} = v$, Lemma 24.39 on page 830 tells us that

$$\mathbf{X}[x_1]\big(\boldsymbol{\alpha}(t)\big) = \mathbf{X}_{\boldsymbol{\alpha}(t)}[x_1] = \boldsymbol{\alpha}'(t)[x_1] = (x_1 \circ \boldsymbol{\alpha})'(t) = u'(t). \tag{26.18}$$

Moreover,

$$\begin{aligned}\mathbf{X}^2[x_1]\big(\boldsymbol{\alpha}(t)\big) &= \mathbf{X}_{\boldsymbol{\alpha}(t)}[\mathbf{X}[x_1]] = \boldsymbol{\alpha}'(t)[\mathbf{X}[x_1]] \\ &= \big(\mathbf{X}[x_1]\circ\boldsymbol{\alpha}\big)'(t) = (x_1\circ\boldsymbol{\alpha})''(t) = u''(t). \qquad (26.19)\end{aligned}$$

Similar equations hold for $\mathbf{X}[x_2]\big(\boldsymbol{\alpha}(t)\big)$ and $\mathbf{X}^2[x_2]\big(\boldsymbol{\alpha}(t)\big)$. Lemma 24.31, page 827, implies that

$$\frac{\partial}{\partial x_1} = \mathbf{x}_u\circ\mathbf{x}^{-1} \qquad \text{and} \qquad \frac{\partial}{\partial x_2} = \mathbf{x}_v\circ\mathbf{x}^{-1}. \qquad (26.20)$$

As a special case of (25.33), page 862, we have

$$\nabla_{\mathbf{X}}\mathbf{X} = \sum_{k=1}^{2}\left\{\mathbf{X}^2[x_k] + \sum_{ij=1}^{2}\mathbf{X}[x_i]\mathbf{X}[x_j]\Gamma_{ij}^k\right\}\frac{\partial}{\partial x_k}. \qquad (26.21)$$

From (26.18)–(26.21) we obtain

$$\begin{aligned}\big(\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'\big)(t) &= \big(\nabla_{\mathbf{X}}\mathbf{X}\big)\big(\boldsymbol{\alpha}(t)\big) \\ &= \big(u'' + \Gamma_{11}^1u'^2 + 2\Gamma_{12}^1u'v' + \Gamma_{22}^1v'^2\big)\,\mathbf{x}_u \\ &\qquad + \big(v'' + \Gamma_{11}^2u'^2 + 2\Gamma_{12}^2u'v' + \Gamma_{22}^2v'^2\big)\,\mathbf{x}_v.\end{aligned}$$

Hence the vanishing of $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ is equivalent to (18.13). ∎

In view of the last result, the existence of geodesics on an abstract surface is guaranteed by Theorem 18.16 on page 567.

Geodesics in the Upper Half-Plane

Consider the upper half-plane with the Poincaré metric, using u, v in place of x, y, so that $E = a^2/v^2 = G$ and $F = 0$. Thus the identity map is a v-Clairaut patch, and equation (18.28) on page 576 becomes

$$\frac{du}{dv} = \frac{c}{\pm\sqrt{\dfrac{a^2}{v^2} - c^2}}, \qquad \text{or equivalently} \qquad \frac{c^2dv}{\sqrt{\dfrac{a^2}{v^2} - c^2}} = \pm c\,du.$$

We integrate this equation in Notebook 26 for $c \neq 0$, and obtain

$$-v\sqrt{\frac{a^2}{v^2} - c^2} = \pm c(u - u_0),$$

whence

$$v^2 + (u - u_0)^2 = \frac{a^2}{c^2},$$

for some constant u_0. Thus the geodesics of the Poincaré upper half-plane are lines (when $c = 0$) or circles that are perpendicular to the horizontal axis.

Geodesics are the analogs for a general surface of straight lines in the plane. The famous parallel postulate of Euclid is equivalent to

> ***Playfair's axiom***[3]: Given a line in the plane and a point not on the line, a unique coplanar line can be drawn through the point parallel to the given line.

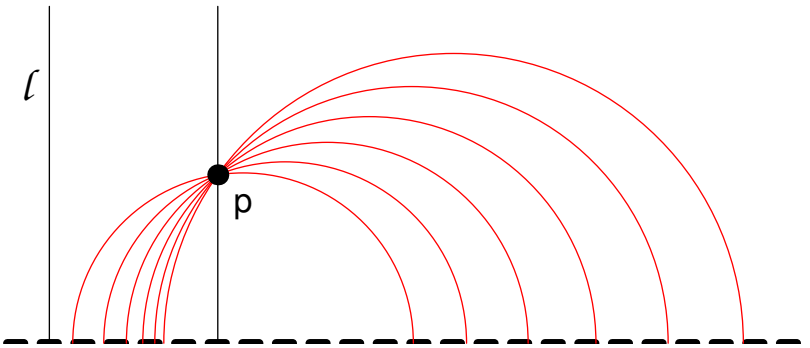


Figure 26.3: Geodesics in the upper half-plane

This axiom clearly does not hold for the Poincaré upper half-plane; in the picture above infinitely many geodesics through the point **p** do not intersect the line ℓ. Many mathematicians had doubts about the parallel postulate, but it was Gauss, Lobachevsky[4] and Bolyai[5] who first observed that the parallel postulate was independent of the other postulates of Euclid. New geometries result when the parallel postulate is replaced by either the

> ***Hyperbolic axiom***: Given a line and a point not on the line, more than one line can be drawn through the point parallel to the given line,

---

[3] John Playfair (1748–1819). Scottish mathematician, and uncle of William Henry Playfair, architect of many prominent Edinburgh buildings.

[4] 

Nikolai Ivanovich Lobachevsky (1792–1856). Russian mathematician, rector of the University of Kazan. His version of non-Euclidean geometry, published in 1829, was the first account of the subject to appear in print.

[5] 

János Bolyai (1802–1860). Hungarian mathematician. Son of Farkas Wolfgang Bolyai. János Bolyai published his version of non-Euclidean geometry as an appendix to his father's book *Tentamen*.

or the

> ***Elliptic axiom***: Given a line in the plane and a point not on the line, there are no lines parallel to the given line.

Almost 40 years elapsed until explicit models of these non-Euclidean geometries were given by Beltrami, Klein and Poincaré. In [Beltr] Beltrami showed by models that an elliptic or hyperbolic geometry is consistent, provided Euclidean geometry is consistent. For more details see [McCl].

26.5 The Exponential Map and the Gauss Lemma

The exponential[6] map provides a clever way using geodesics to map the tangent space at any point of an abstract surface onto the surface itself. Let $\mathcal{M}$ be an abstract surface with $\mathbf{p} \in \mathcal{M}$, and let $\mathbf{v} \in \mathcal{M}_{\mathbf{p}}$. By Theorem 18.16 on page 567, we know that for small $b > 0$ there exists a unique geodesic $\boldsymbol{\alpha}\colon [0, b) \to \mathcal{M}$ such that $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'(0) = \mathbf{v}$.

Definition 26.14. *Let $\mathcal{M}$ be an abstract surface and fix a point $\mathbf{p} \in \mathcal{M}$. The* ***exponential map*** *of $\mathcal{M}$ at $\mathbf{p}$ is the map $\exp_{\mathbf{p}}$ defined on a neighborhood $\mathcal{U}_{\mathbf{p}}$ of $\mathbf{0} \in \mathcal{M}_{\mathbf{p}}$ by*

$$\exp_{\mathbf{p}}(\mathbf{v}) = \boldsymbol{\alpha}_{\mathbf{v}}(1),$$

where $\boldsymbol{\alpha}_{\mathbf{v}}$ is the geodesic in $\mathcal{M}$ with $\boldsymbol{\alpha}_{\mathbf{v}}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'_{\mathbf{v}}(0) = \mathbf{v}$.

Thus

$$\exp_{\mathbf{p}}\colon \mathcal{U}_{\mathbf{p}} \longrightarrow \exp(\mathcal{U}_{\mathbf{p}}) \subseteq \mathcal{M},$$

and $\exp_{\mathbf{p}}(\mathbf{0}) = \mathbf{p}$.

We shall see now what happens when the tangent vector $\mathbf{v}$ is multiplied by a constant.

Lemma 26.15. *For any real number a, the exponential map satisfies*

$$\exp_{\mathbf{p}}(a\,\mathbf{v}) = \boldsymbol{\alpha}_{\mathbf{v}}(at),$$

where $\boldsymbol{\alpha}_{\mathbf{v}}$ is the geodesic in $\mathcal{M}$ with $\boldsymbol{\alpha}_{\mathbf{v}}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}'_{\mathbf{v}}(0) = \mathbf{v}$.

Proof. Both of the curves $\boldsymbol{\alpha}_{a\mathbf{v}}$ and $t \mapsto \boldsymbol{\alpha}_{\mathbf{v}}(at)$ are geodesics with the same initial point and the same initial velocity, namely, $a\mathbf{v}_{\mathbf{p}}$. By Theorem 18.16, they coincide on some interval containing 0. ∎

---

[6]The reason for this name derives from the discussion on page 771. For example, there is a natural Riemannian metric on $\mathrm{SO}(n)$ for which $t \mapsto e^{tA}$ is a geodesic in $\mathrm{SO}(n)$ for any $n \times n$ antisymmetric matrix A.

The next order of business is to show that $\exp_{\mathbf{p}}$ is a diffeomorphism near the origin $\mathbf{0} \in \mathcal{M}_{\mathbf{p}}$, in accordance with the definition on page 818. For this, we need the elementary observation that each tangent space to an abstract surface is itself an abstract surface.

Lemma 26.16. *Let $\mathcal{M}$ be an abstract surface. Then for each $\mathbf{p} \in \mathcal{M}$ there is a neighborhood $\mathcal{O}_{\mathbf{p}}$ of $\mathbf{0} \in \mathcal{M}_{\mathbf{p}}$ such that $\exp_{\mathbf{p}}$ maps $\mathcal{O}_{\mathbf{p}}$ diffeomorphically onto an open subset of $\mathcal{M}$.*

Proof. The differentiability of the exponential map follows from the fact (from the theory of differential equations) that the solutions of the system (18.13) on page 566 depend differentiably on the initial conditions (18.17).

To show that $\exp_{\mathbf{p}}$ is a diffeomorphism near the origin, we need to consider radial curves in the tangent space $\mathcal{M}_{\mathbf{p}}$ emanating from the origin. For $\mathbf{v} \in \mathcal{M}_{\mathbf{p}}$, let $\boldsymbol{\rho}_{\mathbf{v}} \colon \mathbb{R} \to \mathcal{M}_{\mathbf{p}}$ be the curve defined by $\boldsymbol{\rho}_{\mathbf{v}}(t) = t\mathbf{v}$. Then $\boldsymbol{\rho}_{\mathbf{v}}$ is the unique geodesic in the tangent space $\mathcal{M}_{\mathbf{p}}$ (considered as an abstract surface) starting at $\mathbf{0}$ with initial velocity $\mathbf{v}$. Similarly, Lemma 26.15 implies that the curve $t \mapsto \boldsymbol{\xi}_{\mathbf{v}}(t)$ in $\mathcal{M}$ defined by $\boldsymbol{\xi}_{\mathbf{v}} = \exp_{\mathbf{p}} \circ \boldsymbol{\rho}_{\mathbf{v}}$ is the unique geodesic in $\mathcal{M}$ starting at $\mathbf{p}$ with initial velocity $\mathbf{v}_{\mathbf{p}}$.

As an abstract surface, $\mathcal{M}_{\mathbf{p}}$ is canonically identified with its own tangent space $(\mathcal{M}_{\mathbf{p}})_0$ at the origin as follows. For $\mathbf{v} \in \mathcal{M}_{\mathbf{p}}$, let $\mathbf{v}_0 \in (\mathcal{M}_{\mathbf{p}})_0$ be the tangent vector defined by $\boldsymbol{\rho}'_{\mathbf{v}}(0)$. In fact, the tangent map $(\exp_{\mathbf{p}})_* \colon (\mathcal{M}_{\mathbf{p}})_0 \to \mathcal{M}_{\mathbf{p}}$ of $\exp_{\mathbf{p}}$ is this canonical identification because

$$(\exp_{\mathbf{p}})_*(\mathbf{v}_0) = (\exp_{\mathbf{p}} \circ \boldsymbol{\rho}_{\mathbf{p}})'(0) = \boldsymbol{\xi}'(0) = \mathbf{v}.$$

The Inverse Function Theorem now implies that $\exp_{\mathbf{p}}$ maps a neighborhood of $\mathbf{0} \in \mathcal{M}_{\mathbf{p}}$ diffeomorphically onto an open subset of $\mathcal{M}$. ∎

Definition 26.17. *Let $\mathcal{M}$ be an abstract surface and let $\mathbf{p} \in \mathcal{M}$. We say that an open subset $\mathcal{N}_{\mathbf{p}}$ of $\mathcal{M}$ containing $\mathbf{p}$ is a* ***normal neighborhood*** *of $\mathbf{p}$, provided there is an open subset $\mathcal{O}_{\mathbf{p}}$ of the tangent space $\mathcal{M}_{\mathbf{p}}$ such that $\exp_{\mathbf{p}}$ maps $\mathcal{O}_{\mathbf{p}}$ diffeomorphically onto $\mathcal{N}_{\mathbf{p}}$.*

In the special case of the Euclidean plane $\mathbb{R}^2$, any geodesic starting at $\mathbf{p} \in \mathbb{R}^2$ has the form $t \mapsto \mathbf{p} + t\mathbf{v}$. Hence the exponential map at $\mathbf{p}$ is given simply by $\exp_{\mathbf{p}}(\mathbf{v}) = \mathbf{p} + \mathbf{v}$, and any open subset of $\mathbb{R}^2$ containing $\mathbf{p}$ is a normal neighborhood of $\mathbf{p}$.

Next, we use the exponential map to define generalizations of rectangular and polar coordinates.

Definition 26.18. *Let $\mathcal{M}$ be an abstract surface and let $\mathbf{p} \in \mathcal{M}$. Choose an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ of the tangent space $\mathcal{M}_{\mathbf{p}}$, and let $\mathcal{N}_{\mathbf{p}} = \exp_{\mathbf{p}}(\mathcal{O}_{\mathbf{p}})$ be a normal neighborhood of $\mathbf{p}$.*

(i) *A* ***normal patch*** *of* $\mathcal{M}$ *at* $\mathbf{p}$ *(relative to* $\{\mathbf{e}_1, \mathbf{e}_2\}$*) is the map*

$$\mathbf{x}\colon (a,b)\times(c,d) \longrightarrow \mathcal{M} \quad \textit{given by} \quad \mathbf{x}(u,v) = \exp_{\mathbf{p}}(u\mathbf{e}_1 + v\mathbf{e}_2),$$

where the rectangle $\{\, u\mathbf{e}_1 + v\mathbf{e}_2 \mid a < u < b,\ c < v < d\,\}$ *is contained in* $\mathcal{O}_{\mathbf{p}}$.

(ii) *A* ***geodesic polar patch*** *of* $\mathcal{M}$ *at* $\mathbf{p}$ *(relative to* $\{\mathbf{e}_1, \mathbf{e}_2\}$*) is the map*

$$\mathbf{y}\colon (0,c)\times[0,2\pi) \longrightarrow \mathcal{M} \quad \textit{given by} \quad \mathbf{y}(r,\theta) = \exp_{\mathbf{p}}(\mathbf{e}_1 r\cos\theta + \mathbf{e}_2 r\sin\theta),$$

where the punctured disk $\{\, \mathbf{e}_1 r\cos\theta + \mathbf{e}_2 r\sin\theta \mid 0 < r < c,\ 0 \leqslant \theta < 2\pi\,\}$ *is contained in* $\mathcal{O}_{\mathbf{p}}$.

Since a normal patch is given by the exponential map, we have the following fact (see Exercise 11):

Lemma 26.19. *A normal patch is regular near the origin; a geodesic polar patch is normal near the origin except at the origin itself.*

Although regularity of a geodesic polar patch $\mathbf{y}$ fails at the origin, still $\mathbf{y}$ can be extended to a map $\mathbf{y}\colon [0,c)\times(-\infty,\infty) \to \mathcal{M}$ for which both partial derivatives $\mathbf{y}_r, \mathbf{y}_\theta$ are defined at $r = 0$.

Lemma 26.20. *Let* $\mathcal{N}_{\mathbf{p}}$ *be a normal neighborhood on an abstract surface* $\mathcal{M}$ *and let* $\mathbf{q} \in \mathcal{N}_{\mathbf{p}}$. *Then there is one and only one geodesic from* $\mathbf{q}$ *to* $\mathbf{p}$ *that lies entirely in* $\mathcal{N}_{\mathbf{p}}$.

Proof. Any unit-speed geodesic starting at $\mathbf{p}$ is a r-parameter curve of the polar parametrization. Since $\exp_{\mathbf{p}}\colon \mathcal{O}_{\mathbf{p}} \to \mathcal{N}_{\mathbf{p}}$ is a diffeomorphism, one such geodesic reaches $\mathbf{q}$ and is unique. ∎

Next, let us compute the first fundamental form of a geodesic polar patch.

Lemma 26.21. *Let* $\mathbf{y}$ *be a geodesic polar parametrization at a point* $\mathbf{p}$ *on an abstract surface* $\mathcal{M}$. *Then the coefficients of the first fundamental form of* $\mathbf{y}$ *satisfy* $E = 1$ *and* $F = 0$, *so that the metric of* $\mathbf{y}$ *is given by*

$$ds^2 = dr^2 + G(r,\theta)d\theta^2.$$

Proof. Since the r-parameter curves of $\mathbf{y}$ are unit-speed geodesics, we have $E = \langle \mathbf{y}_r, \mathbf{y}_r\rangle = 1$ and $\mathbf{y}_{rr} = 0$. Moreover,

$$F_r = \frac{\partial\langle \mathbf{y}_r, \mathbf{y}_\theta\rangle}{\partial r} = \langle \mathbf{y}_r, \mathbf{y}_{\theta r}\rangle = \langle \mathbf{y}_r, \mathbf{y}_{r\theta}\rangle = \tfrac{1}{2}E_\theta = 0,$$

so that F is constant on each r-parameter curve. Now $\mathbf{y}(0,\theta) = \mathbf{p}$ for all θ so $\mathbf{y}_\theta(0,\theta) = 0$, which implies that $F(0,\theta) = 0$. (F is defined at $(0,\theta)$ by the remark following Lemma 26.19.) Since F is constant on each r-parameter curve, we conclude that F is identically zero. ∎

Let $\mathcal{M}$ be an abstract surface and let $\mathbf{p} \in \mathcal{M}$. It turns out that although the exponential map $\exp_{\mathbf{p}}$ is usually *not* an isometry, it is a partial isometry, as we now explain. Once again, consider $\mathcal{M}_{\mathbf{p}}$ to be an abstract surface; as such, it is isometric to $\mathbb{R}^2$. We call the geodesics in $\mathcal{M}_{\mathbf{p}}$ emanating from $\mathbf{0} \in \mathcal{M}_{\mathbf{p}}$ ***radial***, and we call a tangent vector $\mathbf{v}$ to $\mathcal{M}_{\mathbf{p}}$ ***radial*** if $\mathbf{v}$ is tangent to such a geodesic. Similarly, we call a tangent vector to $\mathcal{M}_{\mathbf{p}}$ at $\mathbf{q} \in \mathcal{M}_{\mathbf{p}}$ ***coradial*** if it is perpendicular to the radial tangent vectors at $\mathbf{q}$. The radial geodesics in $\mathcal{M}_{\mathbf{p}}$ are mapped by $\exp_{\mathbf{p}}$ into the geodesics in $\mathcal{M}$ starting at $\mathbf{p}$. Observe that the notions of "radial" and "coradial" also make sense for tangent vectors to $\mathcal{M}$.

In paragraph 15 of his paper [Gauss2], Gauss states:

> Ductis in superficie curva ab eodem puncto initiali innumeris lineis brevissimis aequalis longitudinis, linea earum extremitates iungens ad illas singulas erit normalis,

which translates as

> If on a curved surface an infinite number of shortest lines [that is, geodesics] of equal length be drawn from the same initial point, the curve joining their extremities will be normal to each of the lines.

(The curve is a geodesic circle.) This is the first enunciation of what is now known as the ***Gauss Lemma***. In terms of the exponential map, this important result is expressed as follows.

Lemma 26.22. (Gauss Lemma) *Let $\mathcal{M}$ be an abstract surface and let $\mathbf{p} \in \mathcal{M}$. Then*:

(i) $\exp_{\mathbf{p}}$ *preserves lengths of radial tangent vectors*;

(ii) $\exp_{\mathbf{p}}$ *preserves orthogonality between coradial and radial tangent vectors.*

Proof. Let $\mathbf{y}$ be a geodesic polar patch on $\mathcal{M}$ at $\mathbf{p}$; then the composition $\mathbf{y}^{-1} \circ \exp_{\mathbf{p}}$ gives the ordinary polar coordinates at $\mathbf{p}$. Let E, F, G denote the coefficients of the first fundamental form of $\mathbf{y}$. Part (i) is a reformulation of the fact from Lemma 26.21 that $E = 1$, and part (ii) is a reformulation of the fact that $F = 0$. ∎

On the other hand, most of the time $\exp_{\mathbf{p}}$ *does* change the lengths and inner products of the coradial vectors; this accounts for the fact that it is almost never an isometry.

Theorem 18.16 gives no information concerning the size of the interval on which the geodesic satisfying given initial conditions is defined. Let $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ be geodesics satisfying the same initial conditions. Since the two geodesics must coincide on some interval, it is clear that there is a geodesic $\boldsymbol{\gamma}_3$ whose domain

of definition contains those of both $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ and coincides wherever possible with the two geodesics. Applying this consistency result, we obtain a single ***maximal geodesic***. The domain of a maximal geodesic may or may not be the whole real line, but in any case we can state

Definition 26.23. *A surface $\mathcal{M}$ is* ***geodesically complete*** *provided that every maximal geodesic is defined on all of $\mathbb{R}$.*

Trivially, a constant curve is a maximal geodesic. Excluding this case, we see that geodesic completeness means that every maximal geodesic is infinitely long in both directions. It is easy to check that spheres and planes are geodesically complete. More generally, any compact surface is geodesically complete; this is a consequence of a famous theorem of Hopf and Rinow, which we shall state at the end of the chapter.

It is easy to find examples of geodesically incomplete surfaces. For example, the punctured plane $\big\{(x, y, 0) \mid x, y \in \mathbb{R},\, (x, y) \neq (0, 0)\big\}$ cannot be geodesically complete, because the geodesic $t \mapsto (t, 0)$ cannot be extended to a map of the whole real line into $\mathbb{R}^2 \setminus (0, 0)$.

Theorem 26.24. *A pseudosphere in $\mathbb{R}^3$ is not geodesically complete.*

Proof. By Corollary 15.23 on page 480, each meridian of a pseudosphere is a tractrix. The curve $\boldsymbol{\alpha}$ given by (2.18), page 52, is a unit-speed parametrization of a tractrix. Hence it gives rise to a geodesic whose trace is a meridian. From (2.18) we see that $\boldsymbol{\alpha}$ is not differentiable at 0, so that $\boldsymbol{\alpha}$ cannot be a geodesic defined on all of $\mathbb{R}$. More generally, no constant speed parametrization of a tractrix can be defined on the whole real line $\mathbb{R}$. ∎

By contrast, the disk and upper half-plane with their Poincaré metrics *are* complete. A more significant example is given in Section 27.3 in the next chapter.

26.6 Length Minimizing Properties of Geodesics

The definition of the distance function $\boldsymbol{\rho}$ given on page 369 for a surface in $\mathbb{R}^n$ also works for an abstract surface $\mathcal{M}$. Thus for $\mathbf{p}, \mathbf{q} \in \mathcal{M}$,

$$\boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) \;=\; \text{the greatest lower bound of the lengths of all piecewise-differentiable curves } \boldsymbol{\alpha} \text{ connecting } \mathbf{p} \text{ to } \mathbf{q}.$$

Definition 26.25. *Let $\mathcal{M}$ be an abstract surface, and let $\boldsymbol{\alpha}$ be a piecewise-differentiable curve connecting $\mathbf{p}, \mathbf{q} \in \mathcal{M}$. We say that $\boldsymbol{\alpha}$* ***minimizes arc length*** *from $\mathbf{p}$ to $\mathbf{q}$ provided*

$$\boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) = \mathsf{length}[\boldsymbol{\alpha}],$$

and any other piecewise-differentiable curve connecting $\mathbf{p}, \mathbf{q}$ and satisfying the same equality is a reparametrization of $\boldsymbol{\alpha}$.

Next, if $\mathbf{p}$ and $\mathbf{q}$ are not too far apart, we find that a curve whose length equals $\boldsymbol{\rho}(\mathbf{p},\mathbf{q})$ is a geodesic:

Theorem 26.26. *For each point $\mathbf{q}$ in a normal neighborhood $\mathcal{N}_\mathbf{p}$ of an abstract surface $\mathcal{M}$, the radial geodesic from $\mathbf{p}$ to $\mathbf{q}$ uniquely minimizes arc length.*

Proof. Let $\mathbf{y}$ be a geodesic polar patch that parametrizes $\mathcal{N}_\mathbf{p}$. We can write $\mathbf{q} = \mathbf{y}(r_0,\theta_0)$; the unique geodesic connecting $\mathbf{p}$ and $\mathbf{q}$ is the radial geodesic $\boldsymbol{\gamma}\colon [0,r_0] \to \mathcal{N}_\mathbf{p}$ defined by $\boldsymbol{\gamma}(r) = \mathbf{y}(r,\theta_0)$. Any other curve $\boldsymbol{\alpha}$ from $\mathbf{p}$ to $\mathbf{q}$ can be parametrized on the same interval $[0,r_0]$. We want then to show that

$$\mathsf{length}[\boldsymbol{\gamma}] \leqslant \mathsf{length}[\boldsymbol{\alpha}]. \tag{26.22}$$

Discarding any loop that $\boldsymbol{\alpha}$ makes at $\mathbf{p}$ only shortens $\boldsymbol{\alpha}$; therefore, we can assume that $\boldsymbol{\alpha}$ never returns to $\mathbf{p}$.

First, consider the case when $\boldsymbol{\alpha}$ never leaves $\mathcal{N}_\mathbf{p}$. Since $\boldsymbol{\alpha}$ never returns to $\mathbf{p}$, there exist differentiable functions $a_1, a_2\colon [0,r_0] \to \mathbb{R}$ such that $\boldsymbol{\alpha}(t) = \mathbf{y}(a_1(t),a_2(t))$. Since $\boldsymbol{\alpha}(0) = \mathbf{p}$ and $\boldsymbol{\alpha}(r_0) = \mathbf{q}$, we have

$$a_1(0) = 0, \quad a_1(r_0) = r_0, \quad a_2(0) = \theta_0 \quad \text{and} \quad a_2(r_0) = \theta_0 + 2n\pi,$$

for some integer n. Since $\mathbf{y}$ is a geodesic polar patch, we have $E = 1$ and $F = 0$, and so the speed of $\boldsymbol{\alpha}$ satisfies

$$\|\boldsymbol{\alpha}'\| = \sqrt{a_1'^2 + G a_2'^2} \geqslant |a_1'| \geqslant a_1'.$$

Hence

$$\mathsf{length}[\boldsymbol{\alpha}] = \int_0^{r_0} \sqrt{a_1'^2 + G a_2'^2}\, dt \geqslant \int_0^{r_0} a_1'\, dt = a_1(r_0) - a_1(0) = r_0. \tag{26.23}$$

The radial geodesic $\boldsymbol{\gamma}$ has unit speed; that is,

$$\mathsf{length}[\boldsymbol{\gamma}] = \int_0^{r_0} dt = r_0. \tag{26.24}$$

Then (26.22) follows from (26.23) and (26.24).

On the other hand, if $\boldsymbol{\alpha}$ does not stay in $\mathcal{N}_\mathbf{p}$, then it must cross the circle $\theta \mapsto \mathbf{y}(r_0,\theta)$; by the proof above, the portion of $\boldsymbol{\alpha}$ inside this circle must have length at least $\mathsf{length}[\boldsymbol{\gamma}]$. It follows that in this case $\mathsf{length}[\boldsymbol{\gamma}] < \mathsf{length}[\boldsymbol{\alpha}]$.

Finally, suppose that $\mathsf{length}[\boldsymbol{\gamma}] = \mathsf{length}[\boldsymbol{\alpha}]$. Then $\boldsymbol{\alpha}$ cannot cross the circle $\theta \mapsto \mathbf{y}(r_0,\theta)$ and (26.23) becomes an equality. Hence

$$\sqrt{a_1'^2 + G a_2'^2} = a_1'. \tag{26.25}$$

Since $G > 0$, (26.25) implies that $a_2' = 0$. Thus a_2 has the constant value θ_0, and so $\boldsymbol{\alpha}(t) = \mathbf{y}(a_1(t),\theta_0)$, showing that $\boldsymbol{\alpha}$ is indeed a reparametrization of $\boldsymbol{\gamma}$. ∎

Lemma 26.27. *Let $\mathcal{M}$ be an abstract surface, and let $\mathbf{p}, \mathbf{q} \in \mathcal{M}$. Suppose that $\boldsymbol{\alpha}\colon [a,b] \to \mathcal{M}$ is a curve on $\mathcal{M}$ such that the length of $\boldsymbol{\alpha}$ is shortest among all curves connecting $\mathbf{p} = \boldsymbol{\alpha}(a)$ to $\mathbf{q} = \boldsymbol{\alpha}(b)$. Then $\boldsymbol{\alpha}$ is a geodesic.*

Proof. Under the assumption that $\boldsymbol{\alpha}$ is not a geodesic, we show that

$$\mathsf{length}[\boldsymbol{\alpha}] > \boldsymbol{\rho}(\mathbf{p}, \mathbf{q}).$$

If $\boldsymbol{\alpha}$ is not a geodesic, then there exists t_0 with $\big(\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}\big)'(t_0) \neq 0$. Since $\nabla_{\boldsymbol{\alpha}'}\boldsymbol{\alpha}'$ is continuous, it is nonzero near t_0; thus without loss of generality, $a < t_0 < b$. There exists $\epsilon > 0$ such that the restriction of $\boldsymbol{\alpha}$ to the interval $[t_0 - \epsilon, t_0 + \epsilon]$ is not a geodesic but lies in a normal neighborhood of $\boldsymbol{\alpha}(t_0)$. Then Theorem 26.26 implies that the length $\mathsf{length}[\boldsymbol{\alpha}\big|[t_0, t_0+\epsilon]]$ of the restriction of $\boldsymbol{\alpha}$ to the interval $[t_0, t_0+\epsilon]$ is strictly greater than $\boldsymbol{\rho}\big(\boldsymbol{\alpha}(t_0), \boldsymbol{\alpha}(t_0+\epsilon)\big)$. But then, using the triangle inequality, we have

$$\begin{aligned}
\mathsf{length}[\boldsymbol{\alpha}] &= \mathsf{length}[\boldsymbol{\alpha}\big|[a,t_0]] + \mathsf{length}[\boldsymbol{\alpha}\big|[t_0, t_0+\epsilon]] + \mathsf{length}[\boldsymbol{\alpha}\big|[t_0+\epsilon, b]] \\
&> \boldsymbol{\rho}\big(\mathbf{p}, \boldsymbol{\alpha}(t_0)\big) + \boldsymbol{\rho}\big(\boldsymbol{\alpha}(t_0), \boldsymbol{\alpha}(t_0+\epsilon)\big) + \boldsymbol{\rho}\big(\boldsymbol{\alpha}(t_0+\epsilon), \mathbf{q}\big) \\
&\geqslant \boldsymbol{\rho}(\mathbf{p}, \mathbf{q}). \ \blacksquare
\end{aligned}$$

An Abstract Surface as a Metric Space

Let us recall an important definition from topology.

