PRELIMINARY COPY

January, 1991

Three-Dimensional Geometry and Topology

William P. Thurston
Princeton University

Contents

List of Figures

Reader's Advisory

This is a draft copy of portions of *Three-dimensional geometry and topology*. It is not for duplication or distribution, as it is still being actively revised.

This book began with notes from a graduate course I gave at Princeton University on the geometry and topology of three-manifolds, over a period of two or three years, starting in 1978. 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. These notes are still available from the Princeton math department.

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.

Much more time has elapsed since than I intended or anticipated. The present text is based roughly on the first seven chapters of the original notes, but with substantial rearrangement and interpolation. Dick Canary, David Epstein, Silvio Levy, and Yair Minsky have contributed extensively to this revision.

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.

% Geometric Manifolds
% Hyperbolic geometry and its friends

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 five, even though it is significant for the material in chapter four.

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 non-linear 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: all but the most elementary of these terms are defined in the Glossary, and the first occurrence of each one is flagged by a dagger (†) for ease of reference.

On the other hand, it may happen that the definition of a word doesn't help you much, because it involves other, still unfamiliar, concepts. In that case you may want to forge ahead and return to the sticky passage later, or consult the reference given for the word in the Glossary to supply yourself with some background.

Any suggestions and corrections, concerning style as well as content, are welcome. Be a critical reader.

William P. Thurston
January 1990

Chapter "What is a manifold?"
two-manifolds
ball
doughnut
pretzel
house
tree
volleyball net
Three-manifolds
alternate universes
solution space
parameter space
concrete descriptions
inherent symmetry

Chapter 1

What is a manifold?

Manifolds are around us in many guises.

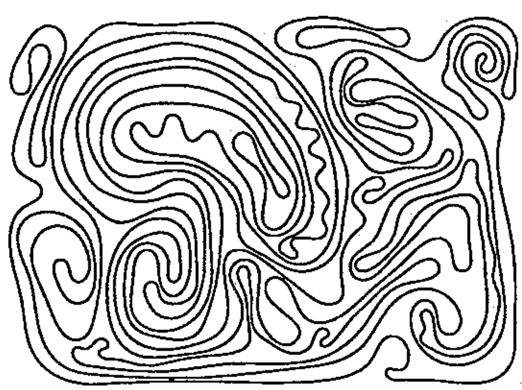
As observers in a three-dimensional world, we are most familiar with two-manifolds: the surface of a ball or a doughnut or a pretzel, the surface of a house or a tree or a volleyball net...

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?



whichmanifold

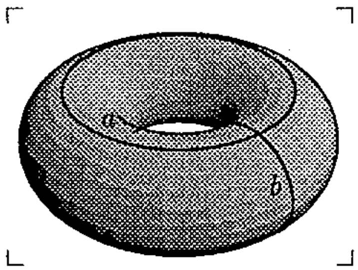
Figure 1.1. Which manifold is this?

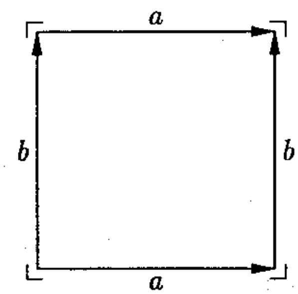
Section "Polygons and surfaces"
sphere
torus
doughnut
symmetries of the torus
Problem "square torus in space"
::one-point compactification
$\widehat{\mathbf{R}^n}$
∞
% sqtorus
% The geometry of the three-sphere
hexagon
% hextorus
Problem "reconciling the symmetries of a torus"

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.

Silvio: add curves





sqtorus

Figure 1.2. 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 1.2 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 $\mathbf{R}^3$.

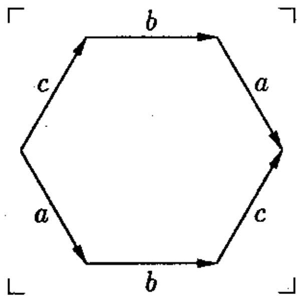
Curiously, a torus is also obtained by identifying parallel sides of a regular hexagon (figure 1.3). This alternate description has six-fold symmetry which is not compatible with the symmetry of the previous description.

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 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?

†embed
†ambient
seven-colored map on the torus
†tiling
†Euclidean
% sqtorus
% hextorus



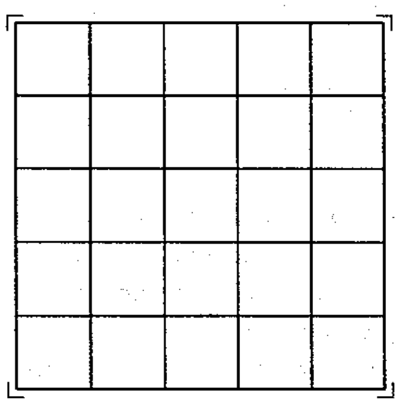
hextorus

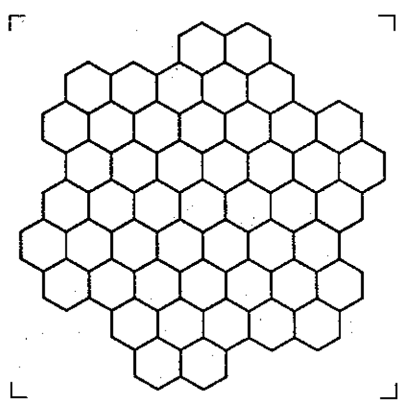
Figure 1.3. The hexagonal torus. Here is another gluing pattern of a polygon which yields a torus. This pattern reveals a different kind of symmetry from the first.

(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 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 1.2, or of hexagons, labeled as in figure 1.3. 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.





torustiles

Figure 1.4. Tiling the plane with toruses. 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 around both axes.

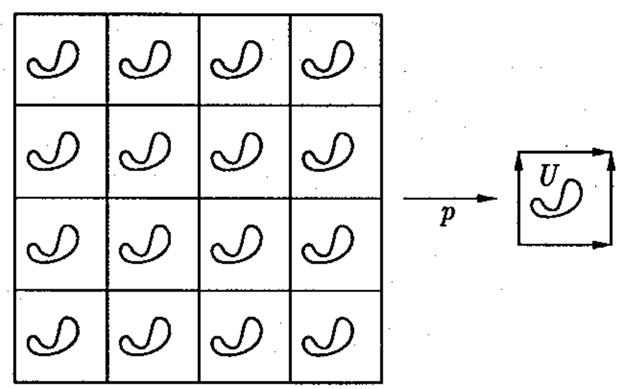
†covering space
†covering map
†simply connected
†universal cover
†covering transformations

†quotient
†action
†covering group
::Euclidean structure
% torusmetric
% Negatively curved surfaces in space

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 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 1.5).

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).



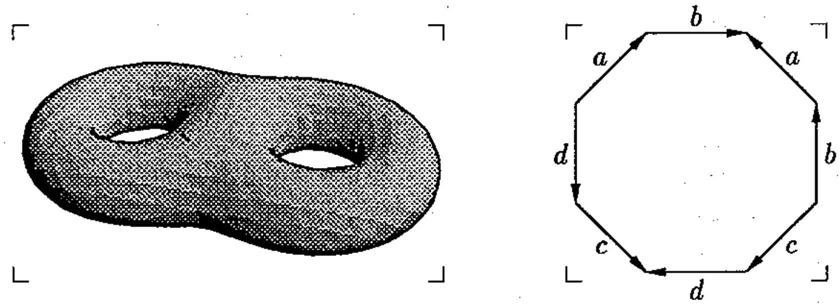
torusmetric

Figure 1.5. 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 not geodesics in the geometry of the torus of revolution.

Section "Hyperbolic surfaces"
::genus-two surface
% The totality of surfaces
% genus2
% genus2
†Euclid's parallel axiom
János Bolyai
Carl Friedrich Gauss
Nikolai Ivanovich Lobachevskii
Lobachevskiian geometry
hyperbolic geometry

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 1.6, 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 $aba^{-1}b^{-1}cdc^{-1}d^{-1}$.



genus2

Figure 1.6. 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.

Silvio: add curves

Are there tilings of the plane by regular octagons, coming from the gluing pattern in figure 1.6? 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

Eugenio Beltrami
Beltrami
model
Georg Friedrich Riemann
hyperbolic space
H^n
dictionary
::Poincaré disk model
Henri Poincaré
% disklines
Definition "inversion in a circle"
::inversion
Exercise "inversions are well-defined"

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.

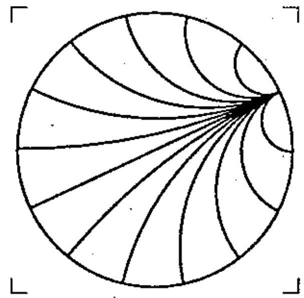
Hyperbolic geometry will be an essential tool for us throughout this book, so it's good to get more or less familiar with it right away, at least in the two-dimensional case. Our initial study of hyperbolic geometry will be based on a particular model, but it's best to keep in mind from the start that 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 1.7). 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.

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.

::power of A with respect to C
% inversion in a circle
% inversion
inversion formula
Proposition "properties of inversions"
::conformal



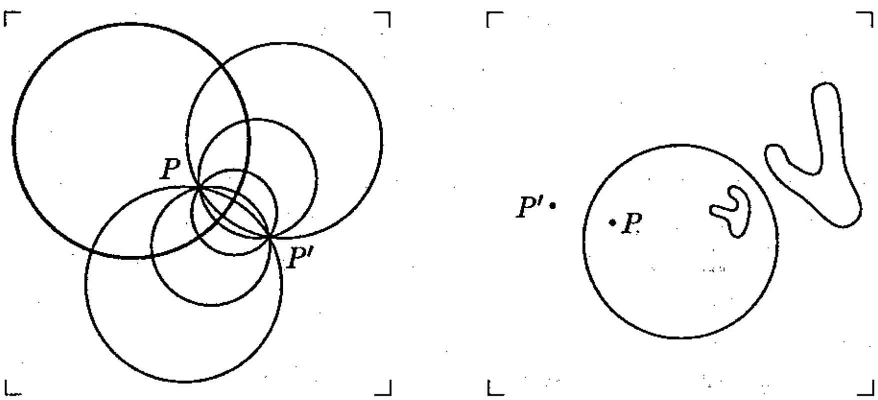
disklines

Figure 1.7. 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.

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 1.8, 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 $OP \cdot OP' = r^2$.



inversion

Figure 1.8. 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.

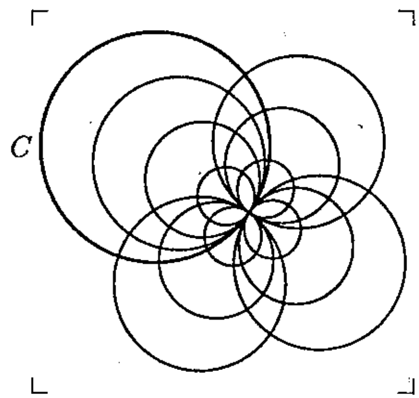
Inversions have lots of neat properties, and none neater than the following, which we will use time after time:

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.*

% orthfamilies
†orthogonal trajectories
†homothety
% inversion formula
Example "mechanical linkages"

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 1.9). Any circle from $\mathcal{F}_T$ meets any circle from $\mathcal{F}_O$



orthfamilies

Figure 1.9. 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.

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 not passing 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.

As a special case, straight lines and circles through the center of C are sent to each other because points closer and closer to the center are sent further and further away. $\boxed{1.2.3}$

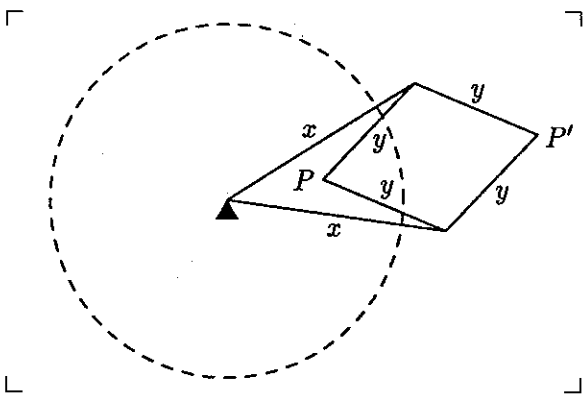
Example 1.2.4 (mechanical linkages). Around the middle of the nineteenth century, with the rapid development of the Industrial Age, there was great interest in the theory of mechanical linkages. An important problem, for a time, was to construct a mechanical linkage that would transform circular motion into straight-line motion, that is, maintain some point on the linkage in

% inversor
% properties of inversions
†Riemannian metric
% The inversive models
% banana

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.

Silvio: check this better.

Exercise 1.2.5. (a) Prove that the linkage of figure 1.10 performs as advertised in its caption.



inversor

Figure 1.10. 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.

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

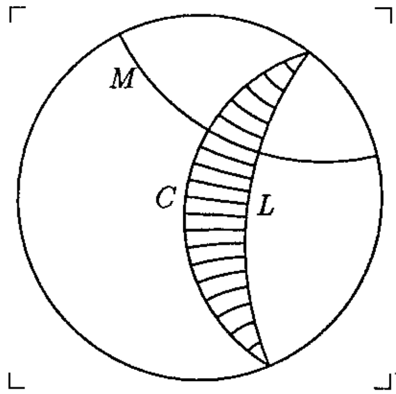
Back to the Poincaré model. If a hyperbolic line appears in the model as (an arc of) a circle orthogonal to ∂D, the hyperbolic reflection in this line appears as (the restriction to D of) the Euclidean inversion in this circle. This is plausible because, by proposition 1.2.3, such an inversion maps D into itself 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 isometries. This is enough to pin down the metric up to a constant factor.

We start with figure 1.11. 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. The

::equidistant curve
% vectorlength
% inversion in a circle

hyperbolic reflection in M leaves L invariant and, by the definition of inversion, also C. 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.



banana

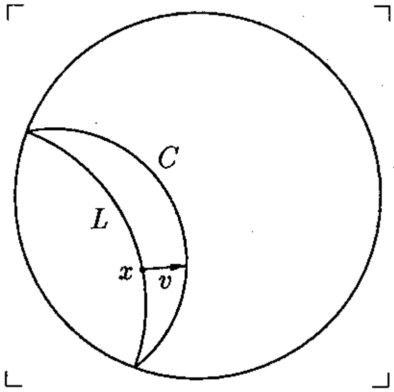
Figure 1.11. Equidistant curve to a line. All points on the arc of circle C lie at the same hyperbolic distance from the hyperbolic line L.

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 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 1.12). 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

†intermediate value theorem
% 3octagons
% genus2
% octatiles

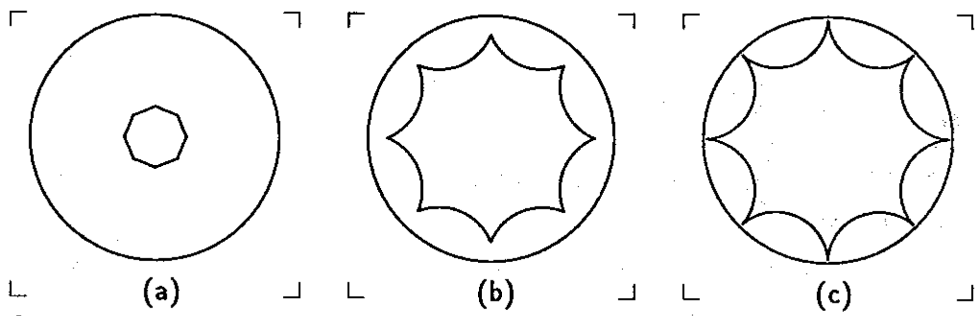


vectorlength

Figure 1.12. 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.

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 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 1.13).



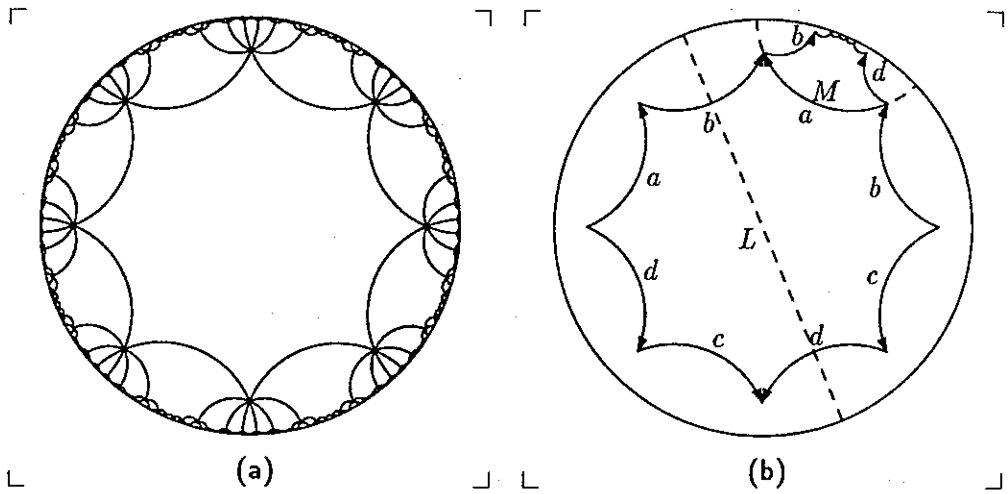
3octagons

Figure 1.13. 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 1.6, to give the pattern shown in figure 1.14(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

% octatiles
::hyperbolic structure
% Polygons and surfaces
% The totality of surfaces
Problem "genus-two symmetry"
octagon
decagon
†order
†immersion
self-intersecting surfaces

obtained from one another by a hyperbolic isometry. For example, the two copies in figure 1.14(b) are mapped to one another by a reflection in L, followed by a reflection in M.



octatiles

Figure 1.14. 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 three-fold 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?

Section "The totality of surfaces"
†oriented
†polygonal region
†induced
::gluing pattern
†orientation-preserving
†orientation-reversing
% sqtorus
% 3gluings
Exercise "homeomorphisms of an interval"
homeomorphisms of an interval
†isotopic
Exercise "gluings in two dimensions"
% homeomorphisms of an interval
two-dimensional gluing is a manifold
orientability of gluings
% 3gluings
torus
Möbius strip
projective plane
computer algorithm
::Euler number

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 1.2 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 1.15.)

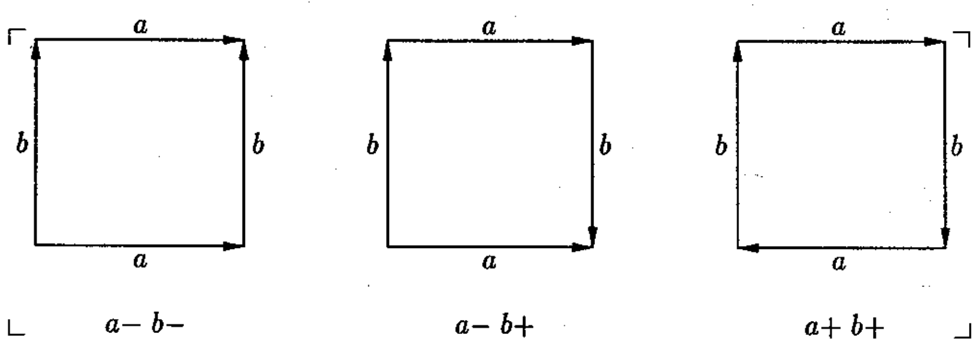
Exercise 1.3.1 (homeomorphisms of an interval). Prove that two homeomorphisms of an interval to itself are †isotopic if and only if they both preserve orientation, or both reverse orientation. (Write down a formula that works.)

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 1.15 shows three gluing patterns that identify opposite edges of a quadrilateral. What two-manifold is obtained according to each of them?



3gluings

Figure 1.15. 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

$\chi(S)$
% sqtorus
% hextorus
::cell
::cell map
::face
::edge
::vertex
::cell division
†differentiable manifold
::differentiable
†differentiable map
% square torus in space
::Euler number
†invariance of domain

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 1.2, 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 1.3, 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 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.

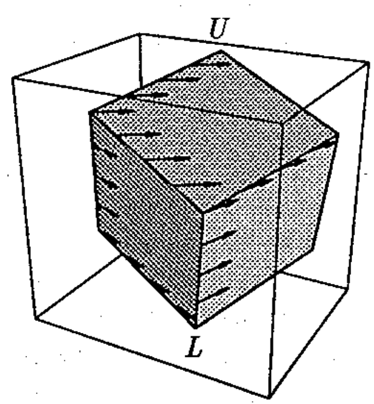
One way to show the invariance of the Euler number would be to prove that these two operations and their inverses are enough to go between any two finite cell divisions. But this approach is not very satisfactory—one can easily

% convexcharges
differentiable surface
::differentiable triangulation
†embedding
†affine map
% triangulation

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, clockwise as viewed from above (figure 1.16).



convexcharges

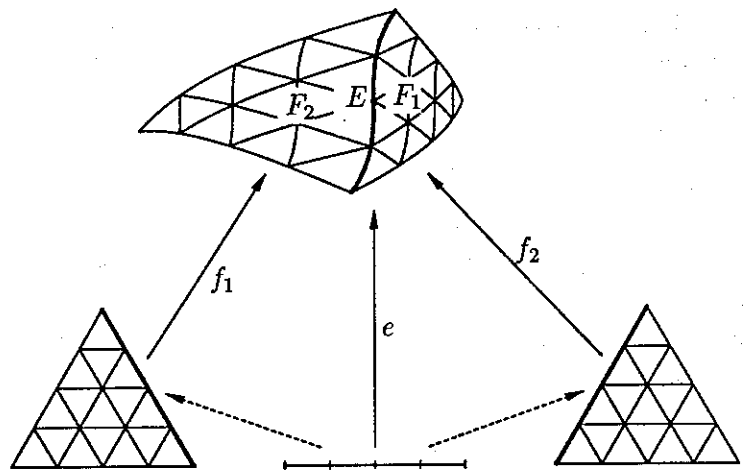
Figure 1.16. Charges on a convex polyhedron. The arrows are part of a horizontal vector field sweeping the surface of this polyhedron; the field is undefined only at the uppermost and lowermost vertices. When the + charges on vertices and the − charges on edges move according to the vector field, they cancel the + charges on the faces.

Silvio: add labels U and L

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 1.17).

% homeomorphisms of an interval
Proposition "nonvanishing vector fields"
::transverse
% transverse triangulation



triangulation

Figure 1.17. 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.

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 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 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

Problem "transverse triangulation"
% nonvanishing vector fields
% triangulation
::star
†similarity
†barycentric subdivision
% torusmetric
% isolated zeros
% nonvanishing vector fields
% transverse triangulation
::index of X at z
$i(X,z)$
Lemma "index independence"
% triangulating an annulus

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 1.17, 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 1.5. 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.

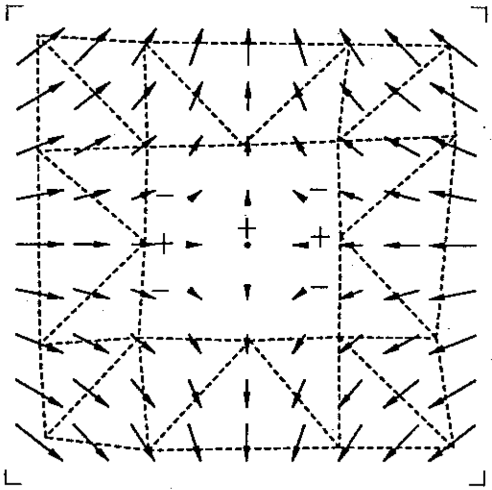
Let X be a vector field on a surface with an isolated zero at a point z. Working as in the proof of problem 1.3.4, construct a small polygon containing z whose edges are transverse to X. Place a $+$ charge on each vertex, a $-$ charge on each edge, and a $+$ charge in the interior of the polygon, and flow the charges off the boundary of the polygon by using X. The *index of X at z*, denoted $i(X,z)$, is the sum of the charges in the interior of the polygon after the operation of the flow.

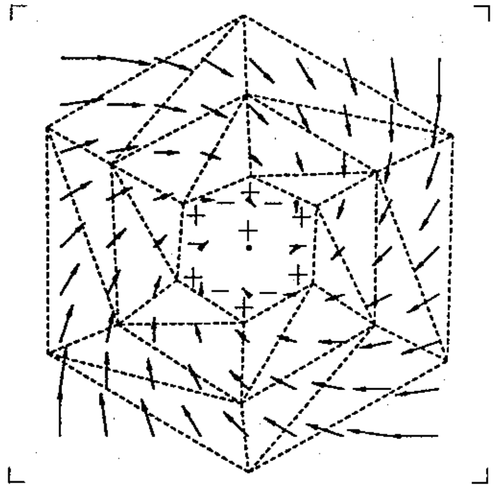
Lemma 1.3.5 (index independence). *If X is a vector field with an isolated zero z, the index of X is independent of the polygon enclosing z; it depends only on the restriction of X to an arbitrarily small neighborhood of z.*

Proof of 1.3.5: Given one polygon containing z with edges transverse to X, we need only show that it gives the same index as another much smaller polygon with the same properties. Using exercise 1.3.6, we subdivide the annulus

Exercise "triangulating an annulus"

::star-shaped





index

Figure 1.18. The index of a vector field. An isolated zero of a vector field leaves a small region where charges don't cancel. For the field on the left, the total charge of the polygon containing the zero—known as the index of the vector field at that point—is -1. Such a zero is called a *saddle*. For the field on the right, the index is 1; an isolated zero with index 1 is called a *sink* or a *source*, depending on whether the field points in or out.

between the two polygons into triangles, and jiggle the vertices to make the edges transverse to X. The flow causes some charge to enter the annulus across the outer boundary and some charge to leave across the inner boundary. The charge left on each triangle of the annulus is zero, and so the charge entering is equal to the charge leaving. 1.3.5

Exercise 1.3.6 (triangulating an annulus). A minor technical detail was suppressed above—how do we triangulate an annulus in the plane? Here we set things up so that the triangulation is easy—with the necessary background, one could instead just quote a theorem about the triangulability of surfaces.

(a) A disk D is said to be *star-shaped* with respect to a point v in its interior if each ray from v to the boundary of D is contained in the interior of D. Show that a star-shaped polygon can be triangulated with v as a vertex.

(b) Given two polygons which are star-shaped with respect to v, triangulate the annulus between their boundaries.

(c) Suppose that we have a polygon which is star-shaped with respect to the isolated zero z of X. Show that the boundary of the polygon can be made transverse to X by jiggling vertices only in the radial direction, and hence that the polygon obtained after jiggling is still star-shaped.

(d) (Harder.) Can you work out a proof without the assumption that the polygons are star-shaped?

The simplest vector fields with isolated zeros are the linear vector fields in the plane, those where the value of the field at a point is obtained by applying

†homotoped
†homotopic
Exercise "index is sign of determinant"
Exercise "isolated zeros"
% index
Proposition "Poincaré index theorem"
% nonvanishing vector fields
% index independence
% triangulating an annulus
% transverse triangulation
% isolated zeros

a linear map to the point. Clearly the origin is a zero of any linear vector field; it is isolated if and only if the linear map has non-zero determinant.

If a vector field with isolated zeros is †homotoped in such a way that the points where it is zero do not change, the indices at the zeros must remain constant since they are integers. This implies that two linear vector fields whose determinants have the same sign must have the same index, since then they are †homotopic through linear vector fields of non-zero determinant.

Exercise 1.3.7 (index is sign of determinant). (a) Sketch enough pictures of linear vector fields that you understand the relationship between determinant and qualitative appearance.

(b) Prove that the index of a linear vector field in the plane is the sign of its determinant.

Exercise 1.3.8 (isolated zeros). Given a finite cell division of a surface, find a way to construct a differentiable vector field on the surface with a source in the middle of each two-cell, a sink at each zero-cell, and a saddle in the middle of each edge (see figure 1.18 for definitions).

Problem 1.3.9. What are the possible values of the $i(X,z)$ when X is a general vector field? Can you find a formula for $i(X,z)$ valid for all vector fields with isolated zeros?

Proposition 1.3.10 (Poincaré index theorem). *If S is a smooth surface and X is a vector field on S with isolated zeros, the Euler number of S is*

$$\chi(S) = \sum_{z \in \text{zeros}} i(X,z).$$

Consequently, the Euler number of a surface does not depend on the cell division used to compute it—it is a topological invariant.

Proof of 1.3.10: Given a finite cell division of S, start by replacing it with a differentiable triangulation, as discussed just before proposition 1.3.3. Subdivide and jiggle the triangulation as necessary to make all the zeros lie inside faces, no more than one to a face. Enclose each zero with a polygon contained in a face and transverse to the field, as explained in the paragraph preceding lemma 1.3.5. Triangulate the annulus formed by taking away the polygon from the face (exercise 1.3.6). Finally, make the rest of the triangulation transverse, again by using the technique in the proof of problem 1.3.4.

Each polygon's contribution to the Euler number is the index of the vector field at the corresponding zero. Each triangle's contribution outside the polygons is zero. This proves the formula.

The last sentence follows because every closed surface admits a vector field with isolated zeros (exercise 1.3.8). 1.3.10

Challenge 1.3.11. Show that this discussion abut the Euler number and the index of isolated zeros can be carried out for a smooth manifold of any dimension.

Problem "classification of surfaces"
two-sphere
% 3gluings
projective plane
torus
Klein bottle
::elementary move
::genus
g-holed torus
::non-orientable genus

If the topology of a closed surface determines its Euler number, could the converse be true as well? This sounds unlikely—so much information in a single number! But amazingly, it's almost true: knowing the Euler number *and* whether or not the surface is orientable is enough to determine the surface. This important result has been known for more than a century; if you are not familiar with it, you are urged to work through the following steps:

Bill: are you sure? Is there an original reference?

Problem 1.3.12 (classification of surfaces). Consider a (connected) surface S obtained by gluing polygons (we allow digons, but not monogons).

(a) S can be obtained by gluing a single polygon.

(b) If the number of vertices in S is greater than one, try to reduce the number by shrinking one of the edges. This is always possible unless the polygon is a digon and the two vertices correspond to distinct points on S, in which case $S = S^2$. Assume from now on that S has one vertex.

(c) Let E be the number of edges of S, so the polygon has $2E$ edges and $\chi(S) = 2 - E$. If $E = 1$, we have the projective plane $\mathbf{RP}^2$, and if $E = 2$, we have the torus or Klein bottle (figure 1.15). Assume from now on that $E \geq 3$.

(d) An *elementary move* in this context consists of cutting a polygon along a diagonal which separates paired edges e_1 and e_2, and then gluing e_1 to e_2. Show that an elementary move does not change the topology of S.

(e) If S is orientable, there exist two pairs of paired edges x, x' and y, y' that separate each other.

(f) Suppose that x, x' and y, y' are pairs of paired edges that separate each other, and that they are paired with reversal of orientation. The remaining edges of the polygon form segments of length m, n, p and q. Using one elementary move, arrange that $m = n = 0$. Follow this with another elementary move, to arrange that $m = n = p = 0$.

(g) If S is orientable, it can be obtained by a gluing of the form

$$a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \dots a_g b_g a_g^{-1} b_g^{-1}.$$

(In this notation, we choose for each edge an orientation that is consistent with the identification, and read the resulting word going once around the polygon.) Therefore, a closed, orientable surface is determined up to homeomorphism by its Euler number, which is any even integer ≤ 2. The number g is called the *genus* of the surface, and the surface is a g-holed torus.

(h) If S is non-orientable, it can be rearranged by one elementary move in such a way that its gluing has two adjacent edges that are paired by a gluing map that preserves orientation.

(i) The non-orientable surface obtained from a hexagon by the gluing $aabbcc$ is homeomorphic to that obtained by the gluing $aabcb^{-1}c^{-1}$.

(j) Any non-orientable surface can be obtained by a gluing of the form

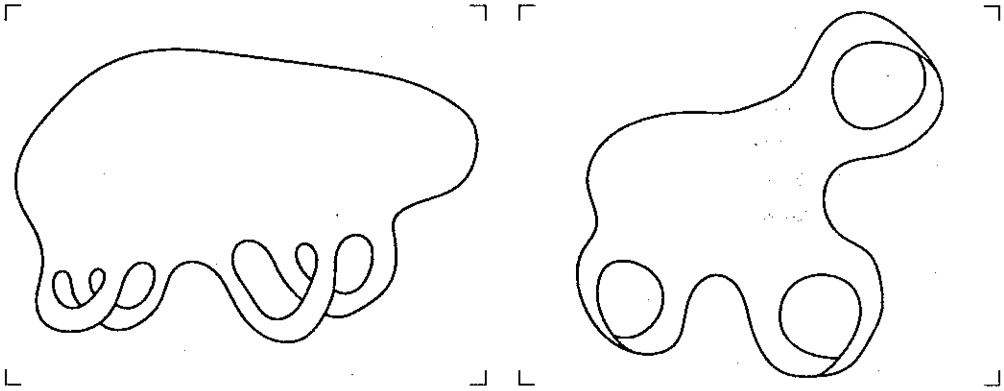
$$a_1 a_1 a_2 a_2 \dots a_g a_g.$$

The number g is called the *non-orientable genus* of the surface. Two closed non-orientable surfaces of the same Euler number are homeomorphic.

Exercise "closed surface minus a disk"
% strips
% Hyperbolic surfaces
% classification of surfaces
% 3octagons
% 3gluings
::elliptic
Problem "Gauss–Bonnet signs"

Exercise 1.3.13. How many gluing patterns are there for a $2n$-gon? How many lead to an oriented surface? How many topological types can be obtained? This shows there is a huge amount of repetition in descriptions of surfaces by gluings.

Exercise 1.3.14 (closed surface minus a disk). Let S be a closed connected surface. Prove that by removing a disk from S we get a disk with attached strips as in figure 1.19. Can you always make all the strips untwisted? All but one?



strips

Figure 1.19. Disks with strips. The result of removing a disk from a closed surface is a disk with strips attached. On the left, we started from an oriented surface of genus two; in the middle, from a surface of non-orientable genus three; on the right, from a projective plane. The figure on the right is the Möbius strip.

In section 1.2 we constructed a hyperbolic structure for the oriented genus-two surface. We can make an analogous construction for any surface of negative Euler number, with the help of problem 1.3.12. In fact, such a surface is obtained by gluing the sides of a $2n$-gon, with $n \geq 3$, in such a way that all vertices are identified to a single vertex. The angle sum of the Euclidean polygon is greater than 2π, so there is a symmetric $2n$-gon of the appropriate size in $\mathbf{H}^2$, as in figure 1.13, whose sides glue up to form a smooth hyperbolic structure on the surface.

Similarly, if a surface has Euler number zero, it can be obtained by gluing sides of a square so that all vertices are identified together. Therefore, it has a Euclidean structure. In the notation of figure 1.15, the surface is either $a-\ b-$ or $a-\ b+$, that is, a torus or a Klein bottle.

