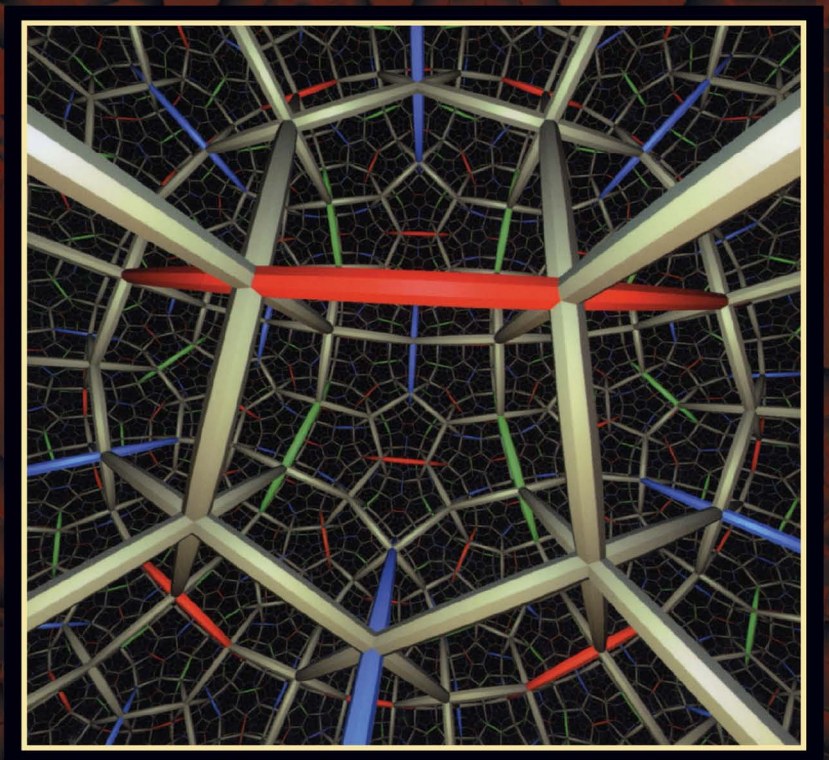
William P. Thurston
Edited by Silvio Levy

Three-Dimensional Geometry and Topology

VOLUME 1



Three-Dimensional Geometry and Topology

Princeton Mathematical Series

EDITORS: LUIS A. CAFFARELLI, JOHN N. MATHER, *and* ELIAS M. STEIN

1. The Classical Groups *by Hermann Weyl*
3. An Introduction to Differential Geometry *by Luther Pfahler Eisenhart*
4. Dimension Theory *by W. Hurewicz and H. Wallman*
8. Theory of Lie Groups: I *by C. Chevalley*
9. Mathematical Methods of Statistics *by Harald Cramér*
10. Several Complex Variables *by S. Bochner and W. T. Martin*
11. Introduction to Topology *by S. Lefschetz*
12. Algebraic Geometry and Topology *edited by R. H. Fox, D. C. Spencer, and A.W. Tucker*
14. The Topology of Fibre Bundles *by Norman Steenrod*
15. Foundations of Algebraic Topology *by Samuel Eilenberg and Norman Steenrod*
16. Functionals of Finite Riemann Surfaces *by Menahem Schiffer and Donald C. Spencer*
17. Introduction to Mathematical Logic, Vol. I *by Alonzo Church*
19. Homological Algebra *by H. Cartan and S. Eilenberg*
20. The Convolution Transform *by I. I. Hirschman and D. V. Widder*
21. Geometric Integration Theory *by H. Whitney*
22. Qualitative Theory of Differential Equations *by V. V. Nemytskii and V. V. Stepanov*
23. Topological Analysis *by Gordon T. Whyburn* (revised 1964)
24. Analytic Functions *by Ahlfors, Behnke, Bers, Grauert et al.*
25. Continuous Geometry *by John von Neumann*
26. Riemann Surfaces *by L. Ahlfors and L. Sario*
27. Differential and Combinatorial Topology *edited by S. S. Cairns*
28. Convex Analysis *by R. T. Rockafellar*
29. Global Analysis *edited by D. C. Spencer and S. Iyanaga*
30. Singular Integrals and Differentiability Properties of Functions *by E. M. Stein*
31. Problems in Analysis *edited by R. C. Gunning*
32. Introduction to Fourier Analysis on Euclidean Spaces *by E. M. Stein and G. Weiss*
33. Étale Cohomology *by J. S. Milne*
34. Pseudodifferential Operators *by Michael E. Taylor*
35. Three-Dimensional Geometry and Topology, Volume 1 *by William P. Thurston. Edited by Silvio Levy*
36. Representation Theory of Semisimple Groups: An Overview Based on Examples *by Anthony W. Knapp*
37. Foundations of Algebraic Analysis *by Masaki Kashiwara, Takahiro Kawai, and Tatsuo Kimura. Translated by Goro Kato*
38. Spin Geometry *by H. Blaine Lawson, Jr., and Marie-Louise Michelsohn*
39. Topology of 4-Manifolds *by Michael H. Freedman and Frank Quinn*
40. Hypo-Analytic Structures: Local Theory *by François Treves*
41. The Global Nonlinear Stability of the Minkowski Space *by Demetrios Christodoulou and Sergiu Klainerman*
42. Essays on Fourier Analysis in Honor of Elias M. Stein *edited by C. Fefferman, R. Fefferman, and S. Wainger*
43. Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals *by Elias M. Stein*
44. Topics in Ergodic Theory *by Ya. G. Sinai*
45. Cohomological Induction and Unitary Representations *by Anthony W. Knapp and David A. Vogan, Jr.*

Three-Dimensional Geometry and Topology

VOLUME 1

William P. Thurston

EDITED BY SILVIO LEVY

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

1997

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data
Thurston, William P.
Three-dimensional geometry and topology / William P. Thurston ;
edited by Silvio Levy.
p. cm. — (Princeton mathematical series ; 35)
Includes bibliographical references and index.
ISBN 0-691-08304-5 (cl : alk. paper)
1. Geometry, Hyperbolic. 2. Three-manifolds (Topology) I. Levy, Silvio.
II. Title. III. Series.
QA685.T49 1997
516'.07—dc21 96-45578

Frontispiece: The sculpture in the photograph is a marble rendition by Helaman Ferguson of the Klein quartic, a genus-three surface with the greatest possible amount of symmetry (336 isometries). It is not possible to represent all this symmetry in space, but the twenty-four heptagons that tile the surface are in fact intrinsically identical. They come from regular heptagons in the hyperbolic plane, meeting three to a vertex; each such heptagon is made of fourteen triangles of the 2-3-7 tiling of Figures 2.10 and 2.14. If you start on any edge of the surface, proceed along it to a fork and turn right, then turn left at the next fork, and keep alternating in this way, you arrive back where you started after eight turns. For this reason the sculpture, which stands at the Mathematical Sciences Research Institute in Berkeley, is named "The Eightfold Way."