Definition 26.28. *A set $\mathscr{M}$ is called a **metric space** if there exists a function $d\colon \mathscr{M} \times \mathscr{M} \to \mathscr{M}$ (called the **distance function** of $\mathscr{M}$) with the properties:*

(i) $d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p})$,

(ii) $d(\mathbf{p}, \mathbf{r}) \leqslant d(\mathbf{p}, \mathbf{q}) + d(\mathbf{q}, \mathbf{r})$,

(iii) $d(\mathbf{p}, \mathbf{q}) \geqslant 0$,

(iv) $d(\mathbf{p}, \mathbf{q}) = 0$ *if and only if* $\mathbf{p} = \mathbf{q}$,

for $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathscr{M}$.

The simplest nontrivial example of a metric space is $\mathbb{R}^n$ with the standard distance function

$$d(\mathbf{p}, \mathbf{q}) = \mathsf{distance}(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|, \qquad \mathbf{p}, \mathbf{q} \in \mathbb{R}^n.$$

However, the notion of metric space is much more general. It is important for the theory of abstract surfaces, because of the following result:

Theorem 26.29. *The distance function $\boldsymbol{\rho}$ of an abstract surface $\mathcal{M}$ gives $\mathcal{M}$ the structure of a metric space.*

Proof. That $\boldsymbol{\rho}$ satisfies (i) in the definition of metric space is clear from the fact that each parametrized curve gives rise to a parametrized curve in the opposite direction, and (iii) is obvious from the fact that the length of any curve is nonnegative. To prove that (ii) is satisfied, let $\epsilon > 0$ and let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be piecewise-differentiable curves on $\mathcal{M}$ connecting $\mathbf{p}$ to $\mathbf{q}$ and $\mathbf{q}$ to $\mathbf{r}$, respectively, such that

$$\mathsf{length}[\boldsymbol{\alpha}] < \boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) + \frac{\epsilon}{2} \qquad \text{and} \qquad \mathsf{length}[\boldsymbol{\beta}] < \boldsymbol{\rho}(\mathbf{q}, \mathbf{r}) + \frac{\epsilon}{2}.$$

We can parametrize the union of the traces of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as a new piecewise-differentiable curve $\boldsymbol{\gamma}$; then

$$\boldsymbol{\rho}(\mathbf{p}, \mathbf{r}) \leqslant \mathsf{length}[\boldsymbol{\gamma}] = \mathsf{length}[\boldsymbol{\alpha}] + \mathsf{length}[\boldsymbol{\beta}] < \boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) + \boldsymbol{\rho}(\mathbf{q}, \mathbf{r}) + \epsilon.$$

Since ϵ is arbitrary, we get (ii).

Finally, to prove that (iv) is satisfied, we proceed as follows. Suppose that $\boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) = 0$, but $\mathbf{p} \neq \mathbf{q}$. Let $\mathcal{N}_{\mathbf{p}}$ be a normal neighborhood of $\mathbf{p}$ with $\mathbf{q} \notin \mathcal{N}_{\mathbf{p}}$; there exists $\epsilon > 0$ such that

$$\{\mathbf{r} \in \mathcal{M} \mid \boldsymbol{\rho}(\mathbf{p}, \mathbf{r}) \leqslant \epsilon\} \subseteq \mathcal{N}_{\mathbf{p}}.$$

By definition of $\boldsymbol{\rho}$, there exists a piecewise-differentiable curve $\boldsymbol{\alpha}\colon [a, b] \to \mathcal{M}$ with $\boldsymbol{\alpha}(a) = \mathbf{p}$ and $\boldsymbol{\alpha}(b) = \mathbf{q}$ such that $\mathsf{length}[\boldsymbol{\alpha}] < \epsilon$. Because $\boldsymbol{\alpha}\big([a, b]\big)$ is connected, there exists t_0 with $a < t_0 < b$ such that $\boldsymbol{\rho}\big(\mathbf{p}, \boldsymbol{\alpha}(t_0)\big) = \epsilon$. Theorem 26.26 then implies that $\mathsf{length}[\boldsymbol{\alpha}] \geqslant \epsilon$. This contradiction shows that $\mathbf{p} = \mathbf{q}$. ∎

There is a notion of completeness for metric spaces.

Definition 26.30. *Suppose that $\mathscr{M}$ is a metric space with distance function d. A sequence of points $\{\mathbf{p}_n\}$ in $\mathscr{M}$ is called a* ***Cauchy sequence****, if for each $\epsilon > 0$ there is an index $N = N(\epsilon)$ such that $d(\mathbf{p}_i, \mathbf{p}_j) < \epsilon$ whenever $i, j \geqslant N$. We say that $\mathscr{M}$ is a* ***complete metric space****, provided every Cauchy sequence converges to some point of $\mathscr{M}$.*

For example, $\mathbb{R}^2 \setminus (0, 0)$ with the standard metric is not complete because the sequence $\{(1/n, 1/n)\}$ converges to no point in $\mathbb{R}^2 \setminus (0, 0)$.

Thus for an abstract surface we have two competing notions of completeness. In fact, they coincide.

Lemma 26.31. *Let $\mathcal{M}$ be a connected abstract surface. If $(\mathcal{M}, \boldsymbol{\rho})$ is a complete metric space, then $\mathcal{M}$ is geodesically complete.*

Proof. Let $\boldsymbol{\beta}\colon [0, b) \mapsto \mathcal{M}$ be a unit-speed geodesic. If $\{t_j\}$ is a sequence in $[0, b)$ with $t_j \to b$, then $\{\boldsymbol{\beta}(t_j)\}$ is a Cauchy sequence in $\mathcal{M}$, since

$$\boldsymbol{\rho}\big(\boldsymbol{\beta}(t_j), \boldsymbol{\beta}(t_k)\big) \leqslant \mathsf{length}[\boldsymbol{\beta}\big|[t_j, t_k]] = |t_j - t_k|.$$

Hence $\{\boldsymbol{\beta}(t_j)\}$ converges to some $\mathbf{q} \in \mathcal{M}$. Let $\mathcal{N}_{\mathbf{q}}$ be a normal neighborhood of $\mathbf{q}$; since $\boldsymbol{\beta}$ is a geodesic in $\mathcal{N}_{\mathbf{q}}$ that reaches $\mathbf{q}$, the exponential map at $\mathbf{q}$ can be used to extend $\boldsymbol{\beta}$ as a geodesic past $\mathbf{q}$. In this way $\boldsymbol{\beta}$ can be extended to a geodesic defined on all of $[0, \infty)$. Similarly, $\boldsymbol{\beta}$ can be extended to a geodesic defined on all of $(-\infty, 0]$, and thus on the whole real line. ▌

Lemma 26.31 has the following generalization:

Theorem 26.32. (Hopf-Rinow) *Let $\mathcal{M}$ be a connected abstract surface. Then the following conditions are equivalent*:

(i) $(\mathcal{M}, \boldsymbol{\rho})$ *is a complete metric space*;

(ii) $\mathcal{M}$ *is geodesically complete*;

(iii) *for each* $\mathbf{p} \in \mathcal{M}$, *the expression* $\exp_{\mathbf{p}}(\mathbf{v})$ *is defined for* every $\mathbf{v} \in \mathcal{M}_{\mathbf{p}}$.

Furthermore, each of (i), (ii), (iii) *implies that for any* $\mathbf{p}, \mathbf{q} \in \mathcal{M}$, *there is a geodesic* $\boldsymbol{\beta}\colon [a, b] \to \mathcal{M}$ *such that* $\boldsymbol{\beta}(a) = \mathbf{p}$, $\boldsymbol{\beta}(b) = \mathbf{q}$ *and* $\boldsymbol{\rho}(\mathbf{p}, \mathbf{q}) = \mathsf{length}[\boldsymbol{\beta}]$.

For a proof [Hicks] or [ON2].

26.7 Exercises

M **1.** Find the curvature of the following metrics:

(a) $ds^2 = (u^m + v^n)du^2 + dv^2$;

(b) $ds^2 = \dfrac{dv^2}{\left(1 - \dfrac{2m}{v} + f(v)\right)} - \left(1 - \dfrac{2m}{v} + f(v)\right) du^2.$

M **2.** A generalization of the flat torus is defined by

$$\mathsf{flattorus}[a, b, c, d](u, v) = \big(a\cos u,\ b\sin u,\ c\cos v,\ d\sin v\big).$$

Show that the coefficients of the induced metric from $\mathbb{R}^4$ are in general nonconstant, but still the Gaussian curvature is zero.

3. A patch on a surface for which the metric has the form

$$ds^2 = (f(u) + g(v))\big(du^2 + dv^2\big)$$

is called a ***Liouville patch***. Compute the Gaussian curvature and Christoffel symbols of a Liouville patch.

4. Let z be a complex number. Show that (or explain why) $\mathfrak{Im}\, z > 0$ if and only if $|z - i| < |z + i|$. Deduce that

$$w = a\frac{z - i}{z + i}$$

lies in $D^2(a)$ if and only if z belongs to $\mathbb{R}^2_+$. Show that the mapping (26.7) corresponds to the inverse function $z = f(w)$ with $w = u + iv$.

5. Show that the substitution $x = v$, $y = e^t$ transforms the Poincaré metric (26.6) into

$$ds^2 = a^2\big(dt^2 \,+\, e^{-2t}dv^2\big).$$

Deduce from Exercise 3 on page 710 that the pseudosphere is locally isometric to (26.6). Further, use the description on page 477 to define an isometry mapping the pseudosphere (minus one meridian) to a subset of the Poincaré upper half-plane $\mathbb{R}^2_+$.

M **6.** An embedding of a Klein bottle into $\mathbb{R}^4$ is given by

$$\mathsf{klein4}[a,b](u,v) \;=\; \Big((a + b\cos v)\cos u,\;(a + b\cos v)\sin u,\\ b\sin v\cos\frac{u}{2},\;b\sin v\,\sin\frac{u}{2}\Big).$$

Show that $\mathsf{klein4}[a,b]$ has nonzero Gaussian curvature. Understand how the four projections of $\mathsf{klein4}[8,3]$ into $\mathbb{R}^3$ were plotted in Notebook 26.

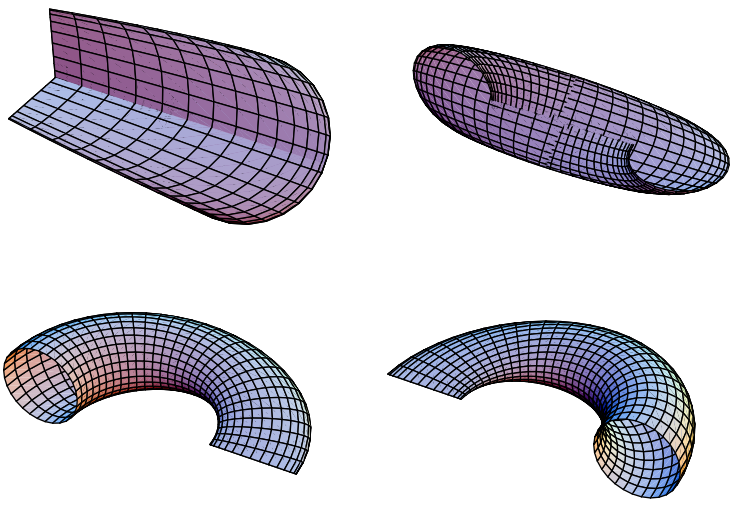


Figure 26.4: Four projections of $\mathsf{klein4}[8,3]$ into $\mathbb{R}^3$

7. Prove the following Gauss equations for abstract surfaces:

Theorem 26.33. *Let* $\mathbf{x}$ *be a regular patch on an abstract surface* $\mathcal{M}$. *Then*

$$\left\{\begin{array}{l} \nabla_{\mathbf{x}_u}(\mathbf{x}_u) \;=\; \Gamma^1_{11}\mathbf{x}_u + \Gamma^2_{11}\mathbf{x}_v, \\ \nabla_{\mathbf{x}_u}(\mathbf{x}_v) \;=\; \Gamma^1_{12}\mathbf{x}_u + \Gamma^2_{12}\mathbf{x}_v \;=\; \nabla_{\mathbf{x}_v}(\mathbf{x}_u), \\ \nabla_{\mathbf{x}_v}(\mathbf{x}_v) \;=\; \Gamma^1_{22}\mathbf{x}_u + \Gamma^2_{22}\mathbf{x}_v. \end{array}\right. \tag{26.26}$$

8. Prove Lemma 27.7.

M **9.** A ***Veronese surface***[7] in $\mathbb{R}^5$ is defined by

$$\mathsf{veronese}[a](u,v) = a(u,\ v,\ u^2,\ uv,\ v^2).$$

Compute the metric, Gaussian curvature and mean curvature vector of $\mathsf{veronese}[a]$. Plot the Gaussian curvature. What is the relation of a Veronese surface to a Whitney umbrella defined on page 307?

M **10.** Compute the Christoffel symbols of the following patches:

$$\mathsf{catenoid}[a], \qquad \mathsf{helicoid}[a,b], \qquad \mathsf{veronese}[a], \qquad \mathsf{klein4}[a,b],$$

whose definitions can all be found in Notebook 26. Display the results in tabular form.

11. Give the details of the proof of Lemma 26.19.

12. Let $\mathbf{p}, \mathbf{v} \in S^2(1) \subset \mathbb{R}^3$ with $\mathbf{p} \cdot \mathbf{v} = 0$, so that $\mathbf{v_p}$ belongs to the tangent space $S^2(1)_{\mathbf{p}}$ to the sphere at $\mathbf{p}$. Find the formula for $\exp_{\mathbf{p}}(t\,\mathbf{v_p})$ for $t \in \mathbb{R}$.

13. Let $\{\mathbf{p}_j\}$ and $\{\mathbf{q}_j\}$ be sequences of points on an abstract surface $\mathcal{M}$ such that in the topology of the surface

$$\lim_{j\to\infty} \mathbf{p}_j = \mathbf{p} \qquad \text{and} \qquad \lim_{j\to\infty} \mathbf{q}_j = \mathbf{q}.$$

Show that

$$\lim_{j\to\infty} \boldsymbol{\rho}(\mathbf{p}_j, \mathbf{q}_j) = \boldsymbol{\rho}(\mathbf{p}, \mathbf{q}),$$

where $\boldsymbol{\rho}$ is defined on page 888.

---

[7] Giuseppe Veronese (1854–1917). Famous Italian algebraic geometer, professor at the University of Padua.

Notebook 26

N26.0

In this notebook, we shall investigate abstract metrics and their curvature. We begin in N26.2 by performing some elementary calculations relevant to the definition of the Poincaré metric on the disk.

Our programs are mainly limited to backing up material in Sections 26.2 and the Exercises. It is useful to begin by recalling the miniprogram from N12.1, which computes the first fundamental form, if only to emphasize that this works perfectly well for surfaces in Euclidean space of dimension greater than 3.

```
metric[x_][u_,v_]:=
Module[{xu,xv,U,V,si},
  si= Simplify[#]/.{U->u,V->v} &;
  xu= D[x[U,V], U];
  xv= D[x[U,V], V];
  ee= xu.xu//si;
  ff=  xu.xv//si;
  gg= xv.xv//si;
  {{ee,ff},{ff,gg}}
]
```

The program brioschi for computing the Gaussian curvature implicitly, when a metric is fed in, is loaded in N26.0.1, togther with other relevant commands.

N26.0.1

N26.2

The code for Figure 26.1 is given below. We must first determine the Cartesian coordinates (x,y,z) of the image of the point (u,v) in the unit disk. So we merely set up the equations and solve them.

```
eq:= {a x==(z+a)u, a y==(z+a)v, x^2+y^2-z^2==-a^2}
```

```
Solve[eq, {x,y}, z]
Solve[eq, z, {x,y}]
```

The output from this cell provides the right-hand side of the next definition that is used to check the solution and (then) to plot Figure 26.1.

```
image[a_][u_,v_]:= a{2a u, 2a v, a^2+u^2+v^2}/(a^2-u^2-v^2)
```

```
F:= image[a][u,v]
F[[1]]^2 + F[[2]]^2 - F[[3]]^2 //FullSimplify
```

The command polarize from N10.4, essentially an abbreviation, was reloaded in N26.0.1. The mathematical construction underlying Figure 26.1 is really the same as that used in Figure 22.2 (stereographic projection from a sphere) .

```
Module[{a,f,P,Q,ta,z,y},
  a= 1.5;
  f= polarize[image[a]];
  P:= f[.45a, π/4]+{0,0,.01};
  Q= {0,0,a};
  ta:= Table[a{Cos[2π k/24], Sin[2π k/24],0}, {k,24}];
  z[2]:= ParametricPlot3D[f[r,θ], {r,0,.6a}, {θ,0,2π}];
  y[3]= {Hue[0], PointSize[.03], Point[P], Point[-Q]};
  y[4]= {Hue[0],Thickness[.01], Line[{P,-Q}]};
  y[5]= {Thickness[.01], Line[{2Q,-2Q}]};
  z[1]= {Polygon[ta]}//Graphics3D;
  z[i_]= Graphics3D[y[i]];
  S[Array[z,5], ViewPoint->{-2,1,1},
    PlotRange->{-1.7a,3a}]
]
```

Figure 26.1

The Lorentzian metric in three dimensions is

```
lor[u_,v_]:= u[[1]]v[[1]]+u[[2]]v[[2]]-u[[3]]v[[3]]//Simplify
```

In order to compute analogs of the first fundamental form coefficients E,F,G, for the hyperboloid in $R^{2,1}$, we need to modify the program metric in N26.0 by replacing the dot product by lor. This is achieved in the next cell.

```
lorentz[a_][u_,v_]:=
Module[{ee,ff,gg},
  xu= D[image[a][U,V], U];
  xv= D[image[a][U,V], V];
  ee:= lor[xu,xu];
  ff:= lor[xu,xv];
  gg:= lor[xv,xv];
  {{ee,ff},{ff,gg}}/.{U->u,V->v}
]
```

The resulting Gaussian curvature of the Poincaré disk is then given by

```
lorentz[a][u,v]
brioschi[lorentz[a]][u,v]
```

A modified metric M3 for the upper half-plane depends on three parameters a,m,n. It turns out that its Gaussian curvature is only constant if m=0, n=2, or n=m+2:

```
M3[x_,y_]:= {{(a/y)^m, 0}, {0, (a/y)^n}}
brioschi[M3][u,v]
```

Here is the Schwartzchild metric and *its* Gaussian curvature:

```
M[m_][u_,v_]:= {{(2m-1)/v, 0}, {0, 1/(1-2m/v)}}
brioschi[M[m]][u,v]
```

N26.2.1 contains some calculations regarding the mapping from the disk to the upper half-plane, and the code for Figure 26.2.

■ N26.2.1

■ N26.4

An integration was needed to find geodesics on the upper half-plane.

```
Integrate[c^2/Sqrt[1/(λ^2 v^2) - c^2], v]//Simplify
```

They are represented diagrammatically by

```
Module[{y,z},
  y[1]={Thickness[.01], Dashing[{.02,.02}],
    Line[{{-2.5,0}, {7,0}}]};
  y[2]= {Line[{{0,0},{0,4}}], Line[{{-2,0},{-2,4}}]};
  y[3]= {Hue[0], Table[Circle[{a,0}, Sqrt[4+a^2], {0,π}],
    {a, .3, 2.8, .5}]};
  y[4]= {txs["l",{-2.3,3}], tx["p",{.3,1.7}]};
  y[5]= {PointSize[.025], Point[{0,2}]};
  z[i_]:= Graphics[y[i]];
  S[Array[z,5], AspectRatio->Automatic, PlotRange->{0,5},
    ImageSize->550]
]
```

Figure 26.3

This used the text-placing commands

```
tx[A_,v_]:= Text[FontForm[A,{"Helvetica", 24}], v]
txs[A_, v_]:= Text[FontForm[A, {"Script", 32}], v]
```

We now skip to Section 26.7 and the chapter's exercises.

■ N26.7

The Klein bottle is embedded into 4-space by means of the patch

```
klein4[a_,b_][u_,v_]:= {(a+b Cos[v])Cos[u], (a+b Cos[v])Sin[u],
  b Sin[v]Cos[u/2], b Sin[v]Sin[u/2]}
```

The next program constructs orthogonal projections from each of the four coordinates in turn. The command Drop is used to delete exactly one component from the chart.

```
orth4d[surf_,{u_,u1_,u2_},{v_,v1_,v2_}]:=
Module[{d,z},
  d[i_]:= Drop[surf, {i}];
  z[i_]:= ParametricPlot3D[d[i]//Evaluate, {u,u1,u2}, {v,v1,v2}];
  GraphicsArray[{{z[1],z[2]},{z[3],z[4]}},
    GraphicsSpacing->{0,-.2}]
]
```

```
S @ orth4d[klein4[8,3][u,v], {u,0,π}, {v,-π,π}]
```

Figure 26.4

It is easy to verify that the flat torus has constant first fundamental form:

```
flattorus[a_,b_,c_,d_][u_,v_]:=
  {a Cos[u], b Sin[u], c Cos[v], d Sin[v]}
```

As explained in N26.0, we are free to use the miniprogram metric to compute its first fundamental form.

```
ft:= flattorus[a,b,c,d]
metric[ft][u,v]
```

The output shows that this is only constant if a=b and c=d. But we can check that the Gaussian curvature is zero regardless:

```
brioschi[metric[ft]][u,v]
```

We project the Veronese surface in various ways to compare it with the Whitney umbrella of N10.4. Add S@ to display the plots.

```
veronese[a_][u_,v_]:= a{u, v, u^2, u v, v^2}
```

```
ve:= veronese[1]
dropve[A_]:= ParametricPlot3D[Drop[ve[u,v], A], {u,-1,1},
  {v,-1,1}]
ga:= {dropve[2], dropve[{2,3}], dropve[{3,4}]}
GraphicsArray[ga]
```

Now we find the first fundamental form of the Veronese surface with a=1:

```
Mve:= metric[ve]
Mve[u,v]
```

We compute and plot its Gaussian curvature:

```
gve:= brioschi[Mve]
gve[u,v]
Plot3D[gve[u,v], {u,-1,1},{v,-1,1}, ColorFunction->Hue]
```

A cosmetic program for displaying the classical first fundamental form is next applied to a number of surfaces, defined in N26.0.1.

```
ds2[x_][u_,v_]:=
Module[{},
  metric[x][u,v];
  ee d[u]^2+2ff d[u]d[v] +  gg d[v]^2
]
```

```
ds2[sphere[a]][u,v]
ds2[hyperbolicparaboloid][u,v]
ds2[monkeysaddle][u,v]
ds2[enneper][u,v]
ds2[catenoid[a]][u,v]
ds2[helicoid[a,b]][u,v]
```

The following definition is the Cartesian version of the generalization of Plücker's conoid, introduced in polar form in N14.1.1.

```
pluecker[m_,n_,a_][u_,v_]:=
  a{u, v, (u+v)^(m/2) Sin[n ArcTan[v/u]]}
```

```
ParametricPlot3D[pluecker[4,2,1][u,v], {u,-1,1}, {v,-1,1}]
```

```
ds2[pluecker[2,2,1]][u,v]
%//FullSimplify
```

Chapter 27

The Gauss-Bonnet Theorem

That the sum of the interior angles of a triangle in the plane equals π radians was one of the first mathematical facts established by the Greeks. In 1603 Harriot[1] showed that on a sphere of radius 1 the area of a spherical triangle (that is, a triangle whose sides are parts of great circles) with angles α, β, γ is given by

$$\alpha + \beta + \gamma - \pi.$$

For example, an octant of the unit sphere $S^2(1)$ has area $\pi/2$, and comprises the interior of a triangle with angles all equal to $\pi/2$. Both the planar and spherical facts are special cases of the Gauss–Bonnet[2] Theorem, which will be proved in this chapter.

The Gauss–Bonnet Theorem consists of a formula for the integral of the Gaussian curvature over all or part of an abstract surface. For example, the curvature and area of a sphere $S^2(a)$ are respectively $1/a^2$ and $4\pi a^2$. Thus

$$\int_{S^2(a)} K\, dA = \int_{S^2(a)} \frac{1}{a^2}\, dA = \frac{1}{a^2} 4\pi a^2 = 4\pi,$$

[1] Thomas Harriot (1560–1621). English mathematician, adviser to Sir Walter Raleigh. He made a voyage to Virginia in 1585–86 and reported on the Native American languages and customs. His telescopic observations were the earliest in England; in particular, he discovered sunspots in 1610 and used them to deduce the period of the sun's rotation.

[2] 

Pierre Ossian Bonnet (1819–1892). French mathematician, who made many important contributions to surface theory, including the Gauss–Bonnet Theorem. Bonnet was director of studies at the École Polytechnique in Paris, professor of astronomy in the faculty of sciences at the University of Paris, and member of the board of longitudes.

showing that the integral of the Gaussian curvature over a sphere is independent of the radius a of the sphere. It is remarkable that for any ellipsoid, or more generally for any closed convex surface $\mathcal{M}$ in $\mathbb{R}^3$, we have

$$\int_{\mathcal{M}} K\,dA = 4\pi.$$

This fact derives ultimately from the presence of K in formulas for area in the image of the Gauss mapping defined on page 333.

Still more generally, for any compact surface $\mathcal{M}$, the Gauss–Bonnet Theorem expresses the integral of the Gaussian curvature in terms of a topological invariant, the *Euler characteristic* $\chi(\mathcal{M})$. This important number is always an integer, and is expressible in terms of the number of vertices, faces, and edges of any triangulation of $\mathcal{M}$.

We begin by generalizing to abstract surfaces the notions of turning angle (given in Section 1.5) and of total signed curvature of a plane curve (given in Section 6.1). This enables us to consider integrals of geodesic curvature in the abstract setting, though the results of this chapter can also be appreciated by using the extrinsic definitions given in Section 17.4.

In Section 27.2, we prove a local version of the Gauss–Bonnet Theorem, using Green's Theorem, which is of independent interest. We improve the results in Section 27.4. A summary of topological facts about surfaces is given in Section 27.5 in order to establish global versions of the Gauss–Bonnet Theorem in Section 27.6. Finally, Section 27.7 is devoted to applications.

27.1 Turning Angles and Liouville's Theorem

Since there is no notion of "horizontal" for a general abstract surface, we must choose a reference vector to measure turning. In the next two lemmas, we suppose that $\mathcal{M}$ is an oriented surface, $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ is a regular curve, and that $\mathbf{X}$ is an *everywhere nonzero* vector field along $\boldsymbol{\alpha}$.

Lemma 27.1. *Fix t_0 with $a < t_0 < b$. Let θ_0 be a number such that*

$$\frac{\boldsymbol{\alpha}'(t_0)}{\|\boldsymbol{\alpha}'(t_0)\|} = \cos\theta_0 \frac{\mathbf{X}(t_0)}{\|\mathbf{X}(t_0)\|} + \sin\theta_0 \frac{J\mathbf{X}(t_0)}{\|\mathbf{X}(t_0)\|}.$$

Then there exists a unique differentiable function $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{X}]\colon (a,b) \to \mathbb{R}$ such that $\boldsymbol{\theta}(t_0) = \theta_0$ and

$$\frac{\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|} = \cos\boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{X}] \frac{\mathbf{X}}{\|\mathbf{X}\|} + \sin\boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{X}] \frac{J\mathbf{X}}{\|\mathbf{X}\|}, \tag{27.1}$$

at all points on the curve. We call $\boldsymbol{\theta}[\alpha, \mathbf{X}]$ the ***turning angle*** *of $\boldsymbol{\alpha}$ with respect to $\mathbf{X}$ determined by θ_0 and t_0.*

Proof. Corollary 1.24 on page 18 implies the existence of a unique function $\boldsymbol{\theta}[\boldsymbol{\alpha}]\colon (a,b) \to \mathbb{R}$ such that $\boldsymbol{\theta}[\boldsymbol{\alpha}](t_0) = \theta_0$ and

$$\begin{cases} \cos\boldsymbol{\theta}[\boldsymbol{\alpha}](t) = \dfrac{\boldsymbol{\alpha}'(t)\cdot\mathbf{X}(t)}{\|\boldsymbol{\alpha}'(t)\|\,\|\mathbf{X}(t)\|}, \\ \sin\boldsymbol{\theta}[\boldsymbol{\alpha}](t) = \dfrac{\boldsymbol{\alpha}'(t)\cdot J\mathbf{X}(t)}{\|\boldsymbol{\alpha}'(t)\|\,\|\mathbf{X}(t)\|}. \end{cases} \tag{27.2}$$

Then (27.2) is equivalent to (27.1). ∎

We can now generalize Lemma 1.26, page 20, to abstract surfaces.