To complete the picture, S^2 and its quotient space $\mathbf{RP}^2 = S^2/\{\pm 1\}$—the surfaces of positive Euler number—have spherical structures. Such a structure is also called *elliptic*.

Might there be some exceedingly clever construction for a hyperbolic structure on the torus or Klein bottle, or a Euclidean structure on the surface of genus two?

Problem 1.3.15 (Gauss–Bonnet signs). (a) Show that the Euler number of a closed surface with a Euclidean structure is always zero.

::connected sum
$M_1 \# M_2$
†frame
% disk embeddings and the frame bundle
†frame
†frame bundle
Exercise "disk embeddings and the frame bundle"

(b) Show that the sum of the (interior) angles of any hyperbolic triangle is always less than π, and the sum of the angles of any spherical triangle is always greater than π. (Hint: use isometries to arrange triangles, to make the comparisons easier.)

(c) Show that the Euler number of a closed surface with a hyperbolic structure is negative, and the Euler number of a closed surface with an elliptic structure is positive.

This is a good place to mention some operations that construct new surfaces from old, and whose higher-dimensional analogues will be important later. We may as well give their definitions in arbitrary dimension.

The *connected sum* of two (connected) n-manifolds M_1 and M_2 is a manifold $M_1 \# M_2$ obtained by removing copies of the n-disk D^n from M_1 and M_2 and gluing the two resulting boundary spheres together.

Exercise 1.3.16. Show that if $S_3 = S_1 \# S_2$ are surfaces, $\chi(S_3) = \chi(S_1)+\chi(S_2)-2$. What happens for manifolds of other dimensions?

This definition can be made more precise: Choose diffeomorphic embeddings $\phi_1 : \bar{D}^n \to M_1$ and $\phi_2 : \bar{D}^n \to M_2$ of the closed n-disk, remove the two images of D^n from the union $M_1 \cup M_2$, and identify the boundaries $\phi_1(\partial D^n)$ and $\phi_2(\partial D^n)$ by the map $\phi_2 \circ \phi_1^{-1}$. To what extent does the topology of the result depend on the choice of ϕ_1 and ϕ_2? Not much, because there is essentially only one way to embed an disk in a connected manifold, up to orientation.

More precisely, if we change ϕ_1 (say) by an isotopy, the topology doesn't change, because any isotopy between embeddings of $\bar{D}^n$ into an n-manifold M can be extended to an isotopy of the whole manifold (see [Hir76, p. 185]). Now associate with an embedding $\phi : D^n \to M$ the †frame at $\phi(0)$ given by the image under $D\phi(0)$ of the canonical basis vectors of $\mathbf{R}^n$. It is easy to see (exercise 1.3.17) that two embeddings are isotopic if and only if their associated †frames can be continuously deformed into one another (that is, they lie in the same connected component of the †frame bundle of M). This means there are two isotopy classes of diffeomorphic embeddings $\bar{D}^n \to M$ if M is orientable, and only one if M is non-orientable. If an orientation is chosen for M, the two classes are determined by whether the embedding preserves or reverses orientation.

Exercise 1.3.17 (disk embeddings and the frame bundle). (a) Show that two embeddings of D^n that map the origin to the same point and have the same derivative there are isotopic.

(b) Show that they're isotopic if they map the origin to the same point and the frames they define at that point lie in the same connected component of $\mathrm{GL}(n,)$.

(c) Show that they're still isotopic if they map the origin to different points, but the frames they define can be continuously deformed into one another.

If the two manifolds are oriented, it makes sense to form the connected sum so that the resulting manifold has an orientation which agrees with the original

% surface semigroup
Exercise "surface semigroup"
†semigroup
Exercise "blowing up a point"
% blowup
::blowing up
Möbius strip
Exercise "autosum"
::autosum

orientations away from the disks which are removed. This condition requires that one of ϕ_1 and ϕ_2, but not both, the orientation-preserving. With this convention, the connected sum of two oriented two-manifolds is a well-defined oriented manifold.

If one of the manifolds is non-orientable, the result of the connected sum does not depend on the choice of orientation for the gluing map, and the result of the operation is again well-defined.

However, when the two manifolds are orientable but not oriented, there is a difficulty: the two possible choices of sign for the gluing map may yield different results. (But exercise 1.3.18 shows that this does not actually happen for surfaces.)

Exercise 1.3.18 (surface semigroup). (a) Show that every closed orientable surface admits an orientation-reversing homeomorphism.

(b) Show that the operation # is a well-defined, commutative, and associative operation on the set of homeomorphism classes of surfaces, making it into a commutative †semigroup.

(c) Show that S^2 acts as an identity element for #.

(d) Show that the torus T^2 and the projective plane $\mathbf{RP}^2$ generate the commutative semigroup of homeomorphism classes of surfaces under #, and that $T^2 \# \mathbf{RP}^2 = \mathbf{RP}^2 \# \mathbf{RP}^2 \# \mathbf{RP}^2$.

(e) Sketch a directed labeled graph whose vertices are in one-to-one correspondence with homeomorphism classes of surfaces, and whose edges show the effect of # with T^2 (if labeled A) and with $\mathbf{RP}^2$ (if labeled B).

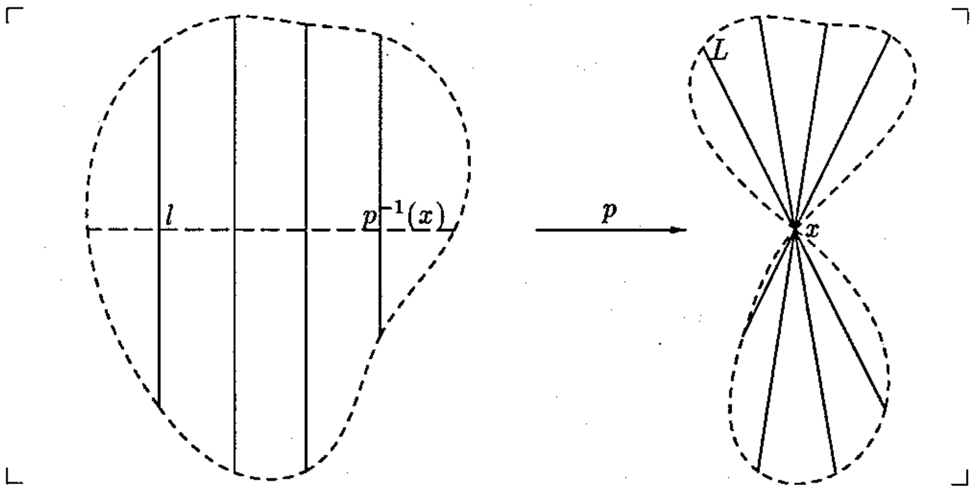
Exercise 1.3.19 (blowing up a point). (a) Let M be a smooth n-dimensional manifold and x a point of M. The operation of replacing x by the set of tangent lines at x, giving the result the topology described in figure 1.20, is called *blowing up* the point x. Show that the resulting topological space is a smooth manifold homeomorphic to $M \# \mathbf{RP}^n$.

(b) If $n = 2$ this amounts to cutting out a disk and gluing in a Möbius strip.

(c) What happens if we do our "blowing up" by using tangent rays instead of tangent lines?

Exercise 1.3.20 (autosum). The connected sum operation has a variation where one removes two disjoint disks on the same manifold, and connects the resulting boundary spheres together. There are two possibilities for the orientation with which the spheres are identified. Analyze what happens to surfaces under this operation of *autosum*.

Exercise 1.3.21. Show that the semigroup of (connected) closed surfaces can be generated from the identity element S^2 by the operations of blowing up and autosum.



blowup

Figure 1.20. Blowing up a point. The blowup of M at x (left) has points of two types: points of $M \setminus \{x\}$, and one-dimensional subspaces of the tangent space $T_x M$. There is a natural map p from the blowup into M, taking each point of the first type to itself and each point of the second type to x. The topology of the blowup is defined by the following conditions: p is a local homeomorphism away from $p^{-1}(x)$; and a neighborhood of a point $l \in p^{-1}(x)$, corresponding to a line L, consists of points corresponding to lines near L, plus points of the first type along those lines and close to x.

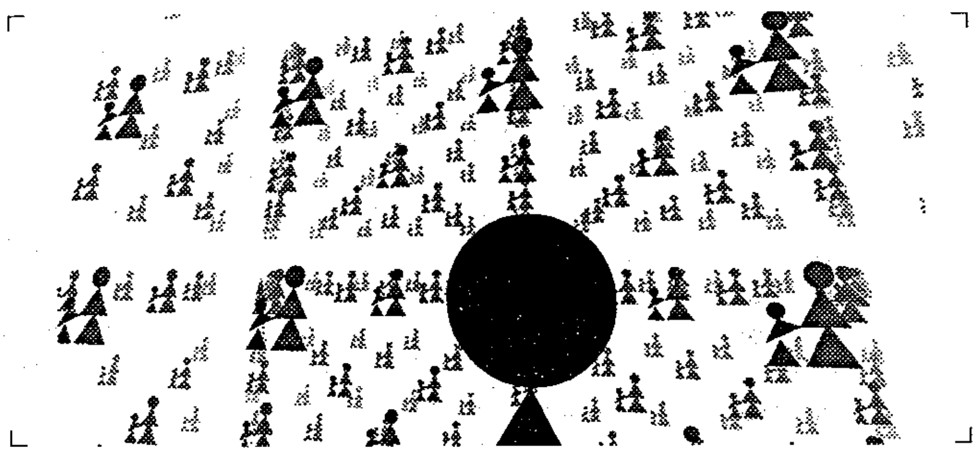
Section "Some three-manifolds"
Example "the three-torus"
% inside3torus
Example "the three-sphere from inside"

1.4. Some three-manifolds

Now that we are well grounded on surfaces, we are ready for a quick flight through a few three-manifolds.

Example 1.4.1 (the three-torus). Probably the easiest three-manifold to understand is the three-torus, which can be obtained by gluing just like the two-torus: start with a cube, and glue each face of the cube to the parallel face, by parallel translation.

To visualize the three-torus, imagine the cube as a rectangular room where you are standing. Imagine what it would be like if the opposite walls were identified with each other, and the floor were identified with the ceiling. When your line of sight arrives at one wall, it continues from the corresponding point on the opposite wall, in the same direction as before. Therefore, if you look straight ahead, you see your back. If you turn to the left, you see your right side. If you look straight down, you see the top of your head. There are six images of yourself which appear to be in immediately neighboring rooms, but there are also rooms which neighbor diagonally, and, really, the lines of sight would continue indefinitely. The appearance would be identical to that of an infinite repeating array of images of yourself (and anybody or anything else in the room) in all three dimensions (figure 1.21). The effect is reminiscent



inside3torus

Figure 1.21. A view from inside the three-torus. If you live in a three-torus, each object appears to be repeated at every point of a three-dimensional lattice.

of a barber shop with mirrors on facing walls, so that you see long lines of repeated images. The difference is that in a barber shop, when you look at an adjacent image, it is facing toward you, while in the torus, it is facing away. In a torus, as you turn toward any image, the image turns away. As you fly toward the image, the image flies away, and you never can meet it.

Example 1.4.2 (the three-sphere from inside). Another easily described three-manifold is the three-sphere S^3. The easiest definition of S^3 is the unit sphere $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ in $\mathbf{R}^4$. Unfortunately, this formula does not antipode

immediately communicate a picture of S^3 to people who are not adept at visualizing four-dimensional space. But there is another way to imagine S^3, from the point of view of an inhabitant.

To prepare the way, think first of what an inhabitant of S^2, the two-sphere, would see. By some mechanism, light rays are supposed to curve around to follow the surface. For instance, you can imagine that the "surface" is really a very thin layer of air between two large concentric glass spheres, which channel light by reflection in much the same way as fiber optics. (Unfortunately, the ecology of this model is not so clear. At best, there is just enough room for one to crawl around on one's stomach.)

Imagine creature A resting at the north pole, and another creature B creeping away. You can work out the visual images in terms of which geodesics (great circles) from the eyes of the A intersect B. As B creeps away, its image as seen by A at first grows smaller, although not quite as fast as it would in the plane. Once B reaches the equator, however, its image grows larger again with continued progress, until at the south pole, its image fills up the entire background of the field of vision of A in every direction.

The same phenomenon would take place in the three-sphere. Let's give ourselves more breathing room than we had in T^3, and imagine that we are in a three-spherical world where a great circle is about two miles in circumference. There is no gravity, and we won't fuss about food, shelter, light or other minor details just now. We have little jets on our backs for flying around wherever we please. If I fly away from you, in any direction, my visual image to you shrinks in size at first fairly rapidly, but as I approach the half-mile mark my visual size changes very slowly: it probably looks to you as though I have stopped making progress. After the half-mile mark, I gradually start to increase in visual size once more. As I approach your antipode, one mile from you, I start to grow rapidly again. When I am three feet from your antipode, the size of my visual image is exactly the same as if I were three feet from you. If I turn around and shout back, it will hurt your ears. We quickly learn that we can carry on a conversation with normal voices, for sound converges again at antipodal points just as light does.

Even though I have the same visual size to you when I hover three feet from your antipode as when I hover three feet from you, there is a difference in my visual image: you see further around to my sides. (There is also a difference in focal distance, but let's put that aside: imagine the light is very bright, so that your pupils are contracted and you don't notice this effect.) The difference becomes very dramatic if I now continue three feet further, so I cover the antipode of your eyes: you now see my image in every direction, and it is as if I were turned inside out onto the inside surface of a great hollow sphere totally surrounding you. You appear to me the same way, as the inside of a hollow sphere surrounding me.

In this description, we have left out an important part of the image. Light does not stop after traveling only a mile, it continues further. When I am a half mile from you, my image to you is as small as possible, but your lines of

Example "elliptic space"
::elliptic space
Example "Poincaré dodecahedral space"
dodecahedron
% poincare12

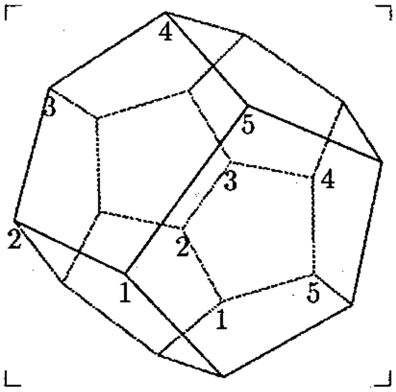
sight continue unimpeded completely around the three-sphere, to arrive back near where they started on yourself. In the background of everything else, you see an image of yourself, turned inside out on a great hollow sphere, with the back of your head in front of you.

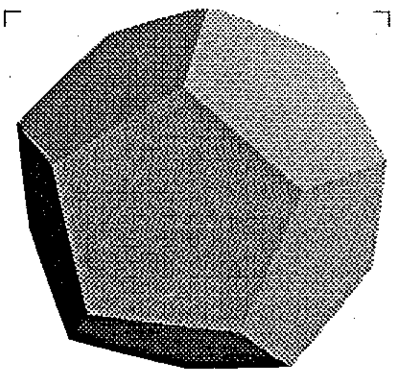
There's another thing we left out: whenever I am at a distance other than one mile from you, you can actually see me in two opposite directions. For instance, when I was three feet from your antipode, had you turned around rather than me, you would have seen a perfectly normal image of me as if I were hovering three feet away and facing you, only slightly faded by the blue haze of the water vapor in the intervening air. You would also appear almost completely normal to me. But if we were to try to shake hands, they would pass through each other.

Example 1.4.3 (elliptic space). Accounts of spherical geometry are marred by exceptions arising from the existence of antipodes—for instance, two points determine a unique line, but only if they're not antipodal. Things work out more cleanly in *elliptic space*, the sphere with antipodal points identified. Topologically, n-dimensional elliptic space is just the projective n-space $\mathbf{RP}^n$, but as the quotient of the sphere by a group of isometries (a very small group, just the identity and the antipodal map), it inherits a geometry, just as the torus inherits a geometry from the Euclidean plane. Geometric assertions can easily be translated back and forth between the sphere and elliptic space, and they're often cleaner in elliptic space: for example, any two points in elliptic space determine a unique line.

We've seen that in the sphere an object moving away from you decreases in apparent size until it reaches a distance of $\pi/2$, then starts increasing again until, when it reaches a distance of π, it appears so large that it seems to surround you entirely. In elliptic space, on the other hand, the maximum distance is $\pi/2$, so that apparent size is a monotone decreasing function of distance. It would nonetheless be distressing to live in elliptic space, since the entire background of your field of view would be filled up with an image of yourself. Looking straight ahead you would see the back of your head, turned upside down and greatly magnified. Everything else would still be visible twice, in opposite directions.

Example 1.4.4 (Poincaré dodecahedral space). This famous example, discovered by Poincaré, is obtained from a dodecahedron by gluing opposite faces. The corners of the pentagons making up opposite faces of a dodecahedron are out of phase: they interleave each other, so there is no question of gluing each face to its opposite using a translation, as in the torus. Consider what happens when we glue them with as little twist as possible: a rotation by 1/10 of a full turn, say in the clockwise direction from front to back, as in a right-handed screw (figure 1.22, left). This prescription is symmetric, so that when we turn the dodecahedron around 180°, the identification appears the same: right-handed screws are right-handed screws in either direction.





poincare12

Figure 1.22. The Poincaré dodecahedral space. Each pentagonal face of a dodecahedron has an opposite face, but the corners of opposite pentagons interleave each other. If a face is glued to its opposite face by one-tenth of a clockwise revolution, the resulting space is the Poincaré dodecahedral space. The edges are glued in triples in this pattern. In order for the gluing to work geometrically, we must start with a spherical dodecahedron, with dihedral angles equal to 120°. This slightly puffy solid is shown on the right under stereographic projection (see exercise 2.2.8)—it is almost indistinguishable from its Euclidean counterpart, which has dihedral angles 116.565°!

A dodecahedron has five edges around each face and each edge is along two faces, so it has $12 \cdot 5/2 = 30$ edges. The face pairings force various identifications of the edges which you can chase through, to check that the edges are glued together in ten groups of three.

To see what happens to the $12 \cdot 5/3 = 20$ vertices of the dodecahedron, consider the spherical triangles obtained by intersecting tiny spheres about the vertices with the dodecahedron. The way these spherical triangles are glued together is such that three triangles come together at each vertex, so the pattern is necessarily that of a tetrahedron. Therefore, the vertices of the dodecahedron are glued together in five groups of four, and the space obtained by this gluing is a manifold since it is locally homeomorphic to Euclidean space, even in neighborhoods of the vertices and neighborhoods of points on the edges.

Can this gluing be done so that the geometry, and not just the topology, is locally Euclidean? If so, the angles between the faces around an edge of the glued-up manifold would add up to 2π, so (assuming everything is symmetric) they would each equal $2\pi/3$ or 120°. But the dihedral angles of a dodecahedron are slightly less—they equal $\arctan -2$, or 116.565°.

Here the geometry of the three-sphere comes to the rescue. Just as the angles of a geodesic triangle on the two-sphere has angles somewhat larger than the angles of a Euclidean triangle, so the dihedral angles of a polyhedron in the three-sphere are somewhat larger than those of a Euclidean polyhedron. A very small regular dodecahedron in the three-sphere has angles very close

% the three-sphere from inside
% poincare12
% dodecahedral tiling
Example "Seifert–Weber dodecahedral space"

% seifertweber

to the Euclidean angles. As the dodecahedron grows, the angles increase. In fact, when the distance from the center to the vertices is $\pi/2$ (or a half-mile, in the scale of example 1.4.2), the dodecahedron is geometrically a two-sphere: the angles are π.

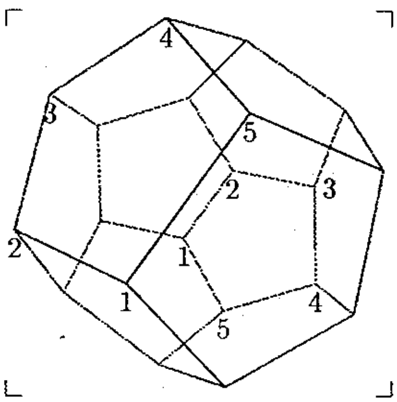
Therefore, somewhere in between there is a dodecahedron with angles exactly $2\pi/3$ radians or 120° (figure 1.22, right). For this dodecahedron, the gluing will work geometrically.

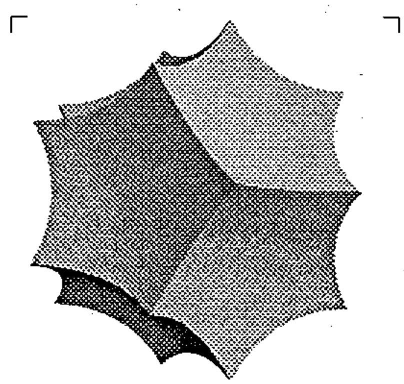
We can unroll the Poincaré dodecahedral space to obtain a tiling of S^3, just as we unrolled T^2 and T^3 to obtain tilings of $\mathbf{E}^2$ and of $\mathbf{E}^3$. This is the same as saying that the Poincaré dodecahedral space, like elliptic space, is a quotient of S^3 by a group of isometries. It would be possible to mark out the tiles exactly using coordinates on $S^3 \subset \mathbf{E}^4$, and it would also be possible to work out the combinatorial structure of the tiling by logic, since we know its local structure. Either approach would involve a fair amount of work at present, but we will return to this question later, in ??. It turns out that it takes 120 dodecahedra to tile the three-sphere. In $\mathbf{E}^4$, this pattern defines a 120-hedron whose faces are regular three-dimensional dodecahedra.

Is this reference going to be there Silvio

Ideas to include: Euler number of various manifolds why using Euler number doesn't g the number of dodecahedra in t as it does in 2D; piecing together neighborhood of vertex; computat of 60° dihedral a in virtual dodecahedron

Example 1.4.5 (Seifert–Weber dodecahedral space). If the opposite faces of a dodecahedron are glued together using clockwise twists by three-tenths of a revolution, instead of one-tenth (figure 1.23), a bit of chasing around the diagram shows that edges are identified in six groups of five. The small spherical triangles around the vertices of the dodecahedron obtained by intersection of the dodecahedron with tiny spheres glue together in such a way that five come together at a vertex. The pattern is that of a regular icosahedron. All twenty vertices are glued together, and the space which results is a manifold known as the Seifert–Weber dodecahedral space.





seifertweber

Figure 1.23. The Seifert–Weber dodecahedral space. If opposite faces of a dodecahedron are glued by three-tenths of a clockwise revolution, the edges are glued in quintuples, and the resulting space is the Seifert–Weber dodecahedral space. The gluing can be realized geometric if we use a hyperbolic dodecahedron with dihedral angles of 72°—the solid shown on the right, in the Poincaré ball model.

::Poincaré ball model
::ideal dodecahedron
::ideal vertices
Example "lens spaces"
::lens space
% lens

The angles of a Euclidean dodecahedron are much larger than the 72° angles needed to do the gluing geometrically. In this case, we can use three-dimensional hyperbolic space $\mathbf{H}^3$, which can be mapped into the interior of a three-dimensional ball, just as in two dimensions. This description in the ball is the *Poincaré ball model* of hyperbolic space. Planes are represented as sectors of spheres orthogonal to the boundary of the ball. The angle between two hyperbolic planes is the same as the angle between the two spheres. Since the spheres intersect the boundary of the ball S^2_∞ perpendicularly, it is also the angle between their circles of intersection with this sphere.

The angles of a regular dodecahedron in $\mathbf{H}^3$ are clearly smaller than they would be in $\mathbf{E}^3$, but can they be as small as 72°? In the limiting case, as a hyperbolic dodecahedron became very large, its vertices would appear to nearly touch S^2_∞. There is a limit, the *ideal dodecahedron*, whose vertices are missing: it has *ideal vertices* on S^2_∞.

By looking at the symmetric placement of three circles on S^2_∞ coming from three adjacent face planes of the ideal dodecahedron which meet at an ideal vertex, it is easy to see that an ideal dodecahedron has 60° dihedral angles.

Therefore, intermediate between a very small hyperbolic dodecahedron with angles approximately 116.565° and a very large dodecahedron with angles tending toward 60° there is a dodecahedron whose dihedral angles are exactly 72°. This dodecahedron can be glued together to make a geometric form of the Seifert–Weber dodecahedral space. It corresponds to a tiling of $\mathbf{H}^3$ by dodecahedra all meeting five to an edge and twenty to a vertex.

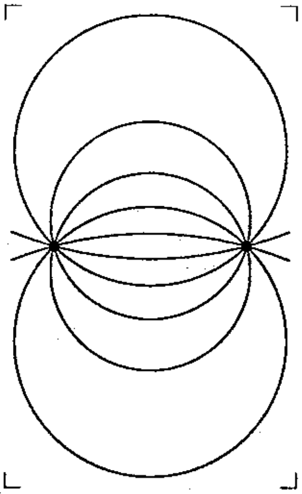
Example 1.4.6 (lens spaces). Consider a ball with its surface divided into two hemispheres along the equator. What happens when we glue one hemispherical surface to the other?

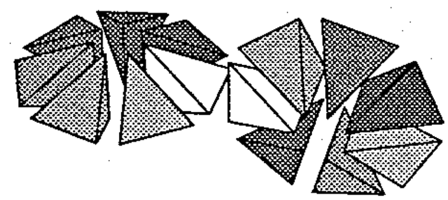
If we glue with no twist at all, so that the identification is the identity on the equator, the resulting manifold is S^3. This is analogous to the way S^2 can be formed by dividing the boundary of a disk into two intervals, and gluing one to the other so as to match each endpoint with itself.

On the other hand, if the hemispheres are glued with a q/p clockwise revolution, where q and p are relatively prime integers, each point along the equator is identified to $p-1$ other points. A neighborhood of such a point in the resulting identification space is like p wedges of cheese stuck together to form a whole cheese, so the identification space is a manifold, called a *lens space* $L_{p,q}$.

To form a geometric model for a lens space, we need a solid something like a lens, where the angle between the upper surface and the lower surface is $2\pi/p$. This is easy to do within S^3. Any great circle in S^3 has a whole family of great two-spheres passing through it. From this family it is easy to choose two that meet at the desired angle $2\pi/p$. Now when the two faces are glued together, neighborhoods of p points on the rim of the lens which are identified fit exactly. This corresponds to a tiling of S^3 by p lenses: see figure 1.24, left.

Problem "reworking lens spaces"
% lens
Example "a knottier example"
% fig8glue
three-dimensional gluing patterns





lens

Figure 1.24. Lens spaces. On the left, S^3 is seen in cross section, tiled with twelve copies of $L_{12,q}$ (we're using stereographic projection: see exercise 2.2.8.) On the right, $L_{7,2}$ is disassembled and reassembled in a different way, showing that it equals $L_{7,3}$.

Problem 1.4.7 (reworking lens spaces). (a) The lens that was glued to form $L_{p,q}$ can be cut up into p tetrahedra, meeting around one edge through its central axis. When this is done, the p tetrahedra can be assembled by gluing first the faces which came from the surface of the lens (figure 1.24, right). What figure does this form? What identities among lens spaces can you construct?

(b) Cut out a solid cylinder around the central axis of the lens used to form $L_{p,q}$. Its upper face is glued to its lower face to form a solid torus, under the identifications. What happens to the rest of the lens when the part of its boundary on the surface of the lens is glued together? Sketch a picture for $L_{3,2}$. Describe how lens spaces can be constructed by gluing together two solid tori.

(c) Show that two lens spaces $L_{p,q}$ and $L_{p',q'}$ are homeomorphic if and only if $p = p'$ and $q' = \pm q \pmod p$ or $qq' = \pm 1 \pmod p$. (This is much harder; see [Bro60] for a proof.)

Example 1.4.8 (a knottier example). Figure 1.25 shows one of the simplest possible three-dimensional gluing patterns. Start with two tetrahedra T and T' with labeled faces and directed edges divided into two types, thick and thin. Then glue faces in pairs, respecting not only face labels but also edge types and directions: for example, face A of T and face A of T' are glued together so that thick edges match thick edges and one thin edge matches the other. In the resulting complex K, all thick edges end up identified, as do all thin edges. Furthermore, all vertices are identified together into one vertex V.

Is K a manifold? We need to check that each point in K has a neighborhood homeomorphic to a 3-ball. This is obvious for points coming from the interiors of the tetrahedra, and for those coming from the interior of a face, where two half-spaces meet. A point in the interior of an edge, too, has a

::link
::cone on a torus
% Identifying faces of polyhedra
figure-eight knot
% fig8diag
% fig8complex
% fig8complex

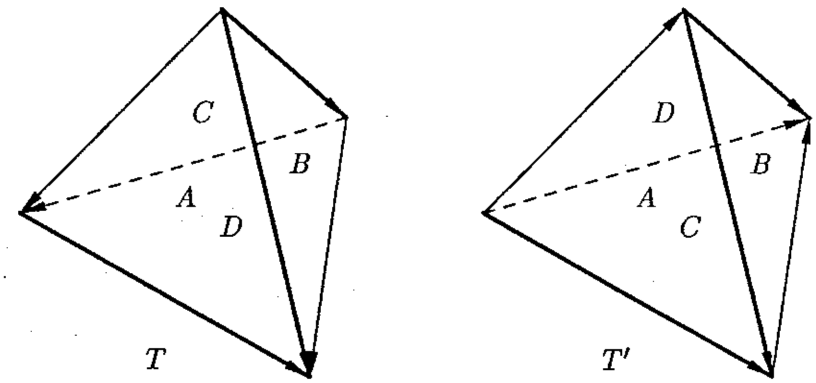


fig8glue

Figure 1.25. A simple three-dimensional gluing pattern. A simple pattern for gluing two tetrahedra. Each face has a label (centered on it), and faces with the same label are identified in a way that is unambiguously determined by the requirement that edge types (thick and thin) and directions match. (In fact, even the pairing of faces could be reconstructed from this requirement.)

neighborhood consisting of several wedge-neighborhoods of its preimages on the edges of T and T', glued together cyclically to make a ball.

But we run into trouble at the vertex V. Imagine a small neighborhood of V that intersects each tetrahedron in small tetrahedral neighborhoods of its vertices. The inside faces of these tetrahedra piece together to form a surface called the *link* of V. One can check explicitly that it is a torus. Alternatively, we can see by counting that the Euler number of the link is 0, and since it is oriented it must be a torus. The neighborhood of V in K is therefore a *cone on a torus*, and in particular V has no neighborhood homeomorphic to a ball and K is not a manifold. In ?? we will treat in more generality and detail the issue of when complexes formed by gluing polyhedra have manifold structures.

Although K is not a manifold, by removing the recalcitrant vertex we get a non-compact manifold $M = K - \{V\}$. Since the gluing map for each pair of faces reverses orientation, M is oriented. By removing an open neighborhood of V, we obtain a compact manifold whose boundary is a torus.

Exercise 1.4.9. Construct the link of V by gluing together the eight triangles which are links of the vertices of the two tetrahedra.

It turns out that M is homeomorphic with the complement (with respect to $S^3 = \mathbf{R}^3 \cup \{\infty\}$) of a figure-eight knot, shown in figure 1.26. To see this, we start by arranging the figure-eight knot along the one-skeleton of a tetrahedron, as in figure 1.27(a). We see that the knot can be spanned by a two-complex, with two edges (the arrows) and four two-cells (the faces of the tetrahedron, each together with a strip and a twisted tab). This complex encloses a compact region of space, whose interior R is homeomorphic to an open ball.

Now imagine the thin edge in figure 1.27(a) coated with red printing ink, the thick edge coated with green, and the knot coated with black. Insert a

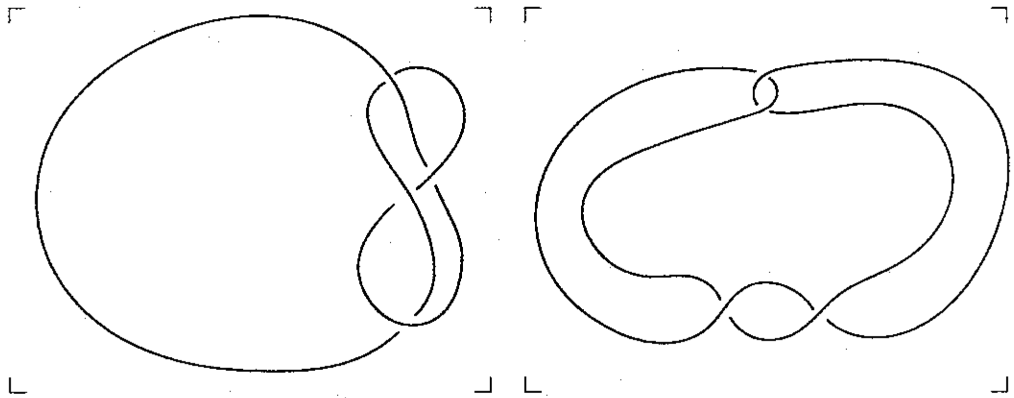


fig8diag

Figure 1.26. Two views of the figure-eight knot. Two views of the figure eight knot, the second most commonly occurring knot in extension cords, dog leashes and garden hoses (after the trefoil).