The publisher would like to acknowledge William P. Thurston and Silvio Levy for providing the camera-ready copy from which this book was printed

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

http://pup.princeton.edu

Printed in the United States of America

5 7 9 10 8 6

ISBN-13: 978-0-691-08304-9 (cloth)

Contents

Preface

This book began with notes from a graduate course I gave at Princeton University on the geometry and topology of three-manifolds, over the period 1978–1980. The notes were duplicated and sent to people who wrote to ask for them. The mailing list grew to a size of about one thousand before a version was frozen. Much of the original draft was written by Steve Kerckhoff and Bill Floyd.

The notes were originally aimed for an audience of fairly mature mathematicians, and presented material not in the standard repertoire. A number of seminars worked through these notes. Some of the feedback from seminars and individuals convinced me that it would be worth filling in considerably more detail and background; there were several places where people tended to get stuck, sometimes for weeks. I embarked on a project of clarifying, filling in and rearranging the material before publishing it.

Far more time (and blood, sweat and tears) has elapsed since the time of the original notes than I intended or anticipated. The present text originated from several chapters of the original notes, but it has undergone deep transmutation.

The ultimate emergence of this book would not have happened without the support of the Geometry Center and the vision of its founding director, Albert Marden. In particular, from 1990 through 1992 the Geometry Center hosted five intense bookwriting workshops centered on drafts of this text, during which both the ideas and the means of communicating them gained greatly from the scrutiny of multiple eyes and the thoughts of many minds. I owe many thanks to Dick Canary, Jim Cannon, David Epstein, Bill Floyd, Steve Kerckhoff, Yair Minsky, who participated in these workshops and contributed to examining and editing the text and graphics on these occasions.

Most of all, I want to say that I have tremendously appreciated and deeply admired working with Silvio Levy, who has stuck with me and with this project through thick and through thin.

William P. Thurston
June 1996

Figure Credits

Almost all the figures in this book were created by Silvio Levy using Mathematica, with labels automatically converted to TEX by Levy's Mathfig utility. Many of these figures were redrawn from originals by the author. The following figures have a different provenance:

The cover image and Figure 2.19 are frames from the video *Not Knot*, produced by the Geometry Center and directed by Charlie Gunn and Delle Maxwell. The scenes were generated by Gunn using custom software and Renderman.

The photograph in Figure 3.14 was taken by Kenneth Olson, with the collaboration of Aribert Munzner. The color photograph showing *The Eightfold Way* was taken by Jon Ferguson. Both were kindly provided by the sculptor, Helaman Ferguson, and are reproduced here with permission.

Figures 4.13, 4.14, and 4.17 were created by the author using Xfig, and Figures 4.22 and 4.23 using Adobe Illustrator. Labels were subsequently retypeset in TEX by Levy.

Reader's Advisory

The style of exposition in this book is somewhat experimental.

The most efficient logical order for a subject is usually different from the best psychological order in which to learn it. Much mathematical writing is based too closely on the logical order of deduction in a subject, with too many definitions before, or without, the examples which motivate them, and too many answers before, or without, the questions they address. In a formal and logically ordered approach to a subject, readers have little choice but to follow along passively behind the author, in the faith that machinery being developed will eventually be used to manufacture something worth the effort.

Mathematics is a huge and highly interconnected structure. It is not linear. As one reads mathematics, one needs to have an active mind, asking questions, forming mental connections between the current topic and other ideas from other contexts, so as to develop a sense of the structure, not just familiarity with a particular tour through the structure.

The style of exposition in this book is intended to encourage the reader to pause, to look around and to explore. I hope you will take the time to construct your own mental images, to form connections with other areas of mathematics and interconnections within the subject itself.

Think of a tinkertoy set. The key is the pieces which have holes, allowing you to join them with rods to form interesting and highly interconnected structures. No interesting mathematical topic is self-contained or complete: rather, it is full of "holes," or natural questions and ideas not readily answered by techniques native to the topic. These holes often give rise to connections between the given topic and other topics that seem at first unrelated. Mathematical exposition often conceals these holes, for the sake of smoothness—

but what good is a tinkertoy set if the holes are all filled in with modeling clay?

In the present exposition, many of the "holes" or questions are explicitly labeled as exercises, questions, or problems. *Most of these are not walking-the-dog exercises* where the dog follows behind on a leash until the awaited event. You may or may not be able to answer the questions, even if you completely understand the text. Some of the questions form connections with ideas discussed more fully later on. Other questions have to do with details that otherwise would have been "left as an exercise for the reader." Still others relate the material under discussion to topics which are neither discussed nor assumed in the main text.

It is important to read through and think about the exercises, questions and problems. It should be possible to solve some of the more straightforward questions. But please don't be discouraged if you can't solve all, or even most, of the questions, any more than you are discouraged when you can't immediately answer questions which occur to you spontaneously.

There are other ways in which the order of development deviates from the order of logical deduction. For instance, manifolds and geometric structures on manifolds are discussed intuitively in the first two chapters, even though the formal definition and basic properties are only presented in Chapter 3. These definitions are somewhat heavy until one has seen some good examples. The concept of an orbifold is defined only in Chapter 5 (to appear in Volume 2), even though it is significant for the material in Chapter 4.

On the other hand, for purposes of reference, the logical order is sometimes chosen over the psychological order. For example, Chapter 2 contains a fuller treatment of hyperbolic geometry than is motivated by the examples and applications which have been given up to that point. The reader may wish to skip some of it, and refer back only as needed for later reference.

Often a beginner gives up reading a book, or parts of it, when he or she hits a morass of unknown terms and notation. Given the nonlinear method of exposition used in this book, it is likely that you'll encounter unfamiliar terms that are not explained in the text. Don't let that discourage you; it may be useful to read ahead and return to the sticky passage later. Some terms are defined in the Glossary, the first occurrence of each being marked with a dagger (†).

Three-Dimensional Geometry and Topology

Chapter 1

What Is a Manifold?

Manifolds are around us in many guises.

As observers in a three-dimensional world, we are most familiar with two-manifolds: the surface of a ball or a doughnut or a pretzel, the surface of a house or a tree or a volleyball net...

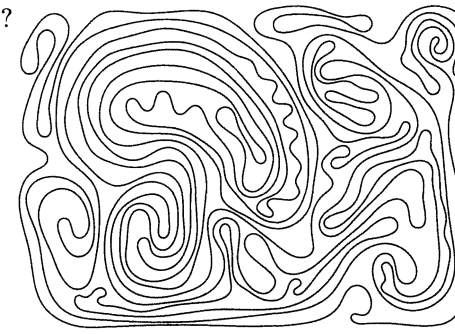
Three-manifolds may seem harder to understand at first. But as actors and movers in a three-dimensional world, we can learn to imagine them as alternate universes.

Mathematically, manifolds arise most often not as physical entities in space, but indirectly: the solution space of some set of conditions, the parameter space for some family of mathematical objects, and so on. Translating such abstract descriptions, where possible, into our concrete imagery of three-dimensional space is generally a big aid to understanding.