Lemma 27.2. *The turning angle* $\boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{X}]$ *is related to the geodesic curvature* $\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]$ *of* $\boldsymbol{\alpha}$ *by the formula*

$$\boldsymbol{\theta}[\alpha,\mathbf{X}]'(t) = \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t)\,\|\boldsymbol{\alpha}'(t)\| - \frac{\langle \nabla_{\boldsymbol{\alpha}'}\mathbf{X},\ J\mathbf{X}\rangle(t)}{\|\mathbf{X}(t)\|^2}. \tag{27.3}$$

Proof. Using the analog of (1.23), we have

$$\left\langle \nabla_{\boldsymbol{\alpha}'(t)}\left(\frac{\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|}\right),\ \frac{J\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|}\right\rangle = \frac{\langle \boldsymbol{\alpha}''(t), J\boldsymbol{\alpha}'(t)\rangle}{\|\boldsymbol{\alpha}'(t)\|^2} = \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t)\|\boldsymbol{\alpha}'(t)\|. \tag{27.4}$$

On the other hand, (writing $\boldsymbol{\alpha} = \boldsymbol{\alpha}(t)$, $\mathbf{X} = \mathbf{X}(t)$ and $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{X}]$) we use (27.1) to compute

$$\begin{aligned} \nabla_{\boldsymbol{\alpha}'}\left(\frac{\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|}\right) &= \nabla_{\boldsymbol{\alpha}'}\left(\frac{\mathbf{X}\cos\boldsymbol{\theta} + J\mathbf{X}\sin\boldsymbol{\theta}}{\|\mathbf{X}\|}\right) \\ &= -\frac{\|\mathbf{X}\|'}{\|\mathbf{X}\|^2}\big(\mathbf{X}\cos\boldsymbol{\theta} + J\mathbf{X}\sin\boldsymbol{\theta}\big) + \left(\frac{-\mathbf{X}\sin\boldsymbol{\theta} + J\mathbf{X}\cos\boldsymbol{\theta}}{\|\mathbf{X}\|}\right)\boldsymbol{\theta}' \\ &\qquad + \frac{\big((\nabla_{\boldsymbol{\alpha}'}\mathbf{X})\cos\boldsymbol{\theta} + (J\nabla_{\boldsymbol{\alpha}'}\mathbf{X})\sin\boldsymbol{\theta}\big)}{\|\mathbf{X}\|} \\ &= -\frac{\|\mathbf{X}\|'}{\|\mathbf{X}\|}\frac{\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|} + \frac{J\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|}\boldsymbol{\theta}' + \frac{\big((\nabla_{\boldsymbol{\alpha}'}\mathbf{X})\cos\boldsymbol{\theta} + (J\nabla_{\boldsymbol{\alpha}'}\mathbf{X})\sin\boldsymbol{\theta}\big)}{\|\mathbf{X}\|}, \end{aligned}$$

so that

$$\left\langle \nabla_{\boldsymbol{\alpha}'}\left(\frac{\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|}\right),\ \frac{J\boldsymbol{\alpha}'}{\|\boldsymbol{\alpha}'\|}\right\rangle = \boldsymbol{\theta}' + \frac{\langle \nabla_{\boldsymbol{\alpha}'}\mathbf{X}, J\mathbf{X}\rangle}{\|\mathbf{X}\|^2}. \tag{27.5}$$

Now (27.3) follows from (27.4) and (27.5). ∎

A practical choice for the vector field $\mathbf{X}$ is $\mathbf{x}_u$, where $\mathbf{x}$ is a patch on an abstract surface. This will lead us to an analog of Euler's Theorem on page 397. Suppose now that the surface $\mathcal{M}$ is oriented, and denote its metric by

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2. \tag{27.6}$$

We shall say (referring to page 879) that a patch $\mathbf{x}$ is ***coherent*** with the orientation of $\mathcal{M}$ if the equation

$$J\mathbf{x}_u = \frac{-F\mathbf{x}_u + E\mathbf{x}_v}{\sqrt{EG - F^2}} \tag{27.7}$$

holds, rather than one with an opposite sign.

Lemma 27.3. *Let $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ be a regular curve whose trace is contained in $\mathbf{x}(\mathcal{U})$, where $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch coherent with the orientation of $\mathcal{M}$. Write $\boldsymbol{\alpha}(t) = \mathbf{x}(u(t), v(t))$ for $a < t < b$. The geodesic curvature of $\boldsymbol{\alpha}$ satisfies*

$$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]\,\|\boldsymbol{\alpha}'\| = \boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{x}_u]' + \frac{-\dfrac{F}{E}(E_u u' + E_v v') - E_v u' + G_u v' + 2F_u u'}{2\sqrt{EG - F^2}}.$$

Proof. We have $\boldsymbol{\alpha}' = u'\mathbf{x}_u + v'\mathbf{x}_v$; thus from (27.7) it follows that

$$\frac{\langle \nabla_{\boldsymbol{\alpha}'}\mathbf{x}_u,\, J\mathbf{x}_u\rangle}{\|\mathbf{x}_u\|^2} = \frac{1}{E}\left\langle u'\,\nabla_{\mathbf{x}_u}\mathbf{x}_u + v'\,\nabla_{\mathbf{x}_v}\mathbf{x}_u,\; \frac{-F\mathbf{x}_u + E\mathbf{x}_v}{\sqrt{EG - F^2}}\right\rangle, \tag{27.8}$$

which equals

$$-u'F\langle \nabla_{\mathbf{x}_u}\mathbf{x}_u, \mathbf{x}_u\rangle + u'E\langle \nabla_{\mathbf{x}_u}\mathbf{x}_u, \mathbf{x}_v\rangle - v'F\langle \nabla_{\mathbf{x}_v}\mathbf{x}_u, \mathbf{x}_u\rangle + v'E\langle \nabla_{\mathbf{x}_v}\mathbf{x}_u, \mathbf{x}_v\rangle$$

divided by $E\sqrt{EG - F^2}$. Since $\langle \nabla_{\mathbf{x}_u}\mathbf{x}_u, \mathbf{x}_u\rangle = \frac{1}{2}E_u$, $\langle \nabla_{\mathbf{x}_v}\mathbf{x}_u, \mathbf{x}_v\rangle = \frac{1}{2}G_u$ and

$$\langle \nabla_{\mathbf{x}_u}\mathbf{x}_u, \mathbf{x}_v\rangle = F_u - \langle \nabla_{\mathbf{x}_v}\mathbf{x}_u, \mathbf{x}_u\rangle = F_u - \tfrac{1}{2}E_v,$$

we see that (27.8) equals

$$\frac{1}{E\sqrt{EG - F^2}}\left(-\frac{u'}{2}\big(FE_u - 2EF_u + EE_v\big) + \frac{v'}{2}\big(-FE_v + EG_u\big)\right).$$

Combining this equation with (27.3) completes the proof. ∎

Corollary 27.4. *With the hypotheses of Lemma 27.3, suppose in addition that $F = 0$ in (27.6). Then*

$$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]\,\|\boldsymbol{\alpha}'\| = \boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{x}_u]' + \frac{1}{2\sqrt{EG}}\,(G_u v' - E_v u'). \tag{27.9}$$

Next, we find a formula for the geodesic curvature of a curve on a patch $\mathbf{x}$ in terms of the geodesic curvatures of the coordinate curves $u \mapsto \mathbf{x}(u, v)$ and $v \mapsto \mathbf{x}(u, v)$. We denote these geodesic curvatures by $(\boldsymbol{\kappa}_g)_1$ and $(\boldsymbol{\kappa}_g)_2$ respectively.

Theorem 27.5. (Liouville[3]) *Let $\mathcal{M}$ be an oriented surface, and suppose that $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ is a regular curve whose trace is contained in $\mathbf{x}(\mathcal{U})$, where $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch coherent with the orientation of $\mathcal{M}$ for which $F = 0$. Then*

$$\boldsymbol{\kappa}_g[\boldsymbol{\alpha}] = (\boldsymbol{\kappa}_g)_1 \cos\boldsymbol{\theta} + (\boldsymbol{\kappa}_g)_2 \sin\boldsymbol{\theta} + \frac{\boldsymbol{\theta}'}{\|\boldsymbol{\alpha}'\|}, \tag{27.10}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}[\boldsymbol{\alpha}, \mathbf{x}_u]$ is the angle between $\boldsymbol{\alpha}$ and $\mathbf{x}_u$.

Proof. Applying (27.9) to the curves $u \mapsto \mathbf{x}(u,v)$ and $v \mapsto \mathbf{x}(u,v)$, we obtain

$$(\boldsymbol{\kappa}_g)_1\sqrt{E} = -\frac{E_v}{2\sqrt{EG}} \qquad \text{and} \qquad (\boldsymbol{\kappa}_g)_2\sqrt{G} = \frac{G_u}{2\sqrt{EG}}.$$

For a general curve $\boldsymbol{\alpha}$, we have

$$u' = \frac{\|\boldsymbol{\alpha}'\|}{\sqrt{E}}\cos\boldsymbol{\theta} \qquad \text{and} \qquad v' = \frac{\|\boldsymbol{\alpha}'\|}{\sqrt{G}}\sin\boldsymbol{\theta},$$

so that

$$-\frac{E_v u'}{2\sqrt{EG}} = \|\boldsymbol{\alpha}'\|(\boldsymbol{\kappa}_g)_1\cos\boldsymbol{\theta} \quad \text{and} \quad \frac{G_u v'}{2\sqrt{EG}} = \|\boldsymbol{\alpha}'\|(\boldsymbol{\kappa}_g)_2\sin\boldsymbol{\theta}. \tag{27.11}$$

Then (27.10) follows from (27.9) and (27.11). ∎

Notice the deceptive similarity between (27.10) and Euler's formula (13.17).

Total Geodesic Curvature

The analog of Definition 6.1 on page 154 for a curve on a surface is

Definition 27.6. *Let $\mathcal{M}$ be an abstract surface. The* ***total geodesic curvature*** *of a curve $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$, denoted by $\mathsf{TGC}[\boldsymbol{\alpha}]$, is the number*

$$\int_{\boldsymbol{\alpha}} \boldsymbol{\kappa}_g[\boldsymbol{\alpha}]\, ds = \int_a^b \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t)\,\big\|\boldsymbol{\alpha}'(t)\big\|\, dt, \tag{27.12}$$

where $\boldsymbol{\kappa}_g[\boldsymbol{\alpha}]$ is the geodesic curvature of $\boldsymbol{\alpha}$, and s denotes arc length.

The abstract definition of the geodesic curvature was given on page 880, though if the surface lies in $\mathbb{R}^3$, one may use the more elementary definition on page 541 to make sense of (27.12).



[3] Joseph Liouville (1809–1882). French mathematician, who worked in number theory, differential equations, dynamics, differential geometry, and complex variables. Liouville played a large role in introducing Gauss's ideas to France.

The generalization of Lemma 6.2 on page 154 is

Lemma 27.7. *The total geodesic curvature of a curve on a surface remains unchanged under a positive reparametrization, but changes sign under a negative reparametrization.*

The proofs of Lemmas 27.7 and 6.2 are essentially the same (see Exercise 8). We also have the following generalization of Lemma 6.3, obtained by integrating (27.3):

Lemma 27.8. *Let $\mathcal{M}$ be an oriented surface, $\boldsymbol{\alpha}\colon (a,b) \to \mathcal{M}$ a regular curve and $\mathbf{X}$ be an everywhere nonzero vector field along $\boldsymbol{\alpha}$. Then the total geodesic curvature can be expressed by*

$$\mathsf{TGC}[\boldsymbol{\alpha}] = \boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{X}](b) - \boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{X}](a) + \int_a^b \frac{\langle \nabla_{\boldsymbol{\alpha}'}\mathbf{X}, J\mathbf{X}\rangle}{\|\mathbf{X}\|^2}\,dt.$$

27.2 The Local Gauss–Bonnet Theorem

First, we need some facts about piecewise-differentiable curves on surfaces. Let $\mathcal{M}$ be an abstract surface, and let $\boldsymbol{\alpha}\colon [a,b] \to \mathcal{M}$ be a piecewise-differentiable curve. We are therefore assuming that $\boldsymbol{\alpha}$ is defined on an open set containing $[a,b]$ and that there is a subdivision

$$a = t_0 < t_1 < \cdots < t_k = b$$

of $[a,b]$ for which $\boldsymbol{\alpha}$ is differentiable and regular on $[t_j, t_{j+1}]$ for $j = 0,\ldots,k-1$. The appropriate left and right derivatives of $\boldsymbol{\alpha}$ are required to exist at the end points of each subinterval. The points $\boldsymbol{\alpha}(t_0),\ldots,\boldsymbol{\alpha}(t_k)$ are called the ***vertices*** of the curve $\boldsymbol{\alpha}$.

The following theorem is an immediate consequence of Corollary 27.4.

Theorem 27.9. *Let $\mathcal{M}$ be an abstract surface, and suppose $\boldsymbol{\alpha}\colon [a,b] \to \mathcal{M}$ is a piecewise-differentiable curve whose trace is contained in $\mathbf{x}(\mathcal{U})$, where $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ is a patch on $\mathcal{M}$ for which $F = 0$. Denote by $\boldsymbol{\alpha}(t_0),\ldots,\boldsymbol{\alpha}(t_k)$ the vertices of $\boldsymbol{\alpha}$. Then*

$$\begin{aligned}\mathsf{TGC}[\boldsymbol{\alpha}] &= \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t)\big\|\boldsymbol{\alpha}'(t)\big\|\,dt \\ &= \sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} \left(\frac{G_u}{2\sqrt{EG}}\frac{dv}{dt} - \frac{E_v}{2\sqrt{EG}}\frac{du}{dt}\right) dt + \sum_{j=0}^{k-1} \big(\boldsymbol{\theta}_{j+1}^- - \boldsymbol{\theta}_j^+\big),\end{aligned}$$

where

$$\boldsymbol{\theta}_j^- = \lim_{t \uparrow t_j} \boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{x}_u]\big(\boldsymbol{\alpha}(t)\big) \qquad \textit{and} \qquad \boldsymbol{\theta}_j^+ = \lim_{t \downarrow t_j} \boldsymbol{\theta}[\boldsymbol{\alpha},\mathbf{x}_u]\big(\boldsymbol{\alpha}(t)\big).$$

The numbers $\boldsymbol{\theta}_j^+$ are only well defined up to integer multiples of 2π, although $\boldsymbol{\theta}_j^+$ completely determines the value $\boldsymbol{\theta}_{j+1}^-$ of the turning angle at the endpoint of the interval $[t_j, t_{j+1}]$. Typically, $\boldsymbol{\varepsilon}_j = \boldsymbol{\theta}_j^+ - \boldsymbol{\theta}_j^-$ is called the ***exterior angle*** of $\boldsymbol{\alpha}$ at $\boldsymbol{\alpha}(t_j)$ (see Figure 27.1 overleaf), and $\boldsymbol{\iota}_j = \pi - \boldsymbol{\varepsilon}_j$ the ***interior angle*** at $\boldsymbol{\alpha}(t_j)$, though we require $0 < \boldsymbol{\iota}_j < 2\pi$ for these definitions to be valid.

In Section 12.4 we defined the notion of area of a closed subset of a surface; that notion admits the following generalization:

Definition 27.10. *Let $\mathcal{M}$ be an abstract surface, $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ an injective regular patch, $\mathcal{S} \subseteq \mathbf{x}(\mathcal{U})$ a closed subset, and $f\colon \mathcal{S} \to \mathbb{R}$ a continuous function. Then the integral of f over $\mathcal{S}$ is given by*

$$\iint_{\mathcal{S}} f \, dA = \iint_{\mathbf{x}^{-1}(\mathcal{S})} (f \circ \mathbf{x}) \sqrt{EG - F^2} \, du dv.$$

We need to show that this definition is geometric. The proof of the following lemma is similar to that of Lemma 12.17 on page 374 (see Exercise 1).

Lemma 27.11. *The definition of the integral of f is independent of the choice of patch.*

Let $\mathcal{S}$ be a bounded subset of the plane $\mathbb{R}^2$ (see page 373). The ***boundary*** of $\mathcal{S}$ is the set, denoted $\partial\mathcal{S}$, of points $\mathbf{p}$ for which every neighborhood of $\mathbf{p}$ contains both a point in $\mathcal{S}$ and a point not in $\mathcal{S}$. In this chapter, by a ***region*** in the plane, we shall mean the union of a bounded connected subset with its boundary. A region $\mathcal{R}$ of $\mathbb{R}^2$ is ***simply-connected*** (a property already encountered on page 788) if and only if its complement $\mathbb{R}^2 \setminus \mathcal{R}$ is connected.

A key tool for proving the Gauss–Bonnet theorem is Green's Theorem. (See, for example, [Buck, page 406].) This can be stated as follows.

Theorem 27.12. (Green[4]) *Let $\mathcal{R} \subset \mathbb{R}^2$ be a simply-connected region, and let $P, Q\colon \mathcal{U} \to \mathbb{R}$ be differentiable functions, where $\mathcal{U}$ is an open set containing $\mathcal{R}$. Then*

$$\iint_{\mathcal{R}} \left(\frac{\partial Q}{\partial u} - \frac{\partial P}{\partial v} \right) du dv = \int_{\partial\mathcal{R}} P \, du + Q \, dv. \tag{27.13}$$

The line integral is performed by parametrizing the boundary in a counterclockwise sense in the plane. To check this is correct, we may merely take $(P, Q) = (0, u)$ and $\mathcal{R} = \{0 \leqslant u \leqslant 1,\ 0 \leqslant v \leqslant 1\}$, so that (27.13) becomes

$$\iint_{\mathcal{R}} 1 \, du dv = \int_{\partial\mathcal{R}} u \, dv,$$

[4]George Green (1793–1841). English mathematician, owner of a windmill in Nottingham. Although he published only 10 papers, he made profound contributions to mathematical physics. He invented Green's functions and coined the term "potential function." Where Green acquired his mathematical skills remains a mystery.

and both sides are equal to 1. More generally, the boundary of a region of an abstract surface will be traversed in the sense determined by the operator J defining the orientation.

We are now ready to prove a first version of the Gauss–Bonnet Theorem.

Theorem 27.13. (Gauss–Bonnet, local version) *Let $\mathcal{M}$ be an oriented abstract surface and let $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ be a regular patch coherent with the orientation of $\mathcal{M}$. Let $\mathcal{R} = \mathbf{x}([a,b]\times[c,d])$, where $[a,b]\times[c,d] \subseteq \mathcal{U}$. Then*

$$\iint_{\mathcal{R}} K\,dA + \int_{\partial\mathcal{R}} \kappa_g\,ds + \sum_{j=0}^{3} \varepsilon_j = 2\pi, \tag{27.14}$$

where $\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3$ are the exterior angles at the vertices $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ of $\mathcal{R}$.

Proof. First, suppose that $\mathbf{x}$ is a patch with $F = 0$. We suppose that the vertices are chosen so that

$$\mathbf{p}_0 = \mathbf{x}(a,c), \quad \mathbf{p}_1 = \mathbf{x}(b,c), \quad \mathbf{p}_2 = \mathbf{x}(b,d), \quad \mathbf{p}_3 = \mathbf{x}(a,d).$$

The portion of $\partial\mathcal{R}$ between $\mathbf{p}_0$ and $\mathbf{p}_1$ is then parametrized by $u \mapsto \mathbf{x}(u,c)$, and we may choose

$$\theta_0^+ = \theta_1^- = 0.$$

We explain next that the hypothesis that $\mathbf{x}$ be consistent with the orientation ensures that the external angle at each vertex lies between 0 and π.

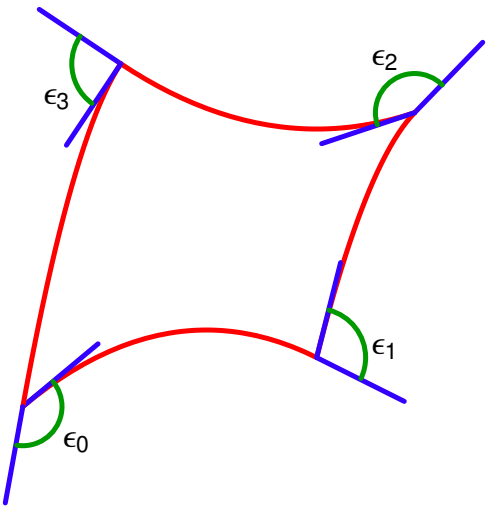


Figure 27.1: A quadrilateral with exterior angles ϵ_i

On the portion of $\partial\mathcal{R}$ between $\mathbf{p}_1$ and $\mathbf{p}_2$, we deduce from (27.1) and (27.7) that the turning angle $\boldsymbol{\theta} = \boldsymbol{\theta}[\mathbf{x}_v, \mathbf{x}_u]$ of the boundary relative to $\mathbf{x}_u$ satisfies

$$\sqrt{E}\sin\boldsymbol{\theta} = \langle \mathbf{x}_v, J\mathbf{x}_u\rangle = \sqrt{EG}.$$

We are therefore free to choose $0 < \sin\boldsymbol{\theta}_1^+ < \pi$, so that the internal angle $\boldsymbol{\iota}_1$ at $\mathbf{p}_1$ also lies between 0 and π. Continuing around the quadrilateral, the portion of $\partial\mathcal{R}$ between $\mathbf{p}_2$ and $\mathbf{p}_3$ is parametrized by $u \mapsto \mathbf{x}(u, d)$, and we need to choose

$$\boldsymbol{\theta}_2^+ = \boldsymbol{\theta}_3^- = \pi \qquad \text{and} \qquad \pi < \boldsymbol{\theta}_3^+, \boldsymbol{\theta}_4^- < 2\pi.$$

The resulting exterior angles are

$$\begin{cases} \boldsymbol{\varepsilon}_1 = \boldsymbol{\theta}_1^+ - \boldsymbol{\theta}_1^- = \boldsymbol{\theta}_1^+, \\ \boldsymbol{\varepsilon}_2 = \boldsymbol{\theta}_2^+ - \boldsymbol{\theta}_2^- = \pi - \boldsymbol{\theta}_2^-, \\ \boldsymbol{\varepsilon}_3 = \boldsymbol{\theta}_3^+ - \boldsymbol{\theta}_3^- = \boldsymbol{\theta}_3^+ - \pi, \\ \boldsymbol{\varepsilon}_0 = \boldsymbol{\theta}_0^+ - \boldsymbol{\theta}_4^- + 2\pi = 2\pi - \boldsymbol{\theta}_4^-. \end{cases}$$

It is necessary to add 2π in the last line so that the internal angle $\boldsymbol{\iota}_0 = \pi - \boldsymbol{\varepsilon}_0$ lies between 0 and 2π (indeed, 0 and π). Therefore,

$$\begin{aligned} \sum_{j=0}^{3} \left(\boldsymbol{\theta}_{j+1}^- - \boldsymbol{\theta}_j^+\right) &= \left(\boldsymbol{\theta}_2^- - \boldsymbol{\theta}_1^+\right) + \left(\boldsymbol{\theta}_4^- - \boldsymbol{\theta}_3^+\right) \\ &= \pi - \boldsymbol{\varepsilon}_2 - \boldsymbol{\varepsilon}_1 + 2\pi - \boldsymbol{\varepsilon}_0 - \boldsymbol{\varepsilon}_3 - \pi = 2\pi - \sum_{j=0}^{3} \boldsymbol{\varepsilon}_j. \end{aligned}$$

By Green's Theorem (Theorem 27.12) and formula (17.6), page 534, we have

$$\begin{aligned} &\int_{\partial\mathcal{R}} \left(\frac{G_u}{2\sqrt{EG}}\frac{dv}{dt} - \frac{E_v}{2\sqrt{EG}}\frac{du}{dt}\right) dt \\ &\qquad = \int_a^b\int_c^d \left(\left(\frac{E_v}{2\sqrt{EG}}\right)_v + \left(\frac{G_u}{2\sqrt{EG}}\right)_u\right) du\,dv \\ &\qquad = -\int_a^b\int_c^d K\sqrt{EG}\,du\,dv = -\iint_{\mathcal{R}} K\,dA. \end{aligned}$$

Substituting the equations above into Theorem 27.9, it follows that the total geodesic curvature of $\partial\mathcal{R}$ is given by

$$\begin{aligned} \int_{\partial\mathcal{R}} \boldsymbol{\kappa}_g\,ds &= \sum_{j=0}^{3} \int_{t_j}^{t_{j+1}} \boldsymbol{\kappa}_g[\boldsymbol{\alpha}](t)\big\|\boldsymbol{\alpha}'(t)\big\|\,dt \\ &= 2\pi - \sum_{j=0}^{3} \boldsymbol{\varepsilon}_j - \iint_{\mathcal{R}} K\,dA. \end{aligned}$$

Since all terms in (27.14) are invariantly defined, the result holds for all patches irrespective of the assumption $F = 0$. ∎

It will be convenient to rewrite (27.14) using interior angles.

Corollary 27.14. *Under the hypotheses of Theorem* 27.13 *we have*

$$\iint_{\mathcal{R}} K\,dA + \int_{\partial\mathcal{R}} \boldsymbol{\kappa}_g\,ds = -2\pi + \sum_{j=0}^{3} \boldsymbol{\iota}_j, \tag{27.15}$$

where $\boldsymbol{\iota}_0, \boldsymbol{\iota}_1, \boldsymbol{\iota}_2, \boldsymbol{\iota}_3$ *are the interior angles at the vertices of* $\mathcal{R}$.

27.3 An Area Bound

This short section provides an independent application of Green's theorem to the area of surfaces of constant negative curvature in $\mathbb{R}^3$, such as the pseudosphere and other examples studied in Chapters 15 and 21. It uses terminology from Section 21.2, in particular Definition 21.11 on page 690.

Lemma 27.15. *Let* $\mathcal{D} = [a,b] \times [c,d]$, *and let* $\mathbf{x}\colon \mathcal{D} \to \mathbb{R}^3$ *be a Tchebyshef principal patch of constant negative curvature* $-a^{-2}$. *Then*

$$\operatorname{area}\big(\mathbf{y}(\mathcal{D})\big) = \iint_{\mathcal{D}} dA \;\leqslant\; 2\pi a^2. \tag{27.16}$$

Proof. Let ω denote the angle function of the associated asymptotic patch

$$\mathbf{y}(p,q) = \mathbf{x}(p+q,\, p-q),$$

described on page 686. Then the principal patch $\mathbf{x}$ has metric

$$ds^2 = a^2(\cos^2\theta\, du^2 + \sin^2\theta\, dv^2),$$

where $\theta = \omega/2$. The infinitesimal area element of $\mathbf{y}$ is given by

$$\begin{aligned} dA \;=\; \sqrt{EG - F^2}\,dp\,dq \;&=\; \sqrt{a^4 - a^4\cos^2\omega}\,dp\,dq \\ &=\; a^2 \sin\omega\, dp\,dq \\ &=\; a^2\omega_{pq}\,dp\,dq, \end{aligned}$$

the last line with the help of Lemma 21.2. By Theorem 27.12, we have

$$\begin{aligned} \operatorname{area}\big(\mathbf{y}(\mathcal{D})\big) \;&=\; a^2 \iint_{\mathcal{D}} \omega_{pq}\,dp\,dq \\ &=\; \frac{a^2}{2}\int_{\partial\mathcal{D}} \big(-\,\omega_p dp + \omega_q dq\big) \\ &=\; a^2\big(\omega(a,c) + \omega(b,d) - \omega(a,d) - \omega(b,c)\big). \end{aligned} \tag{27.17}$$

Since neither $\cos\theta$ nor $\sin\theta$ can vanish, we can assume that $0 < \theta < \pi/2$, so $0 < \omega(p,q) < \pi$ for all p, q. Hence $\operatorname{area}\big(\mathbf{y}(\mathcal{D})\big) \leqslant 2a^2\pi$. ∎

More generally, suppose that we can cover the surface $\mathcal{M}$ by one large rectangle divided into a grid of smaller rectangular Tchebyshef principal patches of the type above. When we sum the contributions arising in (27.17), these cancel out in pairs, and we are left with only external vertices.

Corollary 27.16. *Let $\mathcal{M} \subset \mathbb{R}^3$ be a connected regular surface with constant negative curvature $-a^2$. Then* $\mathsf{area}(\mathcal{M}) \leqslant 2\pi a^2$.

This result extends Theorem 26.24, since it follows that one cannot find any *complete* surfaces of constant negative Gaussian curvature in $\mathbb{R}^3$.

27.4 A Generalization to More Complicated Regions

In order to generalize Theorem 27.13, we introduce the following notion.

Definition 27.17. *Let $\mathcal{M}$ be an abstract surface. A* ***regular region*** *is a compact region whose boundary is the union of a finite number of piecewise-regular simple closed curves that do not intersect each other.*

Simple closed curves were introduced on page 161. Note however that a compact regular surface can be considered as a regular region with empty boundary.

Each of the piecewise-regular simple closed curves that constitute the boundary $\partial\mathcal{Q}$ of a regular region $\mathcal{Q}$ has vertices and exterior angles; therefore, we can speak of the vertices $\mathbf{p}_1, \ldots, \mathbf{p}_m$ and the external angles $\varepsilon_1, \ldots, \varepsilon_m$ of the regular region $\mathcal{Q}$. Let us also call the portion of $\partial\mathcal{Q}$ between consecutive vertices an ***edge*** of $\mathcal{Q}$.

Definition 27.18. *A* ***polygonal region*** *is a regular region with a finite number of vertices, each of which has a nonzero external angle. A* ***triangular region*** *is a polygonal region with exactly* 3 *vertices.*

It will be necessary to decompose regular regions into smaller regions.

Definition 27.19. *A* ***polygonal decomposition*** *of a regular region $\mathcal{Q}$ of an abstract surface is a finite collection $\mathfrak{R}$ of polygonal regions $\mathcal{R}$ called* ***faces*** *such that*

(i) *each $\mathcal{R} \in \mathfrak{R}$ is homeomorphic to a disk*;

(ii) *the union of all regions in $\mathfrak{R}$ is $\mathcal{Q}$*;

(iii) *if $\mathcal{R}_1, \mathcal{R}_2 \in \mathfrak{R}$, then the intersection $\mathcal{R}_1 \cap \mathcal{R}_2$ is either a common edge or a common vertex of $\mathcal{R}_1$ and $\mathcal{R}_2$.*

An ***oriented polygonal decomposition*** *of $\mathcal{Q}$ is a polygonal decomposition in which the faces and edges are assigned orientations as follows*:

(iv) *each face of* $\mathfrak{R}$ *is assigned the orientation of* $\mathcal{M}$;

(v) *each edge of a face* $\mathcal{R} \in \mathfrak{R}$*, when parametrized as a curve* $\boldsymbol{\alpha}$ *with respect to* $\mathcal{R}$*, is such that* $J\boldsymbol{\alpha}'$ *points toward the interior of* $\mathcal{R}$*.*

If all the faces are triangular regions, then $\mathfrak{R}$ *is called a* ***triangulation*** *of* $\mathcal{Q}$*.*

Note that if $\boldsymbol{\gamma}$ is an edge of polygonal regions $\mathcal{R}_1$ and $\mathcal{R}_2$, the orientation of $\boldsymbol{\gamma}$ with respect to $\mathcal{R}_1$ is the opposite of the orientation of $\boldsymbol{\gamma}$ with respect to $\mathcal{R}_2$.