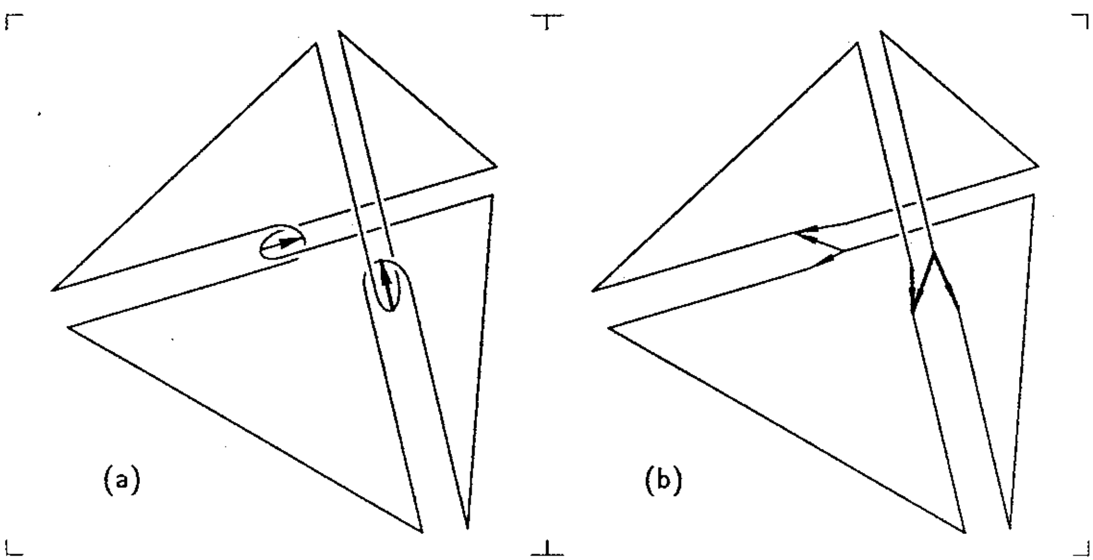


fig8complex

Figure 1.27. A two-complex spanning the figure-eight knot. Left: the figure-eight knot can be spanned by a complex with four two-cells—the faces of a tetrahedron, each augmented by a strip comming off of a vertex and a twisted tab comming off of the opposite edge—and two one-cells, shown as arrows. The region R inside this cell complex is homeomorphic to an open ball. Right: We extend the homeomorphism to a map from the closed ball to the closure of R. Each one-cell is represented by three edges on the boundary, and each two-cell is bounded by three edges and two pieces of the knot. Shrinking the knot pieces to points gives the tetrahedron T from figure 1.25.

balloon
% fig8complex
% fig8glue
% trefcomplex
trefoil knot
% fig8complex

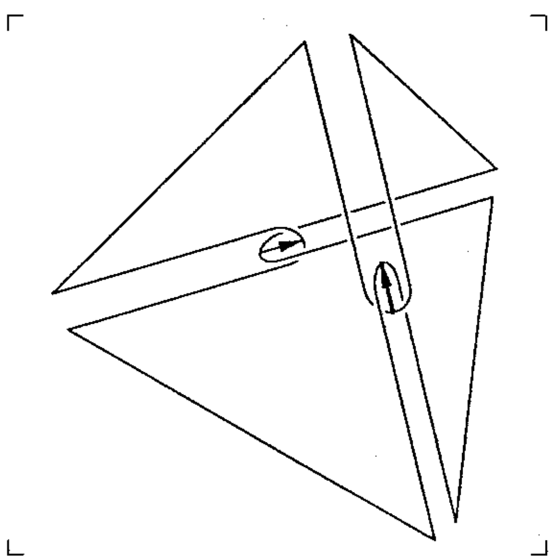
balloon into the region R and inflate it, until it fills up all the nooks and crannies of R. Then remove the balloon and examine it; it will be covered by four regions, corresponding to the four two-cells, and arranged as in figure 1.27(b).

The boundary of each region has five parts: two, shown in the figure without arrows, are colored black and come from the knot. The other three come from the edges of the two-complex: two thin and one thick, or two thick and one thin. Now the knot is not part of the manifold M, so we may expand or contract the black curves without changing M. If we choose to contract each of them to a point, so that the four regions become three-sided, the interior of the balloon, together with the triangular faces, becomes a tetrahedron T exactly as in figure 1.25.

T' is formed similarly, inflating a balloon that contains the point at infinity in $S^3 = \mathbf{R}^3 \cup \{\infty\}$.

We will return to this example later. For now, the difference between the two descriptions—complement of figure-eight knot and union of two tetrahedra—serves to illustrate the need for a systematic way to compare and to recognize manifolds.

Exercise 1.4.10. Suggestive pictures can also be deceptive. Figure 1.28 shows that a trefoil knot, too, can be arranged along the one-skeleton of a tetrahedron, and spanned by a two-complex similar to the one in figure 1.27. Blow balloons in the regions inside and outside this complex, and draw the imprints made by the knot and arrows. Do they give rise to tetrahedra?



trefcomplex

Figure 1.28. A two-complex spanning the trefoil knot. This arrangement for the trefoil knot, although apparently very similar to figure 1.27, does not lead to a decomposition of the complement of the knot into tetrahedra.

Chapter "Hyperbolic geometry and its friends"
Riemannian metric
::isometries
::homogeneity
::isotropy
::frames
sectional curvatures
Euclidean geometry
Spherical geometry
% the three-sphere from inside

Chapter 2

Hyperbolic geometry and its friends

"Geometry" can mean a number of different things in different contexts. In this chapter we will study geometry in the classical sense: our geometries will be analogous to standard Euclidean geometry, with concepts of straight lines or geodesics, angles and planes, a measure of distance, and a large degree of homogeneity.

One way to capture the structure is with a Riemannian metric. Any Riemannian manifold possesses a group of *isometries*, transformations of the space which preserve lines, angles and distances; however, for a typical Riemannian manifold, this group is the trivial group. Here, in contrast, we require *homogeneity* of the geometry, that is, transitivity of its group of symmetries: for any two points of the space there must be at least one isometry taking one to the other, so that each point "looks like" every other point.

In addition to homogeneity, the geometries we will consider in this chapter satisfy another property, *isotropy*. A space is isotropic if it looks the same no matter what position your head is in, or, more precisely, if for any two *frames* (ordered bases of orthonormal tangent vectors) at a point in the space, there is an isometry of the space fixing the point and taking one frame to the other.

Homogeneity and isotropy together are very strong conditions—in particular they implies that sectional curvatures are the same at every point and in every tangent two-plane. There are only three essentially distinct simply connected isotropic geometries in any dimension: with zero sectional curvature, constant positive sectional curvature K, or constant negative sectional curvature $-K$ (after scaling, K may be taken to be 1). They are called Euclidean, spherical and hyperbolic geometry, respectively.

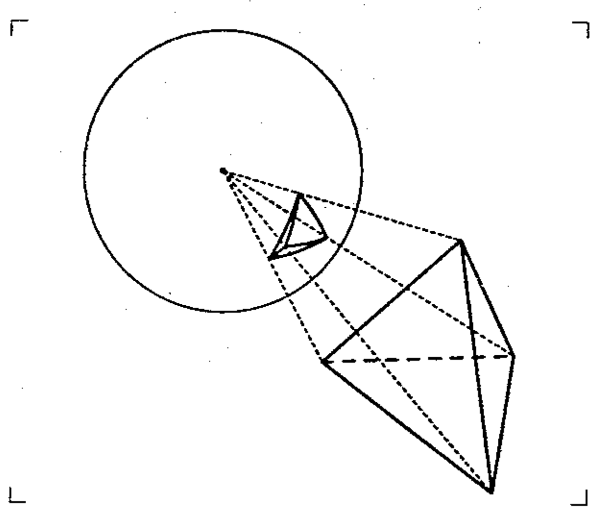
Euclidean geometry is familiar to all of us, since it very closely approximates the geometry of the space in which we live, at least up to fairly large distances. Spherical geometry is the standard geometry of the n-sphere—geodesics are great circles, angles and distances are inherited from $(n + 1)$-dimensional Euclidean space, and so on. We discussed the three-sphere briefly in example 1.4.2, and will return to it at the end of this chapter. Hyperbolic

Hyperbolic geometry
% Geometric Manifolds
::visual sphere
% visualsphere

geometry is the least familiar to most people. It is also the most interesting, and by far the most important for two- and three-dimensional topology. For this reason, we will spend the next several sections giving different pictures of hyperbolic geometry.

In chapter 3 we will extend our analysis to include five other three-dimensional geometries that have significance for three-dimensional topology. These other five geometries are homogeneous but not isotropic: they are the same at every point, but not with your head at any angle. In fact, something like a notion of up and down can be geometrically defined in each of these geometries, and for some of them, a notion of north and south as well.

Two-dimensional geometry can be easily visualized "from the outside", by sketching pictures on paper. In three dimensions and higher, the best insight is often gained by imagining yourself inside the space. To formalize this intuitive concept, we need the idea of a visual sphere. Think of an observer as a point somewhere in an n-dimensional space, with light rays approaching this point along geodesics. Each of these geodesics determines a tangent vector at the point, and the $(n-1)$-sphere of tangent vectors is called the *visual sphere* (see figure 2.1). Alternatively, one can think of a very small sphere centered at the observation point, with each geodesic determining an intersection point with it. An object is perceived as a (segment of a) straight line if its image on the visual sphere is (an arc of) a great circle. The apparent size of an object is determined by the amount of angle it takes up in the visual sphere.



visualsphere

Figure 2.1. The visual sphere

Section "Negatively curved surfaces in space"
differential geometry ::Gaussian curvature ::curvature
intrinsic geometry
extrinsic properties
positive curvature, zero curvature, and negative curvature
rubber ball
sheet of paper
saddle
measurement of the curvature
paraboloid of revolution
% quadposcurv
hyperbolic paraboloid
% quadnegcurv
†quadratic form
Gauss
Theorema Egregium
measurement of curvature

2.1. Negatively curved surfaces in space

Before developing abstract models for hyperbolic geometry, it pays to describe some constructions of a more physical nature. For this we need some basic concepts from the differential geometry of surfaces, which we'll introduce as required. We will not take the time here to develop this beautiful subject in detail, but you are encouraged to consult one of several readily available good sources, for example, [ONe66, Hic65, dC76].

The *Gaussian curvature*, or simply *curvature*, of a surface is a measure of its intrinsic geometry. We use the term "intrinsic" to denote those properties of a surface that are invariant when the surface is bent without being stretched: that is, they depend only on measurement of lengths of curves along the surface itself. By contrast, extrinsic properties depend on the embedding of the surface in space.

There are significant qualitative differences between surfaces with positive curvature, zero curvature, and negative curvature. Near the point of tangency, a surface of positive curvature lies to one side of any of its tangent planes. An example is a rubber ball. A surface of zero curvature has a line along which the surface agrees with its tangent plane. To illustrate this, try holding up a sheet of paper and bending it in different directions, and notice how you can find a straight line on the surface through any point. Surfaces of negative curvature cut through their tangent planes, as in a saddle.

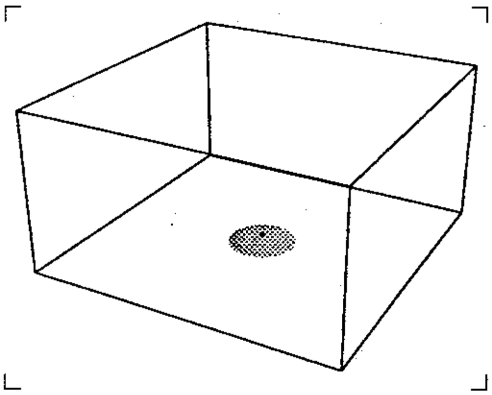
The precise measurement of the curvature depends on the behavior of the surface to second order. One way to perform it is to arrange our coordinate system so that the point under scrutiny on our surface is at the origin, and the tangent plane at that point is horizontal. This means the surface is locally the graph of a function $f(x,y)$ such that f and its first-order partial derivatives f_x and f_y are zero at the origin. The Gaussian curvature at the origin is the determinant of the hessian matrix

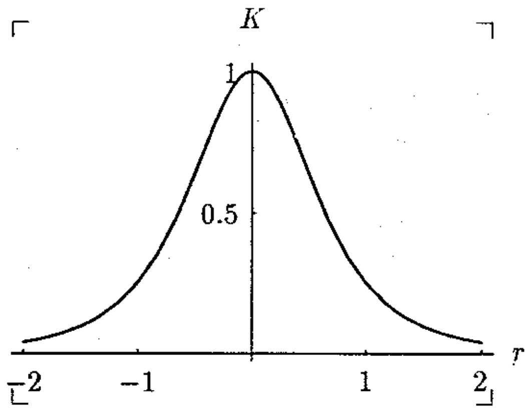
2.1.1. $$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}.$$

For example, the Gaussian curvature of the graph of the polynomial $f(x,y) = Ax^2 + 2Bxy + Cy^2$ at the origin is $K = 4(AC - B^2)$. For a paraboloid of revolution (figure 2.2) this number is positive, whereas for a hyperbolic paraboloid (figure 2.3) it is negative. Notice that here $f(x,y)$ coincides with the †quadratic form given by one-half its Hessian matrix, $f(x,y) = \frac{1}{2}(x,y)H(x,y)^t$. In general, $f(x,y)$ is approximated to second order by the same quadratic form.

The definition given above is based on extrinsic properties. A fundamental theorem, which Gauss called his "Theorema Egregium" or notable theorem, is that the Gaussian curvature is actually an intrinsic invariant of the surface: it does not change when the surface is bent without stretching. Here is an intrinsic way to arrive at the same number: draw a circle of radius r on the surface, centered at the point under scrutiny. If the curvature is positive there,

% quadposcurv
% quadnegcurv
constant positive curvature



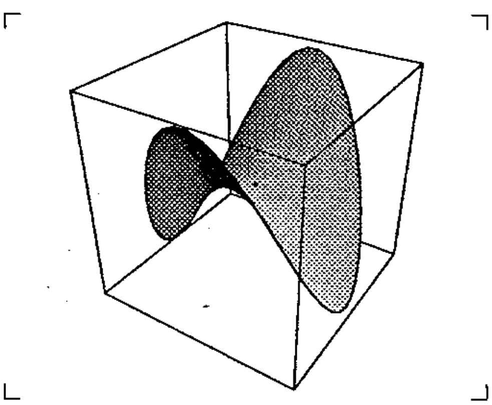


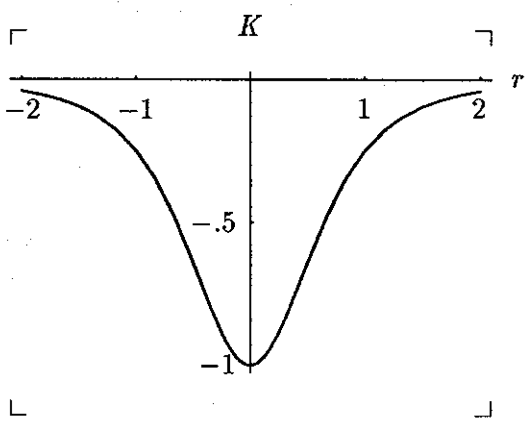
quadposcurv

Figure 2.2. A paraboloid of revolution. This paraboloid has equation $z = -(x^2 + y^2)/2$. As a surface of revolution, its curvature depends only on the distance r from the origin. Computation shows that the curvature at distance r is $(1 + r^2)^{-2}$, whose graph is shown at right.

this circle will have (for small r) a length d smaller than the length $2\pi r$ of a corresponding circle on the plane. Conversely, at a point where the surface is negatively curved, the ratio $d/2\pi r$ is greater than 1. It turns out that the second derivative of the ratio $d/2\pi r$ at $r = 0$ is exactly the negative of the Gaussian curvature as defined above.

The curvature of the surfaces in figures 2.2 and 2.3 is not constant: it goes rapidly to zero away from the origin (keeping the same sign). Neither the intrinsic geometry nor the extrinsic geometry of these surfaces is homogeneous. It is easy to construct a surface in space with constant positive curvature: the sphere. It is both extrinsically and intrinsically homogeneous.



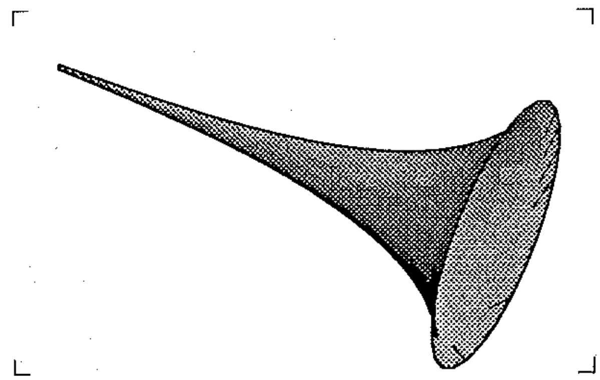


quadnegcurv

Figure 2.3. A saddle-shaped surface. This is the surface $z = (x^2 - y^2)/2$, a hyperbolic paraboloid. It has curvature -1 at the origin, but the curvature falls rapidly toward 0 as the distance r from the origin increases—the surface appears flatter near the edges of the plot. Perhaps surprisingly, the curvature is again a function of r alone.

constant negative curvature
pseudosphere
% pseudosphere
tractrix
% tractrix
Problem "surfaces of revolution of constant curvature"

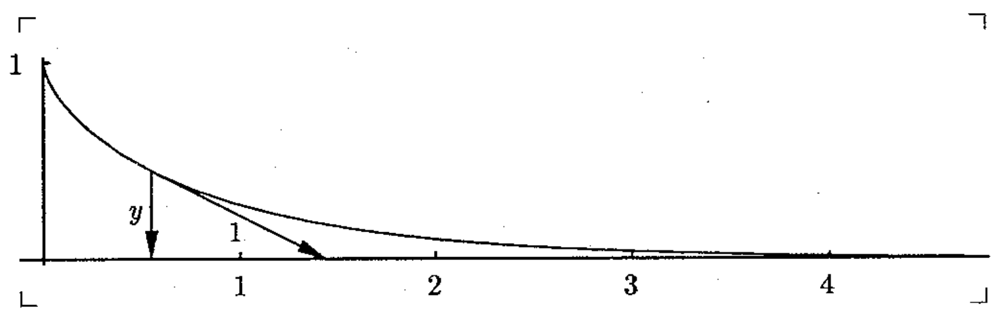
Surfaces of constant negative curvature are less obvious, but some are nevertheless well-known. The simplest of them is the pseudosphere (figure 2.4), the surface of revolution generated by a tractrix (figure 2.5). A tractrix



pseudosphere

Figure 2.4. The pseudosphere. The pseudosphere is the figure obtained by rotating the tractrix about the x-axis. It has constant curvature -1. Any small patch of the surface can be placed isometrically—bending but not stretching—anywhere else on the surface.

is the track of a box of stones that starts at $(0,1)$ and is dragged by means of a chain of unit length by a team of oxen walking along the x-axis. In other words, it is a curve determined (up to translation parallel to the x-axis) by the property that its tangent lines meet the x-axis a unit distance from the point of tangency.



tractrix

Figure 2.5. A tractrix. The tractrix is a curve whose tangent always meets the x-axis a unit distance away from the point of tangency. Therefore the derivative of the y-coordinate with respect to arc length s is $-y$, and $y(s) = e^{-s}$.

Note that the tractrix is not C^2 at the point $(0,1)$: its tangent is turning instantaneously at an infinite rate with respect to arclength. The edge of the pseudosphere is therefore an essential edge, beyond which it cannot be extended smoothly.

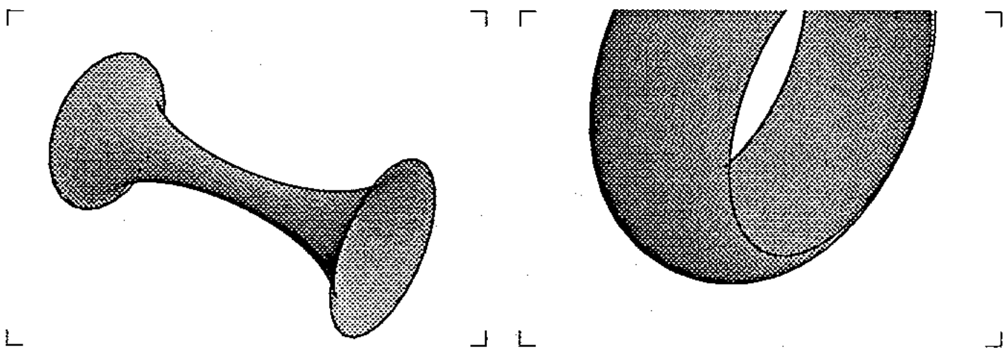
Problem 2.1.2 (surfaces of revolution of constant curvature). Show that the curvature K of a surface of revolution generated by rotating a plane curve

% spindles
pseudosphere, like that of the sphere, is locally homogeneous
locally isotropic
% pseudosphere is locally hyperbolic
% making hyperbolic paper
Exercise "making hyperbolic paper"
% annuli

$\big(x(s), y(s)\big)$, where s is arc length, around the x-axis, is given by

$$K = -\frac{1}{y}\frac{d^2y}{ds^2}.$$

Use this to verify that the curvature of the pseudosphere is -1. Then solve for $y(s)$ when K is 0, -1, and 1, and sketch a few examples of curves $\big(x(s), y(s)\big)$ in each case (see figure 2.6). Show that, for $K = -1$, the curve has a singularity at one or both ends, so the corresponding surface has an essential boundary beyond which it cannot be extended smoothly.



spindles

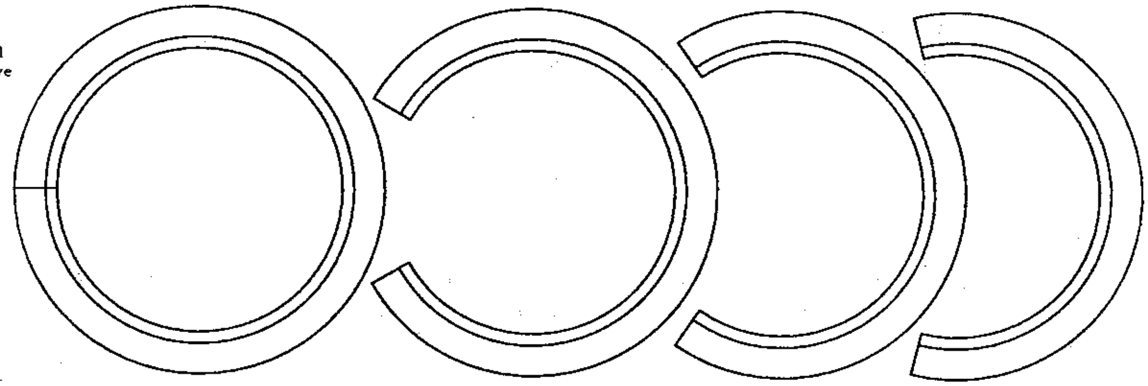
Figure 2.6. Surfaces of revolution of constant curvature. Each of the two planes containing a boundary circle is tangent to the surface along the boundary circle, and the surface has no smooth extension in space.

The intrinsic geometry of the pseudosphere, like that of the sphere, is locally homogeneous: any point has a neighborhood isometric to a neighborhood of any other point. To see this, parametrize the pseudosphere by coordinates (s, θ), where s comes from arc length along the tractrix, and θ is the angle around the x-axis. The defining property of the tractrix implies that the derivative of its y-coordinate with respect to arc length is $-y$, so that $y(s) = e^{-s}$. It follows that for any a and any θ_0, the locally defined map $(s, \theta) \mapsto (s + a, \theta_0 + e^a\theta)$ is an isometry.

The intrinsic geometry of the pseudosphere is also locally isotropic. A small disk on the surface of the pseudosphere can be rotated about its center without stretching. This is a consequence of exercise 2.2.13, but you can also see it directly by tackling exercise 2.1.3.

Exercise 2.1.3 (making hyperbolic paper). (a) Approximate a pseudosphere by a union of truncated cones, each formed from a flat sheet of paper by cutting out a portion of an annulus along two radii (figure 2.7) and joining its radial edges. The radius of an annular segment becomes the distance from the truncated cone to the cone's vertex, and the angle subtended by the segment is proportional to the radius of the truncated cone. Thus the annular segments should all have the same radius—the length of the tangent of the tractrix extended to the x-axis—but varying angles, depending on the y-coordinate of the tractrix.

% negsurface
skirts
Exercise "polyhedral models of negative curvature"



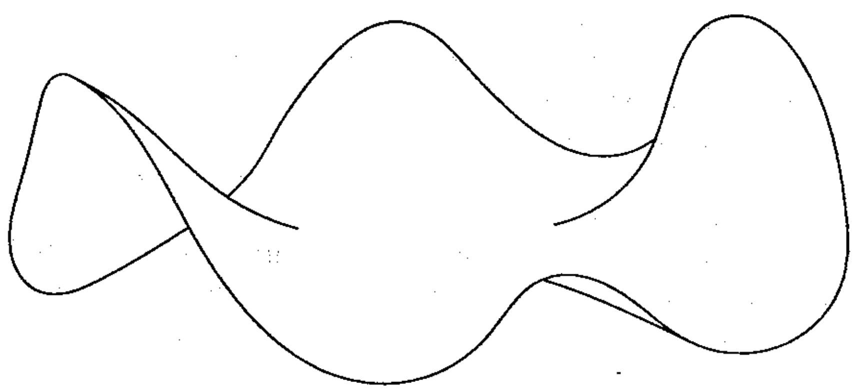
annuli

Figure 2.7. Annuli for making a pseudosphere. Pieces of annuli like this can be cut out and glued together to make model surfaces of negative curvature, like the pseudosphere. The Gaussian curvature will be approximately $1/r^2$, where r is the radius of the circle that bisects the annulus.

It is convenient to photocopy a drawing of several annuli on a sheet of paper. It is also worth marking an extra circle on each annulus, to indicate the extent of overlapping when the annuli are glued together.

(b) Construct a simply connected piece of surface in a similar way, starting with an annular strip that has not been made into a truncated cone. Apply this new surface to the pseudosphere; move it around and turn it. Notice how much intrinsic local homogeneity the pseudosphere has, which is not visible in space.

(c) How far can you extend this piece of hyperbolic paper? Can you get it to look something like figure 2.8?



negsurface

Figure 2.8. A negatively curved surface in space

An alternate medium for this construction is fabric. In fact, skirts with a nice negatively-curved flare can be (and are) made using large segments of annuli, roughly quarter-circles.

Exercise 2.1.4 (polyhedral models of negative curvature). Approximate models of a surface of constant negative curvature can also be constructed by joining triangles.

% making hyperbolic paper
% spindles
Theorem "Hilbert"
†complete
physical surfaces
plants
floppy hats
% making hyperbolic paper
% negsurface
†quadratic form
% Negatively curved surfaces in space
::principal directions
::umbilic
::lines of curvature
% curvaturelines
% Poincaré index theorem

(a) Cut out a number of equilateral triangles from paper, or better, manila folders, and join them so there are seven around each vertex. Alternatively, sew pieces of cloth together.

(b) Compare these models with the paper models of exercise 2.1.3. Can you calculate or estimate what size equilateral triangle would correspond to a given size annulus?

(c) Try making similar models with eight triangles per vertex. What size equilateral triangles would be necessary to make a model comparable to that of (a)?

(d) These polyhedral models have negative curvature concentrated at their vertices. You can make smoother models by diffusing the curvature out along the edges: replace each side of the equilateral triangles by an arc of circle such that the three angles of the resulting curvilinear triangle are $2\pi/7$. (The radius of the circle should be approximately 6.69 times the sides of the triangle.)

(e) Is it possible to make models from congruent triangles that approximate the geometry of a pseudosphere? What about surfaces of revolution as in figure 2.6?

After playing with the paper models above, you may be surprised by the following result:

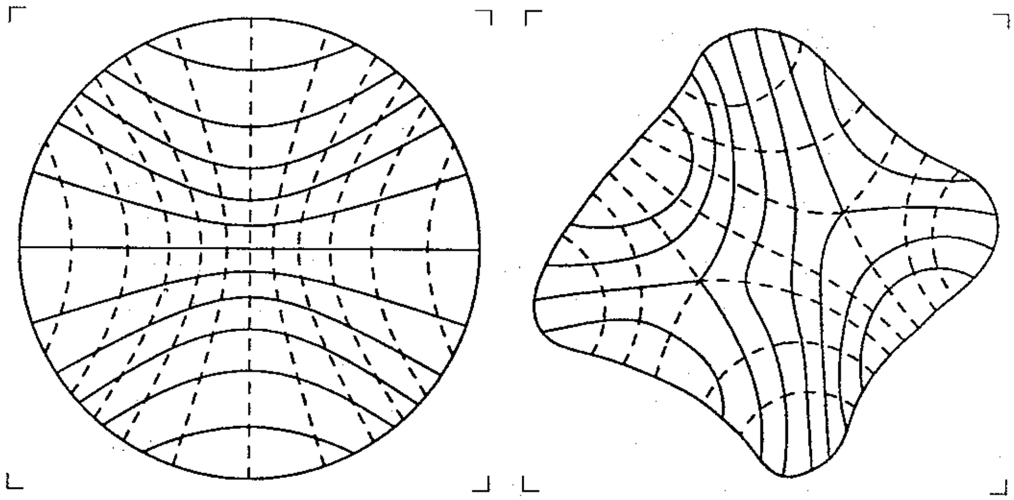
Theorem 2.1.5 (Hilbert). *There is no †complete smooth surface in Euclidean three-space with the local intrinsic geometry of the pseudosphere.*

Actually, we've had hints of this already. We have seen that the pseudosphere and other surfaces of revolution cannot be extended beyond their edges. Other physical surfaces having negative curvature, such as leaves of many kinds of plants, brims of floppy hats, or the paper models of exercise 2.1.3, all develop wavy edges as in figure 2.8, as soon as they grow big enough. This implies that they are not mathematically smooth, but at best of class C^1.

To see this, we need a little more qualitative differential geometry. Imagine you approximate a surface at a point by a †quadratic form as we did on page 37. By a rotation of the xy-plane, this quadratic form can be diagonalized, that is, put in the form $Ax^2 + Cy^2$, with $A \geq C$, say. Then the x-axis is the direction in which the surface curves upward the most (or downward the least), and the y-axis is the direction in which it curves downward the most (or upward the least). These two directions are the *principal directions* of the surface at that point. If $A = C$, the principal directions are undefined, and the point is called an *umbilic.*

For a C^2 surface of negative curvature, the principal directions are defined everywhere and change continuously, so they can be distinguished from one another. In other words, they define two families of curves tangent to them and perpendicular to one another; these are the so-called *lines of curvature* of the surface, and they're illustrated in figure 2.9. For a surface with many waves around the edge, the lines of curvature typically turn through one additional half-revolution for each wave, so they cannot be defined everywhere: they must have branch points in the interior, by a reasoning similar to the one used

distance to the moon
maps of the earth
models of hyperbolic space



curvaturelines

Figure 2.9. Lines of curvature. View from above of the lines of curvature of the hyperbolic paraboloid of figure 2.3 and of the surface in figure 2.8. The surface on the right is seen not to be smooth, because the lines of curvature of a C^2 negatively curved surface cannot branch.

to show the Poincaré index theorem (proposition 1.3.10). Therefore such a surface cannot be of class C^2.

Hilbert's original result in [Hil01] was for real analytic immersions, but the arguments actually work in class C^4. The extension to class C^2 requires some care [?]. In class C^1 an embedding of a complete surface of negative curvature in space is no longer impossible: [Kui55] gives an explicit construction for one.

However, any such embedding would be incredibly unwieldy, and pretty much useless in the study of the surface's intrinsic geometry, as one quickly learns from trying to extend the paper models beyond a certain point. For instance, the length of a circle of radius R feet in a surface of constant curvature $-1/\text{ft}^2$ is $2\pi \sinh R$. (One foot is 30.48 cm.) The circle of radius 1 foot has length 7.38 feet, which is fine: this would be a model of moderate curvature, like a sphere of radius 1 foot. But the lengths grow rapidly. The circle of radius 2 feet has length 22.78 feet, the circle of radius 10 feet has length 13.1 miles, and the circle of radius 20 feet has length 288,673 miles—more than the distance to the moon.

We must therefore resort to distorted pictures of the hyperbolic plane and of hyperbolic space. Just as it is convenient to have different maps of the earth for understanding various aspects of its geometry—for seeing shapes, for comparing areas, for plotting geodesics in navigation—so it is useful to have several maps of hyperbolic space at our disposal.

The several models of hyperbolic space that we shall look at have another important role, besides assisting our imagination. As one of the simple and basic structures of mathematics, hyperbolic geometry shows up in disguise in diverse places. The disguises it wears are usually related to one or another of these models.

Section "The inversive models"
% Hyperbolic surfaces
upper half-space
hemisphere
% inversion in a circle
Definition "inversion in a sphere"
::inversion
% square torus in space
Exercise "lines are circles"
% lines are circles
proper
% properties of inversions
Proposition "properties of inversions II"
% properties of inversions

2.2. The inversive models

One of the models, the Poincaré disk model, we're already somewhat familiar with from section 1.2. In this section we generalize it to arbitrary dimension, and study its first cousins, the upper half-space and hemisphere models. These models share the property that they can be obtained from one another by inversions, so our first task is to make sure we understand inversions in n dimensions. Their definition is essentially the same as definition 1.2.1:

Definition 2.2.1 (inversion in a sphere). If $S \subset \mathbf{E}^n$ is an $(n-1)$-sphere in Euclidean space, the *inversion* i_S in S is the unique map from the complement of the center of S into itself that fixes every point of S, exchanges the interior and exterior of S and takes spheres orthogonal to S to themselves.

As in the two-dimensional case, the image $i_S(P)$ of a point P in a circle S with center O and radius r is the point on the ray $\overrightarrow{OP}$ such that $OP \cdot OP' = r^2$.

It is somewhat annoying that inversion in a sphere in $\mathbf{E}^n$ does not map its center anywhere. We can remedy this by considering the one-point compactification $\widehat{\mathbf{E}^n} = \mathbf{E}^n \cup \{\infty\}$ of $\mathbf{E}^n$ (problem 1.1.1), which is homeomorphic to the sphere S^n. An inversion i_S can then be extended to map the center of S to ∞ and vice versa, so it becomes a homeomorphism of $\widehat{\mathbf{E}^n}$.

Exercise 2.2.2 (lines are circles). Show that the homeomorphism $h : \mathbf{E}^n \cup \{\infty\} \to S^n$ can be chosen in such a way that it maps circles to circles and lines to circles minus $h(\infty)$. (Hint: see stereographic projection, below.)

In view of exercise 2.2.2, it is natural to think of lines and planes as circles and spheres passing through ∞. Many properties of inversions and of the inversive models can be expressed more simply under this convention, so we will use it throughout this section. When we do want to exclude lines and planes, we'll talk about proper circles and spheres. For instance, here's the condensed version of proposition 1.2.3:

Proposition 2.2.3 (properties of inversions II). *If S is an $(n-1)$-dimensional proper sphere in $\mathbf{E}^n$, the inversion i_S is conformal, and takes spheres (of any dimension) to spheres.*

Proof of 2.2.3: For conformality, notice that any two vectors based at a point are the normal vectors to two $(n-1)$-spheres orthogonal to S, so both the angle between them and the angle between their images equal the dihedral angle between the spheres.

The second statement follows from the plane case (proposition 1.2.3) for spheres of codimension one by considering the symmetries around the line joining the centers of the inverted and inverting spheres; and for lower-dimensional spheres because they are intersections of spheres of codimension one. [2.2.3]

Exercise 2.2.4. (a) Since planes are special cases of spheres, what is the natural definition of inversion in a plane?

::Poincaré ball model
::hyperbolic reflections
% minimal hyperbolic properties
% properties of inversions II
% Hyperbolic surfaces totally geodesic
::totally geodesic
% vectorlength
% Hyperbolic surfaces
Exercise "curvature of the Poincaré model"
% measurement of curvature
% Poincaré metric

(b) What do you get when you successively invert in two concentric spheres? What if the spheres are planes?

(c) Show that composition of the inversions in two intersecting Euclidean planes is a Euclidean rotation. How would you define an inversive rotation?

The *Poincaré ball model* of hyperbolic space is what we get by taking the unit ball D^n in $\mathbf{E}^n$ and declaring to be hyperbolic geodesics all those arcs of circles orthogonal to the boundary of D^n. We also declare that inversions in $(n-1)$-spheres orthogonal to ∂D^n are hyperbolic isometries, which we will call *hyperbolic reflections.* According to problem 2.2.17 it would be enough to specify just the geodesics or just the isometries, but we won't take the minimalist approach.