Even when we do this, however, it is often not easy to recognize the identity of a manifold: the same topological object can have completely different concrete descriptions. Furthermore, manifolds may have inherent symmetry that is not apparent from a concrete description.

How can we know a manifold?

Exercise. Which manifold is this?



1.1. Polygons and Surfaces

The simplest and most symmetric surface, next to the sphere, is the torus, or surface of a doughnut. This surface has symmetry as a surface of revolution in space, but it has additional "hidden" symmetry as well. The torus can be described topologically by gluing together the sides of a square. If the square is reflected about its main diagonal to interchange the a and b axes, the pattern of identification is preserved.

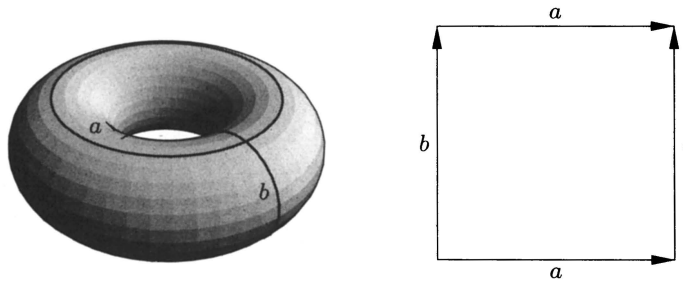


Figure 1.1. The square torus. A torus can be obtained, topologically, by gluing together parallel sides of a square. Conversely, if you cut the torus on the left along the two curves indicated, you can unroll the resulting figure into the square on the right.

Problem 1.1.1 (square torus in space). The *one-point compactification* $\widehat{\mathbf{R}^n}$ of $\mathbf{R}^n$ is the topological space obtained by adding a point ∞ to $\mathbf{R}^n$ whose neighborhoods are of the form $(\mathbf{R}^n \setminus B) \cup \infty$ for all bounded sets B.

(a) Check that the one-point compactification of $\mathbf{R}^n$ is homeomorphic to the sphere S^n.

(b) Consider an ordinary torus in $S^3 = \widehat{\mathbf{R}^3}$, and show that the interchange of curves a and b in Figure 1.1 can be achieved by moving the torus in S^3 (without necessarily preserving its geometric shape). (This question will become much easier after you read Section 2.7.)

(c) Show that this cannot be done in $\mathbf{R}^3$.

Curiously, a torus is also obtained by identifying parallel sides of a regular hexagon (Figure 1.2). This alternate description has six-fold symmetry which is not compatible with the symmetry of the previous description.

Problem 1.1.2 (reconciling the symmetries of a torus). We've seen three concrete descriptions for a torus: as a physical surface in space, as a

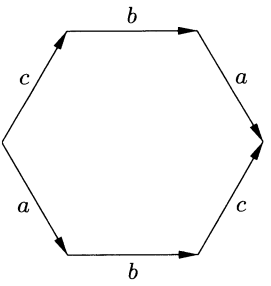


Figure 1.2. The hexagonal torus. Here is another gluing pattern of a polygon which yields a torus. This pattern reveals a different kind of symmetry from the first.

square with identifications, and as a hexagon with identifications. Can you reconcile them to your satisfaction?

(a) Check that gluing the hexagon does yield a torus. Draw the curves needed to cut a torus into a hexagon.

(b) What transformation changes a hexagon with identifications to a quadrilateral with identifications?

(c) Is it possible to †embed the torus in $\mathbf{R}^3$ or in S^3 in such a way that the six-fold symmetry extends to a symmetry of the †ambient space?

(d) One can divide the torus into seven countries in such a way that every country is in one piece and has a (non-punctual) border with every other. In other words, a political map of a torus-shaped world may require up to seven colors. Construct such a seven-colored map. Can it be done symmetrically?

These two descriptions of the torus are closely related to common patterns of †tilings of the †Euclidean plane $\mathbf{E}^2$. Take an infinite collection of identical squares, all labeled as in Figure 1.1, or of hexagons, labeled as in Figure 1.2. Begin with a single polygon, then add more polygons layer by layer, identifying edges of the new ones with similarly labeled edges of the old ones. Make sure the local picture near each vertex looks like the local picture in the original pattern, when the edges of a single polygon were identified: if you follow this rule, each new tile fits in in exactly one way. The result is a tiling of the Euclidean plane by congruent squares or hexagons.

These tilings show that the plane is a †covering space for the torus: the †covering map for the square tiling (say) is the map that identifies corresponding points in each square, taking them all to the same point on the glued-up torus. Since the plane is †simply

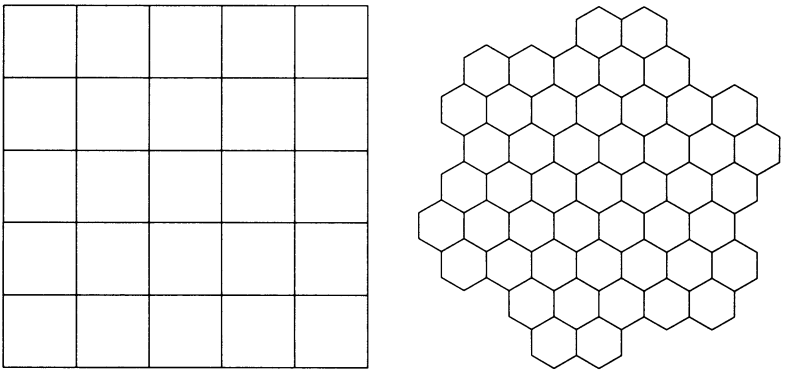


Figure 1.3. Tiling the plane with tori. These tilings of the plane arise from the two descriptions of the torus by gluing polygons. They show the universal covering space of the torus, obtained by "unrolling" the torus.

connected, it is the †universal cover of the torus. The covering map singles out a group of homeomorphisms of the plane, namely those that take any point into another point that has the same image on the torus under the covering map. For the square tiling, for instance, these †covering transformations are the translations that preserve vertices. The torus is the †quotient space of the plane by the †action of this †covering group.

Since the covering transformations are Euclidean isometries, we can give the torus a *Euclidean structure*, that is, a metric that is locally isometric to Euclidean space. This is done as follows: given a point x on the torus, we choose a neighborhood U of x small enough that the inverse image of U in the plane is made up of connected components homeomorphic to U under the covering map p. By shrinking U further, we can make sure that the diameter of these components is less than the distance separating any two of them. Then we declare p to be an isometry between any of these components and U; it doesn't matter which component we choose, because they're all isometric (Figure 1.4).

This locally Euclidean geometry on the torus is not the same as the geometry it has as a surface of revolution in space, because the former is everywhere flat, whereas the torus of revolution has positive Gaussian curvature at some places and negative at others (Section 2.1).