As a first step to generalizing Theorem 27.13, we derive the formula for the integral of the Gaussian curvature over a triangular region.

Corollary 27.20. *Let* $\mathcal{M}$ *be an oriented abstract surface, and suppose that* $\mathcal{T}$ *is a triangular region of* $\mathcal{M}$ *such that* $\mathcal{T} \subset \mathbf{x}(\mathcal{U})$*, where* $\mathbf{x}\colon \mathcal{U} \to \mathcal{M}$ *is a regular patch coherent with the orientation of* $\mathcal{M}$*. Then*

$$\iint_{\mathcal{T}} K\,dA + \int_{\partial\mathcal{T}} \boldsymbol{\kappa}_g\,ds = -\pi + \boldsymbol{\iota}_1 + \boldsymbol{\iota}_2 + \boldsymbol{\iota}_3, \tag{27.18}$$

where $\boldsymbol{\iota}_1, \boldsymbol{\iota}_2, \boldsymbol{\iota}_3$ *are the interior angles at the vertices of* $\mathcal{T}$*.*

Proof. We construct an oriented polygonal decomposition of $\mathcal{T}$, consisting of quadrilaterals, as follows. Choose a point in the interior of the triangular region and points on each of the edges between the vertices. Join the three new points on the edges to the point in the interior. Three quadrilaterals $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$ are formed; we choose them to have their orientations coherent with that of $\mathcal{T}$. Since $\mathcal{T}$ is contained in the trace of the patch $\mathbf{x}$, the three quadrilaterals can be chosen as images under $\mathbf{x}$ of ordinary quadrilaterals in $\mathbb{R}^2$. Therefore, because of additivity of integration over closed subsets of $\mathbb{R}^2$ and (27.15), we have

$$\iint_{\mathcal{T}} K\,dA = \sum_{k=1}^{3} \iint_{\mathcal{R}_k} K\,dA = -6\pi + I - \sum_{k=1}^{3} \int_{\partial\mathcal{R}_k} \boldsymbol{\kappa}_g\,ds, \tag{27.19}$$

where I is the sum of all interior angles.

The boundaries $\partial\mathcal{R}_1, \partial\mathcal{R}_2, \partial\mathcal{R}_3$ contribute six interior edges and six exterior edges. The integrals

$$\int \boldsymbol{\kappa}_g ds$$

cancel in pairs over the interior edges and combine in pairs over the exterior edges; hence

$$\sum_{k=1}^{3} \int_{\partial\mathcal{R}_k} \boldsymbol{\kappa}_g\,ds = \int_{\partial\mathcal{T}} \boldsymbol{\kappa}_g\,ds. \tag{27.20}$$

Included in the sum I are the interior angles $\boldsymbol{\iota}_1, \boldsymbol{\iota}_2, \boldsymbol{\iota}_3$, together with the interior angles at the newly-created vertices. It is easily seen from Figure 27.2 that the interior angles at the newly created vertices add up to 5π, and so

$$I = 5\pi + \boldsymbol{\iota}_1 + \boldsymbol{\iota}_2 + \boldsymbol{\iota}_3. \tag{27.21}$$

Now (27.18) follows from (27.19)–(27.21). ▌

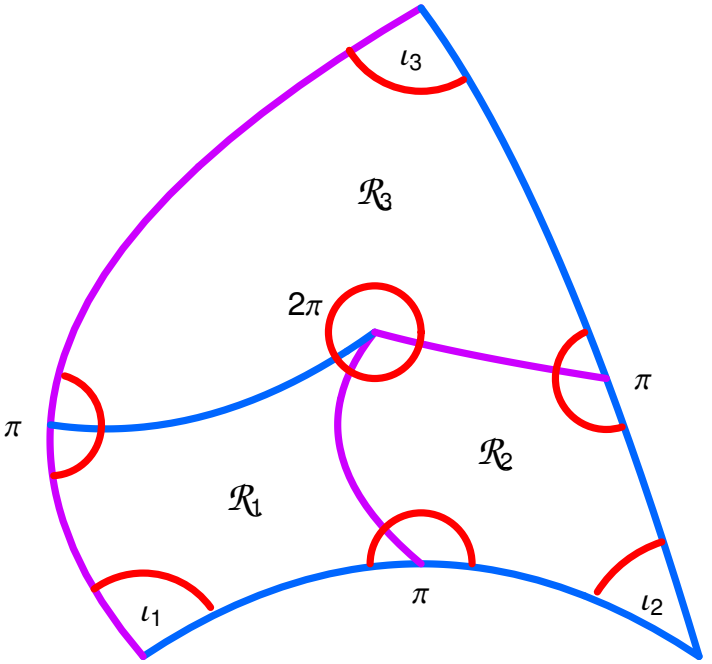


Figure 27.2: A triangle decomposed into three quadrilaterals

Now we can state the original version of the theorem that Gauss proved in [Gauss2, Section 20].

Corollary 27.21. *Suppose that the sides of the region $\mathcal{T}$ in Corollary* 27.20 *are geodesics. Then*

$$\iint_{\mathcal{T}} K\,dA = -\pi + \boldsymbol{\iota}_1 + \boldsymbol{\iota}_2 + \boldsymbol{\iota}_3, \tag{27.22}$$

where $\boldsymbol{\iota}_1, \boldsymbol{\iota}_2, \boldsymbol{\iota}_3$ are the interior angles at the vertices of $\mathcal{T}$.

When $K \equiv 0$, we obtain the familiar theorem that the sum of the angles of a triangle in the plane equals π. More generally:

Corollary 27.22. *Suppose that the sides of the region $\mathcal{T}$ in Corollary* 27.20 *are geodesics, and that K has the constant value $\lambda \neq 0$. Then the area of $\mathcal{T}$ is given by*

$$\mathsf{area}(\mathcal{T}) = \frac{-\pi + \boldsymbol{\iota}_1 + \boldsymbol{\iota}_2 + \boldsymbol{\iota}_3}{\lambda}, \tag{27.23}$$

where $\boldsymbol{\iota}_1$, $\boldsymbol{\iota}_2$, $\boldsymbol{\iota}_3$ are the interior angles at the vertices of $\mathcal{T}$.

When $\lambda = 1$, equation (27.23) reduces to Harriot's formula for the area of a spherical triangle. Thus on a unit sphere, the sum of the interior angles of any geodesic triangle is greater than π, and the excess over π equals the area of the triangle. By contrast, any triangle on a surface of constant Gaussian curvature -1 has area strictly *less* than π. Using the upper half-plane model (Figure 26.3 on page 883), we may choose three vertices on the real axis (thus strictly speaking outside the upper half-plane) and join them by three semicircles. The resulting *asymptotic* geodesic triangle has interior angles all zero and area π relative to the Poincaré metric (26.6).

27.5 The Topology of Surfaces

In this section, we give a rudimentary account of the topology of surfaces. For more information see [AhSa, Chapter 1], [Arm, Chapter 7], [Massey, Chapter 1], [Springer, Chapter 5], and [FiGa]. The most important theorem we need is

Theorem 27.23. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be compact abstract surfaces. The following conditions are equivalent*:

(i) *$\mathcal{M}_1$ and $\mathcal{M}_2$ are homeomorphic*;

(ii) *$\mathcal{M}_1$ and $\mathcal{M}_2$ are diffeomorphic*;

(iii) *$\mathcal{M}_1$ and $\mathcal{M}_2$ are both orientable or both nonorientable and have the same Euler characteristic.*

Suppose that (i) holds. A map $\Phi\colon \mathcal{M}_1 \to \mathcal{M}_2$ defining a homeomorphism will not necessarily be a diffeomorphism; nonetheless the theorem asserts that a diffeomorphism $\Psi\colon \mathcal{M}_1 \to \mathcal{M}_2$ will exist.

Here is the definition of Euler characteristic:

Definition 27.24. *Let $\mathfrak{R}$ be a polygonal decomposition of a regular region $\mathcal{Q}$ of an abstract surface. Put*

$$\begin{cases} \mathsf{V} = \text{the number of vertices of } \mathfrak{R}, \\ \mathsf{E} = \text{the number of edges of } \mathfrak{R}, \\ \mathsf{F} = \text{the number of faces of } \mathfrak{R}. \end{cases}$$

Then $\chi(\mathcal{Q}) = \mathsf{F} - \mathsf{E} + \mathsf{V}$ is called the ***Euler characteristic*** *or* ***Euler number*** *of $\mathcal{Q}$.*

It is implicit in Theorem 27.23 that the Euler characteristic $\chi(\mathcal{M})$ of a compact abstract surtface $\mathcal{M}$ does not depend on the polygonal decomposition chosen. This fact really follows from the array of propositions below, details of which can be found in the books cited above.

Proposition 27.25. *Given a polygonal decomposition of a regular region $\mathcal{Q}$ of an abstract surface, the quantity $\mathsf{F} - \mathsf{E} + \mathsf{V}$ is invariant under the following processes*:

(i) *subdivision of an edge by adding a new vertex at an interior point of the edge*;

(ii) *subdivision of a polygonal region by connecting two vertices by a new edge*;

(iii) *introducing a new vertex into a polygonal region and connecting it by an edge to one of the existing vertices of the region.*

The proof of Proposition 27.25 is instructive and not difficult (see Exercise 2). Clearly, it can be used to show that any two polygonal decompositions of a regular region of an abstract surface have the same Euler characteristic, provided each edge of one of the two decompositions intersects each edge of the other in only a finite number of points and a finite number of closed intervals. However, it can happen that two polygonal decompositions intersect each other in bizarre ways. Nevertheless:

Proposition 27.26. *the quantity $\mathsf{F} - \mathsf{E} + \mathsf{V}$ is independent of the polygonal decomposition used to define it.*

Proposition 27.27. *A regular region of an abstract surface always admits a triangulation.*

If we take these results for granted, the classification described by Theorem 27.23 is best understood by listing examples in the table below. The sphere S^2 and torus T^2 are orientable surfaces in $\mathbb{R}^3$, unlike the projective plane $\mathbb{RP}^2$ and Klein bottle $\mathbb{K}$ that we met in Chapter 11.

A polygonal decomposition of the sphere is obtained by dividing it into eight triangles of the type mentioned on page 901. This results in 6 vertices, 12 edges and 8 regions, whence

$$\chi(S^2) = 6 - 12 + 8 = 2.$$

We can do the same job with far fewer regions, but Proposition 27.26 tells us that the Euler characteristic will not change.

The Euler characteristics of T^2, $\mathbb{RP}^2$, $\mathbb{K}$ can be deduced from the diagrams in Section 11.2, by merely taking $\mathsf{F} = 1$ face, corresponding to the square. The number of vertices and edges will depend on how the perimeter of the square has been identified with itself. For example Figure 11.4 on page 337 represents *both* the torus and Klein bottle with 1 vertex and 2 edges, whence

$$\chi(T^2) = \chi(\mathbb{K}) = 1 - 2 + 1 = 0.$$

The difference of course is that T^2 is orientable, whereas $\mathbb{K}$ is not.

$\mathcal{M}$	orientable	$\chi(\mathcal{M})$
sphere S^2	yes	2
projective plane $\mathbb{RP}^2$	no	1
torus T^2	yes	0
Klein bottle $\mathbb{K}$	no	0
sphere with $g \geqslant 1$ handles	yes	$2-2g$
sphere with $p \geqslant 1$ cross caps	no	$2-p$

It is easy to dream up more general surfaces. Starting from the sphere, one can attach one or more "handles." If the number of handles is denoted by g (g is called the ***genus***), then the resulting surface $\mathcal{M}_g$ is homeomorphic to a torus with g "holes." There is a short cut to computing the Euler characteristic of such a surface: adding a handle or extra hole effectively reduces it by 2, and we obtain the famous formula $\chi(\mathcal{M}_g) = 2-2g$. Note that the sphere has genus 0, and the torus genus 1.

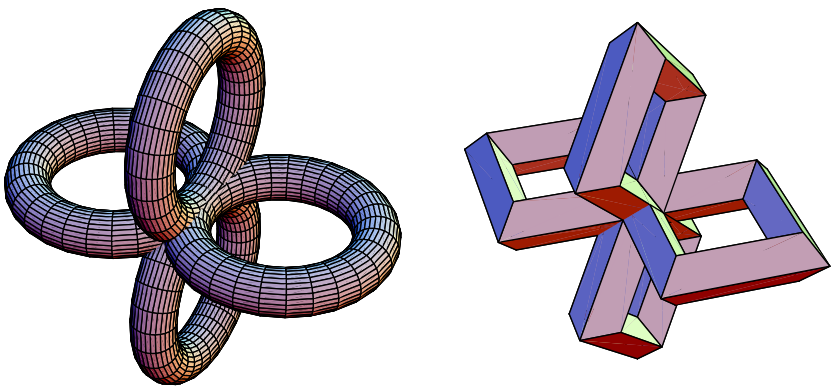


Figure 27.3: Orientable surfaces with $\chi = -6$

Figure 27.3 shows the surface $\mathcal{M}_4$. The two representations are topologically equivalent, and were indeed created by the same program, varying only the number of sample points. We can check the value of $\chi(\mathcal{M}_4)$ by counting the

numbers of vertices, edges and faces on the right. Each of the four polyhedral tori contributes at least 12 vertices, at least 28 edges, and exactly 16 faces. Inspection of the central part from different angles (there is an animation in Notebook 27 to help, using the quaternion techniques of Chapter 23) reveals that the whole surface admits a triangulation with

$$\mathsf{V} = 62, \quad \mathsf{E} = 132, \quad \mathsf{F} = 64,$$

confirming that $\chi = 62 - 132 + 64 = -6$.

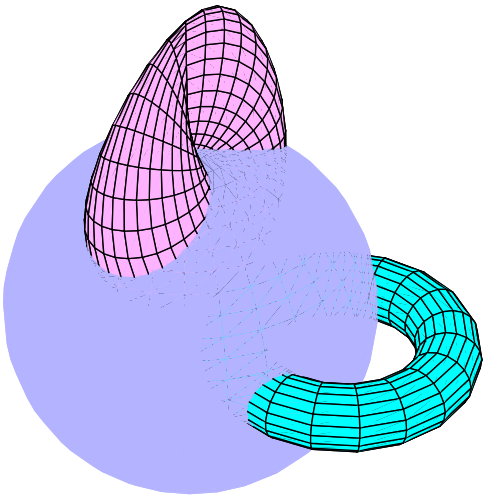


Figure 27.4: A representation of a nonorientable surface with $\chi = -1$

Compact nonorientable surfaces cannot be represented inside $\mathbb{R}^3$ without self-intersections. Nevertheless, it is customary to think of a nonorientable surface as a sphere with a certain number $p \geqslant 1$ of cross caps glued on. The projective plane corresponds to $p = 1$, since we can shrink the sphere so as to obtain the model shown in Figure 11.16 on page 346. Adding a cross cap actually lowers the Euler characteristic by 1, so $\chi(\mathbb{RP}^2) = 2 - 1 = 1$. Figure 27.4 shows a sphere with one cross cap and one handle attached; the result is a surface $\mathcal{S}$ with $\chi(\mathcal{S}) = 2 - 1 - 2 = -1$. Since $\mathcal{S}$ incorporates at least one cross cap it must be nonorientable, and Theorem 27.23 tells us that $\mathcal{S}$ is homeomorphic to a sphere with 3 cross caps attached. Visualizing this equivalence is however best done with planar models of the sort discussed in [FiGa].

The last two results of this section will enable us to globalize the discussion of the Gauss–Bonnet theorem.

Proposition 27.28. *Let $\mathcal{M}$ be an oriented abstract surface, and suppose that $\mathfrak{A} = \{(\mathbf{x}_\alpha, \mathcal{U}_\alpha)\}$ is a family of parametrizations covering $\mathcal{M}$, compatible with the orientation of $\mathcal{M}$. Then there is an oriented triangulation $\mathfrak{T}$ of $\mathcal{M}$ such that every triangular region $\mathcal{T} \in \mathfrak{T}$ is contained in some coordinate neighborhood $\mathcal{U}_\alpha$ with $(\mathbf{x}_\alpha, \mathcal{U}_\alpha) \in \mathfrak{A}$. Furthermore, if the boundary of every triangular region in $\mathfrak{T}$ is positively oriented, then adjacent triangles determine opposite orientations on common edges.*

Proposition 27.29. *Assume that $\mathcal{M}$, $\mathfrak{A}$ and $\mathfrak{T}$ are as in Proposition 27.28, and let $ds_\alpha^2 = E_\alpha du_\alpha^2 + 2F_\alpha du_\alpha dv_\alpha + G_\alpha dv_\alpha^2$ be the expression of the metric of $\mathcal{M}$ in terms of the coordinates given by $\mathbf{x}_\alpha$. Let $\mathcal{Q}$ be a regular region, and let $f\colon \mathcal{Q} \to \mathbb{R}$ be a continuous function. Then*

$$\sum_{\mathcal{T} \in \mathfrak{T}} \iint_{\mathbf{x}_\alpha^{-1}(\mathcal{T})} (f \circ \mathbf{x}_\alpha) \sqrt{E_\alpha G_\alpha - F_\alpha^2}\, du_\alpha\, dv_\alpha$$

is the same for any triangulation $\mathfrak{T}$ of $\mathcal{Q}$ such that every region $\mathcal{T} \in \mathfrak{T}$ is contained in some neighborhood $\mathcal{U}_\alpha$ with $(\mathbf{x}_\alpha, \mathcal{U}_\alpha) \in \mathfrak{A}$.

27.6 The Global Gauss–Bonnet Theorem

The Gauss–Bonnet Theorem is concerned with subsets of abstract surfaces whose boundaries are piecewise-regular curves. More precisely:

Theorem 27.30. (Gauss–Bonnet, global version) *Let $\mathcal{Q}$ be a regular region of an oriented abstract surface $\mathcal{M}$, and let $C_1, \ldots, C_n$ be piecewise-regular curves which form the boundary of $\mathcal{Q}$. Suppose that each C_i is positively oriented. Let $\varepsilon_1, \ldots, \varepsilon_p$ be the external angles of $C_1, \ldots, C_n$. Then*

$$\iint_{\mathcal{Q}} K\, dA + \int_{\partial \mathcal{Q}} \kappa_g\, ds + \sum_{j=1}^{p} \varepsilon_j = 2\pi\chi(\mathcal{Q}). \tag{27.24}$$

Proof. Let $\mathfrak{T}$ be a triangulation of $\mathcal{Q}$ such that every triangular region in $\mathfrak{T}$ is contained in a coordinate neighborhood of one of a family of patches. We give the positive orientation to each triangular region in $\mathcal{T} \in \mathfrak{T}$. In this way, adjacent triangular regions give opposite orientations to their common edge.

We apply Corollary 27.20 to every triangular region in $\mathfrak{T}$ and add up the results. The integrals of the geodesic curvature along the interior edges (that is, those edges that are not contained in one of the boundary curves) cancel one another. We obtain

$$\iint_{\mathcal{R}} K\, dA + \sum_{i=1}^{n} \int_{C_i} \kappa_g\, ds = -\pi\, \mathsf{F} + \sum \iota_{ij}, \tag{27.25}$$

where F is the number of triangular regions in $\mathfrak{T}$ and $\sum \boldsymbol{\iota}_{ij}$ is the sum of all interior angles. Let

$$\begin{aligned} \mathsf{V}_e &= \text{number of external vertices of } \mathfrak{T}, \\ \mathsf{V}_i &= \text{number of internal vertices of } \mathfrak{T}, \\ \mathsf{E}_e &= \text{number of external edges of } \mathfrak{T}, \\ \mathsf{E}_i &= \text{number of internal edges of } \mathfrak{T}. \end{aligned}$$

The angles around each internal vertex add up to 2π; hence the sum of all internal interior angles is $2\pi\mathsf{V}_i$. A similar calculation computes the sum of the external interior angles, and so

$$\sum \boldsymbol{\iota}_{ij} = 2\pi\mathsf{V}_i + \pi\mathsf{V}_e - \sum_{j=1}^{p} \boldsymbol{\varepsilon}_j.$$

Thus (27.25) can be rewritten as

$$\iint_{\mathcal{R}} K\,dA + \sum_{i=1}^{n} \int_{C_i} \boldsymbol{\kappa}_g ds + \sum_{j=1}^{p} \boldsymbol{\varepsilon}_j = \pi(-\mathsf{F} + 2\mathsf{V}_i + \mathsf{V}_e). \tag{27.26}$$

We have the relations

$$\mathsf{V}_e = \mathsf{E}_e, \qquad \mathsf{V} = \mathsf{V}_e + \mathsf{V}_i, \qquad \mathsf{E} = \mathsf{E}_e + \mathsf{E}_i. \tag{27.27}$$

Furthermore, $3\mathsf{F} = 2\mathsf{E}_i + \mathsf{E}_e$, which can be written as

$$-\mathsf{F} = 2\mathsf{F} - 2\mathsf{E}_i - \mathsf{E}_e. \tag{27.28}$$

From (27.26)–(27.28), we compute

$$\begin{aligned} \iint_{\mathcal{R}} K\,dA + \sum_{i=1}^{n} \int_{C_i} \boldsymbol{\kappa}_g\,ds + \sum_{j=1}^{p} \boldsymbol{\varepsilon}_j &= \pi(2\mathsf{F} - 2\mathsf{E}_i - \mathsf{E}_e + 2\mathsf{V}_i + \mathsf{V}_e) \\ &= \pi\big[2\mathsf{F} - 2(\mathsf{E} - \mathsf{E}_e) - \mathsf{E}_e + 2(\mathsf{V} - \mathsf{V}_e) + \mathsf{V}_e\big] \\ &= \pi(2\mathsf{F} - 2\mathsf{E} + 2\mathsf{V}) \\ &= 2\pi\chi(\mathcal{Q}), \end{aligned}$$

proving (27.24). ▮

In the special case of a compact surface we obtain:

Corollary 27.31. *Let $\mathcal{M}$ be a compact orientable abstract surface. Then*

$$\iint_{\mathcal{M}} K\,dA = 2\pi\chi(\mathcal{M}), \tag{27.29}$$

where K is the Gaussian curvature with respect to any metric on $\mathcal{M}$.

Proof. A compact surface has no boundary. Hence (27.24) reduces to (27.29). ▮

In 1828, Gauss proved Theorem 27.30 for the special case of polygonal regions whose sides are geodesics, and Bonnet generalized it to the form above in 1848 (see [Gauss2] and [Bonn1]). The use of Green's Theorem in proving Theorem 27.13 was clarified by Darboux ([Darb2, volume 3, pages 122–128]).

27.7 Applications of the Gauss–Bonnet Theorem

A simple but important application of Corollary 27.31 is the determination of compact orientable surfaces with positive Gaussian curvature.

Theorem 27.32. *A compact orientable surface $\mathcal{M}$ with positive Gaussian curvature is diffeomorphic to a sphere.*

Proof. If $\mathcal{M}$ has positive Gaussian curvature, Corollary 27.31 implies that $\mathcal{M}$ has positive Euler characteristic. Since $\mathcal{M}$ is orientable, it follows from Theorem 27.23 that $\mathcal{M}$ is diffeomorphic to a sphere. ∎

Next, we prove a theorem of Jacobi (see [Jacobi]).

Theorem 27.33. (Jacobi, 1843) *Let $\boldsymbol{\alpha}\colon [a,b] \to \mathbb{R}^3$ be a closed regular curve with nonzero curvature. Assume that the curve described by the unit normal $\mathbf{N}$ (considered as a curve in the unit sphere $S^2(1) \subset \mathbb{R}^3$) is a simple closed curve. Then $\mathbf{N}([a,b])$ divides $S^2(1)$ into two regions of equal areas.*

Proof. Without loss of generality, we can assume that $\boldsymbol{\alpha}$ has unit speed. Let s denote the arc length function of $\boldsymbol{\alpha}$, and let $\tilde{s}$ denote the arc length function of $\mathbf{N}$, considered as a curve on the unit sphere. The second Frenet formula of (7.12), page 197, implies that

$$\frac{d\mathbf{N}}{ds} = -\boldsymbol{\kappa}\,\mathbf{T} + \boldsymbol{\tau}\,\mathbf{B},$$

and so

$$\frac{d^2\mathbf{N}}{ds^2} = -\frac{d\boldsymbol{\kappa}}{ds}\mathbf{T} + \frac{d\boldsymbol{\tau}}{ds}\mathbf{B} - (\boldsymbol{\kappa}^2 + \boldsymbol{\tau}^2)\mathbf{N}.$$

From these two equations we compute the vector triple product

$$\left[\mathbf{N}\ \frac{d\mathbf{N}}{ds}\ \frac{d^2\mathbf{N}}{ds^2}\right] = \boldsymbol{\kappa}\frac{d\boldsymbol{\tau}}{ds} - \boldsymbol{\tau}\frac{d\boldsymbol{\kappa}}{ds},$$

and we also deduce that

$$1 = \frac{d\mathbf{N}}{d\tilde{s}} \cdot \frac{d\mathbf{N}}{d\tilde{s}} = \left(\frac{ds}{d\tilde{s}}\right)^2 (\boldsymbol{\kappa}^2 + \boldsymbol{\tau}^2).$$

We then obtain

$$\boldsymbol{\kappa}_g(\mathbf{N}) = \left\|\frac{d\mathbf{N}}{ds}\right\|^{-3}\left[\mathbf{N}\ \frac{d\mathbf{N}}{ds}\ \frac{d^2\mathbf{N}}{ds^2}\right] = \frac{\boldsymbol{\kappa}\dfrac{d\boldsymbol{\tau}}{ds} - \boldsymbol{\tau}\dfrac{d\boldsymbol{\kappa}}{ds}}{\boldsymbol{\kappa}^2+\boldsymbol{\tau}^2}\,\frac{ds}{d\tilde{s}}$$
$$= \frac{d}{ds}\Big(\arctan\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\Big)\frac{ds}{d\tilde{s}}$$
$$= \frac{d}{d\tilde{s}}\Big(\arctan\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\Big).$$

Let $\mathscr{C}$ denote the curve $\mathbf{N}([a,b])$. Since this is *closed*,

$$\int_{\mathscr{C}} \boldsymbol{\kappa}_g(\mathbf{N})\,d\tilde{s} = \int_{\mathscr{C}} \frac{d}{d\tilde{s}}\Big(\arctan\frac{\boldsymbol{\tau}}{\boldsymbol{\kappa}}\Big)\,d\tilde{s} = 0. \tag{27.30}$$

Let $\mathcal{R}$ be the region enclosed by $\mathscr{C}$; then $\chi(\mathcal{R}) = 1$ because $\mathscr{C}$ is a simple curve. Therefore, from (27.24) and (27.30) we obtain

$$2\pi = 2\pi\chi(\mathcal{R}) = \int_{\mathcal{R}} K\,dA = \mathsf{area}(\mathcal{R}).$$

Since the total area of the sphere is 4π, the theorem follows. ▌

Finally, we state without proof an important generalization of Corollary 27.31 due to Cohn-Vossen [CV].

Theorem 27.34. Let $\mathcal{M}$ be a complete orientable abstract surface. Then

$$\iint_{\mathcal{M}} K\,dA \leqslant 2\pi\chi(\mathcal{M}), \tag{27.31}$$

provided both sides of (27.31) *are finite.*

27.8 Exercises

1. Prove Lemma 27.11 and Corollary 27.14.

2. Prove Proposition 27.25.

3. Let $\mathcal{M} \subset \mathbb{R}^3$ be a compact orientable surface not homeomorphic to a sphere. Show that there are points where the Gaussian curvature is positive, negative, and zero.

4. Let H denote mean curvature. Show that

$$\int_{S^2(a)} H^{2n}\,dA$$

is independent of the radius of the sphere $S^2(a)$ if and only if $n = 1$.

M **5.** Let

$$T(a,b) = \left\{ (x,y,z) \in \mathbb{R}^3 \mid z^2 + \left(\sqrt{x^2+y^2} - a\right)^2 = b^2 \right\}$$

be a torus of wheel radius a and tube radius b (see Section 10.4). Use the parametrization of $T(a,b)$ on page 406 to compute

$$\int_{T(a,b)} K \, dA \qquad \text{and} \qquad \int_{T(a,b)} H^2 \, dA.$$

Do the calculations both by hand and using techniques from Notebook 27. What does the Gauss–Bonnet Theorem tell us about the integral of the Gaussian curvature over $T(a,b)$? Is the integral over a surface of the square of the mean curvature a topological invariant?

6. Let $\mathcal{M}$ be an abstract surface homeomorphic to a cylinder with Gaussian curvature $K < 0$. Show that $\mathcal{M}$ admits at most one simple closed geodesic.

Notebook 27

■ N27.0

This notebook has three aims. One is to construct the two figures in Chapter 27 consisting of quadrilaterals together with angles and text. These, and similar figures in Chapters 12 and 24, were originally produced by the author using a TEX-based program. The editors' exercise for the Third Edition was therefore to convert these pictures into *Mathematica*, if only to bolster the content of the notebooks.

In N27.2, the curves are all parts of parabolas of the type $y=ax^2+bx+c$. In N27.4, we generalize this by allowing parabolas $x=ay^2+by+c$, obtained by rotating the former. For this purpose we load

```
<<Geometry`Rotations`
```

To insert text, we adopt the following abbreviation from N2.4:

```
tx[A_, v_]:= Text[FontForm[A,{"Helvetica",18}], v]
```

The second aim is to construct the figures in Section 27.5, based on the building blocks

```
torus[a_,b_,c_][u_,v_]:=
  {(a+b Cos[v])Cos[u], (a+b Cos[v])Sin[u], c Sin[v]}
crosscap[u_,v_]:=
  {(Sin[u]Sin[2v])/2, Sin[2u]Cos[v]^2, Cos[2u]Cos[v]^2}
torus[a_,b_]:= torus[a,b,b]
ellipsoid[a_,c_]:= torus[0,a,c]
sphere[a_]:= ellipsoid[a,a]
```

The first two definitions were introduced in N7.6 and N11.5 respectively. Another approach will rely on the package

```
<<Graphics`ImplicitPlot3D`
```

The third, and perhaps most important, purpose of this notebook is to verify the Gauss-Bonnet theorem in some simple cases. This we do in N27.8, with the help of the definitions of L, metric, gcurv reloaded in N27.0.1.