Thanks to proposition 2.2.3, we can retrace the arguments in section 1.2 and conclude that the geodesics and isometries define the hyperbolic metric on D^n up to a constant factor, and that the Poincaré model is conformal, that is, hyperbolic angles and Euclidean angles are equal. Furthermore, it is easy to see that spheres of dimension k that meet the boundary orthogonally represent totally geodesic hyperbolic k-planes (a Riemannian submanifold is *totally geodesic* if all geodesics in the big manifold that are tangent to the submanifold are entirely contained in it).

To actually write down a formula for the metric, we look again at figure 1.12, where the hyperbolic length of a vector v is related to the angle of the banana built on it. Because of conformality we can assume that v is orthogonal to a diameter, in which case it is easy to see that the angle is $2/(1-r^2)$ times the Euclidean length of v, in the limit of small v, where r is the distance from the basepoint of v to the origin.

Exercise 2.2.5. Draw a diagram and convince yourself of this formula.

Recall that in this construction we had the choice of a multiplicative factor. The choice we made in section 1.2—setting the hyperbolic length of v equal to the banana angle in the limit when both go to 0—gives the following formula for the hyperbolic metric ds^2 as a function of the Euclidean metric dx^2:

2.2.6.
$$ds^2 = \frac{4}{(1-r^2)^2}dx^2.$$

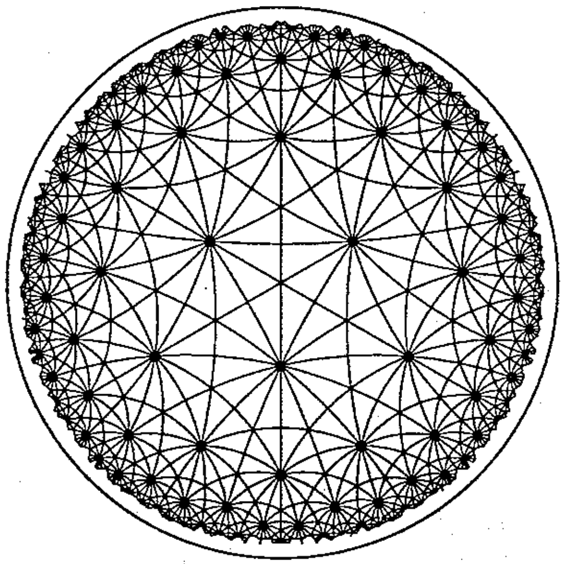
Exercise 2.2.7 (curvature of the Poincaré model). (a) Find the hyperbolic distance from the origin to a point at Euclidean distance r from the origin in the Poincaré model.

(b) Find the hyperbolic length of a circle whose radius is as above.

(c) Find the Gaussian curvature of the Poincaré model at the origin (use the criterion on page 37). The curvature at any other point is the same, since hyperbolic space is homogeneous. So it turns out that our choice of a constant factor in equation 2.2.6 was particularly fortunate.

% poincare237
% octatiles
::sphere at infinity
S^{n-1}_∞
definition of stereographic projection
::stereographic projection
Exercise "stereographic projection"

We see that distances are greatly distorted in the Poincaré model: the Euclidean image of an object has size roughly proportional to its Euclidean distance from the boundary ∂D^n, if this distance is small (figures 2.10 and 1.14). A person moving toward ∂D^n at constant speed would appear to be getting smaller and smaller and moving more and more slowly. She would never get there, of course; the boundary is "at infinity," not inside hyperbolic space.



poincare237

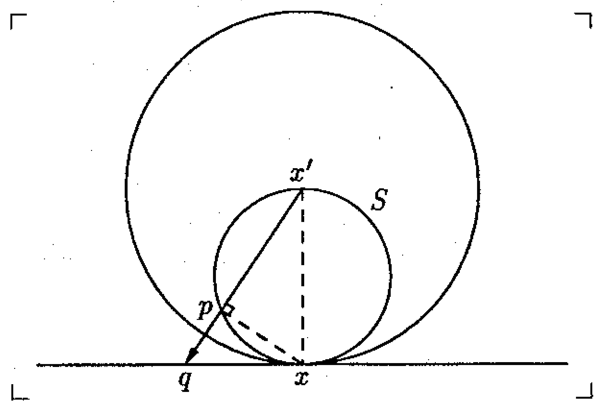
Figure 2.10. Hyperbolic tiling by 2-3-7 triangles. The hyperbolic plane laid out in congruent tracts, as seen in the Poincaré model. The tracts are triangles with angles $\pi/2$, $\pi/3$ and $\pi/7$. Courtesy HWG Homestead Bureau.

Nonetheless, ∂D^n can be interpreted purely in terms of hyperbolic geometry as the visual sphere. For a given basepoint p in D^n, each line of sight, that is, each hyperbolic ray from p, tends to a point on ∂D^n. If q is another point in D^n, each line of sight from q appears, as seen from p, to trace out a segment of a great circle in the visual sphere at p, since p and the ray determine a hyperbolic two-plane. This visual segment converges to a point in the visual sphere of p; in this way, the visual sphere of q is mapped to the visual sphere at p. The endpoint of a line of sight from p, as seen by q, gives the inverse map. In this way the visual spheres of all observers in hyperbolic space can be identified. This construction is independent of the model, and so associates to hyperbolic space $\mathbf{H}^n$ the *sphere at infinity* S^{n-1}_∞.

We now turn to a very useful construction, closely related to inversions. The *stereographic projection* from an n-dimensional proper sphere $S \subset \mathbf{E}^{n+1}$ onto a plane tangent to S at x is the map taking each point $p \in S$ to the intersection q of the line $\overline{px'}$ with the plane, where x' is the point opposite x on S.

% stereoinversion
% hemisphere
dilatation

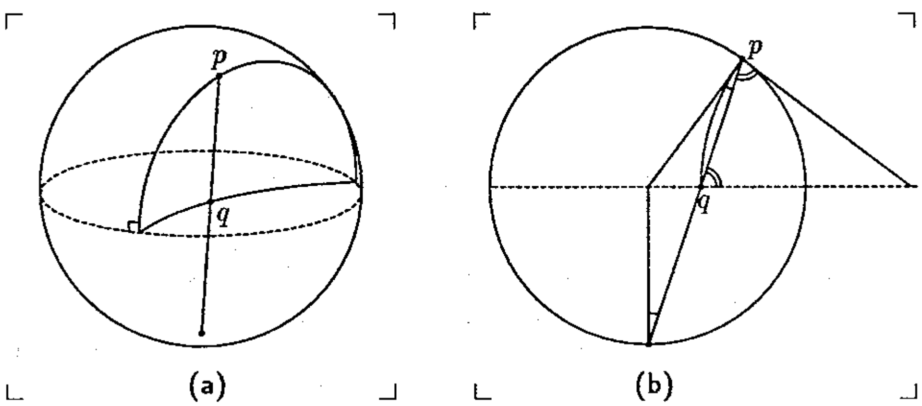
Exercise 2.2.8 (stereographic projection). Show that stereographic projection can be extended to an inversion. (Hint: see figure 2.11.) Consequently, it is conformal, and takes spheres to spheres.



stereoinversion

Figure 2.11. Stereographic projection and inversion. Stereographic projection from a sphere to a plane is identical to inversion in a sphere of twice the radius.

Our next model of $\mathbf{H}^n$ is derived from the Poincaré ball model by stereographic projection. We place the Poincaré ball D^n on the plane $\{x_0 = 0\}$ of $\mathbf{E}^{n+1}$, surrounded by the unit sphere $S^n \subset \mathbf{E}^{n+1}$, and we project from D^n to the northern hemisphere of S^n with center at the south pole $(-1, 0, \ldots, 0)$, as shown in figure 2.12(a). This is an inverse stereographic projection, at



hemisphere

Figure 2.12. The hemisphere model. (a) By stereographic projection from the south pole of a sphere we map the equatorial disk to the northern hemisphere. Transfering the Poincaré disk metric by this map we get a metric on the northern hemisphere whose geodesics are semicircles perpendicular to the equator. (b) The circle going through p and q and orthogonal to the equatorial disk is also orthogonal to the sphere. This shows that the projection of part (a) can also be obtained by following hyperbolic geodesics orthogonal to the equatorial disk.

least up to a dilatation (since the projection plane is equatorial rather than

::hemisphere model
% hemisphere
% uhslines
% Poincaré metric

tangent). In this way we transfer the geometry from the equatorial disk to the northern hemisphere to get the *hemisphere model.* Since stereographic projection is conformal and takes circles to circles, the hemisphere model is conformal and its geodesics are semicircles orthogonal to the equator $S^{n-1} = \partial D^n$.

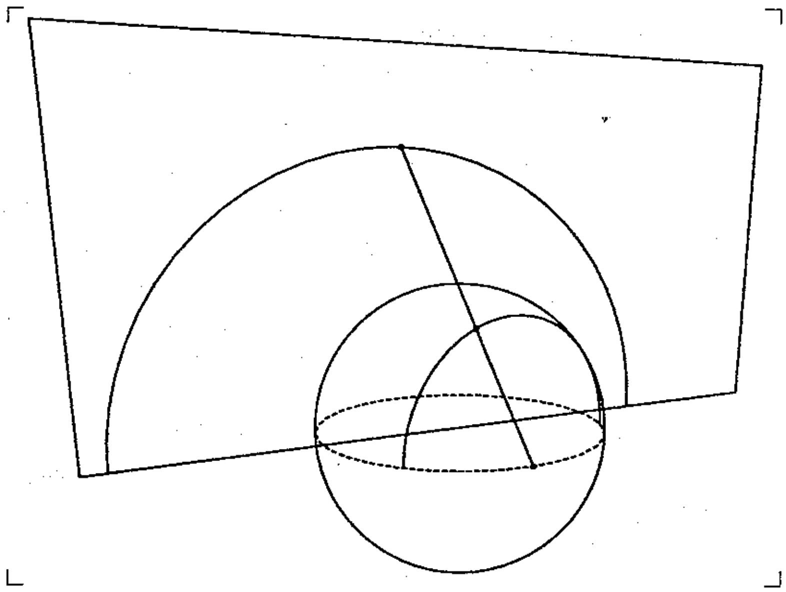
It is easy to see from figure 2.12(b) that for each point $q \in D^n$, the circle orthogonal to the equatorial disk $D^n \subset D^{n+1}$ and to $S^n = \partial D^{n+1}$ meets the northern hemisphere at the same point p as the image of q under the projection above. This means we can interpret this projection purely in terms of hyperbolic geometry: fix a totally geodesic n-space $\mathbf{H}^n$ inside hyperbolic $(n+1)$-space $\mathbf{H}^{n+1}$, and choose one of the half-spaces it determines. For each $q \in \mathbf{H}^n$, the hyperbolic ray from q perpendicular to $\mathbf{H}^n$ and pointing into the half-space we chose converges to a point in the corresponding visual hemisphere, so we get a map $\mathbf{H}^n \to S^n_\infty$. By making $\mathbf{H}^n = D^n$ be the equatorial disk in the Poincaré ball model of $\mathbf{H}^{n+1} = D^{n+1}$, we see that this map $\mathbf{H}^n \to S^n_\infty$ coincides with the projection above.

From the hemisphere model we get the third important inversive model of hyperbolic space, also by stereographic projection. This time we project from a point on the equator, say $(0, \ldots, 0, 1)$, onto a vertical subspace, say $\{x_n = 0\}$, which we identify with $\mathbf{E}^n$. The open northern hemisphere maps onto the open upper half-space $\{x_0 > 0\}$, and the equator—the sphere at infinity of the hemisphere model—maps onto the bounding plane $\mathbf{E}^{n-1} = \{x_0 = 0\}$, except for the center of projection, which is mapped to the point at infinity. In other words, the sphere at infinity here is given by the one-point compactification of the bounding plane, $S^{n-1}_\infty = \mathbf{E}^{n-1} \cup \{\infty\}$. Geodesics are given by semicircles orthogonal to the bounding plane $\mathbf{E}^{n-1}$ (figure 2.13), and hyperbolic reflections are inversions in spheres orthogonal to the bounding plane. Clearly, this model, too, is conformal.

The hyperbolic metric in the upper half-space model has an especially simple form, which is well suited to many computations. We could compute it by writing down an explicit formula for the composition of projections that goes from the Poincaré disk model to the upper half-space model, and then using equation 2.2.6 above; but we can save a lot of energy by observing that this transition map is conformal, so our argument about banana angles applies here too. If we position the banana so that one of its vertices is at infinity, it becomes a solid cone, and then it's easy to compare the angle at the vertex with the Euclidean length of the vector the cone is built on. In the limit when both are small, the Euclidean length is just x_0 times the angle, where x_0 is the Euclidean distance from the vector's basepoint to the bounding plane. In other words, the relation between the hyperbolic metric ds^2 and the Euclidean metric dx^2 is

2.2.9. $$ds^2 = \frac{1}{x_0^2} dx^2.$$

% uhs237
% poincare237



uhslines

Figure 2.13. Geodesics in the upper half-space model. Geodesics in the upper half-space model of hyperbolic space appear as semicircles orthogonal to the bounding plane, or half-lines perpendicular to it.

Thus the Euclidean image of an object has size exactly proportional to its Euclidean distance from the bounding plane $\mathbf{E}^{n-1}$. Figure 2.14 shows the same congrent tracts as figure 2.10, but seen in the upper half-space model.

/u/gt3m/3mbook/pictures/chap2/2/uhs237.ps not found

uhs237

Figure 2.14. Hyperbolic tiling by 2-3-7 triangles. Another view of the hyperbolic world divided into congruent tracts. Upper half-plane projection.

Exercise 2.2.10. Any Euclidean similarity of $\mathbf{E}^{n-1}$ extends in a unique way to a Euclidean similarity preserving upper half-space. Show that such a similarity is a hyperbolic isometry by expressing it as a composition of reflections.

The easy visibility of this significant subgroup of isometries of $\mathbf{H}^n$ is a frequently useful aspect of the upper half-space model.

Exercise 2.2.11. We already know that hyperbolic space is homogeneous and isotropic, but explicit formulas are especially easy to write down in the upper half-space model.

(a) Given two points in upper half-space, write a composition of hyperbolic reflections that will map one to the other.

(b) Let $O(n)$ be the group of isometries of the tangent space to $\mathbf{H}^n$ at a point p. Show that any element of $O(n)$ can be realized by a composition of hyperbolic reflections. (Hint: Show by induction that $O(n)$ is generated by reflections, and that any reflection in $O(n)$ can be realized by a reflection in a sphere orthogonal to $\mathbf{E}^{n-1}$ in the upper half-space model.)

reflections take you everywhere
curve upwards
::horospheres
::horocycles
::tangent
% upper half-space metric
::horoball
% pseudosphere
Exercise "pseudosphere is locally hyperbolic"
% making hyperbolic paper

(c) Show that reflections generate the whole group of isometries of $\mathbf{H}^n$. (Hint: an isometry is determined by its derivative at a point.)

A horizontal Euclidean plane $\{x_0 = c\}$, for $c > 0$, is not a plane in hyperbolic geometry: it lies entirely on one side of a true hyperbolic plane tangent to it, which is a Euclidean sphere. (One can also notice that pushing a horizontal surface up along orthogonal geodesics shrinks the hyperbolic metric, so the surface must curve upwards.) These horizontal surfaces are examples of *horospheres* (or *horocycles* if $n = 2$). Horospheres are characterized by the property that their parallel surfaces are all congruent—in our case, by arbitrary dilatations centered at points on the plane at infinity. Another characteristic property is that a horosphere is orthogonal to all planes passing through a certain point on the sphere at infinity; we say the horosphere is *tangent* to S^{n-1}_∞ at that point.

Problem 2.2.12. Show that these two characterizations are equivalent, and that according to them a horosphere in upper half-space appears as either a horizontal Euclidean plane or a Euclidean sphere tangent to the bounding plane. How does a horosphere appear in the Poincaré disk model?

The intrinsic geometry of a horosphere is Euclidean. This is easiest to see when the horosphere appears as $\{x_0 = c\}$, by examining the form of the hyperbolic metric given by equation 2.2.9. In fact, it follows from the same equation that any Euclidean isometry acting on a horosphere extends to an isometry of $\mathbf{H}^n$ which preserves it and all of its parallel horospheres.

The region $\{x_0 \geq c\}$ above a horosphere is called a *horoball.* If a horosphere appears as a Euclidean sphere tangent to the bounding plane, its corresponding horoball is just the Euclidean ball bounded by that sphere.

Using horocycles we can describe the relation of the hyperbolic plane to the pseudosphere (figure 2.4).

Exercise 2.2.13 (pseudosphere is locally hyperbolic). Consider the map that wraps the region $y \geq 1$ of the upper half-plane around the pseudosphere, taking horocycles $y = C$ to meridians and vertical geodesics $x = C$ to generating curves (tractrixes). Show that, if the map is periodic with period 2π in the x-direction, it is a local isometry. (Hint: arc length along horocycles, measured between fixed vertical geodesics, decreases exponentially with hyperbolic distance from the line $y = 1$, and arc length along meridians on the pseudosphere similarly decreases exponentially with distance from the edge.)

The paper models of exercise 2.1.3 are based on horocycles. In the limit, when the annuli are infinitely thin, the metric becomes the hyperbolic metric, and the circles become horocycles. This comparison demonstrates something that was hard to explain before: the isotropy of the pseudosphere's metric.

We can use the upper half-space model to study the group of isometries of hyperbolic space. Consider a reflection of $\mathbf{H}^n$ given by inversion in an $(n-1)$-sphere S orthogonal to $\mathbf{E}^{n-1}$. The restriction of this inversion to the sphere at infinity $S^{n-1}_\infty = \mathbf{E}^{n-1} \cup \infty$ is just the inversion in the $(n-2)$-sphere

definition of Möbius transformations
::Möbius transformation
::Möbius group
Möb_{n-1}
% reflections take you everywhere
Problem "the Möbius group"
Exercise "Steiner's porism"
Exercise "tangent spheres"
Problem "minimal hyperbolic properties"

$S \cap S^{n-1}_\infty$, and every inversion of S^{n-1}_∞ can be so expressed. A transformation of S^{n-1}_∞ that can be expressed as a composition of inversions is known as a *Möbius transformation*, and the group of all such transformations is the *Möbius group*, denoted Möb_{n-1}. Since, by exercise 2.2.11(c), all hyperbolic isometries can be generated by reflections, it follows that the group of isometries of $\mathbf{H}^n$ is isomorphic to Möb_{n-1}.

Problem 2.2.14 (the Möbius group). Analyze and become familiar with the Möbius group. Show that:

(a) The subgroup of the Möbius group that fixes ∞ is isomorphic to the group of Euclidean similarities.

(b) The subgroup of the Möbius group Möb_n that takes an $(n-1)$-sphere to itself and fixes a point not on that sphere is isomorphic to the group $O(n)$.

(c) For $n > 1$, the Möbius group consists exactly of those homeomorphisms of S^n_∞ that take $(n-1)$-spheres to $(n-1)$-spheres.

(d) Any Möbius transformation that takes a sphere S to a sphere R conjugates i_S to i_R.

(e) The subgroup of the Möbius group that takes a k-sphere to itself is isomorphic to $\text{Möb}_k \times O(n-k)$.

(f) There is a subgroup of the Möbius group isomorphic to $O(n+1)$.

What is the dimension of the Möbius group?

Some geometric problems involving spheres can be greatly simplified by artful application of a Möbius transformation.

Exercise 2.2.15 (Steiner's porism). Suppose you are given an arrangement of circles in the plane consisting of two non-intersecting circles A and B and a chain of circles $X_0, X_2, \ldots, X_{n-1}$, where each X_i is tangent to A, B, and $X_{i+1 \bmod n}$. The circles are disjoint except for tangencies.

Show that if the X_i are erased, and any circle Y_0 tangent to A and to B is constructed, the analogous chain of circles it determines closes up after exactly n circles.

should attribute to Coxeter?

Exercise 2.2.16 (tangent spheres). Let A, B and C be mutually tangent two-spheres. Let X_0 be a fourth sphere tangent to A, B and C. Construct a chain of spheres, beginning with X_0, each sphere being tangent to A, B and C and to its neighbors in the chain.

Show that the chain closes up on the sixth sphere.

Problem 2.2.17 (minimal hyperbolic properties). In the discussion of hyperbolic geometry above, there was no attempt to characterize hyperbolic geometry using a minimal amount of structure. Here are some steps in this direction:

(a) Show that hyperbolic lines can be characterized in terms of the metric as curves that minimize distance between any two points. (Hint: in the upper half-space model, reduce to the case that p and q are on a vertical line.)

(b) Show how to characterize hyperbolic lines directly in terms of the group of isometries, as fixed-point sets.

(c) Show that the only diffeomorphisms of upper half-space to itself that take all hyperbolic lines to hyperbolic lines are the hyperbolic isometries. (This contrasts with the Euclidean case, where affine transformations take lines to lines.)

(d) Conclude that any of these three structures is sufficient as a base structure to define hyperbolic geometry: the set of lines, the group of isometries, or the metric up to a constant multiple.

(e) (Harder.) Show that the measure of angle also is sufficient to define hyperbolic geometry.

2.3. The hyperboloid model and the Klein model

Section "The hyperboloid model and the Klein model"
sphere
::orthogonal transformations
$O(n+1)$
indefinite metric
quadratic form
::Lorentz space
$\mathbf{E}^{n,1}$
special relativity
physicists
time
space
::space-like
::time-like
::light-like
% Lorentz metric
% elliptic space
H^+
::hyperboloid model
::projective model
::Klein model
Beltrami
% Beltrami
% centralprojection

A sphere in Euclidean space with radius r has constant curvature $1/r^2$. By analogy, since hyperbolic space has constant curvature -1, hyperbolic space should be a sphere of radius $i = \sqrt{-1}$.

This seemingly impossible condition can actually be given a reasonable interpretation. Let's see how far analogy can take us. To get an n-sphere, we start with a positive definite quadratic form $Q^+ = x_0^2 + x_1^2 + \cdots + x_n^2$ on $\mathbf{R}^{n+1}$, which gives $\mathbf{R}^{n+1}$ a Euclidean metric $dx^2 = dx_0^2 + dx_1^2 + \cdots + dx_n^2$, making it into $\mathbf{E}^{n+1}$. Restricting to the unit sphere $S = \{Q^+ = 1\}$, we get a Riemannian metric of constant positive curvature 1. The isometries of S come from linear transformations of $\mathbf{E}^{n+1}$ preserving Q^+; the group of these *orthogonal transformations* is denoted $O(n+1)$.

Now let's start instead with the indefinite metric

2.3.1. $$ds^2 = -dx_0^2 + dx_1^2 + \cdots + dx_n^2$$

in $\mathbf{R}^{n+1}$ associated to the quadratic form $Q^- = -x_0^2 + x_1^2 + \cdots + x_n^2$. With this metric, $\mathbf{R}^{n+1}$ is often referred to as *Lorentz space*, and denoted $\mathbf{E}^{n,1}$. When $n = 3$, this is the universe of special relativity, although physicists usually reverse the sign of Q^-. In this interpretation, the vertical direction x_0 represents time, and the horizontal directions represent space. A vector x is *space-like*, *time-like* or *light-like* depending on whether $Q^-(x)$ is positive, negative, or zero. By analogy with the Euclidean case, the length of a vector x is $\sqrt{Q^-(x)}$, so light-like vectors have zero length, and time-like vectors have imaginary length (which we take to be a positive multiple of i).

The sphere of radius i about the origin in $E^{n,1}$ is the hyperboloid $H = \{Q^- = -1\}$. When restricted to this hyperboloid, the indefinite metric ds^2 of equation 2.3.1 becomes a *bona fide*, positive definite Riemannian metric: it is easy to check that any tangent vector to the hyperboloid has real length. Topologically, H is the union of two open disks.

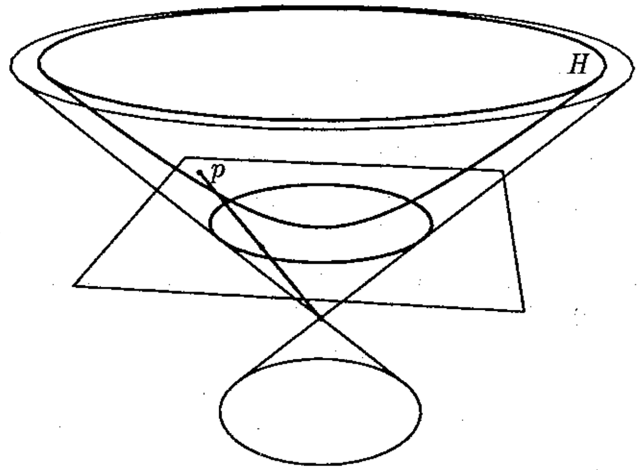
As in the case of the sphere, which in example 1.4.3 is turned into elliptic space, we really want to identify antipodal points of H, to get a subset of projective space $\mathbf{RP}^n$. Unlike the case of the sphere, here antipodal points lie in disjoint components of H, so this subset of $\mathbf{RP}^n$ can be modeled by one component of the hyperboloid, say, the upper sheet H^+, where $x_0 > 0$. This is the *hyperboloid model* of hyperbolic space.

Another way to model the same subset of $\mathbf{RP}^n$ is by inhomogeneous coordinates, where a point $(x_1, \ldots, x_n) \in \mathbf{R}^n$ stands for the line with parametric equation $(t, x_1 t, \ldots, x_n t)$. Then H maps onto the unit disk, giving the *projective model*, or *Klein model* of hyperbolic space (although it was first introduced by Beltrami: cf. page 6). Notice that, although we can look at the Klein disk as embedded in $E^{n,1}$ (figure 2.15), its metric is not the induced metric, but rather the push-forward of the hyperboloid's metric by central projection.

orthogonality
inner product
::orthogonal complement
Exercise "characterization of tangent vectors"
orthogonal transformations

$O(n+1)$
orthogonal transformations are isometries
orthogonal transformations of $\mathbf{E}^{n+1}$

$O(n,1)$
orthogonal transformations of $\mathbf{E}^{n,1}$



centralprojection

Figure 2.15. The hyperboloid model and the Klein model. A point $p = (x_0, x_1, \ldots, x_n)$ on the hyperboloid maps to a point $(x_1/x_0, \ldots, x_n/x_0)$ in $\mathbf{R}^n$, shown here as the horizontal hyperplane $\{x_0 = 1\}$. This transfers the metric from the hyperboloid to the unit disk in $\mathbf{R}^n$, giving the projective model, or Klein model, for hyperbolic space.

How can we see that these are indeed models for n-dimensional hyperbolic space? The best way is not by direct calculation but, as for the inversive models, through the study of lines and isometries.

In $\mathbf{E}^{n,1}$ we still have a notion of orthogonality, given by the inner product $-x_0y_0 + x_1y_1 + \cdots + x_ny_n$. The *orthogonal complement* of any non-zero vector x is an n-dimensional subspace, denoted by $x^\perp$; it contains x if and only if $Q^-(x) = 0$. The orthogonal complement of a subspace is the intersection of the orthogonal complements of its points.

Exercise 2.3.2 (characterization of tangent vectors). If $x \in H$ is a point on the hyperboloid, the tangent space of H at x coincides with $x^\perp$.

We also have the notion of orthogonal transformations, that is, linear transformations of $\mathbf{R}^{n+1}$ preserving Q^-. Just as the group of isometries of S is identical with $O(n+1)$, the group of orthogonal transformations of $\mathbf{E}^{n+1}$, so also the group of isometries of H is $O(n,1)$, the group of orthogonal transformations of $\mathbf{E}^{n,1}$.

Exercise 2.3.3. The proof is basically the same as for $\mathbf{E}^{n+1}$. First an orthogonal transformation of $\mathbf{E}^{n,1}$ clearly induces an isometry on H, non-trivial if the transformation is non-trivial. Next, all isometries of H come from such transformations, that is, $O(n,1)$ acts transitively and isotropically on H:

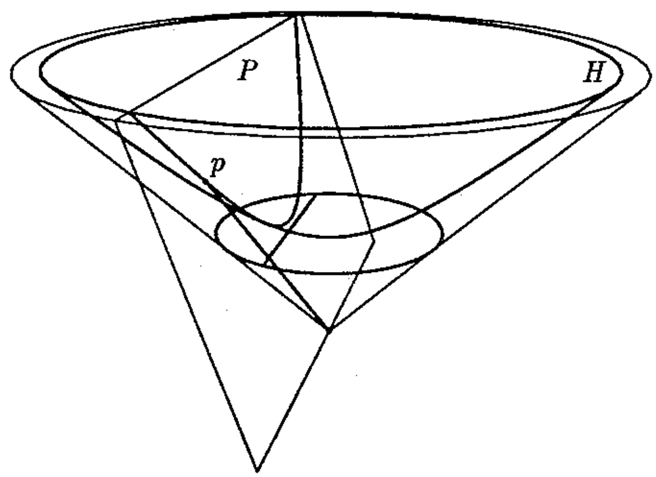
(a) If p and q are points on H, find an element of $O(n,1)$ interchanging p and q. (Hint: consider the orthogonal complement of $p - q$.)

(b) If $p \in H$, show that any isometry of the tangent space of p is induced by an element of $O(n,1)$. (Hint: by part (a), it is enough to consider the case where p is the "north pole" $(1, 0, \ldots, 0)$.)

Exercise "Lorentz transformations" ::Lorentz transformation ::Lorentz group % hyperlines ::light cone Exercise "parametrization of geodesics"

Exercise 2.3.4 (Lorentz transformations). An element of $O(n,1)$ that takes each component of H to itself is called a *Lorentz transformation*. Show that the group of Lorentz transformations, or *Lorentz group*, has index two in $O(n,1)$, and coincides with the group of isometries of H^+.

This gives a way to describe geodesics in the hyperboloid and Klein models: if $p \in H^+$ is a point and v is a non-zero tangent vector to H^+ at p, the geodesic through p in the direction of v is the intersection L of H^+ with the two-plane P determined by p, v and the origin (figure 2.16). The same is true with the Klein disk K instead of H^+. To see this, consider the Lorentz transformation that



hyperlines

Figure 2.16. Geodesics in the hyperboloid and Klein models. In the hyperboloid model geodesics are the intersections of two-planes through the origin with the hyperboloid. In the projective model they're straight line segements.

fixes P and equals -1 on the orthogonal complement of P. The corresponding isometry of H^+ or K fixes exactly those points that lie on L. By uniqueness, the geodesic through p in the direction of v is fixed by this isometry, and so must be contained in L. But L is a connected curve, so it must be the geodesic.

For the hyperboloid model, L is a branch of a hyperbola, whose asymptotes are rays in the *light cone* $\{Q^- = 0, x_0 > 0\}$—so called because, in the relativistic interpretation, it is the union of the trajectories in space-time of light rays emanating at time 0 from a point source at the origin. Rays in the light cone, then, are the points at infinity for this model.

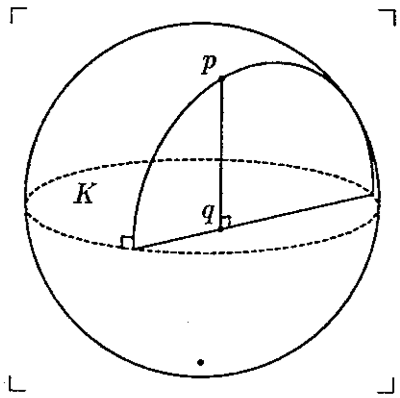
Exercise 2.3.5 (parametrization of geodesics). Show that if v has unit length, L is parametrized with velocity 1 by $p \cosh t + v \sinh t$. What is the analogous formula for the sphere?

For the Klein model, L is a segment of a straight line, meaning that this model is projectively correct: geodesics look straight. This makes the Klein model particularly useful for understanding incidence in a configuration of lines and planes. The sphere at infinity is just the unit sphere S^{n-1}_∞, the image in $\mathbf{RP}^n$ of the light cone. Angles are distorted in the Klein model, but

% Lorentz metric
% The inversive models
% minimal hyperbolic properties
% hemiprojective
% hemisphere
::visual projective space
visual sphere

they can be accurately and conveniently computed in the hyperboloid model if one remembers to use the Lorentz metric of equation 2.3.1, rather than the Euclidean metric.

We now exhibit a correspondence between the Klein model and the hemisphere model of section 2.2 that takes geodesics into geodesics. Since the set of geodesics is sufficient to characterize hyperbolic geometry (problem 2.2.17), we conclude that K, and consequently also H^+, are indeed legitimate models for $\mathbf{H}^n$. If K is placed as the equatorial disk of the unit sphere in $\mathbf{R}^{n+1}$, the correspondence is given by Euclidean orthogonal projection onto K (figure 2.17): geodesics in the hemisphere model are half-circles orthogonal to the equator, so they indeed project to segments of straight lines.



hemiprojective

Figure 2.17. Going from the hemisphere to the Klein model. We get the Klein model of hyperbolic space from the hemisphere model by (Euclidean) orthogonal projection onto the equatorial disk. Compare figure 2.12(a).

Note that this arrangement is similar to the one we made to go from the Poincaré to the hemisphere model, but the projection is different. The composed map, from the Poincaré disk model to the hemisphere back to the Klein model, is a surprising transformation of the unit disk to itself that maps each radius to itself, while simultaneously straightening out every circle orthogonal to the unit sphere into a straight line segment.

Exercise 2.3.6. Find the formula for this Poincaré-to-projective transformation, in polar coordinates (r, θ), where $0 \leq r < 1$ and $\theta \in S^{n-1}$. Find the formula for the inverse transformation.

Exercise 2.3.7. The Poincaré model maps to the hemisphere model by stereographic projection (figure 2.12). Show that the same projection, if extended to H^+, gives a direct correspondence between these two models and the hyperboloid model.

As a subset of projective space $\mathbf{RP}^n$, the Klein model has a natural interpretation in terms of hyperbolic perspective. In fact, it embeds in *visual projective space*, the visual sphere with antipodal points identified. Imagine you are in $\mathbf{H}^{n+1}$, hovering above a copy of $\mathbf{H}^n$. Since light rays coming from a geodesic in $\mathbf{H}^n$ lie in a hyperbolic two-plane, their tangent vectors lie in a

% klein237
::dual hyperspace
% duality
% characterization of tangent vectors
% duality

great circle of your visual sphere, so you see the geodesic as a straight line. The hyperplane $\mathbf{H}^n$ looks like the Klein model! The sphere at infinity of $\mathbf{H}^n$ looks like a sphere to you. In contrast to the situation in Euclidean space, the visual radius of a hyperbolic plane in hyperbolic space is always strictly less than π, provided the plane does not contain the eye. Your image of $\mathbf{H}^n$ shrinks as you move away from it and expands as you move closer. See figure 2.18.