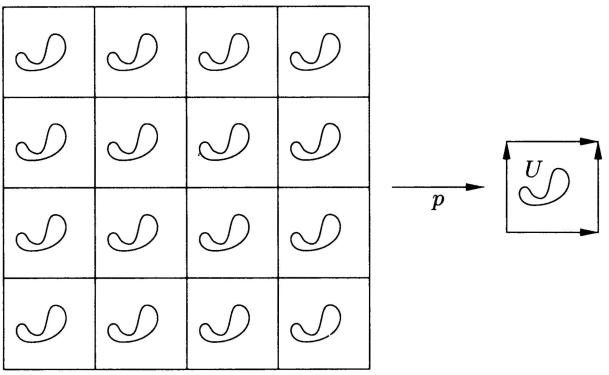


Figure 1.4. Transferring the geometry from the plane to the torus. As a quotient of the plane by a group of isometries, the torus has a locally Euclidean geometry, in which a small open set U is isometric to any component of its inverse image under the covering map p. In this geometry the images of straight lines are †geodesics on the torus. They're usually not geodesics in the geometry of the torus of revolution.

1.2. Hyperbolic Surfaces

Just like the torus, the two-holed torus or *genus-two surface* (Section 1.3) can be obtained by identifying the sides of a polygon. Most familiar is the pattern shown in Figure 1.5, in which we cut along four simple closed curves meeting in a single point, to get an octagon. The four curves can be labeled so that the resulting octagon is labeled $aba^{-1}b^{-1}cdc^{-1}d^{-1}$.

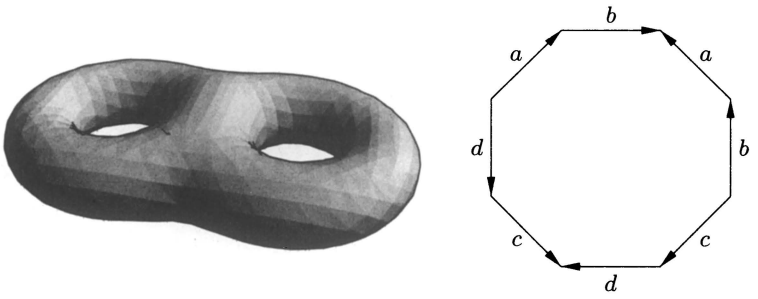


Figure 1.5. A genus-two surface. A two-holed torus, or surface of genus two, can be cut along curves until the result is topologically a polygon. Here we see the most common cutting pattern.

Are there tilings of the plane by regular octagons, coming from the gluing pattern in Figure 1.5? The answer is clearly no in

the Euclidean plane: the interior angle of a regular octagon is $(8-2)\cdot 180°/8 = 135°$, so not even three octagons fit around a vertex, whereas eight would be needed.

But that's not the end of the story: if we don't insist that our plane satisfy †Euclid's parallel axiom, there is nothing to force the sum of the angles of a triangle to be 180°, and we could perhaps choose regular octagons with 45° angles, so they fit nicely eight to a vertex. We will soon describe a concrete construction to do exactly that.

Until the late eighteenth century the validity of the parallel axiom was taken for granted, and in fact much work was invested in frustrated attempts to prove its redundancy by deriving it from Euclid's other axioms and "common notions," all of which seemed much more intuitive. By the 1820s, however, three people had independently come to realize that a self-consistent geometry, with lines, planes, and angles otherwise similar to the usual ones, does not have to satisfy the parallel axiom: they were János Bolyai in Hungary, Carl Friedrich Gauss in Germany and Nikolai Ivanovich Lobachevskii in Russia. Gauss was there first, but he chose not to publish his conclusions, and Bolyai received no recognition until long after his death; and so it is that this non-Euclidean geometry became known as Lobachevskiian geometry until Felix Klein, at the turn of this century, introduced the term hyperbolic geometry, the most current today.

The denial of one of Euclid's axioms remained a profoundly disturbing idea, and although continuing work by Lobachevskii and others not only failed to lead to a contradiction but showed that hyperbolic geometry was remarkably rich, it was a matter of debate throughout most of the nineteenth century whether or not such a geometry could exist. Such doubts lasted until Eugenio Beltrami, in 1868, constructed an explicit model of hyperbolic space—something like a map of hyperbolic space in Euclidean space. Actually, Georg Friedrich Riemann seems to have reached this level of understanding much earlier, for he exhibited a metric for any space of constant curvature in his famous "Lecture on the Hypotheses That Lie at the Foundation of Geometry" (1854), and in general he wrote about lines and planes in such spaces in a manner that indicates a clear grasp of their nature. However, this understanding only entered the general mathematical consciousness with Beltrami's work.

Later other models were introduced, each with its advantages and disadvantages. These models, or maps, are helpful in the same

way that maps of the earth are helpful: they are perforce distorted, but with some imagination one can develop a feeling for the true nature of the landscape by studying them.

Hyperbolic geometry will be an essential tool for us throughout this book, so it's good to get more or less familiar with it right away, at least in the two-dimensional case. Our initial study of hyperbolic geometry will be based on a particular model, but it's best to keep in mind from the start that the same geometric construct—hyperbolic space $\mathbf{H}^n$—can be represented in many different ways. We first give a characterization of hyperbolic lines (geodesics), and of certain line-preserving transformations; from this we derive many other properties of hyperbolic space, including the metric. As we go along we'll develop a "dictionary" to translate between hyperbolic objects and their representations in the model, and as we become fluent we'll start doing this translation automatically.

The hyperbolic plane $\mathbf{H}^2$ is homeomorphic to $\mathbf{R}^2$, and the *Poincaré disk model*, introduced by Henri Poincaré around the turn of this century, maps it onto the open unit disk D in the Euclidean plane. Hyperbolic straight lines, or geodesics, appear in this model as arcs of circles orthogonal to the boundary ∂D of D, and every arc orthogonal to ∂D is a hyperbolic straight line (Figure 1.6). There is one special case: any diameter of the disk is a limit of circles orthogonal to ∂D and it is also a hyperbolic straight line. For simplicity, from now on we will include diameters when talking about arcs orthogonal to ∂D.

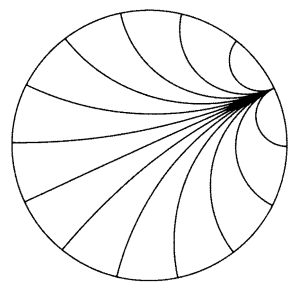


Figure 1.6. Straight lines in the Poincaré disk model. Straight lines in the Poincaré disk model appear as arcs orthogonal to the boundary of the disk or, as a special case, as diameters.

A hyperbolic reflection in one of the lines represented by a diameter of the disk translates, in our model, into a Euclidean reflection

in the same diameter. How about hyperbolic reflections in lines not represented by diameters? They translate into certain Euclidean transformations, called inversions, that generalize reflections:

Definition 1.2.1 (inversion in a circle). If C is a circle in the Euclidean plane, the *inversion* i_C in C is the unique map from the complement of the center of C into itself that fixes every point of C, exchanges the interior and exterior of C and takes circles orthogonal to C to themselves.