- N27.0.1

■ N27.2

Here is a program that constructs a parabola passing through three assigned points.

```
pts2parabola[u_,v_,w_]:=
Module[{f,eq,ta,so},
  f[t_]= a t^2 + b t + c;
  eq[x_]:= f[x[[1]]]==x[[2]];
  ta= eq/@{u,v,w};
  so= Solve[ta, {a,b,c}][[1]];
  {#, f[#]/.so} &
]
```

The output is a parametrization, with parameter t equal to the x-coordinate:

```
u:= {0,0}
v:= {2,1}
w:= {1,-3}
α:= pts2parabola[u,v,w]
α[t]
```

We formulate the plotting command into the following Module. This will draw the parabola from the point u to the point v, and the third argument of edge will often be used merely as a control point to modify the chosen curve through the first two points.

```
edge[u_,v_,w_]:=
Module[{α},
   α= pts2parabola[u,v,w];
   ParametricPlot[α[t], {t, u[[1]], v[[1]]}, Axes->False,
     PlotStyle->{Thickness[.01],Hue[0]}]
]
```

We check that the parabola *does* pass through the assigned points:

```
z0:= {PointSize[.03], Point[u], Point[v], Point[w]}//Graphics
z1:= edge[u,v,w]
S[z0, z1, Axes->True, PlotRange->All]
```

Figure 27.1 also consists of short lines tangent at both ends to the parabola through u,v,w, and their plotting is accomplished by the following miniprogram. Note that the parametrization of the parabola cannot be differentiated directly; this explains the dummy variable in the definition of β.

```
tangents[u_,v_,w_]:=
Module[{α,β,λ},
  α:= pts2parabola[u,v,w];
  β[t_]:= D[α[tt], tt]/.tt->t;
  pm= If[u[[1]]<v[[1]], 1, -1];
  λ[t_]:= {Thickness[.01],Hue[.7],
    Line[{α[t], α[t] + 2pm β[t]/L[β[t]]}]}//Graphics;
  Show[λ[u[[1]]], λ[v[[1]]]]
]
```

We break up the data for Figure 27.1 into separate units. The next cell lists coordinates of the four vertices, and the subsequent one lists the triples of points needed to determine the curves.

```
p[0]:= {0,0}
p[1]:= {6,1}
p[2]:= {8,6}
p[3]:= {2,7}
```

```
pp[0]:= {p[0], p[1], {3,1.5}}
pp[1]:= {p[1], p[2], {6.5,2.8}}
pp[2]:= {p[2], p[3], {4,6}}
pp[3]:= {p[3], p[0], {.5,2.5}}
```

The graphics object z[i] now consists of the curve together with its two tangents:

```
Clear[z]
z[i_]:= {edge@@pp[i], tangents@@pp[i]}
```

```
S @ z[3]
```

Circular arcs indicating the exterior angles are formed from *Mathematica*'s command Circle, introduced in N1.1. The center of the circle y[i] is defined automatically to be p[i]. The extent of the arc is represented by the pair d[i]={a,b}, which defines an arc from aπ to bπ radians. In theory, the value of d[i] could be defined automatically; in practice, all the intervals were added by hand after plotting the edges and inspecting the diagram!

```
Module[{d,r},
  r[i_]= 1;
  d[0]= {-.55,.22}; r[0]=.8;
  d[1]= {-.14,.42};
  d[2]= {.25,1.1}; r[2]=.8;
  d[3]= {.81,1.3};
  y[i_]:= Circle[p[i], r[i], d[i]π]
]
```

The default radius of each circle is 1, but this is modified twice above by adding the required radius by hand after the definition of d[i]. For the final diagram to take account of such modifications, it is essential that the definition of y be made with the symbol := in place of mere equality, as it is sensitive to when evaluation is made. The final ingredient is the text, which needs no comment:

```
txt:= {tx[ϵ0, {1.1,-.6}], tx[ϵ1, {7.4,1.3}], tx[ϵ2, {7.4,7.2}],
  tx[ϵ3, {.7, 6.4}]}//Graphics
```

We put everything together by letting z[i] with i=4,...,7 correspond to the circular arcs. The conditional at the end of line 3 below stops the definition being applied unless i>3, so that z[i] refers to the definition above for i=0,...,3.

```
Module[{hk,ay},
  hk= {RGBColor[0,.6,0], Thickness[.01]};
  z[i_]:= Graphics[Append[hk, y[i-4]]]/; i>3 ;
  z[8]= txt;
  ay:= Flatten[Table[z[i], {i,0,8}]];
  S[ay, AspectRatio->Automatic]
]
```

Figure 27.1

■ N27.4

We can easily generalize the construction of N27.2 to allow quadratic curves that are not necessarily graphs over the horizontal axis. For our purposes, it will be sufficient to include parabolas x=ay^2+by+c, that are obtained by rotating the previous ones by $\pi/2$. If we designate these new ones "type 2," and the previous ones "type 1," we can insert the index k=1,2 as a fourth argument of the new command pts2parabolas. The latter is now plural, as there are two choices; the Off command is used to eliminate the resulting spelling query.

```
Off[General::spell1]
pts2parabolas[u_,v_,w_,k_]:=
Module[{θ,R,f,eq,ta,so},
  θ= If[k==1, 0, π/2];
  R[x_]= Rotate2D[x, θ];
  f[t_]= a t^2 + b t + c;
  eq[x_]:= f[x[[1]]]==x[[2]];
  ta= eq/@R/@{u,v,w};
  so= Solve[ta, {a,b,c}][[1]];
  Rotate2D[{#, f[#]/.so}, -θ] &
]
```

In the plural version of edge, the index k=1,2 is also inserted into the plotting range, so as to determine whether it is the x- or y-coordinate that determines the parameter range. To emphasize the difference between the two types of parabolas, they are also colored slightly differently.

```
Off[General::spell1]
edges[u_,v_,w_,k_]:=
Module[{α},
   α= pts2parabolas[u,v,w, k];
   ParametricPlot[α[t], {t, u[[k]], v[[k]]}, Axes->False,
    PlotStyle->{Thickness[.01],Hue[.4+k/5]}]
]
```

Here is the first example from N27.2.

```
u:= {0,0}
v:= {2,1}
w:= {1,-3}
β:= pts2parabolas[u,v,w,2]
β[t]
```

It is essential to order the arguments of edges correctly in order that the two types of parabolas extend through all three points:

```
z0:= {PointSize[.03], Point[u], Point[v], Point[w]}//Graphics
z1:= edges[u,v,w,1]
z2:= edges[v,w,u,2]
S[z0, z1, z2, Axes->True, PlotRange->All]
```

With these modifications, the remaining code in the construction of Figure 27.2 mimics that of Figure 27.1, and we suppress the details.

- N27.4.1

■ N27.5

We shall need the abbreviation for the identity matrix given in N7.1.

```
id:= IdentityMatrix[3]
```

Here is a program to plot the four tori of Figure 23.3:

```
holes[a_,d_,p_,R_:id]:=
Module[{y,yy,r,s,z,ta},
  y= torus[a,1,1];
  yy[u_,v_]:= RotateRight[y[u,v]];
  r= {d,0,0};
  s= {0,0,d};
  z[X_]:= ParametricPlot3D[R.X, {u,0,2π}, {v,-π/2,3π/2},
    PlotPoints->p];
  ta= {z[y[u,v]+r], z[y[u,v]-r], z[yy[u,v]+s], z[yy[u,v]-s]};
  Show[ta, Boxed->False, Axes->False, PlotRange->All]
]
```

The next two pictures can be seen by adding S@, but similar ones are combined for Figure 27.3.

```
holes[4,4,7]
```

By varying the value of PlotPoints, we can pass from a polyhedral representation of the surface to one that is approximately smooth.

```
holes[5,6,20]
```

```
S @ GraphicsArray[{holes[4,4,30], holes[3,3,5]},
  GraphicsSpacing->-.2]
```

Figure 27.3

The next program computes a rotation matrix associated to a vector x (of norm less than or equal to 1). It is a variant of one defined in N23.1, but incorporates a conditional to avoid evaluating B when x is the null vector.

```
EAif[x_]:=
Module[{l,A,B},
  A[y_]:= Table[Cross[id[[i]],y], {i,3}];
  l= L[x];
  B:= Sin[l] A[x]/l + (1-Cos[l]) A[x].A[x]/l^2;
  id + If[l==0, 0, B]
]
```

We use it to rotate the polyhedral surface. To display the plot, replace Show by S. Without introducing extra small triangular faces, the central part of the picture incorporates 14 vertices and 20 edges.

```
Module[{w,h},
  w[t_]= EAif[.6π{Cos[t], Sin[t], 0}];
  h[t_]:= Show[holes[3,3,5, w[t]],
    PlotRange->{{-6,6},{-6,6},{-6,6}}];
  Do[h[π k/18], {k,36}]
]
```

Here is another approach to topological sketching using an implicit plot of the set of zeros of the *product* of two torus functions:

```
torusimplicit[a_,b_][x_,y_,z_]:= z^2 + (Sqrt[x^2+y^2]-a)^2 - b^2
```

```
f[x_,y_,z_]:= torusimplicit[8,3][x-8,y,z] torusimplicit[8,3][x+8,y,z]
S @ ImplicitPlot3D[f[x,y,z], {x,-20,20}, {y,-12,12},
  {z,-10,10}, PlotPoints->{{7},{3},{2}}, Boxed->False]
```

- N27.5.1

N27.8

The following computations support the exercises to verify the Gauss-Bonnet theorem. First we define the relevant integrand.

```
infarea:= Sqrt[Simplify[ee gg - ff^2]]
```

```
tor:= torus[a,b]
metric[tor][u,v]
```

The Euler characteristic of a torus is zero:

```
Integrate[PowerExpand[gcurv[tor][u,v]infarea], {u,0,2π}, {v,0,2π}]
```

We can also compute the integral of the square of the mean curvature.

```
Integrate[PowerExpand[mcurv[tor][u,v]^2], {u,0,2π}, {v,0,2π}]
```

Finally, we verify the Gauss-Bonnet theorem for an ellipsoid:

```
ell:= ellipsoid[a,b]
metric[ell][u,v];
Integrate[gcurv[ell][u,v]infarea, {u,0,2π}, {v,-π/2,π/2}]
%//PowerExpand
```

Bibliography

[AbSt] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover Publications, New York, 1965.

[Agnesi] M. G. Agnesi, Instituzioni Analitiche ad Uso della Gioventù Italiana, nella Regia-Ducal Corte, Milan, 1748. English by John Colson, Taylor and Wilks, London, 1801.

[Ahlf] L. V. Ahlfors, Complex Analysis, 3rd Edition, McGraw-Hill, New York, 1979.

[AhSa] L. V. Ahlfors and L. Sario, Riemann Surfaces, Princeton University Press, Princeton, NJ, 1960.

[Akh] N. I. Akhiezer, Elements of the Theory of Elliptic Functions, *Translations of Mathematical Monographs* ***79***, American Mathematical Society, Providence, RI, 1990.

[AkGo1] M. A. Akivis and V. V. Goldberg, Introduction to Linear Algebra and Tensors, Dover Publications, New York, 1977.

[Alek] A. D. Aleksandrov, Curves and Surfaces, *Mathematics — Its Content, Methods and Meaning*, Volume 2, MIT Press, Cambridge, MA, 1962, 57–117.

[AnIb] R. L. Anderson and N. H. Ibraigimov, Lie-Bäcklund Transformations in Applications, *SIAM Studies in Applied Mathematics* ***1***, SIAM, Philadelphia, 1979.

[Aoust] L'abbé L. Aoust, Analyse Infinitésimale des Courbes, Gauthier-Villars, Paris, 1869.

[Apéry] F. Apéry, Models of the Real Projective Plane, Friedr. Vieweg & Sohn, Braunschweig, 1987.

[Arch] R. C. Archibald, Curve, *Encyclopedia Britannica*, 14th Edition, New York, 1929.

[Arg] D. E. Arganbright, Practical Handbook of Spreadsheet Curves and Geometric Constructions, CRC Press, Boca Raton, FL, 1993.

[Arm] M. A. Armstrong, Basic Topology, *Undergraduate Texts in Mathematics*, Springer-Verlag, Berlin-New York, 1983.

[Atiyah] M. F. Atiyah, Lectures on Yang-Mills Fields, *Lezioni Fermiane*, Scuola Normale Superiore, Pisa, 1979.

[Bck] A. V. Bäcklund, Om ytori med konstant negativ Krökning, *Lunds Universitets Arsskrift* ***19***, 1883. English translation by E. D. Coddington: Concerning Surfaces with Constant Negative Curvature, New Era Printing Co, Lancaster, PA, 1905.

[Bahd] T. Bahder, Mathematica for Scientists and Engineers, Addison-Wesley, Reading, MA, 1995.

[Baker] A. Baker, A Concise Introduction to the Theory of Numbers, Cambridge University Press, Cambridge, 1994.

[Ball] W. W. Rouse Ball, A Short Account of the History of Mathematics, 4th Edition, MacMillan, London, 1908. Reprint by Dover Publications, New York, 1960.

[Banch] T. Banchoff, Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions, Scientific American Library, New York, 1990.

[BGM] T. Banchoff, T. Gaffney, and C. McCrory, Cusps of Gauss Mappings, *Res. Notes in Math.* ***55***, Pitman, Boston-London, 1982.

[Bar] J. É. Barbier, Note sur le problème du l'aiguille et le jeu du joint couvert, *J. Math. Pures et Appl. (2)* ***5*** (1860), 273–386.

[BaCo] J. L. M. Barbosa and A. G. Colares, Minimal Surfaces in $\mathbb{R}^3$, *Lecture Notes in Mathematics* ***1195***, Springer-Verlag, Berlin-New York, 1986.

[Baud] P. Baudoin, Les Ovales de Descartes et le Limaçon de Pascal, Librairie Vuibert, Paris, 1938.

[Beach] R.C. Beach, An Introduction to Curves and Surfaces of Computer-Aided Design, Van Nostrand Reinhold, New York, 1991.

[Beltr] E. Beltrami, Saggio di interpretazione della geometria non-euclidea, *Giorn. di Mat. Napoli* ***6***, 284–312. Also, *Opere* ***I***, 406–429.

[BeGo] M. Berger and B. Gostiaux, Differential Geometry: Manifolds, Curves, and Surfaces, *Graduate Texts in Mathematics* ***115***, Springer-Verlag, Berlin-New York, 1988.

[Bers] L. Bers, Riemann Surfaces, New York University, Institute of Mathematical Sciences, 1958.

[BDP] J. Bertrand, M. Diguet, and V. Puiseux, Démonstration d'un théorème de Gauss, *Journal de Mathématiques* ***13*** (1848), 80–90.

[Bian] L. Bianchi, Lezioni di Geometria Differenziale, 3a edizione interamente rifatta, Nicola Zanichelli, Bologna, 1927. German translation by M. Lukat: Vorlesungen über Differentialgeometrie, Teubner, Leipzig, 1899, 1910.

[Bieb] L. Bieberbach, Differentialgeometrie, *Teubners Mathematische Leitfäden* ***31***, Teubner, Leipzig, 1932.

[BiGo] R. L. Bishop and S. I. Goldberg, Tensor Analysis on Manifolds, The Macmillan Company, New York, 1968.

[Bj] E. G. Björling, In integrationem aequationis derivatarum partialum superficiei, cujus in puncto unoquoque principales ambo radii curvedinis aequales sunt signoque contrario, *Arch. Math. Phys.* ***(1) 4*** (1844), 290–315.

[Blach] N. Blachman, Mathematica Quick Reference, Version 2, Variable Symbols, Berkeley, CA, 1992.

[Blas1] W. Blaschke, Kreis und Kugel, Veit, Leipzig, 1916.

[Blas2] W. Blaschke, Differentialgeometrie, Volumes 1, 2, Chelsea Publishing Co, New York, 1967, Volume 3, Springer-Verlag, Berlin, 1929.

[Boas] R. P. Boas, Invitation to Complex Analysis, Random House, New York, 1987.

[Bolke] G. Bolke, Die Komplementärflächen der pseudosphährischen Rotationsflächen und ihr Zusammenhang mit allgemeineren pseudosphährischen Flächen, Inaugural Dissertation, Halle, 1901.

[Bolt] V. C. Boltyanski, Envelopes, Pergamon Press, Oxford, 1964.

[Bomp] E. Bompiani, Rappresentazione grafica delle ciclidi di Dupin e delle loro lossodromiche, *Memorie del R. Istituto Lombardo di Scienze e Lettere* ***(3) 21-2*** (1915), 205–241.

[Bonn1] P. O. Bonnet, Mémoire sur la théorie générale des surfaces, *J. Éc. Polytech.* ***19*** (1848), 129–168.

[Bonn2] P. O. Bonnet, Mémoire sur la théorie des surfaces applicables sur une surface donnée, *J. Éc. Polytech.* ***24*** (1865), 209–230.

[Boot] W. M. Boothby, An Introduction to Differentiable Manifolds and Riemannian Geometry, 2nd Edition, Academic Press, New York, 1986.

[BrKn] E. Brieskorn and H. Knörrer, Plane Algebraic Curves, Birkhäuser Verlag, Basel-Boston, 1986.

[Brio] F. Brioschi, Intorno ad alcuni punti della teoria delle superficie, *Annali di Scienze Matematiche e Fisiche* ***3*** (1852), 293–321. Also, *Opere* ***III***, 11–33.

[BrLa] Th. Bröcker and L. Lander, Differentiable Germs and Catastrophes, *London Math. Soc. Lecture Notes* ***17***, Cambridge University Press, Cambridge, 1984.

[BrGi] J. W. Bruce and P. J. Gibbin, Curves and Singularities, Cambridge University Press, Cambridge, 1982.

[BCGH] R. L. Bryant, S. S. Chern, R. B. Gardner, and H. L. Goldschmidt, Exterior Differential Systems, *Mathematical Sciences Research Institute Publications* ***18***, Springer-Verlag, Berlin-New York, 1991.

[Buck] R. C. Buck, Advanced Calculus, 2nd Edition, McGraw-Hill, New York, 1965.

[BuBu] C. Burali-Forti and P. Burgatti, Analisi Vettorale Generale e Applicazioni, Volume 2, Geometria Differenziale, Nicola Zanichelli, Bologna, 1930.

[BuZi] G. Burde and H. Zieschang, Knots, *de Gruyter Studies in Mathematics* ***5***, Walter de Gruyter, Berlin-New York, 1985.

[BFLPU] F. Burstall, D. Ferus, K. Leschke, F. Pedit, and U. Pinkall, Conformal Geometry of Surfaces in S^4 and Quaternions, *Lecture Notes in Mathematics* ***1772***, Springer-Verlag, Berlin-New York, 2002.

[Cajori] F. Cajori, A History of Mathematics, 3rd Edition, Chelsea Publishing Co, New York, 1980.

[dC1] M. P. do Carmo, Differential Geometry of Curves and Surfaces, Prentice Hall, Englewood Cliffs, NJ, 1976.

[dC2] M. P. do Carmo, Superfícies Mínimas, 16*o Colóquio Brasileiro de Matemática*, IMPA, Rio de Janeiro, 1987.

[Cartan] É. Cartan, La Méthode du Repère Mobile, la Théorie des Groupes Continus et les Espaces Généralisés, *Exposés de Géométrie*, V, Hermann, Paris, 1935.

[Casey] J. Casey, Exploring Curvature, Friedr. Vieweg & Sohn, Braunschweig, 1996.

[Cata] E. Catalan, Mémoire sur les surfaces dont les rayons de courbure, en chaque point, sont égaux et de signes contraires, *J. Éc. Polytech.* ***21*** (1858), 129–168.

[CeRy] T. E. Cecil and P. J. Ryan, Tight and Taut Immersions of Manifolds, *Res. Notes in Math.* ***131***, Pitman, Boston-London, 1985.

[Ces] E. Cesàro, Lezioni di Geometria Intrinseca, Tipografia della R. Accademia delle Scienze, Napoli, 1896. German translation by G. Kowalewski: Vorlesungen über Natürliche Geometrie, Teubner, Leipzig, 1901.

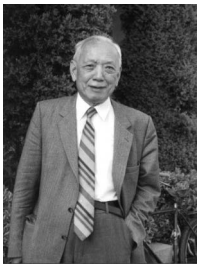
[Chand] K. Chandrasekharan, Elliptic Functions, *Grundlehren der Mathematischen Wissenschaften* ***281***, Springer-Verlag, Berlin-New York, 1985.

[Ch] M. Chasles, Les lignes géodésiques et les lignes de courbure des surfaces du second degré, *Journ. de Math.* ***11*** (1846), 5–20.

[Chern1] S. S. Chern,[1] Introduction to Differential Geometry, Notes, University of California, Berkeley, CA, 1963.

[Chern2] S. S. Chern, Studies in Global Analysis and Geometry, *MAA Studies in Mathematics* ***27***, The Mathematical Association of America, Washington, DC, 1989.

[Chern3] S. S. Chern, Surface theory with Darboux and Bianchi, Miscellanea Mathematica, 59–69, Springer-Verlag, Berlin-New York, 1991.



[1] Shiing-Shen Chern (1911–2004).

[Childe] G. F. Childe, Investigations in the Theory of Reflected Ray-Surfaces, J. C. Juta, Cape Town, 1857.

[Chris] E. B. Christoffel, Über die Transformation der homogenen Differentialausdrücke zweiten Grades, *J. Reine Angew. Math.* (*Crelle's Journal*) ***70*** (1869), 46–70.

[Codaz] D. Codazzi, Sulle coordinate curvilinee d'una superficie dello spazio, *Annali Mat. Pura Appl.* ***(2) 2*** (1868–1869), 101–119.

[CV] S. Cohn-Vossen, Kürzste Wege und Totalkrümmung auf Flächen, *Composito Math.* ***2*** (1935), 69–133.

[Conlon] L. Conlon, Differentiable Manifolds: A First Course, Birkhäuser Verlag, Basel-Boston, 1991.

[CoSm] J. H. Conway and D. A. Smith, On Quaternions and Octonions: Their Geometry, Arithmetic, and Symmetry, A. K. Peters, Natick, MA, 2003.

[Cool1] J. L. Coolidge, A Treatise on the Circle and the Sphere, Oxford, Clarendon Press, 1916. Reprint by Chelsea Publishing Co, New York, 1971.

[Cool2] J. L. Coolidge, The Geometry of the Complex Domain, Oxford, Clarendon Press, 1924.

[Cool3] J. L. Coolidge, A Treatise on Algebraic Plane Curves, Oxford, Clarendon Press, 1931. Reprint by Dover Publications, New York, 1959.

[Cool4] J. L. Coolidge, A History of Geometrical Methods, Oxford University Press, London, 1945. Reprint by Dover Publications, New York, 1963.

[Cool5] J. L. Coolidge, A History of the Conic Sections and Quadric Surfaces, Oxford University Press, London, 1945.

[Costa1] C. J. Costa, Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends, *Bol. Soc. Brasil Math.* ***15*** (1984), 47–54.

[Costa2] S. I. R. Costa, On closed twisted curves, *Proc. Amer. Math. Soc.* ***109*** (1990), 205–214.

[Coxeter] H. S. M. Coxeter, Introduction to Geometry, 2nd Edition, John Wiley and Sons, New York, 1989.

[Crowe] M. J. Crowe, A History of Vector Analysis, Dover Publications, New York, 1994.

[CF] R. H. Crowell and R. H. Fox, Introduction to Knot Theory, Ginn and Company, Boston, 1963. Reprint as *Graduate Texts in Mathematics* ***57***, Springer-Verlag, Berlin-New York, 1978.

[DaKaSa] Y. Dang, L. H. Kauffman, and D. Sandin, Hypercomplex iterations, *Series on Knots and Everything* ***17***, World Scientific, Singapore, 2002.

[Darb1] G. Darboux, Sur une Classe Remarquable de Courbes et de Surfaces Algebriques et sur la Theorie des Imaginaires, Gauthier-Villars, Paris, 1873. Reprint by University Microfilms International, Ann Arbor, MI, 1992.

[Darb2] G. Darboux, Leçons sur la Théorie Générale des Surfaces, 4 volumes, Gauthier-Villars, Paris, 1914, 1915, 1894, 1896. Reprint by Chelsea Publishing Co, New York, 1972.

[Darb3] G. Darboux, Leçons sur les Systèmes Orthogonaux et les Coordonnées, Gauthier-Villars, Paris, 1898.

[Darb4] G. Darboux, Principes de Géométrie Analytique, Gauthier-Villars, Paris, 1917.

[Dark] H. E. Dark, The Wankel Rotary Engine: Introduction and Guide, Indiana University Press, Bloomington, IN, 1974.

[Dedò] M. Dedò, Trasformazioni Geometriche, con un'Introduzione al Modello di Poincaré, Zanichelli, Bologna, 1996.

[De] A. Deprit, Letter to the editor, *Mathematica J.* ***1***, Issue 2, (1990), 4.

[DHKW] U. Dierkes, S. Hildebrandt, A. Küster, and O. Wohlrab, Minimal Surfaces, 2 volumes, Springer-Verlag, Berlin-New York, 1992.

[DoPo] C. T. J. Dodson and T. Poston, Tensor Geometry, 2nd Edition, *Graduate Texts in Mathematics* ***130***, Springer-Verlag, Berlin-New York, 1991.

[Dom] P. Dombrowski, 150 Years after Gauss's "Disquisitiones Generales Circa Superficies Curvas," *Astérisque* ***62***, Société Mathématique de France, Paris, 1979.

[DFN] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, Modern Geometry — Methods and Applications, *Graduate Texts in Mathematics* ***93***, ***104***, Springer-Verlag, Berlin-New York, 1984, 1985.

[DuKo] J. J. Duistermaat and J. A. C. Kolk, Lie Groups, *Universitext*, Springer-Verlag, Berlin-New York, 1999.

[Dupin1] C. Dupin, Développements de Géométrie, Courcier, Paris, 1813.

[Dupin2] C. Dupin, Applications de Géométrie et de Méchanique, Bachelier, Paris, 1822.

[Eells] J. Eells, The surfaces of Delaunay, *Math. Intelligencer* **9** (1987), 53–57.

[Eisen1] L. P. Eisenhart, A Treatise on the Differential Geometry of Curves and Surfaces, Ginn and Company, Boston, 1909. Reprint by Dover Publications, New York, 1960.

[Eisen2] L. P. Eisenhart, Transformations of Surfaces, Princeton University Press, Princeton, NJ, 1923.

[Eisen4] L. P. Eisenhart, Riemannian Geometry, Princeton University Press, Princeton, NJ, 1949.

[Enn1] A. Enneper, Analytisch-geometrische Untersuchungen, *Z. Math. Phys.* **9** (1864), 96–125.

[Erfurt] P. Erfurt, Die Komplementärflächen der pseudosphährischen Schraubenflächen, Inaugural Dissertation, Halle, 1908.

[Euler1] L. Euler, De constructione aequationum ope motus tractorii aliisque ad methodum tangentium inversam pertinentibusm, *Comment. Acad. Petropolit.* **8** (1736), 66–85. Reprinted in *Opera omnia* **(1) 22**, Basel 1926, 83–107.

[Euler2] L. Euler, Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti, *Opera omnia* **(1) 24**, Füssli Turici, Lausanne, 1952. English translation (in part) by D. J. Struik [Stru3, pages 399–406].

[Euler3] L. Euler, Recherches sur la courbure des surfaces, *Mém. de l'Acad. des Sciences, Berlin* **16** (1760), 119–143. Reprinted in *Opera omnia* **(1) 28**, Füssli Turici, Lausanne, 1955, 1–22.

[Euler4] L. Euler, De curvis triangularibus, *Acta Acad. Petropolit.* **2** (1778), 3–30. *Opera omnia* **(1) 28**, Füssli Turici, Lausanne 1955, 298–321.

[Faug] O. Faugeras, Three-Dimensional Computer Vision, The MIT Press, Cambridge MA, 1999.

[Fava] F. Fava, Geometria Differenziale, Levrotto & Bella, Torino, 1991.

[Fed] A. S. Fedenko, A Collection of Problems in Differential Geometry, Nauka, Moscow, 1979. Spanish translation, Mir, 1981. French translation, Mir, 1982.

[Ferg] C. Ferguson and H. Ferguson, Helaman Ferguson: Mathematics in Stone and Bronze, Meridian Creative Group, 1994.

[FeGrMa] H. Ferguson, A. Gray, and S. Markvorsen, Costa's minimal surface via Mathematica *Mathematica in Research and Education* ***5***, (1996) 5–10.

[FeWo] M. Fernández and J. Wolf (editors), Global Differential Geometry: The Mathematical Legacy of Alfred Gray, *Contempopary Mathematics Series* ***288***, American Mathematical Society, Providence, RI, 2001.

[FiGa] P. A. Firby and C. F. Gardiner, Surface Topology, Halsted Press, 1982.

[Fischer] G. Fischer, Mathematical Models, 2 volumes, Friedr. Vieweg & Sohn, Braunschweig, 1986.

[FoTu] A. T. Fomenko and A. A. Tuzhilin, Elements of the Geometry and Topology of Minimal Surfaces in Three-Dimensional Space, *Translations of Mathematical Monographs* ***93***, American Mathematical Society, Providence, RI, 1991.

[Francis] G. K. Francis, A Topological Picture Book, Springer-Verlag, Berlin-New York, 1987.

[Frenet] F. Frenet, Sur les courbes à double courbure, Thèse, Toulouse, 1847. Abstract in *Journ. de Math.* ***17***, 1852.

[Frost] P. Frost, Curve Tracing, MacMillan, London, 1926.

[Fult] W. Fulton, Algebraic Topology, A first Course, *Graduate Texts in Mathematics* ***153***, Springer-Verlag, Berlin-New York, 1995.

[Fuss] N. Fuss, De sphaeris osculantibus, *Mem. Acad. Sci. St. Pétersbourg*, ***7*** (1820).

[Gauss1] C. F. Gauss, Neue allgemeine untersuchungen über Krummen Flächen, *Werke* ***8***, 408-443.[2]

[Gauss2] C. F. Gauss, Disquisitiones Generales Circa Superficies Curvas, *Comm. Soc. Reg. Sc. Gott. Rec.*, ***6*** (1827), 99–146.

[2][Gauss1], which is a preliminary version of [Gauss2], shows the manner in which Gauss's ideas evolved. In particular, [Gauss1] contains material on the curvature of plane curves that is absent from [Gauss2]. In the nineteenth century there were three Latin versions of [Gauss2], together with two translations into French and two translations into German. For English translations of both [Gauss1] and [Gauss2] together with notes and a bibliography of related papers, see [MoHi].