[/u/gt3m/3mbook/pictures/chap2/3/klein237.ps not found]

klein237

Figure 2.18. A view of the hyperbolic plane from a helicopter in hyperbolic space. The same tracts in the hyperbolic plane shown in figures 2.10 and 2.14, this time in the projective disk projection. This picture is in true perspective in hyperbolic three-space. Stare at the plane near its horizon and try to sense the way it slopes away from you and the way the area of the plane grows very rapidly (exponentially) with distance.

Exercise 2.3.8. How can you tell from a distance at what angle two planes meet in $\mathbf{H}^3$?

We now have a geometric interpretation for points in $\mathbf{RP}^n$ that lie in the unit ball and on the sphere. How about points outside the closed ball? If $x \in \mathbf{RP}^n$ is such a point, Q^- is positive on the associated line $X \subset \mathbf{E}^{n,1}$. This means that Q^- is indefinite on the orthogonal complement $X^\perp$ of X, and that the corresponding hyperspace $x^\perp \subset \mathbf{RP}^n$ intersects hyperbolic space. We call $x^\perp$ the *dual hyperspace* of x. The hyperbolic significance of projective duality is that any line from x to $x^\perp$ is perpendicular to $x^\perp$. This is best seen in the hyperboloid model, as shown in figure 2.19(a): if $p \in X^\perp \cap H^+$ represents a point in $x^\perp$ and $v \in X^\perp$ is any tangent vector at p that represents a direction in $x^\perp$, we want to show that v is perpendicular to the tangent vector w that represents the direction from p to X. But w lies in the plane determined by p and X, and, by exercise 2.3.2, is orthogonal to p; since p is orthogonal to X, this implies that w is in fact parallel to X, and consequently orthogonal to v.

Exercise 2.3.9. (a) Prove the assertion in the caption of figure 2.19(b).

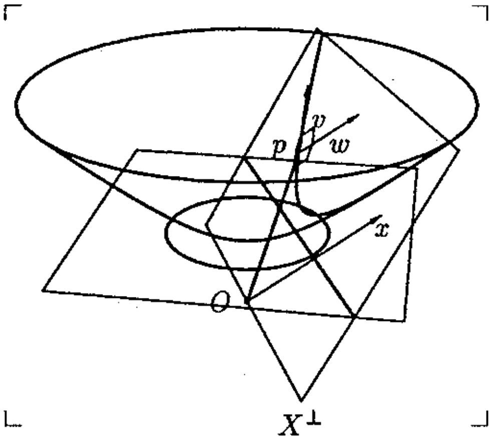
(b) What is the dual of a point in $\mathbf{H}^n$? Of a point on S^{n-1}_∞?

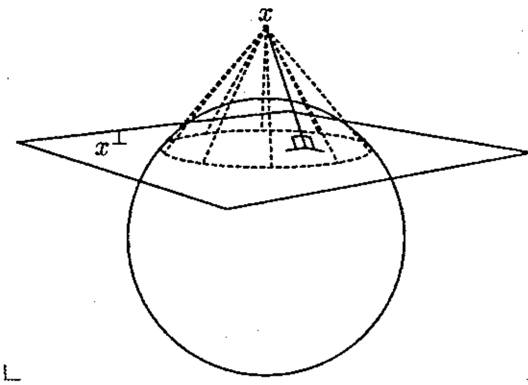
(c) What is the dual of a k-plane?

(d) Show that the dual of a k-plane P is the intersection of the duals of points in P. Write down a dual statement.

In the two-dimensional case the picture is especially simple. Any two lines insersect somewhere in $\mathbf{RP}^2$. If the intersection is inside S^1_∞, the lines meet in the conventional sense, from the point of view of a hyperbolic observer. If the intersection is on S^1_∞, the lines converge together on the visual circle of the observer, and they are called *parallels*. Otherwise, they are called *ultraparallels*, and have a unique common perpendicular in $\mathbf{H}^2$, dual to their intersection point outside S^1_∞.

Exercise "parallelism in hyperbolic space"
% common perpendicular for lines in H^3
Problem "projective transformations of hyperbolic space"
::projective transformation
% minimal hyperbolic properties
Problem "shapes of Euclidean polygons"
†simple





duality

Figure 2.19. Duality between a hyperplane and a point. The dual of a point x outside $\mathbf{H}^n$ is a hyperplane $x^\perp$ intersecting $\mathbf{H}^n$. (a) Lines from $x^\perp$ to x are perpendicular to $x^\perp$, and lines perpendicular to $x^\perp$ go through x (see text for detailed argument). (b) In $\mathbf{RP}^n$, the point x is the vertex of the cone tangent to S^{n-1}_∞ at the $(n-2)$-sphere where $x^\perp$ meets S^{n-1}_∞.

Exercise 2.3.10 (parallelism in hyperbolic space). Extend parallelism and ultraparallelism to k-planes in $\mathbf{H}^n$. Show that ultraparallel $(n-1)$-hyperplanes in $\mathbf{H}^n$ have a unique common perpendicular line. For a related statement about lines in $\mathbf{H}^3$, see proposition 2.5.3.

Problem 2.3.11 (projective transformations of hyperbolic space). A *projective transformation* is a self-map of $\mathbf{RP}^n$ obtained from an invertible linear map of $\mathbf{R}^{n+1}$ by passing to the quotient. An orthogonal transformation of $\mathbf{E}^{n,1}$ clearly gives rise to a projective transformation taking S^{n-1}_∞ to itself; show that the converse is also true.

This implies that any projective transformation of $\mathbf{RP}^n$ that leaves $\mathbf{H}^n$ invariant is an isometry, in contrast with the Euclidean situation, where there are many projective transformations that are not isometries: the affine transformations (compare problem 2.2.17).

Problem 2.3.12 (shapes of Euclidean polygons). The angles of a regular pentagon in plane Euclidean geometry are all 108°, but not all pentagons with 108° angles are regular. Consider the space of (not necessarily †simple) pentagons having 108° angles and sides parallel to the corresponding sides of a model regular pentagon, and parametrize this space by the (signed) side lengths $s_1, \ldots, s_5$.

(a) Show that the the s_i are subject to a linear relation that confines them to a three-dimensional linear subspace V of $\mathbf{R}^5$.

(b) Show that the area enclosed by a pentagon is a quadratic form on V which is isometric to $\mathbf{E}^{2,1}$. How do you measure area for a non-simple pentagon?

(c) Describe a model for the hyperbolic plane in terms of the subset of V consisting of pentagons of unit area. Does your model have a single component?

(d) Show that the space of simple pentagons of unit area is a right-angled pentagon in the hyperbolic plane.

Problem "the paraboloid model" ::paraboloid model

(e) There is an area-preserving "butterfly operation" on pentagons that replaces a side of length s by a side of length $-s$, changing the lengths of the two neighboring sides to make it fit. Interpret this operation in hyperbolic geometry.

(f) Given a non-simple pentagon in V, when is it possible to modify it by a sequence of butterfly moves until it is simple?

(g) Generalize to higher dimensions. Show that the space of simple normalized Euclidean polygons of $n + 3$ sides, having unit area and edges parallel to and in the same direction as some convex model polygon, is parametrized by a convex polyhedron in hyperbolic n-space. Can you describe the three-dimensional hyperbolic polyhedron when the model polygon is a regular hexagon?

Problem 2.3.13 (the paraboloid model). To obtain other projectively correct models of hyperbolic space, one can transform the Klein model by any projective transformation.

(a) Write down a projective transformation that maps the unit sphere of $\mathbf{R}^n$ to the paraboloid $x_n = x_1^2 + \cdots + x_{n-1}^2$. Applying this projective transformation to the Klein model K, we obtain the *paraboloid model* of hyperbolic space.

(b) The paraboloid model is to the Klein model as the upper half-space is to the Poincaré model. More precisely, the upper half-space model singles out a point on the sphere at infinity of hyperbolic space, and the group of hyperbolic isometries preserving this point appears as the group of similarities preserving upper half-space. How does it appear in the paraboloid model?

(c) Show that orthogonal projection from the paraboloid $x_n = x_1^2 + \cdots + x_{n-1}^2$ to the hyperplane $x_n = 0$ in $\mathbf{E}^n$ induces an isomorphism between the group of affine maps preserving the paraboloid and the group of similarities preserving upper half-space.

Problem 2.3.14. We have seen that Lorentz transformations correspond to isometries of $\mathbf{H}^3$, but we did not give a physical interpretation to $\mathbf{H}^3$. Is there any way that people might actually see or experience $\mathbf{H}^3$ in the relativistic universe?

Section "Some computations in hyperbolic space"
% The hyperboloid model and the Klein model
::dual basis
% sphcosines

2.4. Some computations in hyperbolic space

Ultimately, what we seek when we study mathematics is a qualitative understanding. But precise, quantitative manipulations—the nitty-gritty of mathematics—are also important as a way to reach this end, and as a test that our qualitative understanding is correct. The models of hyperbolic space that we developed over the last two sections provide precise representations of hyperbolic objects, but they're intrinsically limited by their lack of symmetry. Informal pictures, on the other hand, can be invaluable as a help to intuition, and with time and experience they become pretty clear and undistorted. However, they are by nature imprecise.

To fill in the gap between these two ways of understanding, it is convenient to know the formulas for measurement in hyperbolic space, so that one can deal with it from the point of view of a surveyor or a builder. Working with formulas tends to be slow and pedestrian, but at least they are precise and the pictures they evoke are undistorted. In this section, then, we develop trigonometric formulas and formulas for the area in the hyperbolic plane.

We work along the lines of section 2.3, investigating the similarities and differences between Lorentz space $\mathbf{E}^{2,1}$ and Euclidean space $\mathbf{E}^3$, and between hyperbolic and spherical geometry. Spherical trigonometry is sometimes presented as an array of easily confused formulas, but these formulas are, in fact, equivalent to statements about dot products of unit vectors in three-space, and can be neatly derived from the formula for inversion of a 3×3 matrix.

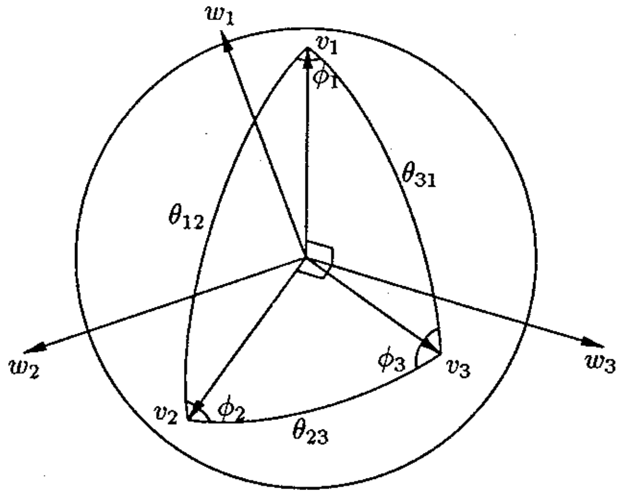
We start with any triple of unit vectors $(v_1, v_2, v_3) \in S^2 \subset \mathbf{E}^3$. If they are linearly independent, so that no great circle, or spherical line, contains all three, they determine a spherical triangle, formed by joining each pair v_i, v_j by a spherical line segment of length $d(v_i, v_j) = \theta_{ij} < \pi$. The *dual basis* to (v_1, v_2, v_3) is another triple (w_1, w_2, w_3) of vectors—but not necessarily unit vectors—in $\mathbf{E}^3$, defined by the conditions $v_i \cdot w_i = 1$ and $v_i \cdot w_j = 0$ if $i \neq j$, for $i, j = 1, 2, 3$. If we let V and W be the matrices with columns v_i and w_i, this can be expressed as $W^t V = I$.

Geometrically, w_i points in the direction of the normal vector of the plane spanned by v_j and v_k, where i, j and k are distinct; it follows that the angle ϕ_i of the spherical triangle $v_1 v_2 v_3$ at v_i is $\pi - \angle(w_j, w_k)$, since the angle between two planes is the complement of the angle between their outward normal vectors (figure 2.20).

To relate all these angles, we consider the matrices $V^t V$ and $W^t W$ of inner products of the two bases, and notice that they are inverse to one another and that

$$V^t V = \begin{pmatrix} 1 & c_{12} & c_{13} \\ c_{12} & 1 & c_{23} \\ c_{13} & c_{23} & 1 \end{pmatrix},$$

::spherical law of cosines
::dual spherical law of cosines
% dualsphcosines



sphcosines

Figure 2.20. Proving the spherical law of cosines

where $c_{ij} = v_i \cdot v_j = \cos\theta_{ij}$. Thus $W^tW = (V^tV)^{-1}$ is a multiple of the matrix of cofactors of V^tV,

2.4.1.
$$W^tW = \frac{1}{\det(V^tV)}\begin{pmatrix} 1-c_{23}^2 & c_{13}c_{23}-c_{12} & c_{12}c_{23}-c_{13} \\ c_{13}c_{23}-c_{12} & 1-c_{13}^2 & c_{12}c_{13}-c_{23} \\ c_{12}c_{23}-c_{13} & c_{12}c_{13}-c_{23} & 1-c_{12}^2 \end{pmatrix}.$$

From this we can easily compute, say,

2.4.2.
$$\cos\phi_3 = -\cos\angle(w_1,w_2) = -\frac{w_1\cdot w_2}{|w_1|\,|w_2|} = \frac{\cos\theta_{12}-\cos\theta_{13}\cos\theta_{23}}{\sin\theta_{13}\sin\theta_{23}},$$

or, in the more familiar notation where A, B, C stand for the angles at v_1, v_2, v_3 and a, b, c stand for the opposite sides,

2.4.3.
$$\cos c = \cos a\cos b + \sin a\sin b\cos C.$$

This is called the *spherical law of cosines.* The *dual spherical law of cosines* is obtained by reversing the roles of (v_1, v_2, v_3) and (w_1, w_2, w_3) with respect to the triangle (figure 2.21): we set the v_i not to the vertices of the triangle, but to unit vectors orthogonal to the planes containing the sides. Then $\phi_k = \pi - \angle(v_i, v_j)$ for i, j, k distinct, and $\theta_{ij} = \angle(w_i, w_j)$. We obtain, for example,

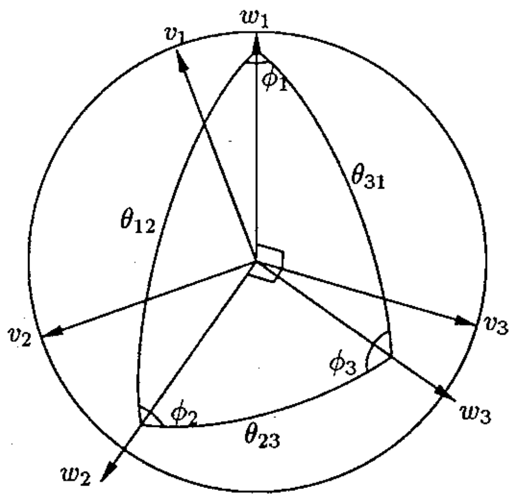
$$\cos\theta_{12} = \cos\angle(w_1,w_2) = \frac{w_1\cdot w_2}{|w_1|\,|w_2|} = \frac{\cos\phi_2\cos\phi_1+\cos\phi_3}{\sin\phi_2\sin\phi_1},$$

or

$$\cos C = -\cos A\cos B + \sin A\sin B\cos c.$$

Exercise 2.4.4. In the limiting case of a very small triangle, we should be able to recover formulas of Euclidean trigonometry. What do you get from the series expansion of the spherical law of cosines when a, b, c are very small? What do you get from the dual spherical law of cosines?

::normalized
% The hyperboloid model and the Klein model
% cases
Proposition "interpretation of the inner product"
imagimag
imagreal
realreal



dualsphcosines

Figure 2.21. The dual spherical law of cosines

We now turn to Lorentz space $\mathbf{E}^{2,1}$. The situation is similar, but somewhat complicated by the fact that non-zero vectors can have real, zero or imaginary length. To cut down the number of cases, we consider only vectors of non-zero length. We may as well assume that they are *normalized* in the sense that they have length 1 or i and their x_0-coordinate is positive if they have length i. We recall from section 2.3 that if a normalized vector $x \in \mathbf{E}^{2,1}$ has length i, it stands for a point on the hyperboloid model H^+ of the hyperbolic plane, just as a unit vector in $\mathbf{E}^3$ gives a point in S^2. If x has length 1, it lies outside the hyperbolic plane, and we denote by $x^\perp$ the trace in the hyperbolic plane of its dual line.

Suppose, then, that x and y are normalized and distinct. The quadratic form Q^-, restricted to the plane spanned by x and y, can have signature $(2,0)$, $(1,0)$, or $(1,1)$, corresponding to the cases where the plane intersects the hyperboloid, is tangent to the cone at infinity, or avoids both (figure 2.22). In each case, we need to interpret the quantity $x \cdot y$, which previously gave the cosine of the angle between two vectors (here the inner product is the one associated with the form Q^-, of course). We may as well do it in arbitrary dimension:

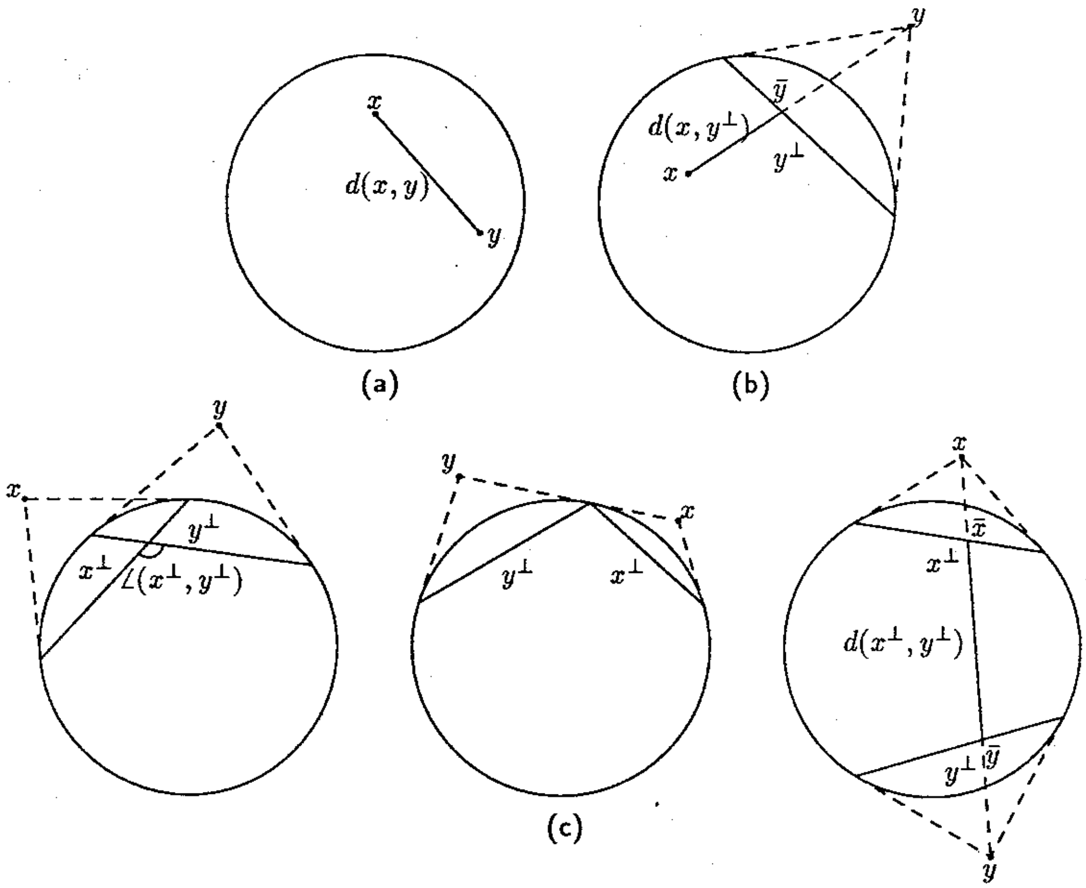
Proposition 2.4.5 (interpretation of the inner product). *If x and y are normalized vectors of non-zero length in $\mathbf{E}^{n,1}$, either*

(a) *$x, y \in H^+$ have length i, and $x \cdot y = -\cosh d(x,y)$; or*

(b) *$x \in H^+$ has length i, y has length 1 and $x \cdot y = \pm \sinh d(x, y^\perp)$; or*

(c) *x and y have length 1, and the hyperplanes $x^\perp, y^\perp \subset \mathbf{H}^n$ are secant, parallel or ultraparallel depending on whether Q^- has signature $(2,0)$, $(1,0)$, or $(1,1)$ on the plane spanned by x and y. In the first case, $x \cdot y = \pm \cos \angle(x^\perp, y^\perp)$; in the second, $x \cdot y = \pm 1$; and in the third, $x \cdot y = \pm \cosh d(x^\perp, y^\perp)$.*

% parametrization of geodesics
% characterization of tangent vectors
% minimum distance implies perpendicularity



cases

Figure 2.22. Interpretation of the inner product for various relative positions of two points. The labels correspond to the cases in proposition 2.4.5; the figures are drawn in the projective model.

Proof of 2.4.5: Let P be the plane spanned by x and y. In cases (a) and (b), P intersects H^+ in a hyperbolic line $L \ni x$, and by exercise 2.3.5 this line is parametrized with velocity 1 by $x \cosh t + v \sinh t$, where v is a unit tangent vector to H^+ at x.

If $y \in H^+$, this implies that $y = x \cosh t + v \sinh t$ for $t = \pm d(x, y)$, depending on the way we chose v. Since x and v are orthogonal (exercise 2.3.2), we get

$$x \cdot y = x \cdot (x \cosh t + v \sinh t) = -\cosh t = -\cosh d(x, y).$$

If, on the other hand, $y \notin H^+$, exercise 2.4.6 shows that the distance $d(x, y^\perp)$ is achieved for the point $\bar{y} = L \cap y^\perp$, because L is the unique perpendicular from x to $y^\perp$. Thus $\bar{y} = x \cosh t + v \sinh t$ for $t = d(x, y^\perp)$, and y, being a linear combination of x and p orthogonal to $\bar{y}$, must be of the form $\pm(x \sinh t + v \cosh t)$. We conclude that

$$x \cdot y = \pm x \cdot (x \sinh t + v \cosh t) = \pm \sinh t = \pm \sinh d(x, y^\perp).$$

The third possibility in (c) is a variation on (a) and (b). Here $L = P \cap H^+$ contains neither x nor y, but we can parametrize it starting at $\bar{x} = L \cap x^\perp$.

% minimum distance implies perpendicularity
Exercise "minimum distance implies perpendicularity"
disambiguating signs
% interpretation of the inner product

Then $x = \pm v$, $\bar{y} = L \cap y^\perp = \bar{x}\cosh t + v \sinh t$ for $t = d(x^\perp, y^\perp)$ (again by exercise 2.4.6), and $y = \pm(\bar{x}\sinh t + v\cosh t)$, so $x \cdot y = \pm\cosh d(x^\perp, y^\perp)$.

We're left with the first two possibilities in (c). If Q^- is positive definite on P, it is indefinite on the orthogonal complement $P^\perp$, so $P^\perp \cap H^+ = x^\perp \cap y^\perp$ is non-empty. Let p be a point in this intersection; to measure $\cos\angle(x^\perp, y^\perp)$ (which is only defined up to sign) it is enough to find tangent vectors to H^+ at p that are normal to $x^\perp$ and $y^\perp$, and take the cosine of their angle. But x and y themselves can serve as such tangent vectors, so $\cos\angle(x^\perp, y^\perp) = \pm x \cdot y$.

If Q^- is positive semidefinite on P, it is also positive semidefinite on $P^\perp$, so $P^\perp \cap H^+ = x^\perp \cap y^\perp$ is empty, but $P^\perp \cap S^{n-1}_\infty$ consists of a single line through the origin. Thus $x^\perp$ and $y^\perp$ meet at infinity—they are parallel. The value of $x \cdot y$ follows from the fact that this case is a limit between the previous two. 2.4.5

Exercise 2.4.6 (minimum distance implies perpendicularity). (a) If $L \subset \mathbf{H}^n$ is a line, $y \in \mathbf{H}^n$ is a point outside L and x is a point on L such that the distance $d(x,y)$ is minimal, the line xy is perpendicular to L.

(b) If $L, M \subset \mathbf{H}^n$ are non-intersecting lines and $x \in L$ and $y \in M$ are points on L and M such that the distance $d(x,y)$ is minimal, xy is perpendicular to L and M.

Exercise 2.4.7. (a) To resolve the ambiguities in signs in proposition 2.4.5, we must assign an orientation to the dual hyperplane of a vector v of real length. This can be done by distinguishing between the two half-spaces determined by $v^\perp$ in $\mathbf{E}^{n,1}$, on the basis of which one contains v; this also distinguishes between the two half-spaces determined by $v^\perp$ in $\mathbf{H}^n$. This done, we can define $d(x, y^\perp)$ in part (b) as a signed quantity; how does the formula read then? What about the various cases in part (c)?

(b) Is there a sensible normalization for vectors of zero length? How would you interpret $v \cdot w$ if either or both vectors have zero length?

Now we can calculate the trigonometric formulas for a triangle in $\mathbf{H}^2$, or, more generally, the intersection with $\mathbf{H}^2$ of a triangle in $\mathbf{RP}^2$. As before, we let (v_1, v_2, v_3) be a basis of normalized vectors in $\mathbf{E}^{2,1}$, forming a matrix V, and we look at its dual basis (w_1, w_2, w_3), whose vectors form a matrix W. Here, V and W are no longer inverse to each other; instead, we can write $W^tSV = I$, where S is a symmetric matrix expressing the inner product associated with Q^- in the canonical basis—here the diagonal matrix with diagonal entries $(-1, 1, 1)$. However, the matrices of inner products, V^tSV and W^tSW, are still inverse to each other:

$$(V^tSV)(W^tSW) = (V^tSV)(V^{-1}W) = V^tSW = (W^tSV)^t = I.$$

Since some of the v_i may have imaginary length, V^tSV no longer has all ones in the diagonal; instead, it looks like this:

$$V^tSV = \begin{pmatrix} \varepsilon_1 & c_{12} & c_{13} \\ c_{12} & \varepsilon_2 & c_{23} \\ c_{13} & c_{23} & \varepsilon_3 \end{pmatrix},$$

% sphericalcofactor
% hypcosines
% imagimag
% realreal
% disambiguating signs
% hyperboliccofactor

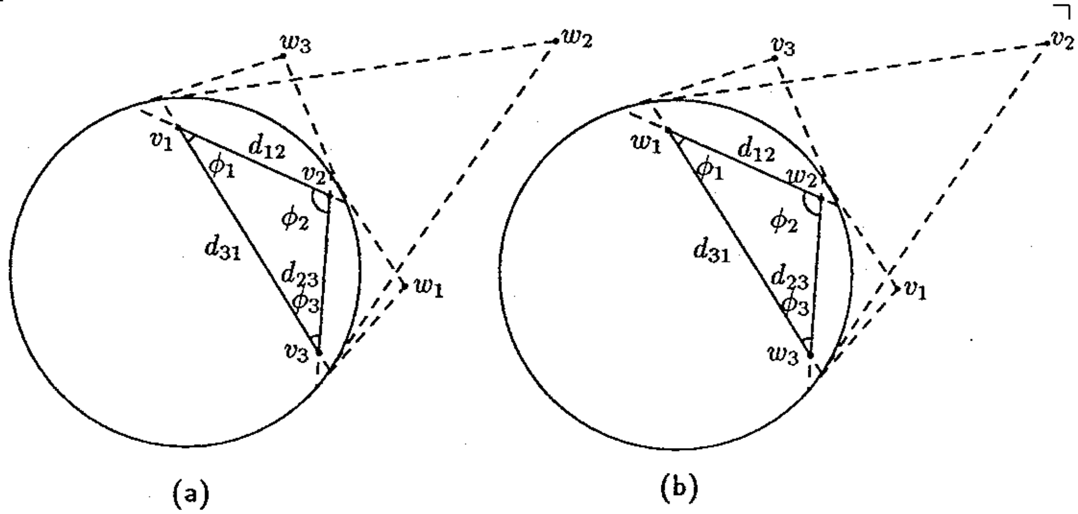
where $\varepsilon_i = v_i \cdot v_i = \pm 1$. It follows, as before, that the matrix of inner products of the w_i is

2.4.8.

$$W^t S W = \frac{1}{\det(V^t S V)} \begin{pmatrix} \varepsilon_2\varepsilon_3 - c_{23}^2 & c_{13}c_{23} - \varepsilon_3 c_{12} & c_{12}c_{23} - \varepsilon_2 c_{13} \\ c_{13}c_{13} - \varepsilon_3 c_{12} & \varepsilon_1\varepsilon_3 - c_{13}^2 & c_{12}c_{13} - \varepsilon_1 c_{23} \\ c_{12}c_{23} - \varepsilon_2 c_{13} & c_{12}c_{13} - \varepsilon_1 c_{23} & \varepsilon_1\varepsilon_2 - c_{12}^2 \end{pmatrix}.$$

This can be used in the same way as equation 2.4.1, but we have to be careful about signs, since normalizing a vector of imaginary length can require multiplication by a negative scalar.

Take the case when v_1, v_2, v_3 all have imaginary length, so they form a triangle with all three vertices in $\mathbf{H}^2$, as in figure 2.23(a). By proposition 2.4.5(a), the interpretation of $c_{ij} = v_i \cdot v_j$ is in terms of the distances d_{ij}, namely, $c_{ij} = -\cosh d_{ij}$. The vectors w_i have real length, and their duals $w_i^\perp$ represent the sides of the triangle; by proposition 2.4.5(c) and exercise 2.4.7(a), $w_i \cdot w_j/(|w_i|\,|w_j|) = -\cos\phi_k$, where i, j, k are distinct and ϕ_k is the interior angle at v_k.



hypcosines

Figure 2.23. Two ways to see a triangle in the hyperbolic plane. We first arrange the basis of unit vectors (v_1, v_2, v_3) to match the triangle's vertices (a), and derive a formula relating side lengths. We then make (v_1, v_2, v_3) correspond to the duals of the edges (b), and obtain a formula relating angles.

Setting all the ε_i to -1 in equation 2.4.8, we see that we must switch the sign of the matrix before normalizing, so the diagonal entries are positive. We then get

$$\cos\phi_3 = -\frac{w_1 \cdot w_2}{|w_1|\,|w_2|} = \frac{c_{13}c_{23} + c_{12}}{\sqrt{c_{13}^2 - 1}\sqrt{c_{23}^2 - 1}} = \frac{\cosh d_{13}\cosh d_{23} - \cosh d_{12}}{\sinh d_{13}\sinh d_{23}}.$$

::hyperbolic law of cosines
% hypcosines
% hyperbolic law of cosines
::hyperbolic Pythagorean theorem
% hyperbolic law of cosines
::hyperbolic law of sines
% hyperboliccofactor
% polygonsines
% interpretation of the inner product

Letting A, B, C stand for the angles at v_1, v_2, v_3 and a, b, c for the opposite sides, we obtain the *hyperbolic law of cosines*

2.4.9. $$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos C.$$

To obtain the dual law, we start with the v_i outside hyperbolic space, dual to the sides of the triangle under consideration, as shown in figure 2.23(b). Then $c_{ij} = -\cos\phi_k$ for i, j, k distinct, and $\left|(w_i \cdot w_j)/(|w_i|\,|w_j|)\right| = \cosh d_{ij}$. Thus we have

$$\cosh d_{12} = \left|\frac{w_1 \cdot w_2}{|w_1|\,|w_2|}\right| = \frac{c_{13}c_{23} + c_{12}}{\sqrt{1-c_{13}^2}\sqrt{1-c_{23}^2}} = \frac{\cos\phi_1\cos\phi_2 + \cos\phi_3}{\sin\phi_1\sin\phi_2},$$

or

$$\cos C = -\cos A \cos B + \sin A \sin B \cosh c.$$

The formulas for a right triangle are worth mentioning separately, since they are particularly simple. Switching A and C (as well as a and c) in the previous formula and setting $C = \pi/2$ we get

$$\cosh a = \frac{\cos A}{\sin B} \qquad \text{if } C = \pi/2$$

(note that $\cos A = \sin B$ in a Euclidean right triangle). From equation 2.4.9 we obtain the *hyperbolic Pythagorean theorem*:

$$\cosh c = \cosh a \cosh b \qquad \text{if } C = \pi/2.$$

From the formula for $\cosh b$ analogous to equation 2.4.9, by making the substitutions $\cosh c = \cosh a \cosh b$ and $\cos B = \cosh b \sin A$ and using the identity $\cosh^2 a = 1 + \sinh^2 a$, we get

$$\sin A = \frac{\sinh a}{\sinh c} \qquad \text{if } C = \pi/2.$$

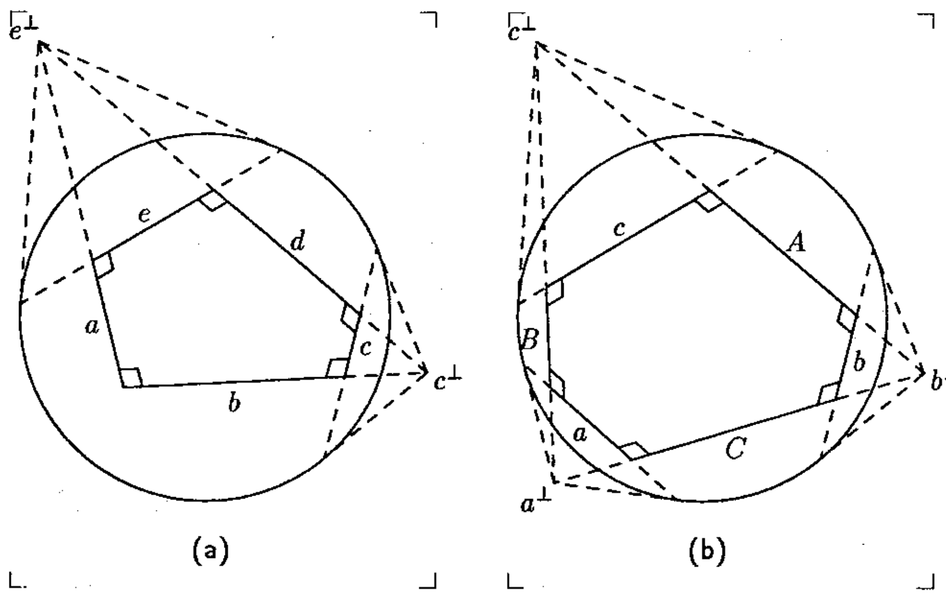
Now given any triangle, the altitude h corresponding to side c satisfies $\sinh h = \sinh a / \sin A$ and also $\sinh h = \sinh b / \sin B$. This proves the *hyperbolic law of sines*, valid for any hyperbolic triangle:

$$\frac{\sinh a}{\sin A} = \frac{\sinh b}{\sin B} = \frac{\sinh c}{\sin C}.$$

So far we've applied equation 2.4.8 to triangles entirely inside hyperbolic space and to their duals. Mixed cases can also be interesting; for example, figure 2.24(a) shows how a pentagon having five right angles can be thought of as a right triangle with two vertices outside the circle at infinity; they are represented in the pentagon by their duals, which form two non-adjacent sides. By using proposition 2.4.5 and keeping track of signs you can convince yourself that the hyperbolic Pythagorean theorem acquires the form

$$\sinh a \sinh b = \cosh d.$$

::hexagonal law of sines
% polygonsines
% The Teichmüller space of a surface
% hyperbolic law of cosines
% spherical law of cosines
% projective transformations of hyperbolic space
Gauss–Bonnet theorem
Gauss



polygonsines

Figure 2.24. Trigonometry in an all-right pentagon and in an all-right hexagon

There follows a *hexagonal law of sines* for all-right hexagons, since we can draw a line perpendicular to two opposite sides to obtain two all-right pentagons:

$$\frac{\sinh a}{\sinh A} = \frac{\sinh b}{\sinh B} = \frac{\sinh c}{\sinh C},$$

in the notation of figure 2.24(b). Right-angled hexagons can also be seen as triangles with all three vertices outside infinity, as shown in the same figure. Such hexagons are useful in the study of hyperbolic structures on surfaces (see section 3.8).