Exercise 1.2.2 (inversions are well-defined). (a) Show the following standard result from Euclidean plane geometry: If A is a point outside a circle C and l is a line through A intersecting C at P and P', the product $AP \cdot AP'$ is independent of l and is equal to AT^2, where $\overrightarrow{AT}$ is a ray tangent to C at T. This product is the *power of A with respect to C*.

(b) Use this to show that Definition 1.2.1 makes sense. (Hint: see Figure 1.7, left.)

(c) Prove that if C has center O and radius r, the image $P' = i_C(P)$ is the point on the ray $\overrightarrow{OP}$ such that $OP \cdot OP' = r^2$.

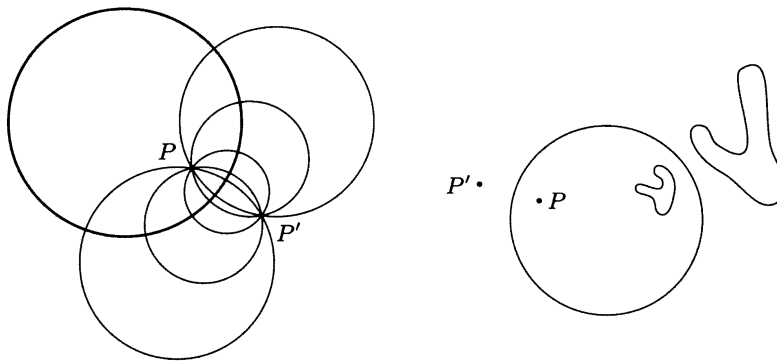


Figure 1.7. Inversion in a circle. All circles orthogonal to a given circle and passing through a given point P also pass through a point P'. We say that P' is the image of P under inversion in the circle. Inversion interchanges the interior and exterior of the circle.

Inversions have lots of neat properties, and none neater than the following, which we will use time after time:

Proposition 1.2.3 (properties of inversions). *If C is a circle in the Euclidean plane, i_C is conformal, that is, it preserves angles.*

Also, i_C takes circles not containing the center of C to circles, circles containing the center to lines, lines not containing the center to circles containing the center, and lines containing the center to themselves.

Proof of 1.2.3. Given two vectors at a point not in C we can construct a circle tangent to each vector and orthogonal to the circle of inversion, and these circles, which are preserved by the inversion, meet at the same angle at their other point of intersection. This shows conformality everywhere but on C. The case of a point on C can be handled by continuity.

Next we show that circles and lines tangent to C are taken to such circles and lines (disrespectively). For the rest of this proof, we let "circle" stand for "circle or line". For any point x on C the plane is filled by a family $\mathcal{F}_O$ of circles orthogonal to C at x and similarly by a family $\mathcal{F}_T$ of circles tangent to C at x (see Figure 1.8). Any

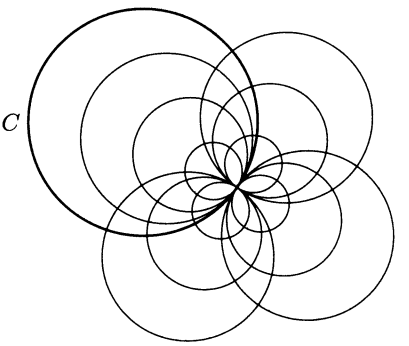


Figure 1.8. Orthogonal families of circles through a point. The circles tangent to a circle C at a given point x are the orthogonal trajectories of the family of circles orthogonal to C at x.

circle from $\mathcal{F}_T$ meets any circle from $\mathcal{F}_O$ perpendicularly at both their intersection points, and thus $\mathcal{F}_T$ forms the set of orthogonal trajectories of $\mathcal{F}_O$. We already know that $\mathcal{F}_O$ is preserved by i_C, so since the inversion is conformal the family of orthogonal trajectories of $\mathcal{F}_O$ is preserved, and thus any circle tangent to C is taken to another circle tangent to C.

Any circle other than a line through the center of C can be blown up or shrunk down to a circle tangent to C by means of a †homothety centered at the center of C. The circle's image under

inversion suffers exactly the opposite fate, by exercise 1.2.2(c): it shrinks down or blows up by the same factor. Since the image of the tangent circle is a circle, so is the image of the original circle. Lines through the center go to themselves by 1.2.2(c). 1.2.3

Notice that circles through the center of C and straight lines are sent to each other because points closer and closer to the center are sent further and further away.

Example 1.2.4 (mechanical linkages). Around the middle of the nineteenth century, with the rapid development of the Industrial Age, there was great interest in the theory of mechanical linkages. An important problem, for a time, was to construct a mechanical linkage that would transform circular motion into straight-line motion, that is, maintain some point on the linkage in a straight line as another point described a circle. In the 1860s Lippman Lipkin and Peaucellier independently found a solution to the problem—the same solution, in fact, involving inversion in a circle. It turned out to be of little use because in practice the relatively large number of moving components (seven bars and six joints) made the linkage wobble more than simpler linkages that, mathematically speaking, only approximated straight-line motion.

Exercise 1.2.5. (a) Prove that the linkage of Figure 1.9 performs as advertised in its caption.

(b) Construct a mechanical linkage that achieves straight line motion.

Back to the Poincaré model. If a hyperbolic line appears in the model as (an arc of) a circle orthogonal to ∂D, the hyperbolic reflection in this line appears as (the restriction to D of) the Euclidean inversion in this circle. This is plausible because, by Proposition 1.2.3, such an inversion maps D into itself pointwise and preserves hyperbolic lines. We will presently see that it also preserves distances, if distances are defined the right way, so the word "reflection" is fully justified.

How then should distances be defined in the hyperbolic plane? Answering this question boils down to describing the †Riemannian metric of the hyperbolic plane, that is, to assigning each point an inner product for the tangent space at that point. In Section 2.2 we will write down a formula, but for now we can learn a whole lot just from geometric constructions. The driving idea is that hyperbolic reflections should preserve distances, that is, they should be

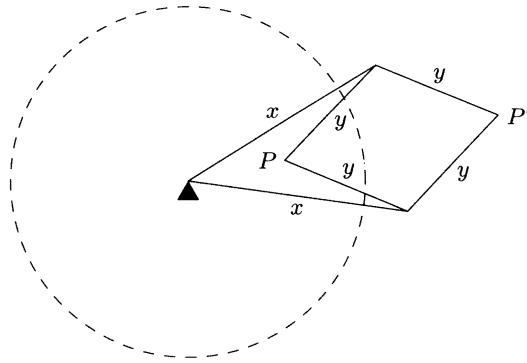


Figure 1.9. A mechanical inversor. This mechanical linkage performs an inversion in the circle of radius $r = \sqrt{(x^2 - y^2)}$. The small triangle near the center of the circle indicates an anchor point for the linkage. If point P is moved around a figure, P' moves around the image of the figure under inversion in the dotted circle.

isometries. This is enough to pin down the metric up to a constant factor.