[GaKaWe] R. J. Gaylord, S. N. Kamin, and P. R. Wellin, Introduction to Programming with Mathematica, Springer-Verlag, TELOS, Berlin-New York, 1996.

[Gomes] F. Gomes Teixeira, Traité des courbes spéciales remarquables planes et gauches, Coimbra, 1908–1915. Reprint by Chelsea Publishing Co, New York, 1971.

[Gott] D. H. Gottlieb, All the way with Gauss-Bonnet and the sociology of mathematics, *Amer. Math. Monthly* ***103*** (1996), 457–469, ***104*** (1997), 35.

[Gour] É. Goursat, Sur un mode de transformation des surfaces minima, *Acta Math.* ***11*** (1887–1888), 135–186 and 257–264.

[Gray] A. Gray,[3] Tubes, 2nd Edition, *Progress in Mathematics* ***221***, Birkhäuser Verlag, Basel-Boston, 2004.

[GrMePi] A. Gray, M. Mezzino, and M. A. Pinsky, Introduction to Ordinary Differential Equations with Mathematica: an Integrated Multimedia Approach, Springer-Verlag, TELOS, Berlin-New York, 1997.

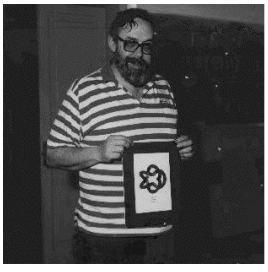
[GrGl2] T. W. Gray and J. Glynn, The Beginner's Guide to Mathematica, Addison-Wesley, Reading, MA, 1992.

[Gugg] H. W. Guggenheimer, Differential Geometry, McGraw-Hill, 1963. Reprint by Dover Publications, New York, 1977.

[Hady] M. A. Hady, Beiträge zur Theorie der W-Flächen, *Bonner Mathematische Schriften* ***16***.

[HaEl] S. W. Hawking and G. R. R. Ellis, The Large Scale Structure of Space-Time, Cambridge University Press, Cambridge, 1973.

[Her] I. N. Herstein, Topics in Algebra, 2nd Edition, John Wiley and Sons, New York, 1975.



[3] The author (1939–1998). In Turin, 1993.

[Hicks] N. J. Hicks, Notes on Differential Geometry, Van Nostrand Reinhold, Princeton, NJ, 1965.

[Higg] A. L. Higgins, The Transition Spiral and Its Introduction to Railway Curves with Field Exercises in Construction and Alignment, Constable, London, 1921.

[Hil] D. Hilbert, Über Flächen von konstanter Gausscher Krümmung, *Trans. Amer. Math. Soc.* ***2*** (1901), 87-99.

[HC-V] D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea Publishing Co, New York, 1952. Anschauliche Geometrie, Zweite Auflage, Springer-Verlag, Berlin, 1996.

[Hille] E. Hille, Analytic Function Theory, 2nd Edition, 2 volumes, Chelsea Publishing Co, New York, 1973.

[Hoff] D. Hoffman, The computer-aided discovery of new embedded minimal surfaces, *Mathematical Intelligencer* ***9*** (1987), 8–21.

[HoMe] D. Hoffman and W. Meeks, A complete minimal surface in $\mathbb{R}^3$ with genus one and three ends, *J. Differential Geometry* ***21*** (1985), 109–127.

[Hopf1] H. Hopf, Über die Drehung der Tangenten und Sehen Ebener Kurven, *Composito Math.* ***2*** (1935), 50–62.

[Hopf2] H. Hopf, Differential Geometry in the Large, Seminar Lectures, New York University 1946 and Stanford University 1956, *Lecture Notes in Mathematics* ***1000***, Springer-Verlag, Berlin-New York, 1983.

[HR] H. Hopf and W. Rinow, Über den Begriff der vollständigen differentialgeometrischen Flächen, *Comm. Math. Helv.* ***3*** (1931), 209–225.

[Jacobi] C. G. J. Jacobi, Ueber einige merkwürdige Curventheoreme, *Astr. Nachr.* ***20*** (1843), 115–120. *Math. Werke* ***7***, 34–39.

[Joach1] M. F. Joachimsthal, Demostrationes theorematum ad superficies curvas spectantium, *J. Reine Angew. Math.* (*Crelle's Journal*) ***30*** (1846), 347–350.

[Joach2] M. F. Joachimsthal, Anwendung der Differential- und Integralrechnung auf die Allegemeine Theorie der Flächen und der Linien Doppelter Krümmung, 3rd Edition, Teubner, Leipzig, 1890.

[Kauf1] L. H. Kauffman, On Knots, *Ann. Math. Stud.* ***115***, Princeton University Press, Princeton, NJ, 1987.

[Kauf2] L. H. Kauffman, Knots and Physics, World Scientific, Singapore, 1991.

[Kelley] J. L. Kelley, General Topology, Springer-Verlag, Berlin-New York, 1975.

[Kenmo] K. Kenmotsu, Surfaces of revolution with prescribed mean curvature, *Tôhoku Math. J. (2) 32* (1980), 147–153.

[Kir] F. Kirwan, Complex Algebraic Curves, *London Mathematical Society Student Texts 23*, Cambridge University Press, Cambridge, 1992.

[Klein] F. Klein, Bemerkungen über den Zusammenhang der Flächen, *Math. Ann. 7* (1874), and *9* (1875/76). *Math. Werke 2*, 63–77.

[Kosz] J. L. Koszul, Homologie et cohomologie des algèbres de Lie, *Bull. Soc. Math. France 78* (1950), 65–127.

[Krey1] E. Kreyszig, Introduction to Differential Geometry and Riemannian Geometry, *Mathematical Expositions No. 16*, University of Toronto Press, 1968. Reprint by Dover Publications, New York, 1991.

[Krey2] E. Kreyszig, On the theory of minimal surfaces, The Problem of Plateau, 138–164, World Scientific, Singapore, 1993.

[Kuen] Th. Kuen, Fläche von constantem negativen Krümmungsmass nach L. Bianchi, *Sitzungsber. d. Bayer. Akad.*, 1884.

[Kuipers] J. B. Kuipers, Quaternions and Rotation Sequences, Princeton University Press, Princeton and Oxford, 2002.

[Lag] J. L. Lagrange, Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies. *Miscellanea Taurinensia 2* (1760–1761), 173–195. Œuvres de Lagrange, Volume 1, Gauthier-Villars, Paris, 1867, 335–362. English translation (in part) by D. J. Struik [Stru3, 406–413].

[Lait] E. V. Laitone, Relation of the conjugate harmonic function to $f(z)$, *Amer. Math. Monthly 84* (1977), 281–283. Review: *Math. Rev. 55*, # 10646.

[Lamb1] G. P. Lamb, Bäcklund transformations at the turn of the century, Bäcklund Transformations, *Lecture Notes in Mathematics 515*, 69–80, Springer-Verlag, Berlin-New York, 1976.

[Lamb2] G. P. Lamb, Elements of Soliton Theory, John Wiley and Sons, New York, 1980.

[Lamé] G. Lamé, Leçons sur les Coordonnées Curvilignes, Mallet-Bachelier, Paris, 1859. Reprint by University Microfilms International, Ann Arbor, MI, 1992.

[Lang] S. Lang, Differential and Riemannian manifolds, 2nd Edition, Springer-Verlag, New York, 1995.

[Law] J. D. Lawrence, A Catalog of Special Plane Curves, Dover Publications, New York, 1972.

[Lee] J. M. Lee, Riemannian Geometry, Springer-Verlag, New York, 1997.

[Lehr] E. Lehr, Über die Kurven, deren Krümmung eine periodische Funktion des Bogens ist, Thesis, Technische Hochschule München, 1932.

[Lem] J. Lemaire, Hypocycloïdes et Épicycloïdes, Librairie Vuibert, Paris, 1938. Reprint by Blanchard, Paris, 1967.

[LeviC] T. Levi-Civita, Lezioni di Calcolo Differenziale Assoluto, Roma, 1925. Reprinted by Dover Publications, New York, 1977 as The Absolute Differential Calculus.

[LCRC] T. Levi-Civita and G. Ricci-Curbastro, Méthodes de calcul différentiel absolu et leurs applications. *Math. Ann.* ***54*** (1901), 125–201.

[Lie] S. Lie, Beiträge zur Theorie der Minimalflächen I, *Math Ann.* ***14*** (1879), 331–416.

[Liou] J. Liouville, Note au sujet de l'article précédent, *J. Math. Pures et Appl.* ***12*** (1847), 265–290.

[Lip] M. M. Lipschutz, Theory and Problems of Differential Geometry, *Schaum's Outline Series*, McGraw-Hill, New York, 1969.

[Lock] E. H. Lockwood, A Book of Curves, Cambridge University Press, Cambridge, 1961.

[Loria1] G. Loria, Curve Piane Speciali, Ulrico Hoepli, Milano, 1930. German translation: Spezielle algebraische und transzendente Kurven, Teubner, Leipzig, 1910.

[Loria2] G. Loria, Curve Sghembe Speciali, 2 volumes, Nicola Zanichelli, Bologna, 1910.

[Ma] R. Maeder, Programming in Mathematica, 2nd Edition, The Advanced Book Program, Addison-Wesley, Redwood City, CA, 1991.

[Main] G. Mainardi, Sulle coordinate curvilinee d'una superficie dello spazio, *Giornale del R. Istituto Lombardo* ***9*** (1856), 385-398.

[Mark] A. I. Markushevich, Remarkable Curves, Mir Publishers, Moscow, 1978.

[Massey] W. S. Massey, Algebraic Topology: An Introduction, Springer-Verlag, Berlin-New York, 1977.

[Max] J. C. Maxwell, On the cyclide, *Quarterly J. of Pure and Applied Math.* ***34*** (1867), 144–159. Also, Collected Works ***2***, 144–159.

[McCl] J. McCleary, Geometry from a Differentiable Viewpoint, Cambridge University Press, Cambridge, 1994.

[Mell] A. P. Mellish, Notes on Differential Geometry, *Ann. of Math. (2)* ***32*** (1931), 181–190.

[Melz] Z. A. Melzak, Companion to Concrete Mathematics, 2 volumes, John Wiley and Sons, New York, 1973, 1976.

[Meu] J. B. Meusnier, Mémoire sur la courbure des surfaces, *Mémoires des savans étrangers* ***10*** (lu 1776) (1785), 477–510.

[MiPa] R. S. Millman and G. D. Parker, Elements of Differential Geometry, Prentice Hall, Englewood Cliffs, NJ, 1977.

[Mil2] J. W. Milnor, Morse Theory, *Ann. Math. Stud.* ***51***, Princeton University Press, Princeton, NJ, 1969.

[Mind] F. A. Minding, Wie sich unterscheiden läßt, ob zwei gegebene krumme Flächen aufeinander abwickelbar sind oder nicht; nebst Bemerkungen über Flächen von unveränderlichem Krümmungsmaß, *J. Reine Angew. Math.* (*Crelle's Journal*) ***19*** (1839), 370–387.

[Monge] G. Monge, Application de l'Analyse à la Géométrie, 5ème édition, revue, corrigée et annotée par M. Liouville, Bachelier, Paris, 1850. Reprint by University Microfilms International, Ann Arbor, MI, 1979.

[MoRo] S. Montiel and A. Ros, Curvas y Superficies,[4] Proyecto Sur de Ediciones, Granada, 1997.

[MoSp] P. Moon and D. Spencer, Field Theory Handbook, Springer-Verlag, Berlin-New York, 1961.

[4] Added in proof: The English edition "Curves and Surfaces," by the same authors, was published by the American Mathematical Society and the Real Sociedad Matemática Española in 2005, as Volume 69 of Graduate Studies in Mathematics.

[MoHi] J. C. Morehead and A. D. Hiltebreitel, Translation with notes and a bibliography of Gauss's General Investigations of Curves and Surfaces of 1827 and 1825, The Princeton University Library, Princeton, 1902.

[MS] G. Mostow and J. H. Sampson, Linear Algebra, McGraw-Hill, New York, 1969.

[Muk] S. Mukhopadhyaya, New methods in the geometry of a plane arc, *Bull. Calcutta Math. Soc.* ***1*** (1909), 31–37.

[Newton] I. Newton, Curves, The Mathematical Works of Isaak Newton, Volume 2, 135–161, Johnson Reprint Corporation, New York-London, 1962.

[NiSh] V. V. Nikulin and I. R. Shafarevich, Geometries and Groups, Springer-Verlag, Berlin-New York, 1987.

[Nits] J. C. C. Nitsche, Lectures on Minimal Surfaces, Volume 1, Cambridge University Press, Cambridge, 1989. English translation (with additions) of the first part of: Vorlesungen über Minimalflächen, *Grundlehren der Mathematischen Wissenschaften* ***199***, Springer-Verlag, Berlin-New York, 1975.

[NoSa] K. Nomizu and T. Sasaki, Affine Differential Geometry, Cambridge University Press, Cambridge, 1994.

[Og1] A. Ogawa, The trinoid revisited, *Mathematica J.* ***2***, Issue 1, (1992), 58–60.

[Og2] A. Ogawa, Helicatenoid, *Mathematica J.* ***2***, Issue 2, (1992), 21–23.

[ON1] B. O'Neill, Elementary Differential Geometry, 2nd Edition, Academic Press, San Diego, 1997.

[ON2] B. O'Neill, Semi-Riemannian Geometry, Academic Press (Harcourt Brace Jovanovich), New York-London, 1983.

[Oprea1] J. Oprea, Differential Geometry and Its Applications, Prentice Hall, Englewood Cliffs, NJ, 1997.

[Oprea2] J. Oprea, The Mathematics of Soap Films, Student Mathematical Library ***10***, American Mathematical Society, Providence, RI, 2000.

[Os1] R. Osserman, A Survey of Minimal Surfaces, Dover Publications, New York, 1986.

[Os2] R. Osserman, The Geometry Renaissance in America: 1938–1988, A Century of Mathematics in America, volume 2, American Mathematical Society, Providence, 1988.

[PaTe] R. S. Palais and C. Terng, Critical Point Theory and Submanifold Theory, *Lecture Notes in Mathematics* **1353**, Springer-Verlag, Berlin-New York, 1988.

[Pen] R. Penrose, The Road to Reality, Jonathan Cape, London, 2004.

[Pson] K. M. Peterson, Über die Biegung der Flächen, Kandidatenschrift, Dorpat, 1853.

[Psen] P. Petersen, Riemannian Geometry, Springer-Verlag, Berlin-New York, 1998.

[PiSt] U. Pinkall and I. Sterling, On the classification of constant mean curvature tori, *Ann. of Math. (2)* **130** (1989).

[Press] A. Pressley, Elementary Differential Geometry *Undergraduate Texts in Mathematics*, Springer-Verlag, Berlin-New York, 2002.

[RaTo1] H. Rademacher and O. Töplitz, Von Zahlen und Figuren, 2nd Edition, Springer-Verlag, Berlin-New York, 1933.

[RaTo2] H. Rademacher and O. Töplitz, The Enjoyment of Mathematics; Selections from Mathematics for the Amateur, A translation from [RaTo1] with Chapters 15 and 28 added by Herbert Zuckerman. Princeton University Press, Princeton, NJ, 1957. Reprint by Dover Publications, New York, 1990.

[Reich] K. Reich, Die Geschichte der Differentialgeometrie von Gauß bis Riemann (1828–1868), *Arch. for the Hist. of Exact Sci.* **11** (1973), 273-382.

[Reid] K. Reidemeister, Knotentheorie, Springer-Verlag, Berlin-New York, 1933. Reprint, 1974. English translation: Knot theory, BCS Associates, Moscow, ID, USA, 1983.

[Remo] M. Remoissenet, Waves Called Solitons: Concepts and Experiments, Springer-Verlag, Berlin-New York, c 1994.

[Reul] F. Reuleaux, Lehrbuch der Kinematik, 2 volumes, Friedr. Vieweg & Sohn, Braunschweig, 1875.

[Ricci] G. Ricci-Curbastro, Résumé de quelques travaux sur les systémes variables de fonctions, *Bulletin des Sciences Mathématiques* **16** (1892), 167–189.

[Riem] B. Riemann, Über die Hypothesen, welche der Geometrie zugrunde liegen, *Abhandlungen der Königlichen Gesellshaft der Wissenschaften zu Göttingen.* ***13*** (1867).

[RoAd] D. F. Rogers and J. A. Adams, Mathematical Elements for Computer Graphics, 2nd Edition, McGraw-Hill, New York, 1990.

[Rolf] D. Rolfsen, Knots and Links, Publish or Perish, Wilmington, DE, 1976.

[Roth] E. H. Roth, Revised Railway Transition Spiral: Text and Tables, J. P. Bell, Lynchburg, VA, 1955.

[Row] T. S. Row, Geometric Exercises in Paper Folding, 2nd Edition, The Open Court Publishing Company, Chigago, 1905. Reprint by Dover Publications, New York, 1963.

[Ryan] P. J. Ryan, Homogeneity and some curvature conditions for hypersurfaces, *Tôhoku Math. J. (2)* ***21*** (1969), 363–388.

[Salmon1] G. Salmon, A Treatise on the Analytic Geometry of Three Dimensions, Longmans, New York, 1928. Reprint by Chelsea Publishing Co, New York, 1965.

[Salmon2] G. Salmon, A Treatise on the Higher Plane Curves, G.E. Stechert, 1934. Reprint by Chelsea Publishing Co, New York, 1960.

[Sav] A. A. Savelov, Ploskie Krivye (Plane Curves), Fizmatgiz, Moscow, 1960.

[Scha] R. Schaper, Grafik mit Mathematica, Addison-Wesley, Bonn, 1994.

[Scherk] H. F. Scherk, Bemerkung über die kleinste Fläche innerhalb gegebener Grenzen, *J. Reine Angew. Math.* (*Crelle's Journal*) ***13*** (1835), 185–208.

[Schill] F. Schilling, Die Pseudosphäre und die Nichteuklidische Geometrie, Teubner, Leipzig, 1935.

[Schwa] H. A. Schwarz, Gesammelte Mathematische Abhandlungen, Volumes 1 and 2, Springer-Verlag, Berlin, 1890. Reprint by Chelsea Publishing Co, New York, 1972.

[Scott] W. R. Scott, Theory of Groups, Prentice Hall, Englewood Cliffs, NJ, 1964.

[Serret1] J. A. Serret, Sur Quelques Formules Relatives à la Théorie des Courbes à Double Courbure, *Journ. de Math.* ***16*** (1851), 193–207.

[Serret2] P. J. Serret, Théorie nouvelle géométrique et mécanique des lignes à double courbure, Mallet-Bachelier, Paris, 1860.

[Serret] J. A. Serret, Lehrbuch der Differential- und Integralrechnung, 3 volumes, German Translation by A. Harnacks, adapted by G. Scheffers, 4th, 5th Editions, Teubner, Leipzig, 1912.

[Sharpe] R. W. Sharpe, Differential Geometry: Cartan's Generalization of Klein's Erlangen Programme, Springer-Verlag, 1997.

[ShTi] W. T. Shaw and T. Tigg, Applied Mathematica, Getting Started, Getting it Done, The Advanced Book Program, Addison-Wesley, Redwood City, CA, 1993.

[Shikin] E. V. Shikin, Handbook and Atlas of Curves, CRC Press, Boca Raton, FL, 1995.

[ShSh] P. A. Shirikov and A. P. Shirikov, Affine Differentialgeometrie, Teubner, Leipzig, 1962.

[Siev] H. Sievert, Über die Zentralflächen der Enneperschen Flächen konstanten Krümmungsmaßes, Dissertation, Tübingen, 1886, Kap VI.

[SmBl] C. Smith and N. Blachman, The Mathematica Graphics Guide Book, Addison-Wesley, Reading, MA, 1994.

[Sny] V. Snyder, Criteria for Nodes in Dupin's Cyclides, with a Corresponding Classification, *Ann. of Math. (1) 11* (1898), 137–147.

[Spiv] M. Spivak, Calculus on Manifolds, Westview Press, 1971.

[Spivak] M. Spivak, A Comprehensive Introduction to Differential Geometry, 2nd Edition, 5 volumes, Publish or Perish, Wilmington, DE, 1979.

[Springer] G. Springer, An Introduction to Riemann Surfaces, Chelsea Publishing Co, New York, 1981.

[Steur] R. Steuerwald, Über Enneper'sche Flächen und Bäcklund'sche Transformation, *Abh. Bayer. Akad. Wiss. (Munich) 40* (1936), 1–105.

[Still1] J. Stillwell, Mathematics and Its History, Springer-Verlag, Berlin-New York, 1989.

[Still2] J. Stillwell, Geometry of Surfaces, Springer-Verlag, Berlin-New York, 1990.

[Stoker] J. J. Stoker, Differential Geometry, John Wiley and Sons, New York, 1969.

[Strub] K. Strubecker, Differentialgeometrie, 2nd Edition, 3 volumes, *Sammlung Göschen 1113/1113a, 1179/1179a, 1180/1180a*, Walter de Gruyter, Berlin, 1964.

[Stru1] D. J. Struik, Outline of a history of differential geometry, *Isis 19* (1933), 92–120, *20* (1933), 161–191.

[Stru2] D. J. Struik, Lectures on Classical Differential Geometry, 2nd Edition, Addison-Wesley, Reading, MA, 1961. Reprint by Dover Publications, New York, 1988.

[Stru3] D. J. Struik, A Source Book in Mathematics, 1200–1800, Harvard University Press, Cambridge, MA, 1969.

[Study] E. Study, Über einige imaginäre Minimalflächen, *Leipz. Akad.-Ber. 63* (1911), 14–26.

[Tcheb] P. L. Tchebyshef, Sur les coupure des vêtments, *Œuvres 2* (1878), 708.

[Ten] K. Tenenblat, Transformations of Manifolds and Applications to Differential Equations, IMPA, Rio de Janeiro, 1996.

[Thom] D'A. W. Thompson, Growth and Form, Cambridge University Press, Cambridge, 1917.

[Thor] J. A. Thorpe, Elementary Topics in Differential Geometry, *Undergraduate Texts in Mathematics*, Springer-Verlag, Berlin-New York, 1979.

[Vais] I. Vaisman, A First Course in Differential Geometry, Marcel Dekker, New York-Basel, 1984.

[Val] G. Valiron, The Classical Differential Geometry of Curves and Surfaces (Translation by James Glazebrook of Chapters XII–XVI of Cours d'Analyse), Math Sci Press, Brookline, MA, 1986.

[Visc] P. Viscensini, Differential geometry in the nineteenth century, *Scientia 107*, 661–696. Translation of La Géométrie Différentiele au XIXème siécle, *Scientia 107*, 617–660.

[vonSeg] D. H. von Seggern, CRC Handbook of Mathematical Curves and Surfaces, CRC Press, Boca Raton, FL, 1990.

[Wagon] S. Wagon, Mathematica in Action, W.H. Freeman and Company, New York, 1991.

[Walk] R. J. Walker, Algebraic Curves, Princeton University Press, Princeton, NJ, 1950. Reprint by Springer-Verlag, Berlin-New York, 1978.

[Warner] F. W. Warner, Foundations of Differentiable Manifolds and Lie Groups, *Graduate Texts in Mathematics*, Springer-Verlag, Berlin-New York, 1983.

[Wea] C. E. Weatherburn, Differential Geometry of Three Dimensions, 2 volumes, Cambridge University Press, Cambridge, 1927, 1930.

[Wei1] K. Weierstrass, Über die geodätischen Linen auf dem dreiachsigen Ellipsoid, *Math. Werke I*, 257–266.

[Wei2] K. Weierstrass, Untersuchungen über die Flächen, deren mittlere Krümmung überall gleich Null ist, *Moatsber. Berliner Akad.* (1866), 612–625. *Math. Werke III*, 39–52.

[Weyl] H. Weyl, The Concept of a Riemann Surface, Addison-Wesley, Reading, MA, 1964. English translation of Die Idee der Riemannschen Flächen, 3rd Edition, Teubner, Stuttgart, 1955.

[Wick] T. Wickham-Jones, Graphics with Mathematica, Springer-Verlag, TELOS, Berlin-New York, 1994.

[Wiel1] H. Wieleitner, Theorie der Ebene Algebraische Kurven höherer Ordnung, *Sammlung Schubert 13*, G. J. Göschen'sche Verlagshandlung, Leipzig, 1908.

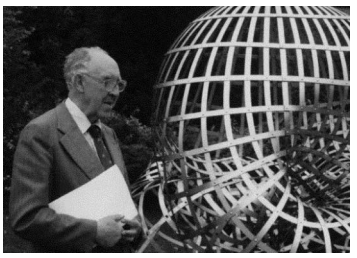
[Wiel2] H. Wieleitner, Spezielle Ebene Kurven, *Sammlung Schubert 56*, G. J. Göschen'sche Verlagshandlung, Leipzig, 1908.

[Will] T. J. Willmore,[5] An Introduction to Differential Geometry, Oxford University Press, London, 1959.

[Wm1] S. Wolfram, The Mathematica Book, 4th Edition, Cambridge University Press, Cambridge, 1999.

[Wm2] S. Wolfram, A New Kind of Science, Wolfram Media, 2002.

---

[5] 

Tom Willmore (1919–2005).
At Oberwolfach, 1998.

[Wood] R. Woodhouse, Mathematica, A History of the Calculus of Variations in the Eighteenth Century, Chelsea Publishing Co, New York, 1964. Reprint of A Treatise on Isoperimetrical Problems and the Calculus of Variations, originally published in 1810.

[Yates] R. C. Yates, Curves and Their Properties, The National Council of Teachers of Mathematics, 1974. Reprint by University Microfilms International, Ann Arbor, MI, 1992.

[Zwi] C. Zwikker, The Advanced Geometry of Plane Curves and their Applications, Dover Publications, New York, 1963.


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Name Index

Agnesi, Maria Gaetana (1718–1799), 58, 931
Ahlfors, Lars Valerian (1907–1996), 931
Airy, Sir George Biddel (1801–1892), 145
Aleksandrov, Aleksandr Danilovich (1912–1999), 931
Aoust, L'abbé Louis (1811–1885), 931
Apollonius of Perga (262–180 BC), 14, 21
Archibald, Raymond Claire (1875–1955), 931
Archimedes of Syracuse (287–212 BC), 83

Bäcklund, Albert Victor (1845–1922), 683, 706
Ball, Walter William Rouse (1850–1925), 932
Barbier, Joseph Émile (1839–1889), 170, 932
Bartels, Karl Eduard (1769–1836), 199
Beltrami, Eugenio (1835–1900), 561, 885
Berkeley, Bishop George (1685–1753), 86
Bernoulli, Daniel (1700–1782), 558
Bernoulli, Jakob (1654–1705), 43, 53
Bernoulli, Johann (1667–1748), 558
Bernoulli, Johann (II) (1710–1790), 558
Bernoulli, Nicolaus (1695–1726), 558
Bertrand, Joseph Louis François (1822–1900), 933
Bessel, Friedrich Wilhelm (1784–1846), 141
Bianchi, Luigi (1856–1928), 386, 483, 683, 933
Bieberbach, Ludwig (1886–1982), 933
Björling, Emanuel Gabriel (1808–1872), 741, 933
Blaschke, Wilhelm Johann Eugen (1885–1962), 386, 933
Bolyai, Farkas Wolfgang (1775–1856), 884
Bolyai, János (1802–1860), 884
Bompiani, Enrico (1889–1975), 933
Bonnet, Pierre Ossian (1819–1892), 624, 901, 921, 934
Bour, Edmond (1832–1866), 752
Bowditch, Nathaniel (1773–1838), 58
Brioschi, Francesco (1824–1897), 533, 934
Burali-Forti, Cesare (1861–1931), 386, 934

Cajori, Florian (1859–1930), 934
Cartan, Élie (1869–1951), 546, 809, 935
Cassini, Gian Domenico (1625–1712), 79
Catalan, Eugène Charles (1814–1894), 508, 510, 935
Cauchy, Augustin Louis (1789–1857), 720, 892
Cayley, Arthur (1821–1895), 114
Cesàro, Ernesto (1859–1906), 40, 139, 935
Chasles, Michel (1793–1880), 134, 935
Chern, Shiing-Shen (1911–2004), 546, 624, 935
Christoffel, Elwin Bruno (1829–1900), 538, 936
Clairaut, Alexis Claude (1713–1765), 199, 573
Codazzi, Delfino (1824–1873), 599, 936
Cohn-Vossen, Stephan (1902–1936), 922, 936, 941

Coolidge, Julian Lowell (1873–1954), 936
Cornu, Marie Alfred (1841–1902), 53
Coxeter, Donald (1907–2003), 936

Darboux, Jean Gaston (1842–1917), 215, 386, 656, 683, 699, 707, 937
Descartes, René du Perron (1596–1650), 1, 24, 76, 932
Dini, Ulisse (1845–1918), 482
Diocles (180 BC), 39
Dupin, Baron Pierre Charles François (1784–1873), 609, 655, 937

Eisenhart, Luther Pfahler (1876–1965), 401, 938
Enneper, Alfred (1830–1885), 378, 509, 938
Euclid of Alexandria (365–300 BC), 2, 884
Euler, Leonhard (1707–1783), 53, 139, 169, 397, 501, 535, 915, 938

Fermat, Pierre de (1601–1665), 84
Frenet, Jean Frédéric (1816–1900), 196, 939
Fresnel, Augustin Jean (1788–1827), 53
Frost, Percival (1817–1898), 939
Fuss, Nicolaus (1755–1826), 939

Gauss, Carl Friedrich (1777–1855), 141, 296, 333, 362, 388, 394, 397, 398, 516, 535, 599, 731, 871, 874, 884, 888, 905, 921, 939, 940
Germain, Sophie (1776–1831), 398
Gibbs, Josiah Willard (1839–1903), 191
Gomes Teixeira, Francisco (1851–1933), 40, 940
Goursat, Édouard Jean Babtiste (1858–1936), 318, 940
Grassmann, Hermann Günther (1809–1877), 191
Graustein, William Caspar (1888–1941), 164
Green, George, (1793–1841), 907

Halley, Edmond (1656–1742), 199
Hamilton, Sir William Rowan (1805–1865), 191, 402
Harriot, Thomas (1560–1621), 901
Henneberg, Ernst Lebrecht (1850–1922), 511
Hesse, Ludwig Otto (1811–1874), 282
Hilbert, David (1862–1943), 602, 941
Hill, Thomas (1818–1891), 139
Hopf, Heinz (1894–1971), 162, 889, 893, 941
Huygens, Christiaan (1629–1695), 76, 116