We are accustomed to the notion that choice of scale in Euclidean space is arbitrary: it does not essentially matter whether we measure in feet or meters. Figures can be scaled up or down arbitrarily.

The same is not true in hyperbolic and spherical geometry. If you double the sides of a hyperbolic or spherical triangle, its angles, given by equations 2.4.9 and 2.4.3, are no longer the same. There is no hyperbolic or spherical analogue for similarity transformations (see problem 2.3.11).

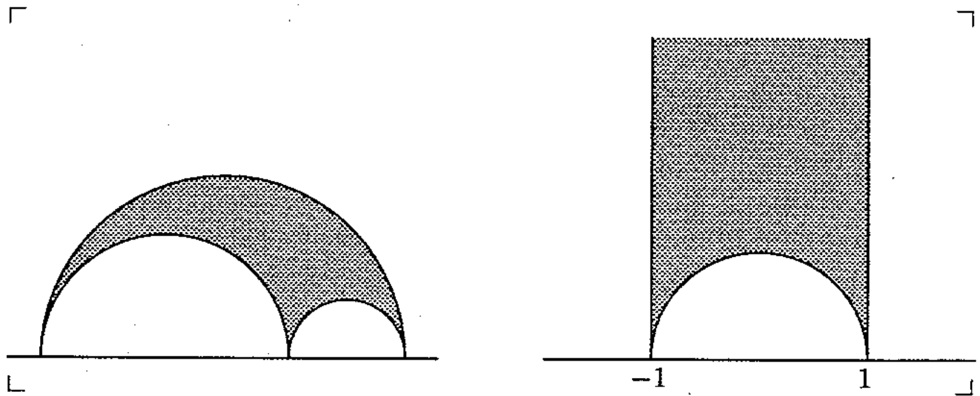
This strong dependency between size and angles can be seen even more clearly in terms of area. For a nice open region in the Euclidean plane—say one with a connected, piecewise smooth boundary—the total amount of curvature of the boundary is always 2π, and in particular the sum of the (signed) exterior angles of a polygon is 2π. For a region of the hyperbolic plane, the total curvature increases with the area; this is a special case of the Gauss–Bonnet theorem, a very general and profound result of differential geometry. We won't state the Gauss–Bonnet theorem in any more generality here; instead we present an elegant method, also due to Gauss, to calculate the area of a hyperbolic triangle using only elementary means. We start with an

Proposition "ideal triangles"
% idealtriangle
% upper half-space metric
Proposition "area of hyperbolic triangles"
% almostideal

ideal triangle, one whose three "vertices" are at infinity. Although unbounded, ideal triangles have finite area:

Proposition 2.4.10 (ideal triangles). *All ideal triangles are congruent, and have area π.*

Proof of 2.4.10: Using the upper half-plane model of $\mathbf{H}^2$, it is easy to see that any ideal triangle can be transformed by isometries so as to match a model triangle with vertices ∞, $(-1,0)$ and $(1,0)$ (figure 2.25).



idealtriangle

Figure 2.25. All ideal triangles are congruent. Given any ideal triangle, we can send one of its vertices to ∞ by inversion, then apply a Euclidean similarity to send the remaining two vertices to $(-1,0)$ and $(1,0)$.

Now let the coordinates of the upper half-plane be x and y, with the x-axis as the boundary. The model triangle is the region given by $-1 \le x \le 1$ and $y \ge \sqrt{1-x^2}$, with hyperbolic area element $(1/y^2)\,dx\,dy$ (equation 2.2.9). Thus the area is

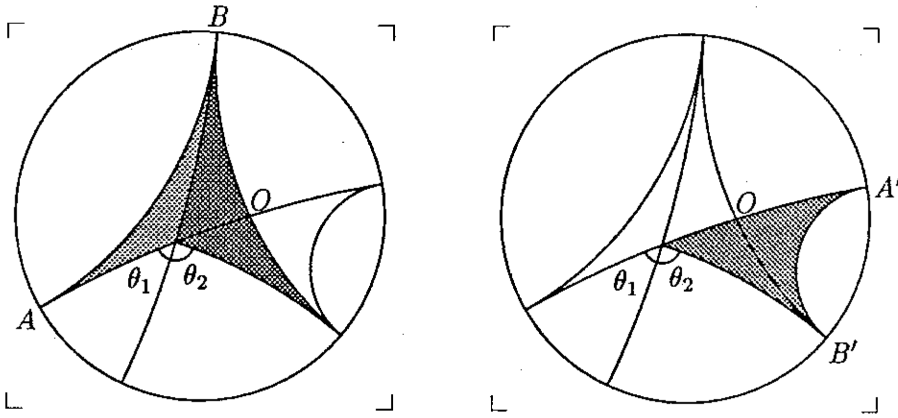
$$\int_{-1}^{1}\int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\,dy\,dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx = \int_{-\pi/2}^{\pi/2} \cos^2\theta\,d\theta = \pi. \qquad \boxed{2.4.10}$$

Proposition 2.4.11 (area of hyperbolic triangles). *The area of a hyperbolic triangle is π minus the sum of the interior angles (the angle being zero for a vertex at infinity).*

Proof of 2.4.11: When all angles are zero we have an ideal triangle. We next look at $\frac{2}{3}$-ideal triangles, those with two vertices at infinity. Let $A(\theta)$ denote the area of such a triangle with angle $\pi - \theta$ at the finite vertex. This is well-defined because all $\frac{2}{3}$-ideal triangles with the same angle at the finite vertex are congruent—the reasoning is similar to that for ideal triangles.

Gauss's key observation is that A is an additive function, that is, $A(\theta_1 + \theta_2) = A(\theta_1) + A(\theta_2)$, for $\theta_1, \theta_2, \theta_1 + \theta_2 \in (0,\pi)$. The proof of this follows from figure 2.26. It follows that A is a $\mathbf{Q}$-linear function from $(0,\pi)$ to $\mathbf{R}$. It is also continuous, so it must be $\mathbf{R}$-linear. But $A(\pi)$ is the area of an ideal

% ideal triangles
% finitetriangle
Exercise "spherical area"
Corollary "area of hyperbolic polygons"

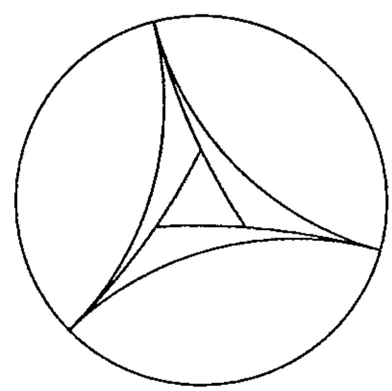


almostideal

Figure 2.26. Area of $\frac{2}{3}$-ideal triangles. By definition, the areas of the shaded triangles on the left are $A(\theta_1)$ and $A(\theta_2)$. Likewise, the area of the shaded triangle on the right is $A(\theta_1 + \theta_2)$. But the shaded areas in the two figures coincide, because the triangles OAB and $OA'B'$ are congruent. Therefore $A(\theta)$ is an additive function of θ; this is used to compute the area of $\frac{2}{3}$-ideal triangles.

triangle, which is π by proposition 2.4.10; it follows that $A(\theta) = \theta$, and the area of a $\frac{2}{3}$-ideal triangle is the complement of the angle at the finite vertex.

A triangle with two or three finite vertices can be expressed as the difference between an ideal triangle and two or three $\frac{2}{3}$-ideal ones, as shown in figure 2.27. You should check the details. [2.4.11]



finitetriangle

Figure 2.27. Area of general hyperbolic triangles. If you subtract a finite hyperbolic triangle from a suitable ideal triangle, you get three $\frac{2}{3}$-ideal triangles. Adding up angles and areas gives proposition 2.4.11.

Exercise 2.4.12 (spherical area). Derive the formula for the area of a spherical triangle by an analogous procedure, starting with 4π as the area of the sphere.

Corollary 2.4.13 (area of hyperbolic polygons). *The sum S of the interior angles of a planar hyperbolic polygon is always less than the sum of the angles of a Euclidean polygon with the same number n of sides, and the deficiency $(n-2)\pi - S$ is the area of the polygon.*

% 3octagons
% shapes of Euclidean polygons

Proof of 2.4.13: Subdivide the polygon into triangles, as in the Euclidean case. 2.4.13

Exercise 2.4.14. (a) What is the area of a surface of genus two made from the regular octagon of figure 1.13(b)?

(b) (Harder.) Show that *any* surface of genus two of constant curvature -1 has the same area.

Exercise 2.4.15. What is a good definition for the area of a non-simple polygon? What is the formula for the area of a non-simple hyperbolic polygon? Compare problem 2.3.12.

Section "Hyperbolic isometries"
% the Möbius group
% the algebraic study of isometries of $\mathbf{H}^n$
::axis
Proposition "axis is unique"
::screw motion
% isometries of $\mathbf{E}^3$
::elliptic
::rotation
::hyperbolic
% banana
::translation
loxodromic

2.5. Hyperbolic isometries

We now turn to the qualitative study of the isometries of hyperbolic space, which, as we saw in problem 2.2.14, form a large group. We could use linear algebra for this purpose, but we'll instead use direct geometric constructions, the better to develop our hyperbolic intuition. The algebraic approach is taken in problem 2.5.23.

We start in three dimensions. Let $g : \mathbf{H}^3 \to \mathbf{H}^3$ be an orientation-preserving isometry other than the identity. An *axis* of g is any line L that is invariant under g and on which g acts as a (possibly trivial) translation.

Proposition 2.5.1 (axis is unique). *A non-trivial orientation-preserving isometry of* $\mathbf{H}^3$ *can have at most one axis.*

Proof of 2.5.1: Suppose that L and M are distinct axes for an orientation-preserving isometry g. If g fixes both L and M pointwise, take a point x on M but not on L. Then g fixes the plane containing L and x, because it fixes three non-collinear points on it. Since g preserves orientation, it is the identity.

If, on the other hand, L is translated by g, we have $d(x, M) = d(g(x), M)$ for $x \in L$, so the function $d(x, M)$ is periodic and therefore bounded. But two distinct lines cannot remain a bounded distance from each other in both directions, since that would imply they have the same two endpoints on the sphere at infinity. $\boxed{2.5.1}$

Exercise 2.5.2. Find a non-trivial orientation-preserving isometry of $\mathbf{H}^3$ that leaves invariant more than one line.

Any orientation-preserving isometry of $\mathbf{E}^3$ is either a translation, a rotation about some axis, or a *screw motion*, that is, a rotation followed by a translation along the axis of rotation. (Exercise 2.5.6 asks you to prove this.) The situation in $\mathbf{H}^3$ is somewhat richer, and has its own special terminology.

If a non-trivial orientation-preserving isometry g of $\mathbf{H}^3$ has an axis that is fixed pointwise, it is called an *elliptic* isometry, or a *rotation* about its axis. In this case the orbit of a point p off the axis—the set of points $g^k(p)$, for $k \in \mathbf{Z}$—lies on a circle around the axis.

If g has an axis that is translated by a non-trivial amount, it is called *hyperbolic*. There are two possibilities here: the orbit of a point off the axis may lie on a plane, always on the same side of the axis; it is in fact contained in an equidistant curve, like the one shown in figure 1.11. In this case we say that g is a *translation*. Alternatively, the orbit can be the vertices of a polygonal helix centered around the axis; in this case g is a screw motion, as can be seen by applying a compensatory translation. (An alternate usage of the word "hyperbolic" specializes it to what we're calling translations, in which case "loxodromic" designates a screw motion. This distinction is not very useful, and we'll not adopt it.)

By the proposition above, no transformation can be at the same time elliptic and hyperbolic. But there are isometries that are neither elliptic nor

::parabolic
% parallelism in hyperbolic space
::parallel
Lemma "common perpendicular for lines in $\mathbf{H}^3$"
% area of hyperbolic triangles
% minimum distance implies perpendicularity
Proposition "finding the axis"

hyperbolic: they are called *parabolic.* For instance, any isometry of $\mathbf{H}^3$ that appears as a Euclidean translation parallel to the bounding plane in the upper half-space model is parabolic.

The proof of the next proposition describes a geometric construction to locate the axis of a non-trivial orientation-preserving isometry of $\mathbf{H}^3$, if it has one. We will need an elementary fact about pairs of lines; recall (cf. exercise 2.3.10) that two lines in $\mathbf{H}^3$ are called *parallel* if they have a common endpoint on S^2_∞, or, equivalently, if the distance between them approaches zero at one end or the other.

Lemma 2.5.3 (common perpendicular for lines in $\mathbf{H}^3$). *Two distinct lines in $\mathbf{H}^3$ are either parallel, or they have a unique common perpendicular.*

Proof of 2.5.3: Let the lines be X and Y, and consider the distance function $d(x,y)$ between points $x \in X$ and $y \in Y$. If the lines are not parallel, this function goes to ∞ as either x or y or both go to ∞; therefore it has a minimum, attained at points x_0 and y_0.

If the minimum is zero, that is, if the lines cross, any common perpendicular must go through the intersection point, otherwise we'd have a triangle with two right angles, which is impossible by proposition 2.4.11. As there is a unique line orthogonal to the plane spanned by X and Y and passing through their intersection point, the lemma is proved in this case. (Notice that this part is false in dimension greater than three.)

If the minimum distance is not zero, the line between x_0 and y_0 is a perpendicular by exercise 2.4.6. If there were another common perpendicular, we'd obtain a quadrilateral in space with all right angles (although conceivably its sides could cross). Subdividing the quadrilateral by a diagonal, we'd get two plane triangles, whose angles add to at least 2π; this is again impossible. $\boxed{2.5.3}$

If L is a line in $\mathbf{H}^3$, we denote by r_L the reflection in L, which is the rotation of π about L.

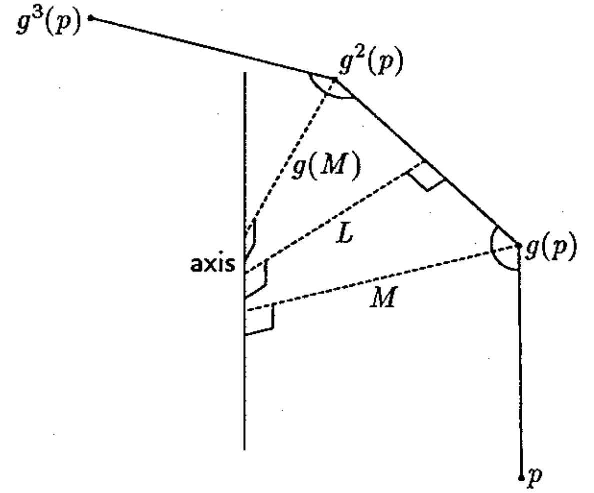
Proposition 2.5.4 (finding the axis). *Any non-trivial orientation-preserving isometry g of $\mathbf{H}^3$ can be written in the form $g = r_L \circ r_M$, where the lines L and M are parallel, secant or neither depending on whether g is parabolic, elliptic or hyperbolic. The axis of g is the common perpendicular of L and M, if it exists.*

Proof of 2.5.4: Take <u>any point</u> p such that $g(p) \neq p$. If $g^2(p) = p$, the midpoint q of the line segment $\overline{p\,g(p)}$ is fixed by g, and the plane through q perpendicular to the line $p\,g(p)$ is invariant. Since g reverses the orientation of this plane, it must act on it as a reflection, fixing a line K. Therefore, $g = r_K$ is elliptic of order two. In this case, one can take L and M to be two orthogonal lines, both orthogonal to K at a point $x \in K$.

If p, $g(p)$ and $g^2(p) \neq p$ are collinear, p is fortuitously on the axis of g, and g is hyperbolic. We can replace p by some other point not on this axis, and reduce to the next case.

% axisconstruction
% common perpendicular for lines in $\mathbf{H}^3$
% finding the axis
Exercise "isometries of $\mathbf{E}^3$"
% finding the axis

In the remaining case, we let M be the bisector of the angle $p\,g(p)g^2(p)$, so that r_M fixes $g(p)$ and interchanges p with $g^2(p)$. To define L, we look at the dihedron with edge $g(p)g^2(p)$ whose sides contain p and $g^3(p)$, respectively, and take L as the line that bisects this dihedron and is also a perpendicular bisector of the segment $\overline{p\,g(p)}$, as shown in figure 2.28. (The dihedral angle



axisconstruction

Figure 2.28. The axis of a three-dimensional isometry. The axis of an isometry g of $\mathbf{H}^3$ may be constructed, in the generic case, by connecting the orbit of a point p in a polygonal path.

along $p\,g(p)$ may be 0 or π, but this doesn't cause problems.) By symmetry, r_L interchanges $g(p)$ with $g^2(p)$, and also p with $g^3(p)$. Therefore, $r_L \circ r_M$ sends p to $g(p)$, $g(p)$ to $g^2(p)$ and $g^2(p)$ to $g^3(p)$. Since $r_L \circ r_M$ and g agree at three non-collinear points, they agree on the whole plane containing these three points. Therefore they agree everywhere, since they both preserve orientation.

The common perpendicular of L and M, if it exists, is the axis of g, because it is invariant under $r_L \circ r_M$, which acts on it as a translation. The sorting into cases now follows from lemma 2.5.3 and from the definitions of parabolic, elliptic and hyperbolic transformations. [2.5.4]

Exercise 2.5.5. Using the decomposition of proposition 2.5.4, show that any parabolic orientation-preserving isometry of $\mathbf{H}^3$ is conjugate to a Euclidean translation of the upper half-space model.

Exercise 2.5.6 (isometries of $\mathbf{E}^3$). Prove that any orientation-preserving isometry of $\mathbf{E}^3$ is either a translation, a rotation or a screw motion. Give a geometric construction for the axis of the transformation, in the latter two cases.

The proof of proposition 2.5.4 also applies to any two-dimensional isometry, whether it preserves or reverses orientation, since such an isometry can be extended to a three-dimensional orientation-preserving isometry. But three-dimensional orientation-reversing isometries, or arbitrary isometries in higher dimension, can have fixed point sets other than lines: for example, a reflection in a plane of any dimension, or a compound rotation about a plane of any

::elliptic
::hyperbolic
::parabolic
Problem "orientation-reversing isometries of $\mathbf{H}^3$"
% the algebraic study of isometries of $\mathbf{H}^n$
::convex function
::strictly convex
::product
Theorem "distance function is convex"
% distinequality
% area of hyperbolic triangles

even codimension. Any transformation with a fixed point in hyperbolic space is still called *elliptic*. An isometry that translates an axis is *hyperbolic*, and as before the axis is unique. Any other non-trivial transformation is *parabolic*.

Problem 2.5.7 (orientation-reversing isometries of $\mathbf{H}^3$). Classify orientation-reversing isometries of $\mathbf{H}^3$. One approach is to modify the construction above, and express an orientation-reversing isometry as the composition of reflection in a plane with reflection in a line. Another approach is to exploit the fact that the square of an orientation-reversing isometry is orientation-preserving. Which isometries of $\mathbf{H}^3$ have orientation-reversing square roots? What are the square roots of the identity? What are the square roots of other isometries which have them?

To study isometries in arbitrary dimension, we continue with the geometric point of view and develop a method, based on the convexity of the distance function, that will be important later, because it can be adapted to other negatively curved metrics on $\mathbf{R}^n$. For another, algebraic, method, see problem 2.5.23.

A *convex function* on a Riemannian manifold is a function f such that, for every geodesic γ, parametrized at a constant speed, the induced function $f \circ \gamma$ is convex. In other words, for every $t \in (0,1)$,

$$f \circ \gamma(t) \leq t f \circ \gamma(0) + (1-t) f \circ \gamma(1).$$

If the inequality is strict for all non-constant γ, we say that f is *strictly convex*.

The *product* of two Riemannian manifolds is the product manifold, with the Riemannian metric that is the sum of the metrics on the factors. It is easy to show that a curve in the product manifold is a geodesic (parametrized at constant speed) if and only if each projection in a factor is also a geodesic (parametrized at constant speed). In particular, a geodesic in $\mathbf{H}^n \times \mathbf{H}^n$ is a curve each of whose projections is a hyperbolic line or a point.

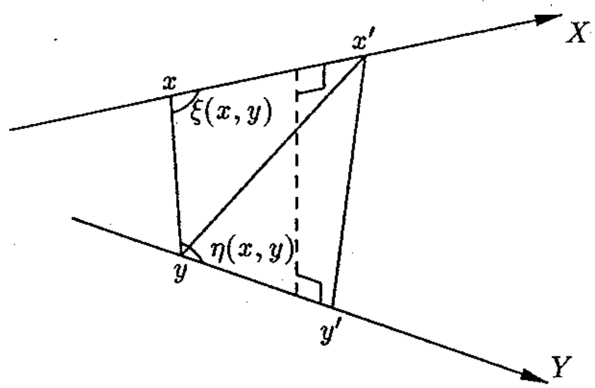
Theorem 2.5.8 (distance function is convex). *The distance function $d(x,y)$, considered as a map $d : \mathbf{H}^n \times \mathbf{H}^n \to \mathbf{R}$, is convex. The composition $d \circ \gamma$ is strictly convex for any geodesic γ in $\mathbf{H}^n \times \mathbf{H}^n$ whose projections to the two factors are distinct lines.*

Proof of 2.5.8: We can assume $n \geq 3$, since $\mathbf{H}^n$ is isometrically embedded in $\mathbf{H}^{n+1}$. We also assume, for now, that the projections X and Y of γ are lines that don't lie on the same plane—in particular, they don't meet, even at infinity.

We parametrize X and Y by arc length, and use x and y to refer to points on X or Y as well as their parameters. Given $x \in X$ and $y \in Y$, we let $\xi(x,y)$ be the angle between the segment $\overline{xy}$ and the positive ray determined on X by x; similarly, $\eta(x,y)$ will be the angle between $\overline{yx}$ and the positive ray determined on Y by y (figure 2.29).

Applying the area formula (proposition 2.4.11) to a triangle with two vertices on X and one on Y, we see that ξ increases monotonically with x when y is fixed. The map $(\xi,\eta) : \mathbf{R}^2 \to (0,\pi) \times (0,\pi)$ is clearly

% area of hyperbolic triangles
% derivative of distance



distinequality

Figure 2.29. Derivative of distance is monotone. Positions along the lines X and Y uniquely determine the angles between the connecting segment and the lines. This can be used to show that the distance function between two lines is convex.

differentiable; we show that it is one-to-one. We do this by looking at $|\xi(x',y')-\xi(x,y)|+|\eta(x',y')-\eta(x,y)|$, where (x,y) and (x',y') are arbitrary pairs of points on X and Y. We can assume that $x' \geq x$. If we also have $y' \geq y$, as in the figure, we can write

$$\begin{aligned}|\xi(x',y')-\xi(x,y)| &+ |\eta(x',y')-\eta(x,y)| \\ &\geq \xi(x',y')-\xi(x,y)+\eta(x',y')-\eta(x,y) \\ &= \big(\xi(x',y')-\xi(x',y)\big)+\big(\xi(x',y)-\xi(x,y)\big) \\ &\quad + \big(\eta(x',y')-\eta(x',y)\big)+\big(\eta(x',y)-\eta(x,y)\big).\end{aligned}$$

Now

2.5.9. $$\xi(x',y')-\xi(x',y) > -\angle yx'y'$$

unless $y=y'$. This is just the triangle inequality for spherical triangles, and the inequality is strict because X and Y are not coplanar. We also have $\eta(x',y')-\eta(x',y)=\angle yx'y'+\text{area}\,\Delta yx'y'$, by the area formula (proposition 2.4.11). Using similar relations for the other differences, we get

2.5.10. $$|\xi(x',y')-\xi(x,y)|+|\eta(x',y')-\eta(x,y)| > \text{area}\,\Delta xyx' + \text{area}\,\Delta yx'y'$$

unless $x=x'$ and $y=y'$.

The case $y' \leq y$ is handled the same way, starting with the inequality $|\xi(x',y')-\xi(x,y)|+|\eta(x',y')-\eta(x,y)| \geq \xi(x',y')-\xi(x,y)-\big(\eta(x',y')-\eta(x,y)\big)$. We conclude that $|\xi(x',y')-\xi(x,y)|+|\eta(x',y')-\eta(x,y)|$ is always positive if $(x,y)\neq(x',y')$, so (ξ,η) is a one-to-one map.

On the other hand, exercise 2.5.13 shows that the function d has gradient $\nabla d=(d_x,d_y)=(-\cos\xi,-\cos\eta)$; it follows that ∇d is also a one-to-one, differentiable map from $\mathbf{R}^2$ to $(-1,1)\times(-1,1)$.

% injective derivative
% spherical triangle inequality
% derivative of distance
% injective derivative

Now the derivative of ∇d is

$$H = \begin{pmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{pmatrix},$$

the hessian of the distance function. A smooth function is convex if and only if its second derivative along any line of the domain is everywhere non-negative, and is strictly convex if the second derivative is everywhere positive. So we have almost arrived at our desired destination. The only thing we still need to do is to show that H is positive definite.

For this we need an infinitesimal version of the argument above. It is tempting to carry this out using the terms in equation 2.5.10 involving the area. However, we avoid doing this, in order to have a valid argument when we are not working in hyperbolic space, but in any space of non-positive curvature. For example, in Euclidean space the area terms would not be present.

Let y_v, y'_v and X_v be the points on the visual sphere at x' that correspond to y, y' and the positive endpoint of X. If d_v denotes distance on the visual sphere, we have $d_v(X_v, y_v) = \xi(x', y)$, $d_v(X_v, y'_v) = \xi(x', y')$ and $d_v(y_v, y'_v) = \angle yx'y'$, and equation 2.5.9 follows from the triangle inequality, as already observed. But, in fact, we can write a stronger inequality, using exercise 2.5.13(c) and the assumption that y_v, y'_v and X_v are not collinear on the sphere. Namely, there exists a positive number ε, depending only on x and y, such that

2.5.11. $$\xi(x', y') - \xi(x', y) + \angle yx'y' \geq \varepsilon|y - y'|.$$

Combining this with a similar equation for η, we get the following counterpart for equation 2.5.10:

2.5.12. $$\left|\xi(x', y') - \xi(x, y)\right| + \left|\eta(x', y') - \eta(x, y)\right| > \varepsilon(|x - x'| + |y - y'|)$$

unless $x = x'$ and $y = y'$.

It follows that the ratio of the image area to the domain area under the differentiable homeomorphism (ξ, η) is at least ε^2. Therefore the hessian of the distance function has a determinant that is always strictly positive or always strictly negative. The distance function is a proper map from $\mathbf{R}^2$ to $[0, \infty)$, and therefore attains its minimum. At the minimum the hessian cannot have negative determinant, so the determinant is everywhere positive. Since the diagonal entries of H are positive, H is positive definite. This completes the proof that if X and Y are not coplanar the distance function is strictly convex.

Now convexity is a closed condition, so it still holds even when the projections X and Y of γ are coplanar, or even when one or both reduce to a point, because all these cases are limits of the case above. But for strict convexity the limiting argument does not work, and indeed the distance is not strictly convex if X or Y is a point, or if they are the same line. So we must find other means to prove strict convexity when X and Y are distinct lines.

At this point we must give up any hope of a proof that works for any space of non-positive curvature—think of parallel lines in $\mathbf{E}^n$. Instead, we're free to

% injective derivative
% infinitesimal injectivity
Exercise "derivative of distance"
Definition "Busemann's definition of negative curvature"
Finsler manifolds
% distance function is convex
% distance function is convex
::translation distance
% distance function is convex

use equation 2.5.10, which still holds with $\geq$ instead of $>$. This inequality implies that equation 2.5.12 holds, so long as $x \notin Y$ and $y \notin X$. Using the gradient ∇d as above we conclude that $d : X \times Y \to \mathbf{R}$ is strictly convex except possibly along two lines of the form $x = x_0$ and $y = y_0$. In particular, $d \circ \gamma : \mathbf{R} \to \mathbf{R}$ is strictly convex except at a maximum of two points; but a convex function that is strictly convex in the complement of a finite set of points is strictly convex everywhere. 2.5.8

Exercise 2.5.13 (derivative of distance). (a) Let x_0 be a fixed point in hyperbolic space. Show that the derivative at x of the distance function $x \mapsto d(x, x_0)$ is the unit vector at x pointing along the geodesic from x_0 to x away from x_0.

Let $\alpha(t)$ be a differentiable curve in hyperbolic space, parametrized by arc length, and suppose that $x_0 \neq \alpha(t_0)$. Prove that the derivative of the distance $d\big(x_0, \alpha(t)\big)$ with respect to t at t_0 is $\cos\theta$, where θ is the angle between the geodesic from x_0 to $\alpha(t_0)$ and $\alpha'(t_0)$.

(b) Show that this is also true in Euclidean space.

(c) Show that it is true in the sphere, provided that x and x_0 are not antipodal.

(d) (For those who know some riemannian geometry.) Show that the same result holds in any riemannian manifold, provided that x does not lie in the cut locus of x_0.

One standard starting point for a study of spaces of negative curvature is to assume the following property as an alternative definition of "negative curvature" [Bus55, chapter 5]:

Definition 2.5.14 (Busemann's definition of negative curvature). Let ABC be a triangle and let X be the midpoint of AB and Y the midpoint of AC. Then XY has length less than half the length of BC.

This definition applies in many situations, including the study of Finsler manifolds, where other definitions from riemannian manifolds do not apply. Theorem 2.5.8 implies that hyperbolic space satisfies this property, as does any simply connected complete riemannian manifolds of strictly negative sectional curvature.

Exercise 2.5.15. Assume that Busemann's definition of negative curvature is satisfied, and do not assume that the distance function is convex. Let AB and CD be geodesics, with midpoints X and Y. Prove that $d(X,Y) < \frac{1}{2}\big(d(A,C) + d(B,D)\big)$, unless all the points lie on the same geodesic. (Hint: draw a diagonal.) What happens when all the points lie on the same geodesic?

Prove that theorem 2.5.8 holds for any space satisfying Busemann's definition of negative curvature.

The *translation distance* of an isometry $g : \mathbf{H}^n \to \mathbf{H}^n$ is the function $d_g(x) = d\big(x, g(x)\big)$. By applying theorem 2.5.8 to to the graph of g, which is a geodesic-preserving embedding of $\mathbf{H}^n$ in $\mathbf{H}^n \times \mathbf{H}^n$, we get:

Corollary "translation distance is convex"
Proposition "classification of isometries of $\mathbf{H}^n$"
% translation distance is convex
% axis is unique
fundamental domain
::hyperbolic mean
::center of mass

Corollary 2.5.16 (translation distance is convex). *For any isometry g of $\mathbf{H}^n$, the translation distance d_g is a convex function on $\mathbf{H}^n$. It is strictly convex except along lines that map to themselves.*

Proposition 2.5.17 (classification of isometries of $\mathbf{H}^n$). *Let g be an isometry of $\mathbf{H}^n$.*

(a) *g is hyperbolic if and only if the infimum of d_g is positive. This infimum is attained along a line, which is the unique axis for g.*

(b) *g is parabolic if and only if the infimum of d_g is not attained. This infimum is then zero; g fixes a unique point p on S^{n-1}_∞, and acts as a Euclidean isometry in the upper half-space model with p at ∞.*

(c) *g is elliptic if and only if d_g takes the value zero. The set $d_g^{-1}(0)$ is a hyperbolic subspace of dimension k, for $0 \le k \le n$.*

Proof of 2.5.17: If d_g attains a positive infimum at some point x, it also attains the infimum at $g(x)$. By convexity, d_g has the same value on the line segment joining x and $g(x)$, so by corollary 2.5.16 the line through x and $g(x)$ is invariant. This line is translated along itself, so g is hyperbolic. The uniqueness of its axis follows just as in the second half of the proof of proposition 2.5.1.

If d_g does not attain an infimum, there is a sequence $\{x_i\}$ such that $d_g(x_i)$ tends toward the infimum. By compactness, we can assume that $\{x_i\}$ converges to a point $x \in S^{n-1}_\infty$, which must be fixed by g. We can take $x = \infty$ in the upper half-space projection, so g acts as a Euclidean similarity. If this similarity has no fixed point on the bounding hyperplane $S^{n-1}_\infty \setminus \{\infty\}$, it is an isometry; therefore d_g goes to zero on any vertical ray, and $\inf d_g = 0$. Also, since g has no axis and no fixed point in $\mathbf{H}^n$, it is parabolic.

If, instead, g does fix a point on $S^{n-1}_\infty \setminus \{\infty\}$, it leaves invariant the vertical line L through that point. If P is a (hyperbolic) hyperplane orthogonal to L, the closed region F between P and $g(P)$ is a fundamental domain for g, that is, for any point $x \in \mathbf{H}^n$, there is some $k \in \mathbf{Z}$ such that $g^k(x) \in P$. In particular, any value of d_g is achieved inside F. Because d_g does not attain its infimum, the compactness argument of the preceding paragraph shows that g fixes a point in $\bar{F} \cap S^{n-1}_\infty$. But if g fixes three points on S^{n-1}_∞, it fixes a whole plane in $\mathbf{H}^n$, contradicting the assumption that $\inf d_g$ is not attained.