We start with Figure 1.10. Consider two orthogonal hyperbolic lines L and M, seen in the model as Euclidean arcs of circles orthogonal to ∂D, and another Euclidan circle C that intersects ∂D in the same two points as L. A simple argument shows that C, too, is orthogonal to M, so that, by the definition of inversion, the hyperbolic reflection in M leaves both L and C invariant. If this reflection is to preserve hyperbolic distances, corresponding points of C on both sides of M must be equally distant from L. In fact, by varying M among the lines orthogonal to L, we see that *all* points of C must be equally distant from L, that is, C must be an *equidistant curve*. The banana-shaped region between L and C can be filled with segments orthogonal to L, all having the same hyperbolic length.

Now apply any hyperbolic reflection to the whole picture. The angle α at the tips of the banana doesn't change, because inversions preserve angles; neither does the width of the banana—the hyperbolic length l of the transversal segments—since we want reflections to be isometries. This means that l should be a function solely of α! Moreover, this function has a finite derivative at $\alpha = 0$, because the Euclidean length l_E of any particular transversal segment is roughly proportional to α for α small, and Euclidean and hyperbolic lengths

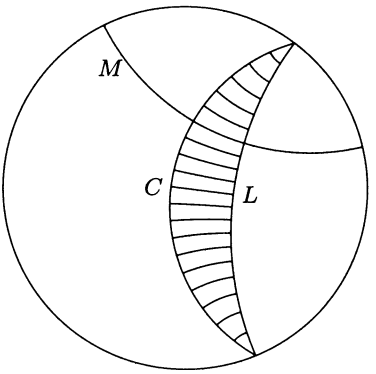


Figure 1.10. Equidistant curve to a line. All points on the arc of circle C lie at the same hyperbolic distance from the hyperbolic line L.

should be proportional to first order. We may take the derivative $dl/d\alpha$ at $\alpha = 0$ to be 1.

We're now equipped to go back and define the Riemannian metric by means of this construction (Figure 1.11). To find the length of a tangent vector v at a point x, draw the line L orthogonal to v through x, and the equidistant circle C through the tip. The length of v (for v small) is roughly the hyperbolic distance between C and L, which in turn is roughly equal to the Euclidean angle between C and L where they meet. If we want an exact value, we consider the angle α_t of the banana built on tv, for t approaching zero: the length of v is then $d\alpha_t/dt$ at $t = 0$.

Actually things are even simpler than that, because two vectors at x having the same Euclidean length also have the same hyperbolic length. This follows from the existence of a hyperbolic reflection fixing x and taking one vector to the other, and from the fact that the derivative of an inversion at a point on the the inversion circle is an orthogonal map (by the remark after Definition 1.2.1). Since Euclidean and hyperbolic vector lengths at a point are proportional, so are the inner products. In particular, the Poincaré model is conformal, because Euclidean and hyperbolic angles are equal.

We can now get back to our tiling of the hyperbolic plane using regular octagons. Remember that we need a tile with hyperbolic angles equal to 45°: but since the Poincaré model is conformal, the Euclidean angle between the arcs that form the edges will be the same. Now imagine a small octagon centered at the origin; since

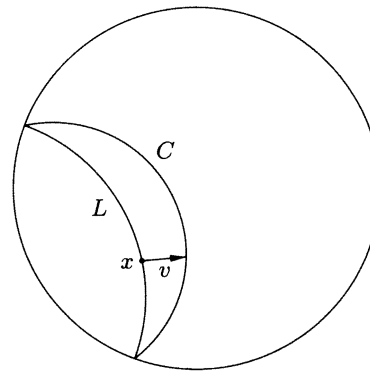


Figure 1.11. Hyperbolic versus Euclidean length. The hyperbolic and Euclidean lengths of a vector in the Poincaré model are related by a constant that depends only on how far the vector's basepoint is from the origin.

its edges (in the model) bend just a little, its angles are close to 135°. By moving the vertices away from the origin we can make the angles as small as we want. By continuity (or, more pedantically, the †intermediate value theorem), there is some octagon in between whose angles are exactly $\pi/4$ (Figure 1.12).

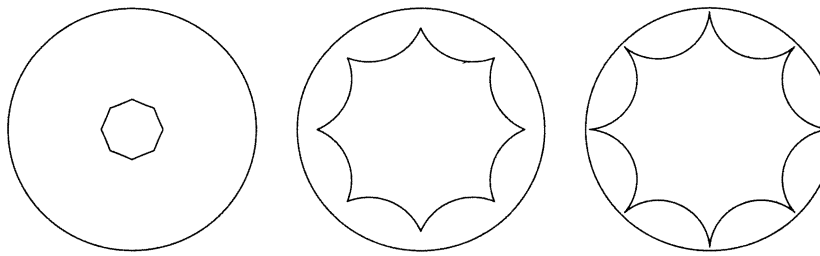
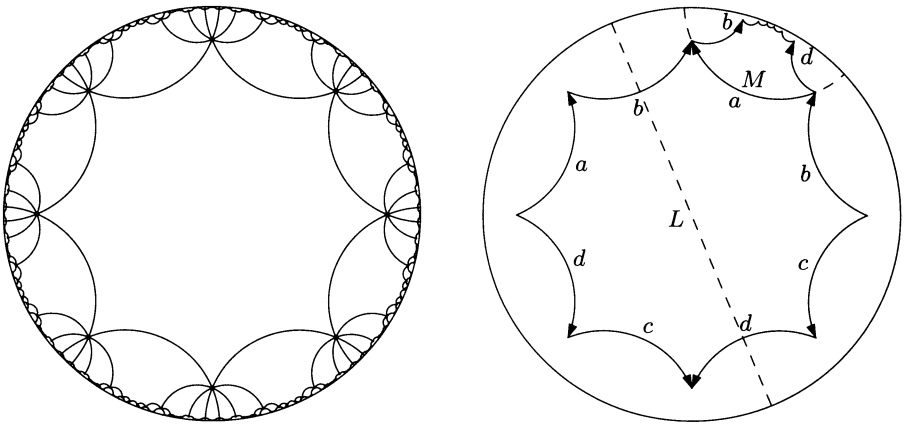


Figure 1.12. Bigger octagons in hyperbolic space have smaller angles. Between a tiny, Euclidean-like octagon with large angles (a) and a very large one with arbitrarily small angles (c) there must be one with angles exactly $\pi/4$ (b).

Once we've found the octagon we want, we take identical copies of it and place them on the hyperbolic plane respecting the identifications prescribed by Figure 1.5, to give the pattern shown in Figure 1.13(a). The copies look different depending on where they

are in the model—in particular, they quickly start looking very small as we move away from the origin—but they can all be obtained from one another by hyperbolic isometries. For example, the two copies in Figure 1.13(b) are mapped to one another by a reflection in L, followed by a reflection in M.



(a) (b)

Figure 1.13. A tiling of the hyperbolic plane by regular octagons. (a) A tiling of the hyperbolic plane by identical regular octagons, seen in the Poincaré disk projection. (b) To get the small octagon from the big one, reflect in L, then in M.

This tiling of the hyperbolic plane shows that a genus-two surface can be given a *hyperbolic structure*, a geometry such that the surface looks locally like the hyperbolic plane. The construction exactly parallels the one we saw for the torus at the end of Section 1.1: again we have a covering space, the hyperbolic plane, with a Riemannian metric preserved by all the covering transformations, so we can transfer that metric to the quotient space.