Jacobi, Carl Gustav Jacob (1804–1851), 271, 941
Joachimsthal, Ferdinand (1818–1861), 466, 941

Kästner, Abraham Gotthelf (1719–1800), 20
Kepler, Johannes (1571–1630), 85
Klein, Christian Felix (1849–1925), 338, 885, 942
Kowalewski, Gehrhard (1876–1950), 935
Kronecker, Leopold (1823–1891), 824
Kummer, Ernst Eduard (1810–1893), 318

L'Hôpital, Guillaume Francois Antoine Marquis de (1661–1704), 816
Lacroix, Sylvestre Francois (1765–1843), 139
Lagrange, Joseph Louis (1736–1813), 193, 502
Lamé, Gabriel (1795–1870), 518, 594, 943
Laplace, Pierre Simon, Marquis de (1749–1827), 58
Legendre, Adrien Marie (1752–1833), 462, 749
Leibniz, Baron Gottfried Wilhelm (1646–1716), 45, 819
Leonardo da Vinci (1452–1519), 54
Levi-Civita, Tullio (1873–1941), 848, 943
Lie, Marius Sophus (1842–1899), 683, 728

Liebmann, Heinrich (1874–1939), 473, 603
Liouville, Joseph (1809–1882), 558, 655, 893, 905, 943
Lissajous, Jules Antoine (1822–1880), 58
Lobachevsky, Nikolai Ivanovich (1792–1856), 884
Loria, Gino (1862–1954), 40, 943

Maclaurin, Colin (1698–1746), 86
Mainardi, Gaspare (1800–1879), 599, 944
Maupertuis, Pierre Louis Moreau de (1698–1759), 199
Maxwell, James Clerk (1831–1879), 655, 656, 944
Mellish, Arthur Preston (1905–1930), 171, 944
Mercator, Gerardus (1512–1594), 317
Meusnier, Jean Babtiste (1754–1793), 501, 944
Minding, Ernst Ferdinand Adolf (1806–1885), 470, 944
Möbius, August Ferdinand (1790–1868), 337
Monge, Gaspard (1746–1818), 302, 535, 944

Newton, Sir Isaac (1642–1727), 14, 39, 48, 945
Nielsen, Niels (1865–1931), 142
Nuñez Salaciense, Pedro (1502–1578), 250

Pascal, Blaise (1623–1662), 58
Pascal, Étienne (1588–1651), 58, 932
Peterson, Karl Mikhailovich (1828–1881), 599
Plateau, Joseph Antoine Ferdinand (1801–1883), 501
Playfair, John (1748–1819), 884
Plücker, Julius (1801–1868), 435
Poincaré, Jules Henri (1854–1912), 876, 885

Rademacher, Hans (1892–1969), 946
Reidemeister, Kurt Werner Friedrick (1893–1971), 946
Reuleaux, Franz (1829–1905), 172, 946
Ricci-Curbastro, Gregorio (1853–1925), 836, 943, 946
Richmond, Herbert William (1863–1948), 740
Riemann, Georg Friedrich Bernhard (1826–1866), 847, 947
Rinow, Willi (1907–1979), 889, 893, 941
Rolle, Michel (1652–1719), 109

Saccheri, Giovanni Girolamo (1667–1733), 561
Salmon, George (1819–1904), 947
Scheffers, Georg (1866–1945), 948
Scherk, Heinrich Ferdinand (1798–1885), 513, 947
Schwarz, Herman Amandus (1843–1921), 741, 947
Schwarzschild, Karl (1873–1916), 879
Senff, Karl Eduard (1810–1849), 199
Serret, Joseph Alfred (1819–1885), 199, 947, 948
Serret, Paul Joseph (1827–1898), 199, 948
Snyder, Virgil Snyder (1869–1950), 948
Steiner, Jakob (1796–1863), 345
Stoker, James Johnston (1905–1988), 948
Strubecker, Karl (1904–1991), 949
Struik, Dirk Jan (1894–2000), 40, 949
Study, Eduard (1862–1930), 949
Syamadas Mukhopadhyaya (1867–1937), 168, 945

Töplitz, Hans (1881–1940), 946
Tchebyshef, Pafnuty Lvovich (1821–1894), 683, 949
Terguem, Olry (1782–1862), 466
Thompson, D'Arcy Wentworth (1860–1948), 949
Tschirnhausen, Ehrenfried Walter (1651–1708), 59

Veronese, Giuseppe (1854–1917), 895
Viviani, Vincenzo (1622–1703), 207

Wallis, John (1616–1703), 450
Wankel, Felix Heinrich (1902–1988), 173
Weatherburn, Charles Ernest (1884–1974), 950
Weber, Wilhelm Eduard (1804–1891), 296
Weierstrass, Karl Theodor Wilhelm (1815–1897), 719, 950
Weingarten, Julius (1836–1910), 394
Weyl, Hermann Klaus Hugo (1885–1955), 809, 950
Whitney, Hassler (1907–1989), 74, 164, 307
Wieleitner, Heinrich (1874–1933), 40, 950
Willmore, Thomas James (1919–2005), 950
Woodhouse, Robert (1773–1827), 951

Yates, Robert Carl (1904–1963), 951

Subject Index

Absolute value, 4
Abstract surface, 531, 858, 871, 902
Acceleration
  of a curve, 6, 204, 280, 559
    normal component of, 387
    tangential component of, 882
    with respect to a connection, 853
Addition
  of functions, 815
  of vectors, 2, 821
Agnesi, curve of, 58
Airy functions, 145
Algebra
  Lie, 832
  of differentiable
      functions, 815
Algebraic curve, 142
Analytic
  function, complex, 720
  function, real, 810
Angle
  between
    tangent vectors to $\mathbb{R}^n$, 265
    tangent vectors to a surface, 371
    vectors in $\mathbb{R}^n$, 3
  Euler, 791
  exterior, 907
  function
    between curves in $\mathbb{R}^2$, 18
    of a principal patch, 690
    of an asymptotic patch, 685
  interior, 907
  of a rotation
    in $\mathbb{R}^2$, 135
    in $\mathbb{R}^3$, 773, 782
  oriented, 5, 373, 397, 547
  tangent-radius, 24, 28
  turning, 19, 902
Anticomplex map, 516
Antipodal map, 311, 343
Antisymmetric bilinear form, 854
Arc length, 12, 202
  of a polar parametrization, 82
  of an ellipse, 477, 486
Archimedes, spiral of, 83
Area
  element of, 372, 373
  of a sphere, 901
  of a subset of a patch, 374, 907
  of a surface, 372
Associated family
  of minimal patches, 721, 728
Associative law, 131, 769
Astroid, 27, 252, 485
  3-dimensional, 214
Astroidal
  ellipsoid, 407, 415
  sphere, 407
Asymptotic
  curve, 390, 432, 559
    differential equation for an, 559
  direction, 390, 558
  patch, 561
  vector, 390
Atlas, 811
  complete, 812
Axis
  of a generalized helicoid, 470
  of revolution, 462

Bäcklund transform, 683, 706
Ball in $\mathbb{R}^3$, 783

Barrel type, 475
Base curve, 432, 571
Basis
  orthonormal, 132
  Theorem, 823
Beltrami–Enneper Theorem, 561
Bernoulli, lemniscate of, 43, 44, 79
Bessel function, 141, 215
Bianchi identity, 864
Bianchi transform, 698
Bicylinder, 214, 215
  tangent developable to, 451
Binormal vector field to a curve in $\mathbb{R}^3$, 196
Björling
  curve, 741
  formula of, 741
Bohemian dome, 417
Bonnet
  theorem on parallel surfaces, 624
  theorem with Gauss, 901
Boundary of a set, 907
Bounded set, 373
Bour's minimal surface, 752
Bowditch curve, 58, 349
Bracket, 831, 832
Brioschi, formula of, 533, 874

C^∞ functions, 810
Canal surface, 646
Capture point, 55
Cardioid, 45, 81, 213, 252, 485, 754
  evolute of a, 115
  parallel curve to a, 115
  pedal curve to a, 114
Cassini, oval of, 79
Catalan
  constant of, 510
  minimal surface of, 510, 744
  theorem of, 508
Catenary, 27, 45, 102, 103, 116, 464
  involute of a, 116
  natural equation of a, 138, 139
  pedal curve to a, 116
Catenoid, 464, 505–507, 547, 737, 753
Cauchy–Riemann equations, 701, 720
Cauchy–Schwarz inequality, 3, 243, 264, 684, 741
Cayley, sextic of, 114
Cayley–Hamilton Theorem, 402, 772
Center
  curve of a canal surface, 646
  of curvature, 99
Chain rule, 7, 267, 280, 825
  for curves on a manifold, 831
Change of coordinates, 300, 811
Christoffel symbols, 538–540, 544, 547, 566, 861, 873, 882
Circle, 7, 40, 45, 48, 80, 105, 396
  central, 339, 341
  geodesic, 571
  involute of a, 105, 106, 115
  natural equation of a, 138
  osculating, 100, 107, 108, 239
    to a cubic parabola, 108
    to a logarithmic spiral, 108
    to a parabola, 110
  pedal curve to a, 114
Cissoid, 47, 485
  evolute of a, 103
  pedal curve to a, 114
C^k functions, 810
Clairaut
  patch, 573
  relation of, 574
Closed curve, 155
  rotation index of a, 160
  simple, 161
Clothoid, 53, 133, 140
  natural equation of a, 138
Coefficients of the
  first fundamental form, 362, 394, 402
  second fundamental form, 394, 402, 724, 731
Coherently oriented, 335
Compact
  subset, 373
  surface, 341
Complementary surface, 698
Complete

atlas, 812
elliptic integral of the second kind, 477
geodesically, 889
metric space, 892
Complex
analytic function, 720
conjugate, 4
derivative of a patch, 726, 728
Euclidean n-space, 193
manifold, 812
map, 516
structure, 3, 133, 265, 277, 331, 332, 880
Composition
of maps, 131
of rotations, 782
Cone
circular, 450, 485
generalized, 438
vertex of a, 438
Confocal surfaces, 609, 616
Conformal
diffeomorphism, 370
map, 370, 518, 663
Conic type, 478
Conical edge of Wallis, 450
Conjugate
harmonic, 721
in a group, 771
minimal patches, 728
of a quaternion, 777
Weierstrass patch, 736
Conjugation
complex, 194, 725
quaternion, 776
Connected subset, 334, 598
Connection
Levi-Civita, 858
natural, on $\mathbb{R}^n$, 850
on a differentiable manifold, 849
Riemannian, 857
torsion of a, 867
torsion-free, 858
Contact of order k, 236
Contrapedal, 117
Convex plane curve, 164
Coordinate
function, 811
neighborhood of a patch, 811
Coordinates
change of, 300, 811
rectangular to polar, 303
elliptic, 611
natural, of $\mathbb{R}^n$, 301
parabolic, 616
polar, 303, 364, 577
Coradial tangent vector, 888
Corkscrew
rule, 780
surface, 472
Cornu, spiral of, 53
Cosine integral, 143
Costa
minimal
curve, 749
surface, 746
Covariant
derivative, 849
tensor, 837
tensor field, 836
Critical point, 282
Cross
cap, 346
product, 192, 332, 773
Cubic
curve
nonsingular, 77
equation, 47
twisted, 202, 451, 794
Curtate cycloid, 41, 57
Curvature
center of, 99
Gaussian, 397, 398, 532, 540, 874
of a minimal curve, 731
geodesic, 541, 881
mean, 397, 398, 417, 520
vector field, 753
normal
of a curve, 541

of a surface, 385, 389, 397
of a cardioid, 45
of a circular helix, 201
of a curve in $\mathbb{R}^2$, 14, 136
of a curve in $\mathbb{R}^3$, 203
of a curve in $\mathbb{R}^n$, 195, 196
of a cycloid, 40
of a lemniscate, 43
of a parallel curve, 112
of a polar parametrization, 82
of Viviani's curve, 244
principal, 390, 394, 397
radius of, 99, 139
scalar, 865
signed, of a curve in $\mathbb{R}^2$, 14, 196
tensor
Ricci, 865
Riemann, 864
total, geodesic, 905
total, signed, 154
unsigned geodesic, 541
Curve
acceleration of a, 6, 204, 280, 387, 559
algebraic, 142
asymptotic, 390, 432, 559, 566
base, of a ruled surface, 432
bicylinder, 214, 451
Bowditch, 58, 349
center, of a canal surface, 646
closed, 155
length of a, 155
period of a, 155
rotation index of a, 160
simple, 161
convex, 164
cubic, 794
devil's, 85
differentiable, 5
director, of a ruled surface, 432
geodesic, 565, 566, 575, 577, 881, 883
curvature of a, 541
maximal, 568, 889
torsion of a, 545
helicoidial over
a cardioid, 213
a cycloid, 213
a figure of eight, 213
a logarithmic spiral, 206
a plane curve, 206
implicitly defined in $\mathbb{R}^2$, 74
in $\mathbb{R}^n$, 5
integral, of a vector field, 851
isotropic, 728
length of a, 9
Lissajous, 58, 349
minimal, 728, 741
Bour's, 752
regular, 728
negative reparametrization of a, 7
nonparametric form of a, 7, 74
normal
curvature of a, 541
line to a, 101
of constant slope, 246
of pursuit, 54
on a differentiable manifold, 830
on a surface, 302
on one side of a line, 164
parallel to a, 111
parametric form of a, 5, 7, 74
pedal to a, 113
piecewise-differentiable, 5
plane
arc length of a, 12
positive reparametrization of a, 7
pregeodesic, 511, 568
principal, 390, 465, 545, 595, 609
radius of curvature of a, 14
regular, 6
at a point, 26
rotation, 787
signed curvature of a, 14
simple closed, 161
sine oval, 166
space
arc length of, 202
speed of a, 6
striction, 445, 446

strophoid, 85
tangent line to a, 101
total
geodesic curvature of a, 905
signed curvature of a, 154
trace of a, 6
tube around a, 209
u-parameter, 293
unit-speed, 6
unsigned geodesic
curvature of a, 541
v-parameter, 293
vector field along a, 8
velocity of a, 6, 204
Viviani's, 207, 208, 215, 244, 440, 451, 795
curvature of, 244
torsion of, 244
Curvilinear patch, 606, 620
Cusp, 48, 102, 116
Cyclide
horn, 659
ring, 659
spindle, 659
Cyclide of Dupin, 655, 659
Cycloid, 40, 57, 115, 213, 252, 485, 744
curtate, 41, 42, 57
curvature of a, 40
evolute of a, 115
prolate, 41, 42, 57
Cylinder, 337, 450
circular, 291, 485, 536, 572
elliptical, 572
generalized, 438
over a figure of eight, 439

Darboux
frame field, 546
vector field, 215
Deformation
from a helicoid to a catenoid, 505
isometric, 723, 728
Degree
of a continuous function of a circle onto itself, 160
of a tensor field, 836
Deltoid, 57, 485
involute of a, 173
parallel curve to a, 115
Derivation, 831
Derivative
complex, of a patch, 726, 728
covariant, 849
directional, 266, 267
of a vector field, 275, 277
Descartes
folium of, 75, 76
Determinant
of a linear transformation, 397
of an orthogonal matrix, 770
Developable surface, 437
Devil's curve, 85
Diagonalizable
matrix, 772
Diffeomorphism
between
manifolds, 818
open sets of $\mathbb{R}^n$, 268
regular surfaces, 309
conformal, 370
local, 818
Differentiable
curve
in $\mathbb{R}^n$, 5
on a regular surface, 302
on manifold, 830
function, 301
between $\mathbb{R}^n$ and $\mathbb{R}^m$, 268
between regular surfaces in $\mathbb{R}^n$, 309
manifold, 809, 811
map between manifolds, 818, 829
real-valued function, 809
vector field on $\mathbb{R}^n$, 272
Differential
1-form, 838
form, 373
of a function, 363, 364, 839
of a map, 268
Dimension, 811

Dini, surface of, 482
Diocles, cissoid of, 47
Direction
  asymptotic, 390, 558
  on a surface, 389
  principal, 390
Directional derivative, 266
Directrix, 432
  of a ruled surface, 432
Distance
  between points in $\mathbb{R}^n$, 3
  function, 369, 891
  infinitesimal, 363
  intrinsic, 369
Distribution parameter, 446
Divergence, 279, 282
Division ring, 777
Domain
  fundamental, 747
Dot product, 2, 194
Doubly periodic function, 747
Doubly ruled surface, 432
Dual 1-form of a vector field, 856
Dupin's Theorem, 609

Edge, 911
  of regression, 440
Eigenvalues and eigenvectors, 396, 770
Eight
  figure of, 44, 105, 213, 252, 349, 486
    cylinder over, 439
    involute of a, 105, 106, 116
Eight surface, 378
Element of area, 372, 373
Ellipse, 21, 110, 111
  arc length of an, 486
  curve parallel to an, 112
  evolute of an, 100
  focus of an, 22
  pedal curve to an, 116
Ellipsoid, 415, 581, 593, 612, 615
  astroidal, 407, 415
  Gaussian curvature of an, 414
  Mercator, 317
  nonparametric form of an, 312
  of revolution, 312, 627
  parametric form of an, 313
  stereographic, 313
  support function of an, 414
Elliptic
  coordinate patch, 611
  integral of the second kind, 477
  point, 398, 558
Elliptical
  helicoid, 562
  torus, 210, 305
Enneper
  minimal surface of, 378, 416, 418, 509, 547, 734, 752
Envelope, 659
  of curves, 175
  of surfaces, 646
Epicycloid, 144, 158
  natural equation of an, 144
Epitrochoid, 157
Equation
  differential, of a geodesic, 566
  natural, 138
    of a catenary, 138, 139
    of a circle, 138
    of a clothoid, 138
    of a hypocycloid, 144
    of a logarithmic spiral, 138
    of an epicycloid, 144
    of an involute of a circle, 138, 142
Equations
  Bianchi–Darboux, 699, 707
  Cauchy–Riemann, 701, 720
  Gauss, 538, 599, 711
    for abstract surfaces, 895
  Peterson–Mainardi–Codazzi, 593, 600, 601, 625, 848
  Weingarten, 394, 467, 522, 600, 711
Euclidean
  motion of $\mathbb{R}^n$, 130, 230, 388
  n-space, 2, 264, 809, 811
    complex, 193
Euler
  characteristic, 915
  constant of, 143

Theorem of, 397
Euler angles, 791
Evolute
of a cissoid, 103
of a cycloid, 115
of a logarithmic spiral, 115
of a plane curve, 99, 100, 108
of a space curve, 241
of a tractrix, 103
of an ellipse, 100
Exponential map, 773, 885
Exponentiation, 771
Exterior angle, 907

Face, 911
Factorization
of quaternions, 792
Figure eight, 342
First fundamental form, 362, 401, 402
coefficients of the, 362, 394, 402
Flat, 398
generalized helicoid, 481
ruled surface, 437
torus, 879
Focal
interval, 639
patch, 641
point, 639
set, 640
Focus
of a lemniscate, 43
of an ellipse, 22
Folium
of Descartes, 75, 76
perturbed, 76
of Kepler, 85
Form
bilinear
antisymmetric, 854
signature of a, 855
symmetric, 854
symmetric nondegenerate, 855
symmetric positive definite, 855
differential, 838
first fundamental, 401, 402
second fundamental, 401, 402
third fundamental, 402
Formula
Björling's, 741
Brioschi's, 533, 874
Frame field
Darboux, 546
Frenet, 196, 560, 794
natural, of $\mathbb{R}^n$, 273
Freeway exit, 53
Frenet
formulas, 197–201, 203, 204, 213, 545, 560, 561
frame field, 196, 560, 794
on Viviani's curve, 208, 795
Fresnel integral, 53
Function
Airy, 145
arc length, 12, 202, 363
Bessel, 141, 215
complex analytic, 720
coordinate, of a patch, 811
differentiable, 301, 809, 814
differential of a, 363, 364, 839
distance, 369, 891
square of, 301
doubly periodic, 747
Gamma, 143
harmonic, 520, 721
conjugate, 721
height, 301
holomorphic, 626, 720, 812
Kronecker delta, 824
meromorphic, 737
natural coordinate, of $\mathbb{R}^n$, 273, 301, 810
real analytic, 810
scaling, 518
support
of a hyperboloid, 414
of a plane curve, 176
of a surface in $\mathbb{R}^3$, 413
of an ellipsoid, 414
Fundamental
domain, 747

form
first, 401, 402
second, 401, 402
third, 402
theorem of plane curves, 136, 137
theorem of space curves, 231

Gamma
constant γ, 143
function Γ, 143
Gauss
equations, 538, 599, 711
for abstract surfaces, 895
Lemma, 888
map, 296, 316, 333, 388, 464, 753
of a minimal curve, 729
of a minimal surface, 516
notation, 362, 394
Theorema Egregium, 536, 537, 875
Gauss map
higher order, 794
Gaussian curvature, 397, 398, 406, 532, 540, 874
of a minimal curve, 731
General relativity, 865, 879
Generalized
cone, 438
cylinder, 438
parabola, 237
Genus, 917
Geodesic, 511, 565, 853, 881
circle, 571
curvature, 541, 881
differential equation of a, 566
maximal, 568, 889
on a surface of revolution, 577
on the plane, 577
on the Poincaré half-plane, 883
polar metric, 887
radial, 888
slant on a Clairaut patch, 575
torsion, 545
Geodesically complete surface, 889
Gibbs–Heaviside notation, 386
Gimbal lock, 793
Globalization, 817
Gradient, 274, 282, 311, 857
Great circle, 250, 463
Group, 131
of rotations, 779
of unit quaternions, 779

Harmonic
function, 721
patch, 721
Height function, 301
Helicoid, 418, 451, 505, 537, 595, 737, 753
circular, 448
Gaussian curvature of a, 579
mean curvature of a, 579
elliptical, 562, 563
generalized, 470, 471, 483, 485
flat, 481
geometric definition of a, 376
principal patch on a, 596
Helix, 200, 794
circular, 200, 201, 206, 451
curvature of a, 201
torsion of a, 201
elliptical, 215
incline of a, 200
over a curve, 206
radius of a, 200
spherical, 249
Henneberg
minimal surface of, 511, 738
Hessian, 282
Holomorphic
function, 720, 812
map, 516
Homothety, 142
Homotopy, 161, 789
Horn
cyclide, 659
torus, 305
Hundekurve, 51
Hyperbola, 754
Hyperbolic
paraboloid, 296, 449

point, 398, 558
Hyperboloid
elliptic, 464
of one sheet, 314, 415, 433, 612, 615
Gaussian curvature of a, 414
support function of a, 414
of revolution, 464, 627
of two sheets, 415, 612, 615
Gaussian curvature of a, 414
support function of a, 414
type, 478
Hyperplane, 129, 236
Hypersphere, 236
radius of a, 236
Hypocycloid, 144, 158
natural equation of a, 144
Hypotrochoid, 157

Identity
Bianchi, 864
element, 131, 815
Jacobi, 831
Lagrange, 193, 194
map, 4, 129, 813
matrix, 128, 769
Imaginary quaternion, 779
Implicitly defined hypersurface, 236
Incline
of a helix, 200
Inequality
Cauchy–Schwarz, 3, 243, 264, 684, 741
triangle, 3, 11
Infinitesimal
arc length, 363
distance, 363
volume element, 372
Injective patch, 291, 814
Inner product, 856
Inside-out torus, 305
Integers, sum of four squares of, 778
Integral
curve of a vector field, 851
elliptic of the second kind, 477
complete, 477
Interior angle, 907
Intrinsic distance, 369
Inverse, 131
function theorem, 828
patch, 665
Inversion, 347, 662
center, 662
pole of an, 662
radius, 662
Involute, 104
of a catenary, 116
of a circle, 105, 106, 115
natural equation of an, 138, 142
of a cubic parabola, 116
of a deltoid, 173
of a figure of eight, 105, 106, 116
Isolated singularity, 737
Isometric deformation, 723, 728
Isometry, 130, 366
local, 366
Isothermal patch, 518, 522, 736
Isotropic curve, 728

Jacobi identity, 831
Jacobian matrix, 271, 289, 334, 344, 374, 612, 617, 828
Joachimsthal–Terquem Theorem, 466, 467
Joachimsthal–Terquem, Theorem of, 647
converse to, 483, 485, 570

Kepler, folium of, 85
Klein bottle, 338, 341, 349, 351, 378, 419, 894
Knot
eight, 216
torus, 211, 215
trefoil, 211
Kronecker delta function, 824
Kuen's surface, 484

L'Hôpital's rule, 816
Lagrange identity, 193, 194
Laplacian, 279, 520
Legendre's relation, 749
Leibnizian property, 820, 832

Lemma, Gauss, 888
Lemniscate, 754
  curvature of a, 43
  of Bernoulli, 43, 44, 79, 115, 485
    parallel curve to a, 115
Length
  of a closed curve, 155
  of a curve, 9, 136
Level surface, 311
Lie
  algebra, 773, 832
  bracket, 831
    of vector fields, 861, 864
Limaçon, 58, 81, 114
Line
  in $\mathbb{R}^n$, 7
  normal to a curve in $\mathbb{R}^2$, 110
  tangent to a curve in $\mathbb{R}^2$, 110
Linear map, 3
Linearity property, 3, 128, 820, 832
Lissajous curve, 58, 349
Local
  diffeomorphism, 309, 818
  homeomorphism of a patch, 811
  isometry, 366
  surface in $\mathbb{R}^n$, 288
  system of coordinates
    of a surface, 297
Logarithmic spiral, 23, 206, 485
Lower g ij's, 860
Loxodrome, 250

Maclaurin, trisectrix of, 86
Manifold, 809
  complex, 812
  differentiable, 809, 811
  product, 813
  pseudo-Riemannian, 856
  Riemannian, 856
  topological, 811
Map
  affine, 130
  anticomplex, 516
  antipodal, 311, 343
  complex, 516
  conformal, 370, 663
  differentiable, 268
    between manifolds, 818, 829
  exponential, 885
  Gauss, 296, 316, 333, 388, 464, 753
    of a minimal curve, 729
    of a minimal surface, 516
  holomorphic, 516
  identity, 4, 129, 813
  linear, 3, 128
  of surfaces, 309
  orientation-preserving, 128, 130
  orientation-reversing, 128, 130
  orthogonal, 133
  projection, of the tangent bundle, 833
  secant, 163
  stereographic, 810
  tangent, 268, 289, 310, 388, 826
Matrix
  associated with a linear transformation, 271
  diagonalizable, 772
  identity, 128, 769
  Jacobian, 271, 289, 334, 344, 374, 612, 617, 828
  of maximum rank, 272
  orthogonal, 769, 782
    determinant of a, 770
  rank of a, 272, 290
  rotation, 773
  skew-symmetric, 772, 798
  trace of a, 397, 771
Maximal geodesic, 568, 889
Mean curvature, 397, 398, 406, 417, 520
  vector field, 753
Mercator ellipsoid, 317
Meridian, 250, 374
  of a generalized helicoid, 471, 483
  of a sphere, 250
  of a surface of revolution, 463, 570
Meromorphic function, 737
Metric, 855
  components of a, 859
  of $\mathbb{R}^2$

in Cartesian coordinates, 364
in polar coordinates, 364
Poincaré
on the disk, 876
on the half-plane, 877, 883, 894
pseudo-Riemannian, 855
Riemannian, 362, 366, 855
Schwarzschild, 879
space, 891
complete, 892
Minimal
curve, 728
Costa, 749
Weierstrass, 736
isothermal patch, 721
surface, definition of, 398
Möbius strip, 337, 339, 340, 349, 378, 419, 432
as a ruled surface, 432
Gaussian curvature of a, 437
mean curvature of a, 437
Monge patch, 302, 418, 513, 579
Monkey saddle, 303, 392, 403
area of a, 377
asymptotic curves of a, 564
Gauss map of a, 319
Gaussian curvature of a, 404
generalized, 305
mean curvature of a, 404
metric of a, 377
normal sections of a, 393
Monotone, 164
Moving frame of a principal patch, 702
Multiplication
of functions, 815
of quaternions, 775

Natural
connection on $\mathbb{R}^n$, 850
coordinate functions of $\mathbb{R}^n$, 273, 810
equation, 138
of a hypocycloid, 144
of an epicycloid, 144
of a catenary, 138, 139
of a circle, 138
of a clothoid, 138
of a logarithmic spiral, 138
of an involute of a circle, 138, 142
frame field of $\mathbb{R}^n$, 273
Nephroid, 485
Nielsen, spiral of, 142
Noncylindrical ruled surface, 444
Nondegenerate symmetric bilinear form, 855
Nonparametric
form of a curve, 7, 74
representation of a surface, 287
Nonregular points, 301
Norm, 194
in $\mathbb{R}^n$, 3
of a quaternion, 776
Normal
curvature
of a curve, 541
of a surface, 385, 389, 397
neighborhood, 886
patch, 887
section, 391
to a patch, 294
to a regular surface in $\mathbb{R}^n$, 309
variation of a surface, 502, 503
North pole, 250, 729, 813
Notation
Gauss, 362, 394
Gibbs–Heaviside, 386

Open
set of $\mathbb{R}^n$, 268
submanifold, 812
subset of a regular surface in $\mathbb{R}^n$, 298
Opposite point, 169, 172
Order
of an algebraic curve, 142
of contact, 236
Orientable, 332
abstract surface, 880
Orientation, 334
preserving
curvilinear patch, 606

transformation, 128, 770
Oriented, 332, 880
angle, 5, 373, 397, 547
coherently, 335
polygonal decomposition, 912
Orthogonal
matrix, 769, 782
Orthogonal system of curves, 626
Orthonormal basis, 132
Osculating
circle, 100, 107, 108, 239
sphere, 240
Oval, 169
of Cassini, 79
of constant width, 169
spread of an, 169
width of an, 169