Finally, if d_g takes the value zero, g is by definition elliptic, and its zero-set is a k-dimensional subspace because the entire line joining any two fixed points is fixed. 2.5.17

The convexity of $d(x, y)$ as a function of one variable alone is enough to define a *hyperbolic mean*: a rule that gives, for any finite collection of points in $\mathbf{H}^n$ with associated weights, or masses, its *center of mass*. The center of mass should be preserved under hyperbolic isometries, it should depend continuously on the points and their weights, and if there is only one point it should be the point itself.

Corollary "finite hyperbolic group has a fixed point"
†measure
Exercise "the hyperbolic median"
::median
Problem "other hyperbolic means"
Problem "the algebraic study of isometries of $\mathbf{H}^n$"
% Lorentz transformations

One of the characterizations of the mean in Euclidean space works here too: it associates with the collection $\{(p_i, m_i)\}$ of points p_i with mass m_i the point x that minimizes the function $\sum_i m_i d^2(p_i, x)$. Although $d(p_i, x)$ is not a strictly convex function of x, its square is. The sum is therefore strictly convex, and since it is unbounded on any line, there is a unique point where it attains its minimum.

Corollary 2.5.18 (finite hyperbolic group has a fixed point). *A finite group of isometries of* $\mathbf{H}^n$ *has at least one fixed point.*

Proof of 2.5.18: Let $x \in \mathbf{H}^n$ be any point. If F is a finite group of isometries of $\mathbf{H}^n$, the center of mass of the orbit of x, with each point equally weighted, is fixed by F. 2.5.18

Exercise 2.5.19. Show that any action of any compact group G on $\mathbf{H}^n$ has a fixed point. You may use the existence of a Haar †measure, that is, a measure μ on G such that $\mu(gA) = \mu(A)$ for $A \subset G$ and $g \in G$.

Exercise 2.5.20 (the hyperbolic median). A *median* for a weighted collection of points in $\mathbf{E}^1$ is any point that minimizes the weighted sum of the distances to the points. The set of medians is either a single point, or an interval.

This definition generalizes word for word to $\mathbf{E}^n$ and to $\mathbf{H}^n$. Analyze the existence and uniqueness of the Euclidean and hyperbolic median when $n = 2$ (a typical case), and describe its qualitative properties. Give a geometric characterization of the median of three points.

Problem 2.5.21 (other hyperbolic means). There is a very simple definition of a hyperbolic mean using the hyperboloid model: Given a collection $\{(p_i, m_i)\}$, treat the p_i as vectors in $H^+ \subset \mathbf{E}^{n,1}$, take their mean, and multiply by a scalar to put it back on the hyperboloid.

What is the relation between this mean and the square-of-distance mean? Is it the same? Is it expressible as the minimum of a linear combination of convex functions of distance?

Problem 2.5.22. The Brouwer fixed-point theorem asserts that any continuous map $D^n \to D^n$ has a fixed point. Use this theorem, together with an understanding of isometries of S^{n-1} and $\mathbf{E}^{n-1}$, to classify the isometries of $\mathbf{H}^n$, at least in the cases $n = 2$ and $n = 3$.

Problem 2.5.23 (the algebraic study of isometries of $\mathbf{H}^n$). By exercise 2.3.4, isometries of $\mathbf{H}^n$ are in one-to-one correspondence with linear maps of $\mathbf{R}^{n+1}$ that leave invariant each component of the set $\{Q^- = -1\}$, where Q^- is a quadratic form of type $(n, 1)$. We can use this correspondence to obtain information on the properties of hyperbolic isometries.

(a) As a warm-up exercise, let V be a two-dimensional real vector space, Q a (possibly degenerate) quadratic form on V, and $A : V \to V$ a linear map preserving Q. Describe the relationship between Q and the eigenvectors and eigenvalues of A.

% the paraboloid model

(b) Let V be an $(n+1)$-dimensional real vector space and Q a quadratic form of type $(n, 1)$ on V. Show that there is an isomorphism from V into $\mathbf{R}^{n+1}$ that transforms Q into the form $Q^- = -x_0^2 + x_1^2 + \cdots + x_n^2$. This means that we can just study $\mathbf{E}^{n,1}$.

(c) From now on, assume that A is a linear transformation of $\mathbf{E}^{n,1}$ preserving Q^-. If W is a minimal invariant subspace of A, show that W has dimension one or two. (This is true for any transformation of any non-trivial vector space.) If the dimension is two, Q^- is positive definite on W.

(d) Factor the characteristic polynomial p of A into irreducible quadratic and linear factors over $\mathbf{R}$. For any irreducible quadratic factor q of p, the subspace annihilated by $q(A)$ must be positive definite, so the roots of q are on the unit circle.

(e) If p has any roots that are not on the unit circle, there are precisely two such roots, λ and λ^{-1}, and the isometry of hyperbolic space induced by A is hyperbolic. (We sometimes say that A itself is hyperbolic.)

(f) If A fixes some vector v with $Q^-(v) < 0$, its induced isometry is elliptic; it fixes some totally geodesic subspace $P \subset \mathbf{H}^n$ isometric to $\mathbf{H}^k$, for some $0 \le k \le n$, and "rotates" the normal space of P.

(g) If all characteristic roots are on the unit circle and A fixes no vectors v with $Q^-(v) < 0$, A fixes a non-zero v with $Q^-(v) = 0$. All such fixed vectors are multiples of one another. The isometry induced by A is parabolic. The hyperplane $P = \{w : v \cdot w = 1\}$, where $\cdot$ is the inner product associated with Q^-, is invariant by A. It intersects the region $Q^- < 0$ in the projective paraboloid model of problem 2.3.13. The quadratic form Q^- is degenerate on the tangent space of P, and it induces a Euclidean metric on the quotient space $P/\mathbf{R}v$. The transformation induced by A on $P/\mathbf{R}v$ is an Euclidean isometry without fixed points.

Section "Complex coordinates for hyperbolic three-space"
::inhomogeneous coordinate
$t \notin \mathbf{C}^* = \mathbf{C} \setminus \{0\}$
::homogeneous coordinates
::point at infinity
∞
% square torus in space
::Riemann sphere
% projective transformations of hyperbolic space
::projective transformation
% linear fractional transformation
::linear fractional transformation
::fractional linear transformation
Exercise "linear fractional transformations are Möbius transformations"
linear fractional transformations take circles to circles
% properties of inversions
% linear fractional transformations are Möbius transformations
% the action of GL(2, C) at infinity

2.6. Complex coordinates for hyperbolic three-space

Three-dimensional hyperbolic geometry is intimately associated with complex numbers, and many geometric relations in $\mathbf{H}^3$ can be elegantly described through this association.

The complex plane $\mathbf{C}$ embeds naturally in the complex projective line $\mathbf{CP}^1$, the set of complex lines (one-dimensional complex subspaces) of $\mathbf{C}^2$. The embedding maps a point $z \in \mathbf{C}$ to the complex line spanned by $(z, 1)$, seen as a point in $\mathbf{CP}^1$; we call z the *inhomogeneous coordinate* for this point, while any pair $(tz, t) \in \mathbf{C}^2$, with $t \notin \mathbf{C}^* = \mathbf{C} \setminus \{0\}$, is called a set of *homogeneous coordinates* for it. The remaining point in $\mathbf{CP}^1$, namely the subspace spanned by $(1, 0)$, is the *point at infinity*; we can make ∞ its "inhomogeneous coordinate".

Topologically, $\mathbf{CP}^1$ is the one-point compactification $\widehat{\mathbf{C}}$ of $\mathbf{C}$ (cf. problem 1.1.1), so we can extend the usual identification of $\mathbf{E}^2$ with $\mathbf{C}$ to ∞. This shows that $\mathbf{CP}^1$ is a topological sphere, called the *Riemann sphere*.

As in the real case (problem 2.3.11), a *projective transformation* of $\mathbf{CP}^1$ is what you get from an invertible linear map of $\mathbf{C}^2$ by passing to the quotient. Projective transformations are homeomorphisms of $\mathbf{CP}^1$. If a projective transformation A comes from a linear map with matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, its expression in inhomogeneous coordinates is

2.6.1.
$$A(z) = \frac{az + b}{cz + d}$$

(naturally, this should be interpreted as giving a/c for $z = \infty$ and ∞ for $z = -d/c$). A map $A : \mathbf{CP}^1 \to \mathbf{CP}^1$ of the form 2.6.1 (with $ad - bc \neq 0$) is called a *linear fractional transformation* (or *fractional linear transformation*). Linear fractional transformations behave in a familiar way:

Exercise 2.6.2 (linear fractional transformations are Möbius transformations). (a) Show that, under the usual identification $\mathbf{CP}^1 = S^2$, any linear fractional transformation is a Möbius transformation, that is, a composition of inversions. (Hint: look at $z \mapsto 1/z$ first.)

(b) Show that any *orientation-preserving* Möbius transformation of S^2 is a linear fractional transformation.

Sometimes Möbius transformations of $\widehat{\mathbf{C}}$ are considered to be just the linear fractional transformations. To avoid confusion, we won't use this convention.

Problem 2.6.3. It follows from proposition 1.2.3 and exercise 2.6.2 that linear fractional transformations map circles (including lines) into circles. This can also be proved directly, by applying such a transformation to the general equation of a circle.

Is there a conceptual way to explain why circles go to circles? We'll take up this question again in problem 2.6.5.

GL(2, C)
PGL(2, C) = GL(2, C)/C*
PSL(2, C)
SL(2, C)
% definition of Möbius transformations
% linear fractional transformations are Möbius transformations
% The hyperboloid model and the Klein model
†nullspace
% orthogonal transformations are isometries
Problem "the action of GL(2, C) at infinity"

Two non-singular linear maps of $\mathbf{C}^2$ give the same projective transformation of $\mathbf{CP}^1$ if and only if one is a scalar multiple of the other. Thus, identifying together scalar multiples in the linear group $\mathrm{GL}(2,\mathbf{C})$ gives the group of projective transformations of $\mathbf{CP}^1$, which we naturally denote by $\mathrm{PGL}(2,\mathbf{C}) = \mathrm{GL}(2,\mathbf{C})/\mathbf{C}^*$. This group is also known as $\mathrm{PSL}(2,\mathbf{C})$, because it can be obtained by identifying together scalar multiples in the special linear group $\mathrm{SL}(2,\mathbf{C})$, consisting of linear transformations of $\mathbf{C}^2$ with unit determinant.

As we saw on page 51, a Möbius transformation of S^{n-1}_∞ can be extended to a unique isometry of $\mathbf{H}^n$. Since $\mathrm{PGL}(2,\mathbf{C})$ acts on S^{n-1}_∞ by Möbius transformations (exercise 2.6.2), this action can be extended to all of $\mathbf{H}^3$, providing the first link between hyperbolic geometry and the complex numbers:

Theorem 2.6.4. *The group of orientation-preserving isometries of* $\mathbf{H}^3$ *is* $\mathrm{PGL}(2,\mathbf{C})$, *identified via the action on* $S^2_\infty = \mathbf{CP}^1$.

This was first proven by Poincaré, who followed essentially the reasoning above. Silvio: reference (1883?)

There is another, more intrinsic way to obtain the action of $\mathrm{PGL}(2,\mathbf{C})$ on $\mathbf{H}^3$. Consider the real vector space V of Hermitian forms on $\mathbf{C}^2$, that is, of maps $H : \mathbf{C}^2 \times \mathbf{C}^2 \to \mathbf{R}$ linear in the second variable and satisfying

$$H(w,v) = \overline{H(v,w)}$$

for $v, w \in \mathbf{C}^2$. In the canonical basis, a Hermitian form has matrix $\begin{pmatrix} r & z \\ \bar z & s \end{pmatrix}$ with $r, s \in \mathbf{R}$ and $z \in \mathbf{C}$, so V has dimension four.

The determinant of a form represented by $\begin{pmatrix} r & z \\ \bar z & s \end{pmatrix}$ is $rs - z\bar z$, so the function det is a quadratic form on V. The signature of det is $(1,3)$, because $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, for example, form an orthonormal basis. Thus V, with the quadratic form $-\det$, is isomorphic to $\mathbf{E}^{3,1}$; by section 2.3, this means that the set of definite Hermitian forms in $\mathbf{C}^2$, up to multiplication by (non-zero real) scalars, forms a model for $\mathbf{H}^3$!

The sphere at infinity S^2_∞ in this description consists of Hermitian forms of rank one, up to scalars. To each such form we associate its †nullspace, which is a point in $\mathbf{CP}^1$; the nullspace determines the form up to multiplication by a scalar (check this), so we get a canonical identification of S^2_∞ with $\mathbf{CP}^1$.

Now define an action of $\mathrm{GL}(2,\mathbf{C})$ on V as follows: for any $A \in \mathrm{GL}(2,\mathbf{C})$ and $H \in V$, let $A(H)$ be such that $A(H)(v,w) = H(Av, Aw)$. The action of A preserves the form $-\det$ up to a scalar—the determinants of all elements of V get multiplied by $|\det A|^2$—so the induced projective transformation in $V/\mathbf{R}^*$ gives an isometry of $\mathbf{H}^3$, as discussed on page 54. This isometry remains the same when we multiply A by a scalar, so in effect we have an action of $\mathrm{PGL}(2,\mathbf{C})$ on $\mathbf{H}^3$ by isometries. We also have an action of $\mathrm{PGL}(2,\mathbf{C})$ on $S^2_\infty = \mathbf{CP}^1$.

Problem 2.6.5 (the action of GL(2, C) at infinity). (a) Show that the action of $A \in \mathrm{GL}(2,\mathbf{C})$ on $\mathbf{CP}^1$, in terms of the identification described above, is

% linear fractional transformations take circles to circles
Exercise "classification by the trace"
Problem "complex trigonometry"
% classification by the trace
% finding the axis
% classification by the trace
% interpretation of the inner product
% Some computations in hyperbolic space

given (in homogeneous coordinates) by multiplication by A^{-1}. Are all linear fractional transformations obtained in this way?

(b) Use this to show that $\mathrm{PGL}(2, \mathbf{C})$ coincides with the group of orientation-preserving isometries of $\mathbf{H}^3$. (Hint: Orientation-preserving isometries of $\mathbf{H}^3$ act simply transitively on the set of triples of distinct points at infinity. The same is true for the action of projective transformations of $\mathbf{CP}^1$, as you should check by writing an explicit formula).

(c) If $H \in V$ represents a point of $V/\mathbf{R}^*$ outside the sphere at infinity, the signature of H is $(1,1)$, that is, H is indefinite. What is the zero-set of H in $\mathbf{C}^2$? Show that, passing to the quotient, the zero-set becomes a circle in $\mathbf{CP}^1$. Can every circle in $\mathbf{CP}^1$ be obtained in this way? Show that, under the identification of $\mathbf{CP}^1$ with the set of forms of rank one in V, all forms in the zero-set of H are orthogonal to H (for the inner product associated with $-\det$). Thus the zero-set of H is the intersection of the polar plane of H with the sphere at infinity.

(d) How do zero-sets transform under the action of $\mathrm{PGL}(2, \mathbf{C})$? (Compare part (a)). Can you answer problem 2.6.3 now?

Exercise 2.6.6. Show in two ways that the group of orientation-preserving isometries of $\mathbf{H}^2$ is $\mathrm{PGL}(2, \mathbf{R})$. (Hints: (1) Use the fact that $\mathrm{PGL}(2, \mathbf{R})$ is the subset of $\mathrm{PGL}(2, \mathbf{C})$ that preserves the real axis. (2) In the discussion above, replace Hermitian forms on $\mathbf{C}^2$ by quadratic forms on $\mathbf{R}^2$.)

Exercise 2.6.7 (classification by the trace). Classify orientation-preserving isometries of $\mathbf{H}^3$ into hyperbolic, elliptic and parabolic according to the quantity $\mathrm{tr}^2 A/\det A$, where $A \in \mathrm{GL}(2, C)$ represents the isometry. (Hint: you can assume that $\det A = 1$. By conjugation, you can also assume that A is in Jordan normal form, so it is either a diagonal matrix or $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$.)

Show how to find the axis of a hyperbolic or elliptic isometry, and the angle of rotation of an elliptic isometry, in terms of a representative $A \in \mathrm{GL}(2, C)$. What is the trace of a rotation of order two?

Silvio: check ref. Troels Jorgensen

Problem 2.6.8 (complex trigonometry). The set of oriented lines in $\mathbf{H}^3$ has the structure of a complex two-manifold, namely $(\mathbf{CP}^1)^2 \setminus \Delta$, where Δ is the diagonal $\{(z, z) : z \in \mathbf{CP}^1\}$. Derive formulas for the geometry of lines in $\mathbf{H}^3$, as follows:

(a) Let W be the vector space of linear transformations of $\mathbf{C}^2$ with trace 0. There is a one-to-one correspondence between unoriented lines in $\mathbf{H}^3$ and one-dimensional subspaces of W whose non-zero representatives are in $\mathrm{GL}(2, \mathbf{C})$. (Hint: see exercise 2.6.7.)

(b) Consider the inner product associated with the quadratic form $-\det$ on W: it is given by $A \cdot B = \frac{1}{2} \mathrm{tr}\, AB$. Using proposition 2.5.4 and exercise 2.6.7, interpret the inner product $A \cdot B$ of two vectors of unit length in terms of the geometry of the pair of lines given by A and B (compare proposition 2.4.5).

(c) Derive a formula for the relationship between a triple of lines in $\mathbf{H}^3$ and the dual triple of common orthogonals to pairs of these lines. This formula should generalize the squared form of many formulas from section 2.4—for instance, the spherical formulas come from the situation that three lines intersect in a point in $\mathbf{H}^3$.

Problem "trace relations" % complex trigonometry

(d) Interpret $W \cap \mathrm{SL}(2,\mathbf{C})$ as the set of *directed* lines in $\mathbf{H}^3$. Refine parts (b) and (c) to take into account the extra information.

Problem 2.6.9 (trace relations). (a) Show that any two elements A and B of $\mathrm{SL}(2,\mathbf{C})$ satisfy

$$\operatorname{tr} AB + \operatorname{tr} AB^{-1} = \operatorname{tr} A \operatorname{tr} B.$$

(Hint: use the fact that A satisfies its characteristic polynomial.)

(b) Show that the trace of every element of the group generated by A and B is expressible as a polynomial in $\operatorname{tr} A$, $\operatorname{tr} B$, and $\operatorname{tr} AB$.

(c) Prove that for any complex numbers x, y, and z, such that the symmetric matrix

$$\begin{pmatrix} 1 & x & z \\ x & 1 & y \\ z & y & 1 \end{pmatrix}$$

is nonsingular, there exist A and B in $\mathrm{SL}(2,\mathbf{C})$ such that $\operatorname{tr} A = x$, $\operatorname{tr} B = y$ and $\operatorname{tr} AB = z$, and the subgroup of $\mathrm{SL}(2,\mathbf{C})$ generated by A and B is unique up to conjugacy.

(Hint: Interpret this matrix as the matrix of inner products for a triple of unit vectors in the space W of problem 2.6.8, representing order-two rotations a, b and c. If the matrix is non-singular, the three vectors form a basis; use this to reconstruct the group G generated by a, b and c. Then set $A = bc$ and $B = ca$, and show that the group generated by A and B has index two in G, and can also be determined up to conjugacy.)

(d) Compare the results when this analysis is applied to two subgroups of $\mathrm{SL}(2,\mathbf{C})$ that have the same image in $\mathrm{PSL}(2,\mathbf{C})$.

(e) What interpretation can be given when the matrix of (c) is singular?

(f) Show that the trace of every element of the group generated by A, B and C is expressible as a polynomial in the traces of A, B, C, AB, BC, CA and ABC. In fact, the first six traces are almost enough to express everything: show, by expanding the trace of $(ABC)^2$, that the trace of ABC satisfies a quadratic equation in terms of the other six traces, the other root of which is the trace of ACB.

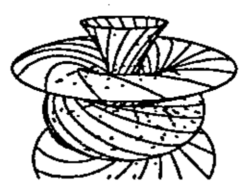
Section "The geometry of the three-sphere"
% definition of stereographic projection
::Hopf circle
::fibration base
::Hopf map
% hopfcircles

2.7. The geometry of the three-sphere

Just like the circle and the two-sphere, the three-sphere is very round. But there are some beautiful, classical aspects to its roundness that are not easy to guess from its lower-dimensional sisters.

The most direct way to define S^3 is as the unit sphere $\{x_1^2+x_2^2+x_3^2+x_4^2 = 1\}$, but visualizing this set directly is hard for most people. Stereographic projection (page 46) gives a picture of S^3, as $\widehat{\mathbf{R}^3} = \mathbf{R}^3 \cup \{\infty\}$, that makes the sphere much more tangible, and preserves some aspects of its geometry: for example, the roundness of circles and spheres. But this picture suffers from a loss of symmetry: $\widehat{\mathbf{R}^3}$ is not as round as it should be, and objects the same size in S^3 are not the same size in $\widehat{\mathbf{R}^3}$. To overcome this drawback, it helps to practice imagining the rigid motions of S^3 as they appear in $\widehat{\mathbf{R}^3}$.

Identifying $\mathbf{R}^4$ with $\mathbf{C}^2$, we obtain new pictures. If the coordinates in $\mathbf{C}^2$ are z_1 and z_2, the equation of a unit sphere becomes $|z_1|^2 + |z_2|^2 = 1$. Each complex line (one-dimensional subspace) in $\mathbf{C}^2$ intersects S^3 in a great circle, called a *Hopf circle*. Since exactly one Hopf circle passes through each point of S^3, the family of Hopf circles fills up S^3—we say it forms a *fibration* of S^3 by circles. The base of the fibration—what you get when you collapse each fiber to a point—is the set of complex lines of $\mathbf{C}^2$, that is, the Riemann sphere $\mathbf{CP}^1 = \widehat{\mathbf{C}}$. In this way we get a map $S^3 \to S^2$, called the *Hopf map*.



hopfcircles

Figure 2.30. The Hopf fibration in S^3. The three-sphere $|z_1|^2 + |z_2|^2 = 1$ fibers over the Riemann sphere: each fiber is a great circle, the locus of $z_1/z_2 =$ constant. Each torus $\{|z_2| = a\}$, for $0 < a < 1$—which can also be expressed as $\{|z_1| = a\}$—divides S^3 into two solid tori, one of them containing the point at infinity in the picture.

Figure 2.30 shows what the Hopf fibration looks like under stereographic projection. In this figure the vertical axis is the intersection of S^3 with the complex line $z_1 = 0$, and the horizontal circle is the intersection with $z_2 = 0$. The locus $\{|z_2| \le a\}$, for any $0 < a < 1$, is a solid torus neighborhood of the $z_2 = 0$ circle. Its boundary $\{|z_2| = a\}$, a torus of revolution, is filled up by Hopf circles, each winding once around the z_1-circle and once around the

::Villarceau circles
Exercise "the Hopf fibration"
Hopf flow
::quaternions
::real
::pure quaternions
::conjugate
::absolute value
::norm
::unit quaternion
::skew field

z_2-circle. Any torus of revolution in $\mathbf{R}^3$ can be transformed into a torus of this form by a similarity. The fact that any torus of revolution has curves winding around in both directions that are geometric circles is quite surprising, so much so that such circles on the torus have a special name: *Villarceau circles.*

Exercise 2.7.1 (the Hopf fibration). (a) Show that the maps $g_t : S^3 \to S^3$ given by multiplication by e^{it}, for $t \in \mathbf{R}$, are isometries, and that they leave the Hopf fibration invariant. Thus S^3 has isometries that don't have an axis: the motion near any point is like the motion near any other point. This is one way in which S^3 seems "rounder" than S^2. The one-parameter family $\{g_t\}$ is called the Hopf flow.

(b) Show that two Hopf circles C and C' are parallel in the sense that any two points in C' are at the same distance from C (in the metric of S^3).

(c) Show that the metric in $\mathbf{CP}^1 = S^2$ induced by the Hopf map is the standard, round metric. What are the images of great circles?

The three-sphere, like the circle, is a topological group. This is easiest to see using *quaternions*, which are, so to speak, an extension of complex numbers. We spend some time here investigating their properties, because they turn out to be useful in several ways.

The space $\mathcal{H}$ of quaternions is simply $\mathbf{R}^4$, together with a certain non-commutative multiplication $\mathcal{H} \times \mathcal{H} \to \mathcal{H}$. This multiplication is bilinear over $\mathbf{R}$, so in order to define it we can just specify its effect on a basis of $\mathcal{H}$. The basis is traditionally denoted $\{1, i, j, k\}$, and the action is as follows: 1 is the identity, and

$$i^2 = j^2 = k^2 = -1,$$
$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik.$$

The subspace spanned by 1 is identified with $\mathbf{R}$, and its elements are called *real*; the subspace $\mathbf{R}i + \mathbf{R}j + \mathbf{R}k$ is the space of *pure quaternions.* It is easy to see that the quaternion product is associative, but not commutative; that $\mathbf{R}$ is the center of $\mathcal{H}$ (so the structure of $\mathcal{H}$ as a ring determines its structure as a real vector space), and that a quaternion is pure if and only if its square is a non-positive real number (so the ring structure also determines the decomposition into real and pure subspaces).

Exercise 2.7.3. Show that any ring automorphism of $\mathbf{H}$ is a linear map. This differs from the situation in $\mathbf{C}$, which has many automorphisms.

The *conjugate* of a quaternion $q = a + bi + cj + dk$ is $\bar{q} = a - bi - cj - dk$. The product $q\bar{q}$ is positive real, and its square root is the *absolute value*, or *norm*, of q, denoted by $|q|$. This coincides with the norm of q as a vector in $\mathbf{E}^4$. A quaternion of norm 1 is a *unit quaternion.* Any non-zero quaternion q has an inverse $q^{-1} = |q|^{-2}\bar{q}$; thus $\mathcal{H}$ is a *skew field.* If p and q are quaternions, we have $\overline{pq} = \bar{q}\bar{p}$ and $|pq| = |p|\,|q|$. Since the set of unit quaternions is the unit three-sphere $S^3 \subset \mathbf{E}^4$, we have proved the first statement in the following theorem:

Theorem "the group structure of S^3"
% rotating S^2 with quaternions
Exercise "rotating S^2 with quaternions"
% rotating S^2 with quaternions
% hopfcircles
::unitary transformations

Theorem 2.7.4 (the group structure of S^3). *The three-sphere has the structure of a non-commutative group, with center $\{\pm 1\}$. Left or right multiplication gives a self-action of S^3 by orientation-preserving isometries. Conjugation gives a self-action by isometries which, in addition, takes any two-sphere with center* 1 *onto itself. The quotient $S^3/\{\pm 1\}$ is isomorphic to the group* SO(3) *of orientation-preserving isometries of S^2.*

SO(2)
†simply transitively

Proof of 2.7.4: The distance along S^3 between two points depends solely on the norm of the difference between them; since the norm is preserved by left or right multiplication by a unit quaternion, these map are isometries. They're orientation-preserving by continuity, because S^3 is connected.

Conjugation is also an isometry; since it fixes 1, it leaves invariant the sets of points at constant distance from 1, which are two-spheres.

The action of S^3 on these two-spheres by conjugation defines a homomorphism $\rho : S^3 \to \mathrm{SO}(3)$, whose kernel is the center of S^3. Surjectivity is shown in exercise 2.7.5. 2.7.4

Exercise 2.7.5 (rotating S^2 with quaternions). Let S^2 be the sphere of pure quaternions of unit norm. Show that conjugation by a unit quaternion $r + p$, where r is real and p is pure, rotates S^2 around the axis in the direction of p by an angle $2\arctan(|p|/|r|)$. (Hint: assume first that $r = \cos\theta$, $p = i\sin\theta$. Extend to arbitrary p by showing that i can be taken to any point on S^2 by conjugation.)

The descriptions of S^3 via quaternions and via complex numbers can be combined. If we look at $\mathcal{H}$ as a complex vector space, multiplication on the left by the quaternion i is the same as multiplication by the complex number i, so the vector field $X_i(p) = ip$, for $p \in S^3$, induces the Hopf flow on S^3. There is added symmetry in this description, because i plays no special role among the quaternions: any pure quaternion of unit norm can be used in lieu of i to impart a complex vector space structure to $\mathbf{R}^4$ (compare exercise 2.7.5). Thus there are many Hopf flows and many Hopf foliations on S^3. In particular, we can take three mutually orthogonal vector fields X_i, X_j and X_k to get three mutually orthogonal families of Hopf circles.

Exercise 2.7.6. Using stereographic projection as in figure 2.30, with the identity element $1 \in \mathcal{H}$ at the origin and i, j and k on the three coordinate axis, try to get an idea of what the three orthogonal Hopf foliations look like.

There is another way to describe the group structure on S^3, via *unitary transformations* of $\mathbf{C}^2$, that is, complex linear transformations that preserve the standard Hermitian form $(z_1, z_2) \cdot (w_1, w_2) = z_1\bar{w}_1 + z_2\bar{w}_2$ on $\mathbf{C}^2$. Here again the case of the circle is analogous: the group SO(2) of orthogonal linear transformations of the Euclidean plane having determinant 1 acts †simply transitively on S^1, so fixing a point $x \in S^1$ provides an identification $\mathrm{SO}(2) \cong S^1$ that takes each $g \in \mathrm{SO}(2)$ to $g(x) \in S^1$.

Similarly, the group SU(2) of unitary transformations of $\mathbf{C}^2$ of determinant 1 acts simply transitively on S^3, so by fixing a point $x \in S^3$ we get an

% the group structure of S^3

identification $SU(2) \cong S^3$. If $x = (1,0) \in S^3 \subset \mathbf{C}^2$, any point $(z_1, z_2) \in S^3$ gets identified to

$$\begin{pmatrix} z_1 & -\bar{z}_2 \\ z_2 & \bar{z}_1 \end{pmatrix}.$$

Exercise 2.7.7. Consider $\mathcal{H}$ with the complex structure defined by multiplication on the left by a given unit quaternion. Show that multiplication on the right by any quaternion is a complex linear map. Show that multiplication on the right by a unit quaternion is a unitary map. This again provides an identification $S^3 \cong SU(2)$, which makes S^3 the group that preserves, simultaneously, *all* the complex structures defined by multiplication on the left by unit quaternions.

The connection between the group structure on S^3 and the geometry of S^2, too, can be recovered from this point of view. Any element of SU(2) acting as an isometry S^3 takes fibers to fibers, so its action passes to the quotient, giving a projective transformation of $\mathbf{CP}^1$ that is also an isometry of S^2. As a map of S^2, a projective transformation of $\mathbf{CP}^1$ is orientation-preserving, so we get a homomorphism $SU(2) \to SO(3)$. An element of SU(2) induces the identity map on $\mathbf{CP}^1$ if and only if it is a scalar multiple of the identity map on $\mathbf{C}^2$; since the only two such elements are $\pm\,\mathrm{Id}$, we get an injective homomorphism $PSU(2) \to SO(3)$, where $PSU(2) = SU(2)/\{\pm\,\mathrm{Id}\}$. You should check that this map is also surjective, and consequently an isomorphism.

Exercise 2.7.8. Topologically, PSU(2) is just $S^3/\{\pm 1\} = \mathbf{RP}^3$. Describe a direct correspondence between $\mathbf{RP}^3$ and SO(3), by realizing $\mathbf{RP}^3$ as the unit ball in $\mathbf{R}^3$ with antipodal points on its boundary identified. (Hint: for a point $r, \theta \in \mathbf{R}^3$, with $\theta \in S^2$ and $r \in [0,1]$, consider the rotation through an angle πr around the axis that contains θ.)

Another intriguing aspect of S^3 is the structure of its own group of rigid motions. The two actions of S^3 on itself, by left and right multiplication, commute, so we can define a homomorphism $\tau : S^3 \times S^3 \to SO(4)$ by

$$\tau(g,h)(x) = gxh^{-1}.$$

By theorem 2.7.4 some transformation $\tau(h,h)$ will get us from any orthonormal frame at 1 to any other such frame, so further composition with a transformation of the form $\tau(gh^{-1}, 1)$ will get us to any orthonormal frame at any point. This shows that τ is surjective, and also that the kernel of τ is $\{\pm(1,1)\}$, so we get an isomorphism

$$SO(4) = (S^3 \times S^3)/\mathbf{Z}_2.$$

How can the group of isometries of a space so round as S^3 be almost a product? To understand this better, consider more carefully the nature of right and left multiplication.

If $g \in S^3$ is not ± 1, the transformation $x \mapsto gx$ fits into the unique Hopf flow generated by the Hopf field $x \mapsto px$, where p is the unit quaternion in the

% hopfcircles

direction of the g purely imaginary component of g. Similarly, when $h \in S^3$ is not ± 1, the transformation $x \mapsto xh$ fits into a unique Hopf flow, generated by a vector field $x \mapsto xp$. The two kinds of Hopf flows are distinguished as left-handed or right-handed: the circles near a given circle wind around it in a left-handed sense or in a right-handed sense (as the threads of a common screw or jar lid). As we have seen algebraically, any right-handed Hopf flow commutes with any left-handed Hopf flow.

Problem 2.7.9. The commutation of right-randed and left-handed Hopf flows can also be seen geometrically.

(a) Describe the sensation (i.e., the motion and rate of turning) of a person in S^3 when being left-multiplied by a one-parameter subgroup of S^3.

(b) Let X and Y be right- and left-handed Hopf fields. Show that, by starting at any point in S^3 and moving in the direction orthogonal to both X and Y, you reach a point where $X = Y$. Consequently, there is an entire circle C^+ where $X = Y$. Similarly, there is a circle C^- where $X = -Y$, and that $d(C^+, C^-) = \pi$.

(c) Show that the locus of points that lie at a distance a from C^+ is a torus invariant under X and Y. This is also the locus of points at a distance $\pi - a$ from C^-. Show that X and Y act as rigid motions of these tori, and consequently commute. Compare figure 2.30.

(d) Interpret C^+ and C^- algebraically, in terms of the element of SO(4) associated with multiplication on the left by an element that generates X and on the right by an element that generates Y.

(e) Describe the sensation of a person in S^3 being acted upon by a general one-parameter subgroup of SO(4).

Problem 2.7.10. Give a geometric description of the homomorphism $SO(4) \to SO(3) \times SO(3)$. (Hint: Consider the space of right-handed Hopf flows and the space of left-handed Hopf flows.)

Chapter "Geometric Manifolds"

Chapter 3

Geometric Manifolds

In chapter 1 we looked at a good number of manifolds. In doing this, we relied more on intuition and common sense than on definitions. It is now time to study manifolds a bit more systematically.

Manifolds come to us in nature and in mathematics by many different routes. Very frequently, they come naturally equipped with some special pattern or structure, and to understand the manifold we need to "see" the pattern. At other times, a manifold may come to us naked; by finding structures that fit it, we can gain new insight, relate it to other manifolds, and take better care of it.