It is not an accident that we were able to cover the torus with the Euclidean plane, and the genus-two surface with the hyperbolic plane: all surfaces can be given simple geometric structures, as we'll see in Section 1.3.

Problem 1.2.6 (genus-two symmetry). How much symmetry does a surface of genus two have?

(a) Show how to embed a genus-two surface in space so as to have threefold symmetry.

(b) Show that a surface of genus two may be obtained from either a regular octagon or a regular decagon by identifying parallel sides. Draw pictures of the curves needed to cut the surface into an octagon, or a decagon.

(c) Let T_8 be the rotation of †order 8 of the octagon, and T_{10} the rotation of order 10 of the decagon. These transformations go over to homeomorphisms of order 8 and 10 of the surface of genus two. How many fixed points do T_8 and T_{10} have? How many periodic points of order less than 8 or 10?

(d) Is it possible to embed a genus-two surface in space so as to admit a symmetry of order 8? of order 10? What if you consider †immersions instead of embeddings, that is, if you allow self-intersections?

1.3. The Totality of Surfaces

Gluing edges of polygonal regions in dimension two, in the way we have been doing, always gives rise to a two-dimensional manifold. To be precise, let $F_1, \ldots, F_k$ be †oriented †polygonal regions, and suppose that the total number of boundary edges is even. Give the edges the orientations induced by the orientations of the regions. A *gluing pattern* consists of a pairing of edges and, for each such pair, a choice of $+$ or $-$ indicating whether the pair should be identified by an †orientation-preserving or †orientation-reversing homeomorphism. For example, the gluing pattern shown in Figure 1.1 pairs opposite edges and assigns to each pair the symbol $-$, since both gluing maps reverse orientation. (Notice that the arrows in the figure indicate matching directions, rather than edge orientations. See also Figure 1.14.)

Exercise 1.3.1 (homeomorphisms of an interval). Prove that two homeomorphisms of an interval to itself are †isotopic if and only if they both preserve orientation, or both reverse orientation. (Write down a formula that works.)

Exercise 1.3.2 (gluings in two dimensions). (a) Using Exercise 1.3.1, show that a gluing pattern determines a unique topological space.

(b) Show that this space is always a two-dimensional manifold.

(c) Show that the manifold is oriented if the gluing pairs edges with opposite orientations.

(d) Figure 1.14 shows three gluing patterns that identify opposite edges of a quadrilateral. What two-manifold is obtained according to each of them?

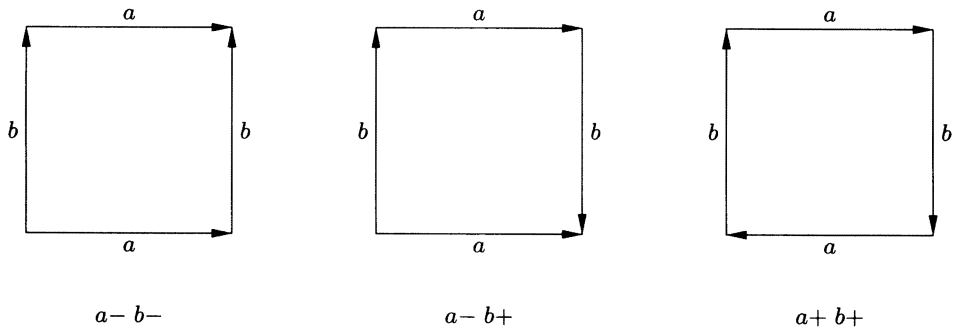


Figure 1.14. Three square gluings. Three possible gluings for a square. The signs associated with each pair of edges indicate whether the gluing map preserves or reverses orientation; the arrows convey the same information.

(e) Explain a necessary and sufficient condition for the two-manifold to be orientable, that could be read off by a computer from the gluing pattern.

It is not easy to visualize directly what surface is obtained when you glue a many-sided polygon, or several polygons, in a given pattern. However, there is an easily computed numerical invariant, the *Euler number* of a surface, that enables one to recognize surfaces quickly. If F is the number of constituent polygons (or faces) and E and V are the numbers of their edges and vertices *after* identification, the Euler number $\chi(S)$ of the glued-up surface S is given by $F - E + V$.

For instance, when a torus is formed by gluing a square as in Figure 1.1, we have one polygon, two edges (the sides of the square identified in pairs) and one vertex (all four vertices identified into one). Therefore $\chi(T^2) = 1 - 2 + 1 = 0$. For the hexagonal torus of Figure 1.2, we get three edges and two vertices, since the vertices are identified in triples; so again $\chi(T^2) = 1 - 3 + 2 = 0$. The sphere S^2 can be divided into four triangles, to form a tetrahedron with six edges and four vertices, so $\chi(S^2) = 4 - 6 + 4 = 2$; if it is divided into six squares to form a cube, the computation is $\chi(S^2) = 6 - 12 + 8 = 2$.

Cutting up a surface into polygons and their edges and vertices, as above, is an example of cell division. A *cell* is a subset $C \subset X$, where X is any Hausdorff space, homeomorphic to an open disk of some dimension, with the condition that the homeomorphism can be extended to a continuous map from the closed disk into X, called the *cell map*. A *face* is a two-cell, an *edge* is a one-cell, and a *vertex* is a zero-cell. A *cell division* of X is a partition of X into cells, in

such a way that the boundary of any n-cell is contained in the union of all cells of dimension less than n.

If X is a †differentiable manifold, we will generally assume that our cell divisions are *differentiable*: this means that, for each cell C, the cell map can be realized as a differentiable map from a convex polyhedron onto the closure of C, having maximal rank everywhere. (The idea of differentiability at a point requires that the map be defined on a neighborhood of the point in $\mathbf{R}^n$. So if $X \in \mathbf{R}^n$ is not an open set, we say that a map $X \to \mathbf{R}^m$ is differentiable if it is the restriction of a †differentiable map on an open neighborhood of X.)

Differentiable cell divisions correspond closely to our intuitive idea of cutting a surface into polygons. Occasionally we will encounter cell divisions that are not of this type—for example, in Problem 1.1.1(a) a sphere is expressed as a union of a vertex and a face.

The *Euler number* of a space X having a finite cell division is defined as the sum of the numbers of even-dimensional cells, minus the sum of the numbers of odd-dimensional cells. It is natural to ask: Is the Euler number independent of the cell division? The answer is yes, and we'll prove it for differentiable surfaces.

A surface can have cells of dimension at most two, by the theorem on the †invariance of domain. Let's check what happens to the Euler number when a two-cell or a one-cell is further subdivided. If an edge is divided into two, by placing a new vertex in its middle, this adds one edge and one vertex. They contribute with opposite signs, so they cancel. If a two-cell is subdivided into two, by means of a new edge between two of its existing vertices, this adds one two-cell and one one-cell. These also contribute with opposite signs, so they cancel.