Parabola, 109
cubic, 108
involute of a, 116
generalized, 237
osculating circle to a, 110
pedal curve to a, 114
semicubical, 25, 86
Parabolic
coordinate patch, 616
point, 398, 558
Paraboloid, 303, 415, 597
circular, 378
elliptic, 615, 617
hyperbolic, 296, 392, 415, 449, 562, 615, 617
asymptotic lines of a, 391
general, 434
generalized, 579
principal curves of a, 392
nonparametric form of a, 303
Parallel
curve
curvature of a, 112
to a cardioid, 115
to a curve in $\mathbb{R}^2$, 111
to a deltoid, 115
to a lemniscate, 115
to an ellipse, 112
helical curve of a generalized helicoid, 471
of a sphere, 250
of a surface of revolution, 463, 577
patch, 621
surface, 620
shape operator of a, 622
Parametric form of a curve, 7
Parametrization
by arc length, 13
pedal, 176
rational, 77
Pascal's snail, 58
Patch, 288, 295, 811
Clairaut, 573
coordinate
function of a, 811
neighborhood of a, 811
curvilinear, 606, 620
orientation preserving, 606
regular, 606
elliptic coordinate, 611
geodesic polar, 887
harmonic, 721
conjugate, 721
injective, 291, 814
inverse, 665
isothermal, 518, 522, 736
minimal isothermal, 721
Monge, 302, 418, 513, 579
normal, 887
on a topological space, 810
parabolic coordinate, 616
parallel to a patch, 621
principal, 467, 596
regular, 291
at a point, 292
ruled, 432
trace of a, 288
triply orthogonal, 607
unit normal of a, 607
vector perpendicular to, 294
Weierstrass, 736
conjugate, 736

Pedal
curve, 113
to a catenary, 116
to a circle, 114
to a cissoid, 114
to a parabola, 114
to an ellipse, 116
parametrization, 176
Pendulum, 116
Period
of a closed curve, 155
Peterson–Mainardi–Codazzi equations, 593, 600, 601, 625, 687, 848
Piecewise-differentiable curve in $\mathbb{R}^n$, 5
Pinch point, 307
Piriform, 178
Pitch of a spherical loxodrome, 250
Planar point, 398, 558
Plane, 2, 485, 536
geodesic on the plane, 577
Plücker, conoid of, 435
generalization of, 436
Poincaré metric
on the disk, 876
on the upper half plane, 877, 894
Point
elliptic, 398, 558
focal, 639
hyperbolic, 398, 558, 559
nonregular, 301
of application of a tangent vector to $\mathbb{R}^n$, 264
of inflection, 116
opposite, 169
parabolic, 398, 558
pinch, 307
planar, 398, 558
ridge, 642
singular, 301
umbilic, 597, 639
Polar
coordinates, 364, 577
Polar parametrization, 81
arc length of a, 82
curvature of a, 82
radius function of a, 81
Pole
north, 813
of an analytic function, 737, 738, 747
of an inversion, 662
south, 813
Polygonal region, 911
Positive definite symmetric bilinear form, 855
Pregeodesic, 511, 568, 743
Principal
curvature, 390, 394, 397
curve, 390, 465, 545, 609
differential equation for a, 595
direction, 390
normal vector field to a curve in $\mathbb{R}^3$, 196
patch, 467, 596
radius of curvature, 639
vector, 390
Product
by scalars, 2, 194
dot, 2
interior, 856
manifold, 813
of functions, 815
scalar, 2, 264, 362
symmetric, 859
vector cross, 192, 194, 265, 277, 332, 773
vector triple, 193, 405, 532, 606, 921
Profile curve
of a generalized helicoid, 470
of a surface of revolution, 462, 697
Projection
map of the tangent bundle, 833
stereographic, 313, 730
Projective space, real, 784, 799, 840
Prolate cycloid, 41, 57
Pseudo cross cap, 351, 378
Pseudo-Riemannian
manifold, 856
metric, 855
Pseudosphere, 477, 480, 486

metric of a, 710, 894
Pullback, 837, 872
Pursuit curve, 54
Pythagorean Theorem, 48, 361, 362

Quadrilateral, 908
Quaternion
factorization, 792
group, 798
imaginary, 775, 779
notation, 191
unit, 776, 779, 782

Radial
geodesic, 888
tangent vector, 888
Radius
function of a polar parametrization, 81
inversion, 662
of a canal surface, 646
of a helix, 200
of a hypersphere, 236
of a Tchebyshef patch, 684
of a Tchebyshef principal patch, 690
of curvature of a curve in $\mathbb{R}^2$, 14, 99, 139
tube, of a torus, 305
wheel, of a torus, 305
Railroad construction, 53
Rank of a matrix, 272, 290
Rational parametrization, 77
Real projective
plane, 338, 343
space, 784, 799, 840
Reflection
of $\mathbb{R}^n$, 128
Region, 907
polygonal, 911
regular, 911
triangular, 911
Regression, edge of, 440
Regular
curve, 6
at a point, 26
curvilinear patch, 606, 793
isothermal patch, 727
map between manifolds, 793, 829
minimal curve, 728
patch in $\mathbb{R}^n$, 291
region, 911
surface in $\mathbb{R}^n$, 297
Regular points, 301
Reparametrization, 7
of a geodesic, 568
positive, 7
unit-speed, 203
Reuleaux polygon, 172
Rhumb line, 250
Ricci curvature, 865
Richmond
minimal surface of, 740
Ridge point, 642
Riemann
curvature tensor, 864
sum approximation, 12
surface, 768
Riemannian
connection, 857
manifold, 856
metric, 362, 855
Right conoid, 436, 449
Ring
cyclide, 659
Rotation, 773
angle, 135
curve, 787
group, 779
index, 160
of $\mathbb{R}^2$, 135
of $\mathbb{R}^4$, 797
of $\mathbb{R}^n$, 128
sequence, 791
Ruled
patch, 432
surface, 432
Rulings, 432, 449, 571

Saddle surface, 303
Scalar
curvature, 865

multiplication, 2
product, 2
Scalar multiplication, 821
Scale factor, 370
Scaling function, 518
Scherk
fifth minimal surface of, 521, 522
minimal surface of, 513, 514, 753
generalization of, 522
implicit plot of, 516
Schwarzschild
metric, 879
surface, 879
Scissors trick, 789
Screw motion, 470
Sea shell, 319
Secant map, 163
Second fundamental form, 401, 402
coefficients of the, 394, 402, 724, 731
Section
normal, 391
of the tangent bundle, 833
Semicubical parabola, 25, 86
Set
bounded, 373
compact, 373
of zeros, 74
Shape operator, 386, 388
definition of the, 386
eigenvalues and eigenvectors of the, 396
Sievert's surface, 486
Sign
of a function, 468
of a number, 15
Signature, 855
Signed curvature of a curve in $\mathbb{R}^2$, 14
Simple curve, 161
Sine
–Gordon equation, 691
integral, 143
oval curve, 166
surface, 319, 785
Sine surface, 781
Singular curve at a point, 26
Singular points, 301, 306
of the sine surface, 785
on Steiner's surface, 346
Skew-symmetric matrix, 772
Slant
of a generalized helicoid, 470
of a geodesic in a v-Clairaut patch, 575
Slope, 246
Soliton, 683
South pole, 813
Space
Euclidean, 2, 264, 809, 811
complex, 193
metric, 891
complete, 892
tangent, 880
to a differentiable manifold, 821, 855
Speed
of a curve in $\mathbb{R}^n$, 6
ratio, 55
Sphere, 288, 299, 315, 406, 415, 472, 475, 536, 547, 813
3-dimensional, 779
area of a, 374
astroidal, 407
Gaussian curvature of a, 406
metric of a, 374
osculating, 240
twisted, 472
Spherical
helix, 249
spiral, 216
triangle, 901
Spindle
cyclide, 659
torus, 305
Spiral
logarithmic, 23, 82, 108, 115, 485
evolute of a, 115
natural equation of a, 138
Nielsen's, 142
of Archimedes, 83
of Cornu, 53

spherical, 216
Spread of an oval, 169
Square
of the distance function, 301
Steiner, Roman surface of, 345, 346, 378, 419
Stereographic
ellipsoid, 313
injection, 813
map, 250, 371, 810
projection, 250, 313, 730, 813, 841
Striction curve, 445, 446
Strophoid, 85, 116
Subgroup, 131
Sum of vectors, 264
Superquadric, 415
Support function
of a plane curve, 176
of a surface in $\mathbb{R}^3$, 413
Surface
abstract, 531, 858, 871, 902
orientable, 880
canal, 646
compact, 341
complementary, 698
corkscrew, 472
cross cap, 378, 419
developable, 437
Dini's, 482
eight, 306, 378
funnel, 537
generalized helicoid
axis of a, 470
meridian of a, 471
profile curve of a, 470
geodesically complete, 889
Kuen's, 484
level, 311, 334
local, 288
mapping, 309
minimal
Bour's, 752
Catalan's, 510, 744
catenoid, 464, 505–507, 547
Costa's, 746
definition of, 398
Enneper's, 378, 416, 418, 509, 547, 734
helicoid, 376, 505
Henneberg's, 511, 738
Richmond's, 740
Scherk's, 513, 514, 522, 753
Scherk's fifth, 521, 522
trinoid, 753
with one planar end, 740
monkey saddle, 303, 392, 403, 537
nonorientable, 336, 338, 340, 341
nonparametric form of a, 311, 411
normal
curvature of a, 385, 389, 397
to a patch in $\mathbb{R}^3$, 295
variation of a, 502, 503
of revolution, 462
axis of revolution of a, 462
barrel type, 475
conic type, 478
flat, 485
football type, 475
generalized, 486, 797
geodesic on a, 577
hyperboloid type, 478
meridian of a, 463, 570
parallel of a, 463, 577
principal curves of a, 466
profile curve of a, 462, 697
standard parametrization of a, 462
orientable, 332
oriented, 332, 880
parallel, 620
shape operator of a, 622
regular, 814
in $\mathbb{R}^n$, 297, 814, 858
parametric representation of a, 311
Riemann, 768
ruled, 432
base curve of a, 432
director curve of a, 432
directrix of a, 432
doubly, 432
flat, 437

Gaussian curvature of a, 437
noncylindrical, 444
rulings of a, 432, 571
Schwarzschild, 879
Sievert's, 486
sine, 319, 781, 785
sphere, 406
Steiner's Roman, 345, 346, 378, 419
tangent developable, 438
torus, 305, 317, 337, 377, 406, 416, 418, 547, 579
translation, 416
twisted, 348
vector perpendicular to, 309
Veronese, 895
Symmetric
bilinear form, 854
product, 859
System
of local coordinates, 297, 811
orthogonal, of curves, 626
triply orthogonal, of surfaces, 604

Tangent
bundle, 833, 841
projection of the, 833
developable, 438
map, 268, 310, 388, 826
space
of $\mathbb{R}^n$, 264
to a differentiable manifold, 821
to a surface, 307
vector
coradial, 888
radial, 888
to $\mathbb{R}^n$, 264
to a manifold, 820
to a patch, 293
Tangent bundle
of a sphere, 787
Tangent-radius angle, 24, 28
Tchebyshef
net, 683
parametrization of a pseudosphere, 695
patch, 684
principal patch, 690
Tensor
covariant, 837
field, 836
vectorvariant, 837
Theorem
basis, 823
Beltrami–Enneper, 561
Bonnet's, 624
Cayley–Hamilton, 402, 772
Dupin's, 609
Euler's, 397
fundamental
of plane curves, 136, 137
of space curves, 231
inverse function, 292, 310, 828
Joachimsthal–Terquem, 466, 467, 647
converse to, 483, 485, 570
mean value, 10
of Catalan, 508
Pythagorean, 48, 361, 362
Theorema Egregium, 536, 537, 875
Third fundamental form, 402
Topological manifold, 811
Torsion
geodesic, 545
of a circular helix, 201
of a connection, 867
of a curve in $\mathbb{R}^3$, 197, 203, 561
of Viviani's curve, 244
Torus, 317, 337, 377, 406, 416, 418, 547, 579
crossed, 351
elliptical, 210, 305
flat, 879, 893
Gaussian curvature of a, 406
horn, 305
knot, 211
tube around a, 213, 215
mean curvature of a, 406
ring, 305
spindle, 305
Total
geodesic curvature of a curve on a

surface, 905
signed curvature of a curve in $\mathbb{R}^2$, 154
Trace
of a curve in $\mathbb{R}^n$, 6
of a linear transformation, 397
of a matrix, 771
of a patch, 288
Tractoid, 480
Tractrix, 50, 51, 53, 116, 482, 889
differential equation of a, 51
evolute of a, 103
Transformation
affine, 130, 278
by reciprocal radii, 663
orthogonal, 128, 133
Translation
of $\mathbb{R}^n$, 130
surface, 416
Trefoil knot, 211
Triangle
asymptotic geodesic, 914
equilateral, 485
inequality, 3, 11
spherical, 901
Triangular region, 911
Triangulation, 912
Trinoid, 753
Triple product, 193, 405, 532, 606, 921
Triply orthogonal
patch, 607
system of surfaces, 604
Trisectrix of Maclaurin, 86
Tube
around
a curve, 209
a torus knot, 213
Viviani's curve, 215
as a canal surface, 649
radius of a torus, 305
Turning angle
of a curve in $\mathbb{R}^2$, 19
with respect to a vector field, 902
Twisted
cubic, 202, 451, 794
surface, 348

u-parameter curve, 293
Umbilic point, 597, 639
Unit
normal
to a patch in $\mathbb{R}^3$, 295
normals of a triply orthogonal patch, 607
quaternion, 782
speed curve in $\mathbb{R}^n$, 6
tangent vector field to a curve in $\mathbb{R}^n$, 196
Unsigned geodesic curvature, 541
Upper g ij's, 860

v-parameter curve, 293
Vector
cross product, 192, 194, 332, 773
field
along a curve, 8, 196, 280, 852
binormal, 196
differentiable, 272
dual to a 1-form, 856
image by a diffeomorphism, 278
mean curvature, 753
normal to a surface, 294
on $\mathbb{R}^n$, 272
on a differentiable manifold, 832
on a patch, 294
on a surface, 294, 309
principal normal, 196
tangent to a surface, 294
unit tangent to a curve, 196
velocity of a curve in $\mathbb{R}^n$, 280
in $\mathbb{C}^n$, 193
in $\mathbb{R}^n$, 263
normal
to a patch, 294
to a surface, 309
part of a tangent vector to $\mathbb{R}^n$, 264
perpendicular
to a patch, 294
position, 2
principal, 390

tangent, 820
point of application of a, 264
to a patch, 293
vector part of a, 264
triple product, 193, 405, 532, 606, 921
velocity, 830
Vectorvariant
tensor, 837
tensor field, 836
Velocity, 6, 204, 280
initial, 830
of a curve on a differentiable manifold, 830
Veronese surface, 895
Versiera, 58
Vertex
of a cone, 438
of a parabola, 109
of a piecewise-differentiable simple closed curve, 906
of a plane curve, 167
Viviani
curve of, 207, 208, 215, 244, 795
curvature of, 244
normal surface to, 440
tangent developable to, 451
torsion of, 244
tube around, 215

Wallis, conical edge of, 450
Weierstrass
$\wp$ function, 747
ζ function, 748
minimal curve, 736
patch, 736
Weingarten
equations, 394, 467, 522, 600, 711
surface, 394
Wheel radius of a torus, 305
Whitney umbrella, 307, 378, 895
Width of an oval, 169


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Notebook Index

This index lists the principal definitions visible in the printed part of the notebooks.

A[x_] 23.1
agnesi[a][t_] 2.9, 3.0
analytic2minimal[Ψ_][t_][u_,v_] 22.3
ar[v_,w_] 6.0
ar[{v_, w_}] 11.1
archimedesspiralpolar[n_,a_][θ_] 3.5
arclengthprime[α_][t_] 1.3
arclength[α_][t_] 1.3, 22.0
area[x_][u_,v_] 12.4
astell[a_,b_,c_][u_,v_] 13.5
astroid[a_][t_] 1.8, 22.0
astroidcat[u_,v_] 22.6
asymeq1[x_][u_,v_] 18.1

bianchi[x_,θ_,ϕ_][u_,v_] 21.5
bianchips 21.7
binormal[α_][t_] 7.2, 14.2
binormalsurf[α_][u_,v_] 14.2
bjmincurve[α_][z_] 22.6
bjoerling[α_,γ_][z_] 22.6
boat[c_,d_,e_,f_] 2.8
bohdom[a_,b_,c_,d_][u_,v_] 13.8
bourmincurve[m_][z_] 22.8
bourpolar[m_][t_][r_,θ_] 22.8
brioschi[M_][u_,v_] 17.1

cardioid[a_][t_] 2.3
cassiniimplicit[a_,b_][x_,y_] 3.3
catalan[a_][u_,v_] 16.4, 22.0
catalandef[a_][t_][u_,v_] 22.4
catenary[a_][t_] 1.8, 2.4, 15.1, 20.0
catenoid[c_][u_,v_] 15.1, 16.0, 17.0, 20.0, 22.0
christoff1[x_][u_,v_] 17.3
christoff2[M_][u_,v_] 17.3

christoff3[M_][u_,v_] 25.4
circ[a_][t_] 1.3, 15.0
circleinvolute[a_][t_] 1.8
circularcone[a_,b_][u_,v_] 18.0, 20.0
circularcylinder[a_][u_,v_] 20.0
cissoid[a_][t_] 2.5, 3.0
clairautgeodesic[x_,c_][t_] 18.5
clairautv[pm_,c_,u0_][x_][u_,t_] 18.5
clothoid[n_,a_][t_] 2.7
clothoidnds[min_,max_][n_,a_][s_] 2.7
clothoidprime[n_,a_][t_] 2.7
cnccoprofile[pm_,a_,b_][t_] 15.6
cnccotcheb[a_,b_][u_,v_] 21.4
cnchyprofile[pm_,a_,b_][t_] 15.6
cnchytcheb[a_,b_][u_,v_] 21.4
col[X_] 12.0
compare[f_,a_,b_] 5.4
complete[seq_, opt___] 18.4
cone[p_, γ_][u_,v_] 14.2, 20.3
coortang[x_, r_] 13.1
costa[u_,v_] 22.7
cpcprofile[pm_,a_,b_][t_] 15.5
cpcsurfrev[a_,b_][u_,v_] 15.5
cr 22.1
crosscap[u_,v_] 11.5, 27.0
crosscapmap[{x_, y_, z_}] 11.5
curveonsurf[surf_,α_][t_] 8.6
cycgen[a_,c_,k_,rau_,rav_, opts___] 20.3
cycloid[a_,b_][t_] 2.1, 22.0
cylinder[d_, γ_][u_,v_] 12.5, 14.2

deform[t_][x_,y_][u_,v_] 22.1
deltoid[a_][t_] 2.9
devilimplicit[a_,b_][x_,y_] 3.6
diff[f_][u_,v_] 22.2
diffc[f_][u_,v_] 22.2
dini[a_, b_][u_, v_] 15.7
dinitcheb[a_,c_] 21.8
dinitcheb[a_,b_,c_][u_,v_] 21.8
dip[ins_][g_] 22.7

divergence[vf_, coor_] 9.5
ds2[x_][u_,v_] 26.8

EA[x_] 23.1
EAif[x_] 27.5
edge[u_,v_,w_] 27.2
edges[u_,v_,w_,i_] 27.4
eight[t_] 4.3, 15.0
eightcylinder[u_,v_] 10.3
ellhypcyclide[a_,c_,k_][u_,v_] 20.3
ellipse[a_,b_] 5.0
ellipse[a_,b_][t_] 1.3, 2.4, 22.0
ellipsemin[a_,b_][t_][u_,v_] 22.6
ellipsemincurve[a_,b_][z_] 22.6
ellipsoid[a_,c_] 27.0
ellipsoid[a_,b_,c_][u_,v_] 10.6, 18.0
ellipticalcone[a_,b_,c_][u_, v_] 10.6
ellipticalhyperboloidminus[a_,b_,c_][u_,v_] 14.1
ellipticalhyperboloidplus[a_,b_,c_][u_,v_] 10.6, 14.1
ellipticcoor[d_,e_,f_][a_,b_,c_][w_][u_,v_] 19.6, 20.0
ellipticparaboloid[a_,b_][u_,v_] 19.2, 20.0
enneper[u_, v_] 12.6, 16.4, 17.0, 22.0
ennepermincurve[n_][z_] 22.3
epicycloid[a_,b_][t_] 5.5, 19.0
epitrochoid[a_,b_,h_] 6.2
Eq[f_,g_] 8.2
evolute[α_,n_] 4.2
evolute[α_][t_] 4.1
evolute3d[γ_][t_] 8.3

flattorus[a_,b_,c_,d_][u_,v_] 26.8
foc[i_,x_][u_,v_] 20.2
focal[i_,x_][u_,v_] 20.1
folium[t_] 3.1
frenetframe[α_][t_] 7.5
from[q_] 23.2
funnel[a_,b_,c_][u_,v_] 17.2

gausscurv[x_][u_,v_] 13.4
gcurv[x_][u_,v_] 13.5, 17.1
gcurvsq[k_][f1_,f2_,f3_][u1_,u2_,u3_] 13.7

genhel[α_,c_][u_,v_] 15.4
genhypparab[a_,b_][u_,v_] 14.1
gensurfrev[α_,γ_][u_,v_] 15.8
geodesiccurv[x_,α_][t_] 17.4
geodesictorsion[x_,α_][t_] 17.5
grad[f_, coor_] 9.4, 10.0

harmonic2analytic[x_][z_] 22.4
hel2cat[t_][u_,v_] 16.2
helical[α_,c_][t_] 7.5
helicoid[a_,b_,c_][u_,v_] 12.5, 22.0
helicoid[a_,c_][u_,v_] 16.0, 17.0, 18.0, 20.0
helicoid1[a_,b_,c_][u_,v_] 22.4
helix[a_,b_][t_] 7.3, 10.7, 14.0, 20.0
henneberg[u_,v_] 16.4
hennebergconj[u_,v_] 22.4
highlight[α_, h_:1/20, k_:1/100][t_] 7.3
highlight[α_, h_:1/20, k_:3/200][t_] 13.0
holes[a_,d_,p_,R_:id] 27.5
hyperbola[a_,b_][t_] 2.4, 20.0
hyperbolicparaboloid[u_,v_] 13.0, 20.0
hyperboloid1[a_,b_,c_][u_,v_] 10.6, 15.0
hyperboloid2[a_,b_,c_][u_,v_] 10.6
hypocycloid[a_,b_][t_] 5.5
hypotrochoid[a_,b_,h_] 6.2

id 7.1, 23.1, 27.5
image[a_][u_,v_] 26.2
infarea 27.8
inscurve[n_,a_,α_][t_] 14.2
intrinsic[f_,a_:0,{c_:0,d_:0,e_:0},{min_:-10, max_:10},opts___][t_] 5.3
intrinsic3d[κ_,τ_,c_,{min_:-10,max_:10}][u_] 8.2
inversion[ρ_][p_] 20.4
inversion[q_List,ρ_][p_] 20.4
invertedcurve[q_, ρ_][α_][t_] 20.4
invertedsurf[q_,ρ_][surf_][u_,v_] 20.4
invertedtorus[{p1_,p2_,p3_},ρ_,a_,b_,c_][u_,v_] 20.5
invev[α_,r_][t_] 4.5
involute[α_,c_:0][t_] 4.3
invertedprincurv[q_,ρ_][x_][u_,v_] 20.5

J[{x_,y_}] 1.1, 2.0, 11.0
jacobian[x_][u_, v_] 10.1

κ[α_][t_] 7.4
κ2[α_][t_] 1.4, 2.0
κ2implicit[g_][x_,y_] 3.3
κ2polar[ρ_][t_] 3.4
keplerimplicit[a_,b_][x_,y_] 3.6
klein4[a_,b_][u_,v_] 26.8
kleinbottle[a_][u_, v_] 11.4
kuen[a_][u_,v_] 15.7
kuentcheb[a_,b_][u_,v_] 21.7
kummer[x_,y_,z_] 10.7

L[v_] 1.1, 2.0
laplacian[f_, coor_] 9.5
lemniscate[a_][t_] 2.2, 3.0, 11.6, 13.8, 22.0
length[a_,b_][α_] 1.3
lengthn[a_,b_][α_] 1.3
limacon[a_,b_][t_] 2.9
limaconpolar[n_,a_,b_][θ_] 3.4
linearpursuit[a_,k_][t_] 2.8
lissajous[n_,d_,a_,b_][t_] 2.9
logspiral[a_,b_] 5.0
logspiral[a_,b_][t_] 1.6
logspiralpolar[a_,b_][θ_] 3.5
lor[u_,v_] 26.2
lorentz[a_][u_,v_] 26.2

mcurv[x_][u_,v_] 13.5
mcurvsq[k_][f1_,f2_,f3_][u1_,u2_,u3_] 13.7
meancurv[x_][u_,v_] 13.4
mercator[a_,b_,c_][u_,v_] 16.7
meridians[gl_] 8.4
metric[x_][u_,v_] 12.1, 26.0
moebiusstrip[a_][u_, v_] 11.3, 14.0
monge[h_][u_,v_] 10.2
monkeysaddle[u_,v_] 10.1, 13.0, 17.0, 18.0, 20.0
movingframe[x_][u_,v_] 21.6

nephroid[a_][t_] 19.0
nielsenspiral[a_][t_] 5.5

normal[α_][t_] 7.2, 14.2
normal2ellipse[s_] 4.5
normalcurv[x_,α_][t_] 17.4c
normalline[α_][s_,t_] 4.5
normalsurf[α_][u_,v_] 14.2

octdown[x_] 19.7
orth4d[surf_, {u_,u1_,u2_}, {v_,v1_,v2_}, opts___] 26.8
osculatingcircle[α_][t_] 4.4
oval[p_][ψ_] 6.8

pacman[n_][θ_] 3.4
para[a_,A__] 23.4
parabola[a_] 5.0
parabola[a_][t_] 1.3, 2.4
paraboliccoor[d_,e_,f_][a_,b_,c_][w_][u_,v_] 19.7
parabolicplot[a_,col_,opts___][p_,q_,r_,s_] 19.7
paraboloid[a_,b_][u_,v_] 10.4, 13.0, 18.0
parallelsurf[x_,r_][u_,v_] 19.8
param2imp[α_][x_,y_] 3.1
parcurve[α_,r_][t_] 4.5
pedal[α_,v_:{0,0}][t_] 4.6
piriform[a_,b_][t_] 6.9
plane[a1_,a2_,b1_,b2_,c1_,c2_][u_,v_] 19.5
plotintrinsic3d[κ_,τ_, a_,{min_:-10,max_:10}, opts___] 8.2
plotpatch2d[F_, {x_,x1_,x2_}, {y_,y1_,y2_}, {nx_,ny_}] 19.10
pluecker[u_,v_] 14.1
pluecker[m_,n_,a_][u_,v_] 26.8
polarize[f_][r_,θ_] 10.4
polarplane[u_,v_] 18.5
PolarPlot[r_, {t_,min_,max_},opts___] 3.4
prcurv[x_][u_,v_] 20.2
princurv[x_][u_,v_] 13.5, 20.1
pseudosphere[a_][u_, v_] 15.6, 18.0, 20.0
pseudospheregeodesic[a_,c_,u0_,pm_][t_] 18.5
pseudospheretcheb[a_][u_,v_] 21.4
pts2parabola[u_,v_,w_] 27.2
pts2parabolas[u_,v_,w_,i_] 27.4

R[q_] 23.3
refl[α_,v_][t_] 5.1
reuleauxpolygon[n_,a_][t_] 6.7

revcatenary[a_][t_] 2.4
rexp[t_][z_] 22.1
ricci[M_][j_,l_][u_,v_] 25.5
ride[c_,s_] 23.5
riemann[M_][i_][j_,k_,l_][u_,v_] 25.5
riemann[M_][i_,j_,k_,l_][u_,v_] 25.5
roll[s_] 5.5
roman[u_,v_] 11.5
romanmap[{x_, y_, z_}] 11.5
rot[α_,θ_][t_] 5.1
ruled[α_,γ_][u_,v_] 14.1

S[X__] 6.0, 9.0
sc[t_] 1.3
scalar[M_][u_,v_] 25.5
scherk[a_][u_,v_] 16.5
scherk5[a_,b_,c_][u_,v_] 16.8
scherk5implicit[x_,y_,z_] 16.8
seashell[γ_,r_][t_, θ_] 10.7
selectgraphics3d[obj_,b_] 19.1
semicubicimplicit[a_,b_,c_,d_][x_,y_] 3.6
sfera[gl_] 23.5
sg 18.6
shape[x_][u_,v_] 13.3
singordon[f_][u_,v_] 21.3
sinoval[n_,a_][t_] 6.5
sinsurface[u_,v_] 10.7, 23.0
sinsurfaceimplicit[a_][x_,y_,z_] 10.7
solvegeo[x_,{u0_,v0_},b_,m_] 18.3
sphere[a_] 17.0, 18.0, 27.0
sphere[a_][u_,v_] 8.4, 10.1, 13.0, 23.0
sphericalhelix[a_,b_][t_] 8.5
sphericalloxodrome[a_,b_][t_] 8.6
spiral[a_] 1.6
splot[σ_, {a_,b_}, gl_:3/4] 8.5
stereographicellipsoid[a_,b_,c_][u_,v_] 10.6
stereographicsphere[u_,v_] 8.6
strophoid[a_][t_] 3.6
surfacejacobian[x_][u_,v_] 10.1
surfrev[α_][u_, v_] 15.1, 19.0, 20.2
system 17.3

τ[α_][t_] 7.4, 18.1
tandev[α_][u_,v_] 14.3
tangent[α_][t_] 7.2, 14.2
tangentline[α_][d_,t_] 2.6, 4.5
tangents[u_,v_,w_] 27.2
torus[a_,b_] 27.0
torus[a_,b_,c_][u_,v_] 7.6, 10.4, 13.0, 17.0, 18.0, 20.0, 27.0
torusknot[a_,b_,c_][p_,q_][t_] 7.7
torusimplicit[a_,b_][x_,y_,z_] 27.5
tractrix[a_][t_] 2.6, 15.0
tran[α_, v_][t_] 5.1
trisectrixpolar[a_][θ_] 3.6
tsc[α_,{a_,b_}] 6.1
tubecurve[γ_][r_][t_,θ_] 7.6
turningnumber[α_,{a_,b_},opts___] 6.2
twicubic[t_] 7.3
twisphere[a_,b_][u_,v_] 15.4
twist[α_,a_,b_][u_, v_] 11.6
tx[A_, v_] 2.4, 26.4, 27.0
tx[A_,v_,o_] 4.0
txb[A_, v_] 9.0
txs[A_, v_] 24.2, 26.4

umbilic[x_][u_,v_] 19.2
unitnormal[x_][u_,v_] 10.2, 17.4

V[f_] 8.2
veronese[a_][u_,v_] 26.8
viviani[a_][t_] 7.5, 14.0, 23.0

wallis[a_,b_,c_][u_,v_] 14.5
weier[f_,g_][z_] 22.5
weiercurv[f_,g_][z_] 22.5
weingarten[x_][u_,v_] 13.3
WP[z_] 22.7
WZ[z_] 22.7

z[i_] 27.2
zcprofile[pm_, a_, c_][t_] 15.7