One way to show the invariance of the Euler number would be to prove that these two operations and their inverses are enough to go between any two finite cell divisions. But this approach is not very satisfactory—one can easily get lost in the technical details, and fail to see what's really going on. A more insightful idea is to relate the Euler number to something that clearly doesn't depend on any cell division: vector fields on the surface.

Let's look at a simple example first, before tackling the problem in full generality. Consider the sphere S^2, carrying a cell division that is realized as a convex polyhedron in $\mathbf{E}^3$. Arrange the polyhedron in space so that no edge is horizontal—in particular, so there is exactly one uppermost vertex U and lowermost vertex L.

Put a unit + charge at each vertex, a unit − charge at the center of each edge, and a unit + charge in the middle of each face. We will show that the charges all cancel except for those at L and at U. To do this, we displace the vertex and edge charges into a neighboring face, and then group together all the charges in each face. The direction of movement is determined by the rule that each charge moves horizontally, counterclockwise as viewed from above (Figure 1.15).

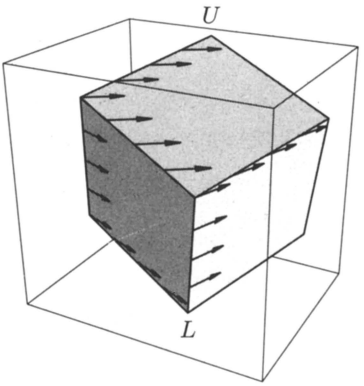


Figure 1.15. Charges on a convex polyhedron. The arrows are part of a horizontal vector field sweeping the surface of this polyhedron; the field is undefined only at the uppermost and lowermost vertices. When the + charges on vertices and the − charges on edges move according to the vector field, they cancel the + charges on the faces.

In this way, each face receives the net charge from an open interval along its boundary. This open interval is decomposed into edges and vertices, which alternate. Since the first and last are edges, there is a surplus of one −; therefore, the total charge in each face is zero. All that is left is +2, for L and for U.

We now generalize this idea to any differentiable surface with a *differentiable triangulation*. This means a differentiable cell division where the faces are modeled on triangles, in such a way that the cell map for any face is an †embedding taking each side of the model triangle onto an edge of the cell division, and the cell map for this edge is compatible with the cell map for the face (that is, they differ by an affine map between domains: see Figure 1.16).

The assumption that the surface is triangulated isn't really restrictive: if we start with a differentiable cell division that is not a triangulation, we can subdivide edges and faces so each face is

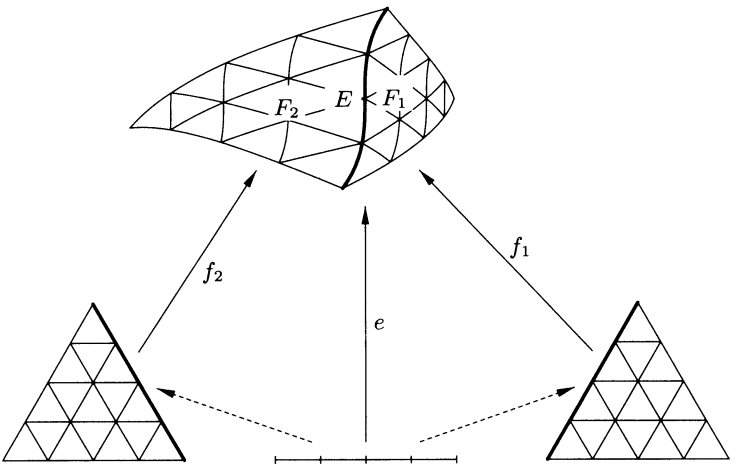


Figure 1.16. Compatibility condition for a triangulation. The cell map f_1 from the model triangle onto face F_1, composed with an affine map from the model interval to the appropriate side of the model triangle, agrees with the cell map e from the model interval onto edge E. Similarly, f_2 agrees with e after composition. This means we can refine the triangulation by subdividing the model triangle and the model interval.

modeled on a triangle with embedded sides. We have already seen that this process doesn't change the Euler number. We also have to adjust the cell maps, by a process similar to that of Exercise 1.3.1, so that edge maps are compatible with face maps.

Proposition 1.3.3 (nonvanishing vector fields). *If a differentiably triangulated closed surface admits a nowhere zero tangent vector field, its Euler number is zero.*

Proof of 1.3.3. Suppose first that the vector field is everywhere *transverse* to the triangulation, that is, nowhere tangent to an edge. In Problem 1.3.4 you're asked to show that we can arrange for that to be the case, by subdividing the triangulation and otherwise adjusting it, without changing the Euler number. By subdividing we can also make the field nearly constant within each triangular face: more precisely, for each face, in some coordinate chart, the direction of the field should change by at most ε, and the direction of the edges should change by at most ε along the edge.

Given such a transverse triangulation, we apply the idea of moving vertex and edge charges in the direction given by the vector field.

If a vertex's charge moves into a face, so do the charges for the two adjacent edges around the face. This means that in each face either exactly one edge charge gets pushed in or two edge charges and one vertex charge; the case of three or zero edge charges being pushed in is ruled out because it cannot occur for a constant field, and our field is nearly constant. In both allowable cases, the face is left with a total charge of zero, so the Euler number is zero. 1.3.3

Problem 1.3.4 (transverse triangulation). To complete the proof of Proposition 1.3.3, we must show the triangulation can be changed so as to become transverse to the field and so the field and edge directions are nearly constant within each face.

(a) Cover the surface with a finite number of coordinate patches. By drawing equally spaced lines parallel to each edge as in Figure 1.16, subdivide the triangulation so finely that the *star* of each vertex v—that is, the union of edges and faces incident on v—lies in a single coordinate patch, and that the direction of each edge and of the field in the star of v, measured in these coordinates, changes by no more than ε.

(b) Imagine the sets of directions of the edges and of the field as intervals on the circle, of length bounded by ε. Show that you can make the intervals of directions of the edges avoid the interval of directions of the field by moving v a little bit in the appropriate direction, and extending the movement to each edge incident on v by means of a Euclidean †similarity (with respect to the patch coordinates) that keeps the other endpoint of the edge fixed.

(c) Now extend this process to all vertices simultaneously. First show that we can assume that the vertices can be colored red, green and blue, so that no two vertices of the same color are joined by an edge. (Hint: use †barycentric subdivision.) Adjust all red vertices at once, then all green vertices. This leaves all edges transversal.

The torus T^2 has nowhere zero vector fields: consider a uniform field on $\mathbf{E}^2$ and take the quotient as in Figure 1.4. So its Euler number is zero.

What about other surfaces? Most of them do not admit a nowhere zero vector field. The best we can do is to find a vector field that is zero at isolated points (see Exercise 1.3.8). The proof of Proposition 1.3.3 suggests that charges cancel in regions away from the zeros of such a field, so we now need to study its behavior near its zeros.

Let X be a vector field on a surface with an isolated zero at a point z. Working as in the proof of Problem 1.3.4, construct a small