The fact that there are all these different grades, or flavors, of manifolds was not clearly understood during the early development of topology. Different constructions were perceived more as alternative technical contexts for doing topology than as building blocks for essentially different structures. One of the remarkable achievements of topologists over the past forty years has been to come to grips with these distinctions, which are, contrary to intuition, substantive: for example, topological, piecewise linear and differentiable manifolds are inequivalent in dimensions four and higher, although in dimensions two and three the distinctions collapse.

Most of the myriad other possible structures—complex structures, foliations, hyperbolic structures, and so on—are considerably more restrictive than differentiable structures. These more restrictive structures can have great power in dimensions two and three.

Section "Basic definitions"
::manifold
Definition "pseudogroup"
::pseudogroup
::local
Definition "$\mathcal{G}$-manifold"
::$\mathcal{G}$-manifold
::$\mathcal{G}$-atlas
::$\mathcal{G}$-compatible coordinate charts
::local coordinate systems
::transition map
::coordinate change
::compatible

3.1. Basic definitions

A *manifold* is a topological space that is locally modeled on $\mathbf{R}^n$.

What it means to be locally modeled on $\mathbf{R}^n$ depends on what property, or pattern, of $\mathbf{R}^n$ we want to capture. The idea is to patch the manifold together seamlessly from small pieces of fabric with the pattern. A pattern is described operationally, in terms of the transformations that preserve it; by allowing chunks of $\mathbf{R}^n$ to be glued together only according to these transformations, we get a manifold with the desired pattern. The set of allowed gluing maps should satisfy some natural properties:

Definition 3.1.1 (pseudogroup). A set $\mathcal{G}$ of homeomorphisms between open sets of a topological space X is a *pseudogroup* if it satisfies the following conditions:

(a) The restriction of an element $g \in \mathcal{G}$ to any open set in its domain is also in $\mathcal{G}$.

(b) The composition $g_1 \circ g_2$ of two elements of $\mathcal{G}$, when defined, is in $\mathcal{G}$.

(c) The inverse of an element of $\mathcal{G}$ is in $\mathcal{G}$.

(d) The property of being in $\mathcal{G}$ is *local*, that is, if $U = \bigcup_\alpha U_\alpha$ and g is a local homeomorphism $g : U \to V$ whose restriction to each U_α is in $\mathcal{G}$, then $g \in \mathcal{G}$.

The basic example is the pseudogroup Top of *all* homeomorphisms between open subsets of $\mathbf{R}^n$. A topological manifold is a space that has the local topological pattern, or structure, of $\mathbf{R}^n$. More generally, a $\mathcal{G}$-manifold is a topological space covered by local coordinate systems such that the coordinate change maps belong to $\mathcal{G}$. Here is a more precise statement:

Definition 3.1.2 ($\mathcal{G}$-manifold). An n-dimensional *$\mathcal{G}$-manifold M* is a topological space X with a *$\mathcal{G}$-atlas* on it. A $\mathcal{G}$-atlas is a collection of *$\mathcal{G}$-compatible coordinate charts*, or *local coordinate systems*, whose domains cover X. A coordinate chart is a pair (U_i, ϕ_i), where U_i is open in X and $\phi_i : U_i \to \mathbf{R}^n$ is a homeomorphism onto its image. Compatibility means that, whenever two charts (U_i, ϕ_i) and (U_j, ϕ_j) intersect, the *transition map* or *coordinate change*

$$\gamma_{ij} = \phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$$

is in $\mathcal{G}$.

We make a standing assumption that manifolds are Hausdorff except where otherwise noted, but the definition of a $\mathcal{G}$-manifold also applies to non-Hausdorff manifolds.

Of course, we don't really regard the charts as an essential part of the structure of M. Two $\mathcal{G}$-atlases A_1 and A_2 for X are *compatible* if their union is also a $\mathcal{G}$-atlas. Two compatible atlases are considered to define the same $\mathcal{G}$-manifold structure for X.

Example "differentiable manifolds"
differentiable manifold
diffeomorphism
% Smoothings
Example "real analytic manifolds"
†real analytic
::real analytic manifold
Example "foliations"

Mathematicians have a loathing for ambiguity and indefiniteness, so often a manifold is defined by one of two polite fictions: an equivalence class of $\mathcal{G}$-atlases, or a maximal $\mathcal{G}$-atlas. Either object is so huge and complicated as to be unimaginable, in contrast with individual $\mathcal{G}$-atlases which are generally not so hard to construct and deal with.

Just as important as the class of manifolds defined by a $\mathcal{G}$-structure is the class of maps that preserve that structure. The simplest case is when two $\mathcal{G}$-manifolds are taken to one another by a homeomorphism that, when expressed in terms of local charts, is given by elements of $\mathcal{G}$ (on a small enough neighborhood of each point). Such a map is called a $\mathcal{G}$-isomorphism, and we have every right to consider the two manifolds related by it as being equivalent, or "identical" in some sense.

We won't give a complete treatment of $\mathcal{G}$-structures and the variety of beautiful results and questions relating to them: in fact, we will soon focus on a small class of rigid pseudogroups. But here is a short list of examples, most of which will play a part later in the book. For a more thorough discussion, you can consult [Hae58].

Example 3.1.3 (differentiable manifolds). If C^r is the pseudogroup of C^r diffeomorphisms between open sets of $\mathbf{R}^n$, for $r \geq 1$, a C^r-manifold is called a differentiable manifold (of class C^r), or C^r-manifold. A C^r-isomorphism is called a diffeomorphism.

It is often convenient to consider only C^∞, or smooth, maps and manifolds. This doesn't cause any loss in generality, because a C^r-structure on a manifold determines a unique C^∞-structure [Whi36]. In other words, every C^r atlas on a manifold is C^r-compatible with a C^∞ atlas, and any two C^∞ atlases compatible with the same C^r atlas are isomorphic via a C^∞-diffeomorphism close to the identity.

Silvio: explain this better

The situation is not so nice for topological manifolds. In 1956, Milnor proved the surprising result that there are several inequivalent differentiable manifolds homeomorphic to S^7 [Mil56]. After these exotic spheres, other manifolds having inequivalent differentiable structures, or no differentiable structure, were found.

In low dimension, as we've mentioned, those distinctions collapse: every two- or three-dimensional topological manifold has a differentiable structure unique to diffeomorphism. See section 3.3 and [Mun60].

Example 3.1.4 (real analytic manifolds). Let C^ω be the pseudogroup of local †real analytic diffeomorphisms of $\mathbf{R}^n$. A C^ω-manifold is called a *real analytic manifold.* Real analytic diffeomorphisms are uniquely determined, using analytic continuation, by their restriction to any open set. Every smooth structure on a manifold admits a real analytic structure compatible with it [Whi36].

reference to Morrey?

Example 3.1.5 (foliations). Write $\mathbf{R}^n$ as the product $\mathbf{R}^{n-k} \times \mathbf{R}^k$ and let $\mathcal{G}$ be the pseudogroup of local diffeomorphisms ϕ that take horizontal factors to

::foliation
::leaves
% hopfcircles
Exercise "irrational foliation on the torus"
% hopfcircles
% foliatetorus
Exercise "the Reeb foliation"
% foliatetorus
% hopfcircles

horizontal factors, that is, that have the form

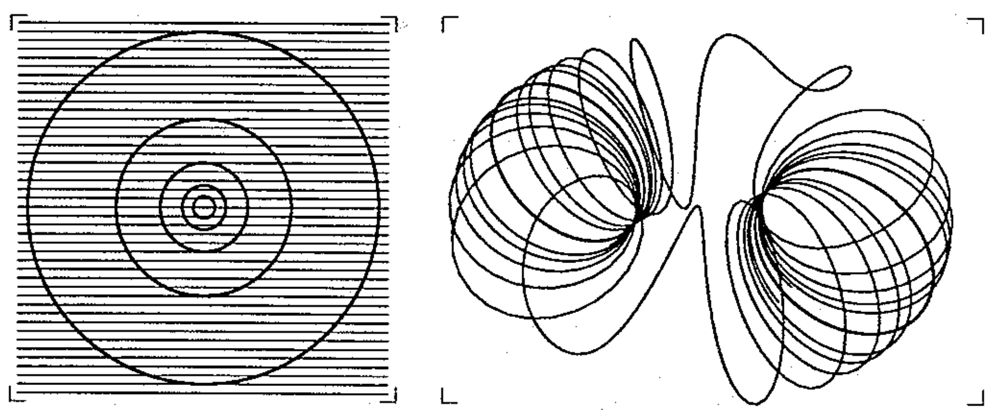
$$\phi(x,y) = \big(\phi_1(x,y), \phi_2(y)\big),$$

for $x \in \mathbf{R}^{n-k}$ and $y \in \mathbf{R}^k$. A $\mathcal{G}$-structure is called a *foliation* of codimension k (or dimension $n-k$). To visualize a foliation one should think not of local coordinate charts, but of the inverse images of factors $\mathbf{R}^{n-k} \times \{y\}$ under the charts, which piece together globally to give the *leaves* of the foliation.

One-dimensional foliations exist on many manifolds: any nowhere vanishing vector field has an associated foliation, obtained by following the flow lines. One example is the Hopf fibration of S^3 illustrated in figure 2.30. In this case all the leaves close upon themselves, but in general the leaves may spiral around the manifold in a complicated way:

Exercise 3.1.6 (irrational foliation on the torus). For a torus $T^2 = \mathbf{R}^2/\mathbf{Z}^2$, consider the constant vector field $X(x,y) = (1,\alpha)$. The flow lines of X are the images of the straight lines $y = \alpha x + y_0$ on T^2, under the projection map. Show that these leaves are topological circles if α is rational, and dense subsets of T^2 if α is irrational. What value of α gives the circles of figure 2.30?

The torus has many other foliations. For example, the quotient of $\mathbf{R}^2 \setminus \{0\}$ by a homothety centered at the origin is a topological torus. Since the foliation of $\mathbf{R}^2 \setminus \{0\}$ by horizontal lines is preserved by homothety, we get a foliation on the torus by passing to the quotient: see figure 3.1.



foliatetorus

Figure 3.1. A foliation on the torus. (a) The horizontal foliation is preserved by the homothety $x \mapsto 2x$. (b) A foliation is obtained on the quotient manifold, a torus. Two of the leaves are circles, hence compact; these are darker in the picture. All other leaves, of which two are shown, spiral around, accumulating onto the circles.

Exercise 3.1.7 (the Reeb foliation). Apply the process of figure 3.1 in one higher dimension, that is, starting from a foliation by horizontal planes in $\mathbf{R}^3 \setminus \{0\}$. What is the quotient manifold? Describe the quotient foliation qualitatively; show that one of its leaves is a torus. Cutting the manifold open along this torus gives two solid tori, which can be glue back to form the three-sphere, as shown in figure 2.30. How smooth is the resulting foliation of S^3?

Example "complex manifolds"
::complex manifold
::uniformization theorem
†Riemann mapping theorem
Poincaré
Klein
::manifold-with-boundary

This discovery of this foliation [Ree52] was something of a surprise, for before then, no foliation of S^3 was known. Later it became clear that codimension-one foliations are very common: Lickorish [Lic65] and Novikov [Nov65] showed that they exist for every oriented three-manifold, and Wood [Woo69] extended this to non-orientable three-manifolds. More generally, a manifold of any dimension has a codimension-one foliation if and only if its Euler number is zero [Thu76]. (See [Hae58] and [?] for related results for foliations of codimension greater than one.)

Euler number has not been defined in general

On the other hand, Haefliger [Hae58] proved that no foliation of S^3 can be real analytic, and Novikov [Nov65] proved that every foliation of S^3 or $S^2 \times S^1$, as well as many other three-manifolds, has a closed leaf that is a torus. Gabai [?, ?] has constructed foliations with no closed torus leaves on many three-manifolds outside the scope of Novikov's theorem. These results have interesting consequences for the topology of three-manifolds.

define total foliation; Detlef/Hardop: all threem have total foliations

Example 3.1.8 (complex manifolds). When n is even, $\mathbf{R}^n$ can be identified with $\mathbf{C}^{n/2}$. Let Hol be the pseudogroup of local biholomorphic maps of $\mathbf{C}^{n/2}$, that is, holomorphic maps that have holomorphic local inverses. (It turns out that this is always the case for holomorphic local homeomorphisms.) A Hol-manifold is called a *complex manifold* of (complex) dimension $n/2$.

When $n = 2$, a map is holomorphic if and only if it is conformal and preserves orientation. Therefore an orientation-preserving isometry of the Poincaré disk model for $\mathbf{H}^2$ is biholomorphic, and every orientable hyperbolic surface inherits the structure of a complex manifold.

Stereographic projection from the unit sphere to $\mathbf{C}$ is a conformal map. A collection of maps obtained by rotating the sphere and then mapping by stereographic projection to $\mathbf{C}$ constitutes an atlas for a complex structure on S^2 (provided they don't all omit the same point); this is the complex structure of $\mathbf{CP}^1$, the Riemann sphere.

Similarly, orientation-preserving isometries of $\mathbf{E}^2$ are holomorphic. It follows that a surface having a hyperbolic, Euclidean or elliptic structure—that is, a metric locally isometric to hyperbolic, Euclidean or elliptic space—also has a derived complex structure. In particular, all orientable surfaces can be made into complex manifolds.

This easy observation has a converse: every complex structure on a closed surface comes from a hyperbolic, Euclidean, or elliptic structure. The converse is a celebrated result known as the *uniformization theorem.* It is closely related to the †Riemann mapping theorem. The uniformization theorem was the subject of much attention (and contention) by Poincaré, Klein, and others in the latter part of the nineteenth century.

Bill: this seems unsatisfactory. Where did the contention come from? Silvio

We close this section with terminology that, although standard, is potentially quite confusing.

A *manifold-with-boundary* is, in general, not a manifold; it is a space locally modeled on the half-space $\mathbf{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbf{R}^n : x_n = 0\}$. This means

::boundary
::closed manifold
::open

that each point p in a manifold-with-boundary has a neighborhood that looks like a neighborhood of some point in $\mathbf{R}^n_+$, either in the interior or on the boundary. The set of points that don't have neighborhoods like those of an interior point of $\mathbf{R}^n_+$ form the *boundary*.

Again, by specializing to a pseudogroup of local homeomorphisms of $\mathbf{R}^n_+$, we get various kinds of manifolds. For instance, if we use the pseudogroup of local homeomorphisms of $\mathbf{R}^n_+$ that are restrictions of Euclidean isometries, we obtain Euclidean manifolds with geodesic boundary. Hyperbolic manifolds with geodesic boundary and elliptic manifolds with geodesic boundary are defined similarly.

A manifold (with no boundary) that is compact is called a *closed manifold*, and a non-compact one is sometimes called *open*. This is sometimes in conflict with the usage of "open" and "closed" in general topology, but there isn't much we can do about it.

Section "Gluings and piecewise linear manifolds"
% $\mathcal{G}$-manifold
% The totality of surfaces
% Some three-manifolds
::n-simplex
†convex hull
†affinely independent
†affine space
::face
::subsimplex
::edge
::vertex
::facet
::interior
†affine hull
::simplicial complex
†locally finite
::polyhedron
::k-skeleton
::convex polyhedron
::simplicial
::subdivision

3.2. Gluings and piecewise linear manifolds

As with most raw definitions, direct use of definition 3.1.2 is rare. Atlases are an inconvenient way to describe a manifold, and normally one uses some "higher-level" operation, such as gluing together pieces of $\mathbf{R}^n$ or taking the quotient of an existing manifold by a group, to construct new manifolds. We've used gluings many times already in this book, especially in section 1.3—where we saw that one always gets a manifold by gluing edges of polygonal regions in pairs—and in section 1.4.

As long as we were dealing with surfaces, our intuition was a pretty sound guide. However, some of the conclusions we might reach from this experience break down in high dimensions. In three dimensions, most of these difficulties can still be circumvented, so that throughout most of this book, we will discuss polyhedra and gluings and three-manifolds in a rather intuitive way. But in this section and the next we will work through some of the relevant technical issues so that we can comfortably ignore them subsequently.

We consider first gluings of simplices, then gluings of convex polyhedra, which are more common in practice. Recall that an *n-simplex* σ is the †convex hull of $n+1$ †affinely independent points $v_0, \ldots, v_n$ (in some †affine space, necessarily of dimension at least n). The convex hull of a subset of $\{v_0, \ldots, v_n\}$ is a *face*, or *subsimplex*, of σ. As usual, a two-dimensional face is called an *edge* of σ, and a one-dimensional face (or the unique point in it) is a *vertex*. A face of dimension $n-1$ will be called a *facet*. The *interior* of a simplex is what's left when you take away its proper faces; it can also be defined as the topological interior of the simplex considered as a subset of its †affine hull.

A *simplicial complex* is a †locally finite collection Σ of simplices (in some affine space), satisfying the following two conditions: any face of a simplex in Σ is also in Σ, and the intersection of two simplices in Σ is either empty or a face of both. The union of all simplices in Σ is called the *polyhedron* of Σ, and denoted by $|\Sigma|$. The *k-skeleton* of Σ is the subcomplex consisting of simplices of dimension k or lower.

If a simplicial complex has finitely many simplices, whose union forms a convex set, we recover the familiar notion of a *convex polyhedron*, defined, for instance, as the convex hull of a finite set of points. When we talk about a convex polyhedron we'll be referring to this notion, even if finiteness is not explicitly mentioned.

Given a map from the vertices of a simplicial complex to an affine space, there is a unique way to extend it to the complex's polyhedron so that the map is affine within each simplex. Such a map is called *simplicial*. A *subdivision* of a complex Σ is any complex Σ' having the same polyhedron as Σ, and such that every simplex in Σ is contained in some simplex in Σ'. The importance of simplicial complexes as a technical tool is that an arbitrary continuous map from a polyhedron (into, say, an affine space) can be approximated by a simplicial map in the same homotopy class, if we subdivide the domain appropriately. reference

::piecewise linear equivalent
::piecewise linear
::piecewise linear manifold
% $\mathcal{G}$-manifold
::triangulated
::triangulation
% The totality of surfaces
% triangulation
Problem "triangulating piecewise linear manifolds"
::link
cone on link
::cone
Proposition "spherical links imply manifold"
% cone on link

Naturally, we want to consider two complexes isomorphic if there is a simplicial homeomorphism between the two. But in practice, a somewhat weaker concept of equivalence turns out to be equally important: two complexes are called *piecewise linear equivalent* if they can be subdivided to yield isomorphic complexes.

More generally, a map defined on a subset of an affine space is called *piecewise linear* if it can be extended to a simplicial map on the polyhedron of some simplicial complex. Piecewise linear homeomorphisms between open subsets of $\mathbf{R}^n$ form a pseudogroup, which we call PL. We thus obtain the important notion of a *piecewise linear manifold*, namely, a PL-manifold in the sense of definition 3.1.2.

Problem 3.2.1. Show that PL really is a pseudogroup.

proving property (d) of definition 3.1.1 seems very hard.

The polyhedron of a simplicial complex, or a space homeomorphic to it, is said to be *triangulated* by the complex. More exactly, a *triangulation* of a topological space X is a simplicial complex Σ, together with a homeomorphism $|\Sigma| \to X$; the various concepts defined for simplicial complexes can be transferred to X by the homeomorphism. We've already used triangulations in the context of surfaces, in section 1.3 (see especially figure 1.17).

Problem 3.2.2 (triangulating piecewise linear manifolds). Fix a locally finite atlas for a piecewise linear manifold. Prove that the manifold can be triangulated in such a way that each coordinate chart maps each simplex linearly into $\mathbf{R}^n$. In particular, any piecewise linear manifold can be triangulated.

J. Stallings, Lectures on polyhedral topology, Lecture Notes at Tata Institute, 1968; J. F. P. Hudson, PL Topology, Mathematical Lectu[res] Notes, Chicago, 1966-1967

When is a triangulated space a manifold? To study this question, we need the notion of the *link* of a simplex. If σ is a simplex in a simplicial complex Σ, let $\tau_1, \ldots, \tau_k$ be the simplices of Σ containing σ. For each τ_i, let σ_i be the simplex "opposite" σ in τ_i, in the sense that $\sigma \cap \sigma_i = \emptyset$ and τ_i is the convex hull of $\sigma \cup \sigma_i$. The link $\operatorname{link}(\sigma, \Sigma)$ of σ is the simplicial complex consisting of the σ_i.

Exercise 3.2.3. (a) Show that if Σ' is a subdivision of Σ and σ is a simplex in both Σ and Σ', the links of σ in Σ and Σ' are piecewise linear equivalent.

(b) The *cone* on a topological space X, which we denote by CX, is the product $X \times [0,1]$ with $X \times \{1\}$ identified to a point. Show that a point in the interior of a p-simplex σ has a neighborhood homeomorphic to $D^p \times C(\operatorname{link}(\sigma, \Sigma))$.

Proposition 3.2.4 (spherical links imply manifold). *Let X be a triangulated space. If the link of every simplex dimension p is homeomorphic to an $(n-p-1)$-sphere, X is a topological manifold.*

In fact, it's enough to check that the links of vertices are $(n-1)$-spheres; this implies the condition on links of arbitrary simplices.

Proof of 3.2.4: By exercise 3.2.3(b), every point in X has a neighborhood of the form $D^p \times CS^{n-p-1}$, which is homeomorphic to $D^p \times D^{d-p}$, since the cone on the sphere is a ball. These neighborhoods cover X, so X is a manifold. $\boxed{3.2.4}$

†homology
Example "manifold does not imply spherical links"
::suspension
Poincaré dodecahedral space
% Poincaré dodecahedral space
::gluing
% gluings in two dimensions
Proposition "manifolds have spherical links in dimension three"
::simply connected at x

It would be natural to suppose that the converse is true as well, but this is not the case if $n \geq 5$. The right statement is this: the polyhedron of a simplicial complex is a topological manifold if and only if the link of each cell has the †homology of a sphere, and the link of every vertex is simply-connected. The proof is far beyond the scope of the current discussion, but we mention one famous example due to Edwards:

Supply reference and check correctness: Edwards or Cannon

Example 3.2.5 (manifold does not imply spherical links). The *suspension* of a topological space X, which we denote by ΣX, is the cone on X with $X \times \{0\}$ identified to a point. For example, the suspension of the n-sphere is the $(n+1)$-sphere.

Now if we take the Poincaré dodecahedral space of example 1.4.4, and call it P, the suspension ΣP is not a manifold, because the link of $P \times \{0\}$ is P itself, which is not a sphere. But if we take the double suspension $\Sigma^2 P$ (that is, the suspension of ΣP), this obstruction disappears. Now there is an edge with link P, which is not a sphere, but has the homology of a sphere. It turns out that $\Sigma^2 P$ is homeomorphic to S^5, and is therefore a topological manifold.

Now consider a finite set of n-simplices. A *gluing* is a choice of pairs of simplex facets (each facet appearing in exactly one of the pairs), together with simplicial identification maps between the facets of each pair. We are interested in the identification space, which is the quotient of the union of the simplices by the equivalence relation generated by the identification maps.

Exercise 3.2.6. Show that, given a finite set of simplices and a gluing, the resulting space is homeomorphic to the polyhedron of a simplicial complex.

Exercise 3.2.7. In a gluing of a collection of three-dimensional simplices, each edge enters into exactly two gluings, one for each of its faces. Composing these, one gets a cycle of gluings that eventually must return to the original edge. Suppose that the composition of gluings around the cycle reverses the edge's orientation. Describe a neighborhood of the fixed point of the return map of the edge to itself, in the resulting identification space.

Exercise 3.2.8. Show that a manifold obtained by gluing is orientable if and the faces of each simplex are oriented consistently, and all face identifications are orientation-reversing (cf. exercise 1.3.2).

The following result is a particular case of Edwards' criterion, but we will prove it directly:

Proposition 3.2.9 (manifolds have spherical links in dimension three). *The space obtained by a gluing of three-dimensional simplicies is a three-manifold if and only if the link of every vertex is homeomorphic to S^2.*

Proof of 3.2.9: We need some test that non-manifolds will fail, and we choose local simple connectivity. A topological space X is *simply connected at x*, where $x \in X$, if for any neighborhood U of x there is a smaller neighborhood $x \in V \subset U$ such that any closed loop in the punctured neighborhood $V \setminus \{x\}$

% two-dimensional gluing is a manifold
% a knottier example
Proposition "gluing is manifold iff $\chi = 0$"
% cone on link

is homotopic to a point within $U \setminus \{x\}$. Manifolds of dimension greater than two clearly are simply connected everywhere.

If X is a polyhedron and v is a vertex of X, it is certainly the case that follows that X minus its vertices is a manifold, because the link of any edge is automatically a circle (cf. exercise 1.3.2(b)). Therefore the link of any vertex of X. But it is easy to see that X is simply-connected at v if and only if link(v, X); since the only simply conneced surface is S^2, we conclude that if the link of v is not a sphere, X is not simply connected at v, and therefore is not a manifold. **3.2.9**

Even when a space obtained by gluing three-dimensional simplices is not a manifold, it can be made into one by removing the vertices whose links are not spheres. This is what we did in example 1.4.8. Alternatively, we can remove an open neighborhood of each bad vertex, to obtain a compact manifold with boundary.

Proposition 3.2.10 (gluing is manifold iff $\chi = 0$). *If X is obtained by gluing simplices, X is a three-manifold if and only if its Euler number is zero. In general, if X has k vertices $v_1, \dots, v_k$, we have*

$$\chi(X) = k - \frac{1}{2}\sum_{i=1}^{k} \chi\bigl(\operatorname{link}(v_i, X)\bigr).$$

Proof of 3.2.10: Let e, f and t be the number of edges, two-faces, and tetrahedra in X. Then $f = 2t$, since each face lies on two tetrahedra and each tetrahedron has four faces. We also have

$$\sum_{i=1}^{k} \chi(\operatorname{link}(v_i, X)) = 2e - 3f + 4t,$$

since each edge accounts for two vertices in links of vertices, each face accounts for three edges and each tetrahedron for four faces. The desired equality follows.

Since the Euler number of the two-sphere is 2, and the Euler number of every other closed surface is less than 2, we get $\chi(X) \geq 0$, with equality if and only if X is a manifold. **3.2.10**

Now let X be a space obtained by gluing n-simplices. Since the gluing maps are linear, X can be immediately given the structure of a piecewise linear manifold in the complement of the $(n-2)$-skeleton. Now consider an $(n-2)$-simplex: its link is a circle. By reasoning as in exercise 3.2.3(b), one sees that the interior of the simplex has a neighborhood that is piecewise linear isomorphic to a neighborhood in $\mathbf{R}^n$. Thus, the piecewise linear structure extends to the complement of the $(n-3)$-skeleton.

In general, if the link of every cell of dimension greater than k is isomorphic a piecewise linear sphere, it follows that the link of each k-cell has the structure

Proposition "piecewise linear manifolds have spherical links"
% spherical links imply manifold
::Hauptvermutung
% manifold does not imply spherical links
::vertices
::face
::combinatorially equivalent
::combinatorial class

of a piecewise linear manifold, and one can ask whether this piecewise linear manifold is a standard piecewise linear sphere. With this understanding, the statement of the next proposition makes sense:

Proposition 3.2.11 (piecewise linear manifolds have spherical links). *A space X obtained by gluing simplices is a piecewise linear manifold if and only if the link of each cell is piecewise linear homeomorphic to the standard piecewise linear sphere.*

Proof of 3.2.11: If the link of each cell is a standard piecewise linear sphere, the construction of proposition 3.2.4 gives a piecewise linear atlas.

A triangulation of a piecewise linear manifold is not well defined up to piecewise linear homeomorphism, but nonetheless, the link of any piecewise linear embedded k-cell can be defined, and it is always a $n-k-1$ sphere. (Show how to make sense of the link, show that it is well-defined, and that it is always an $(n-k-1)$-sphere.) This gives the converse. 3.2.11

There are many other interesting problems involving triangulations of manifolds: When can a topological manifold be triangulated, and do any two triangulations of a manifold admit isomorphic subdivisions? This latter question is the famous *Hauptvermutung*, or fundamental conjecture, and it is clearly not true in general: the triangulation of S^5 described in example 3.2.5 is not piecewise linear equivalent to the standard triangulation. The first question has also been answered in the negative for general manifolds: Kirby, Siebenmann and Wall showed there exist topological manifolds of dimension 6 that do not admit any combinatorial structure, and there are non-standard structures for combinatorial manifolds of dimension 5. But in dimension three or lower all topological manifolds are triangulable, and all triangulations are piecewise linear equivalent.

R. Kirby and L. Siebenmann, Bull. Amer. Math. Soc. 75 (1969)

T. Rad'o, Abh. Math. Sem. Univ. Hamburg, 11 (1935); C Papakyriakopoulos and E. Moise, Ann of Math (2) 56 (1952); R. Bing, Ann. of Math. (2) 69 (1959)

So far we've only considered gluings of simplices, and only by affine maps. What happens when we allow gluings of arbitrary convex polyhedra, by arbitrary maps? Is the topology of the identification space determined by just the gluing pattern, or does it depend on the particular choice of a gluing?

The set of *vertices* of a convex polyhedron is the unique minimal set of points whose convex hull is the polyhedron; a *face* is an intersection of the convex polyhedron with a half-space, having dimension less than the dimension of the polyhedron.

Two convex polyhedra are *combinatorially equivalent* if there is a one-to-one correspondence of their faces of every dimension that preserves incidence. Thus the vertices are in one-to-one correspondence; each edge of one polyhedron corresponds to an edge of the other polyhedron with corresponding vertices; and so on.

If α is a combinatorial equivalence between polyhedra P and Q, a homeomorphism $h : P \to Q$ is in the *combinatorial class* of α if h sends each face of every dimension to the face specified by α. It is easy to see that every combinatorial equivalence can be realized by a homeomorphism in its combinatorial

::gluing pattern
::gluing compatibility condition
::barycentric subdivision
Problem "uniqueness of gluings"
The Alexander trick
icosahedron

class, but of course not all homeomorphisms are in the combinatorial class of a combinatorial equivalence.

Now consider a finite set of combinatorial classes of n-dimensional convex polyhedra. A *gluing pattern* is a choice of polyhedron facets (each facet appearing in exactly one of the pairs), together with a combinatorial equivalence between the facets in each pair. A *gluing* realizing this pattern is a choice of actual convex polyhedra of the given combinatorial types, and a choice of actual homeomorphisms between the faces in the given combinatorial equivalence classes.

Not any choice of homeomorphisms will do, however: there is a compatibility condition. Any face β of dimension $n-2$ or less is a face of two or more facets, so it enters into two or more pairings. Compositions of the various identifications will result in the identification of β with possibly many other faces. A bit of reflection shows that there will always be chains of these identifications that take β to itself. The compatibility condition is this: For any face β and any chain of identifications whose composition takes β back to itself in the combinatorial class of the identity, the composed identification must actually be the identity.

One way to choose compatible gluing homeomorphisms makes use of the *barycentric subdivision* of a polyhedron, which is defined just like the barycentric subdivision of a simplex. Given a combinatorial equivalence between two polyhedra, we form their barycentric subdivisions, map corresponding vertices to corresponding vertices, and extend the correspondence to an affine map on each simplex. It is easy to see that if a gluing map is built in this way, it satisfies the compatibility condition.

Problem 3.2.12 (uniqueness of gluings). (a) (The Alexander trick.) Show that a homeomorphism of the unit ball in $\mathbf{R}^n$ that is the identity on its boundary is isotopic to the identity. (Hint: comb all the tangles to a single point.)

(b) Show that any two gluings realizing a given gluing pattern yield homeomorphic identification spaces.

Problem 3.2.13. (a) Give a definition for the the link of a cell in a space obtained by gluing polyhedra.

(b) Extend the various results of this section to this case.

The number of possible gluings for polyhedra grows very fast with the number of facets.

Exercise 3.2.14. (a) In how many ways can one pair each face of a cube with its opposite, so as to produce an oriented three-manifold?

(b) Show that the number of gluing patterns for an octahedron is $8! \cdot 3^4/(4! \cdot 2^4) = 8505$. Can you estimate how many of these yield manifolds? How many yield orientable manifolds?

(c) Same question for an icosahedron.

Example "piecewise projective manifolds"
::piecewise projective
::piecewise projective manifold
Problem "piecewise linear and piecewise projective structures"
::refinement
Example "piecewise integral projective manifolds"

Because of symmetries, many gluing patterns give obviously homeomorphic results—for example, in the case of an icosahedron, this reduces the number of different gluing diagrams by a factor of about 120. (Since some gluings have a certain amount of symmetry, the reduction is not actually quite this much.) This still leaves a huge number which are not obviously homeomorphic. It seems unlikely that the same phenomenon occurs, as in dimension two, when very large numbers of different gluing patterns give homeomorphic manifolds, but how can we tell? We need more tools to be able to name and recognize a three-manifold when it is described in different ways.

We conclude this section with some further examples of $\mathcal{G}$-structures. Although related to piecewise linear structures, these notions will not be used later.

Example 3.2.15 (piecewise projective manifolds). A homeomorphism $\mathbf{R}^n \to \mathbf{R}^n$ is *piecewise projective* if the domain is a union of n-simplices such that the map is a projective transformation on each simplex. A PP-manifold, where PP is the pseudogroup of local piecewise projective homeomorphisms, is called a *piecewise projective manifold.*

If F is a finite set of points in $\mathbf{R}^{n+1}$ which do not all lie in a hyperplane, then boundary of the convex hull of F is of course homeomorphic to a sphere. The sphere automatically has a piecewise projective structure, where coordinate patches are defined by projection of a piece of the boundary from a point inside the convex hull to a plane. When the point changes and the plane changes, the transition functions are piecewise projective.

Problem 3.2.16 (piecewise linear and piecewise projective structures). The piecewise linear pseudogroup is contained in the piecewise projective pseuodgroup, so a piecewise linear manifold automatically has a piecewise projective structure. The piecewise linear structure is a *refinement* of the piecewise projective structure it defines.

(a) Show that every piecewise projective structure can be refined to give a piecewise linear structure.

(b) Show that for any two refinements of a piecewise projective structure on a manifold, there is a piecewise projective transformation of the manifold to itself that carries on to the other. Thus, the refinement is not canonical, but it is canonical up to piecewise projective equivalence.

(c) Show that the boundary of the convex hull of F, as above, has a natural piecewise linear structure. Does the link of a vertex in a piecewise linear-manifold have a canonical piecewise linear-structure, or only a canonical piecewise projective structure?

Not all interesting pseudogroups act transitively. Here is one of the most intriguing examples:

Example 3.2.17 (piecewise integral projective manifolds). The group $\mathrm{PGL}(n+1, \mathbf{R})$ of projective transformations of $\mathbf{RP}^n$ contains the discrete

::piecewise integral projective
Problem "piecewise linear and piecewise integral projective structures"
% The Teichmüller space of a surface

subgroup $\mathrm{PGL}(n+1, \mathbf{Z})$. The pseudogroup PIP of *piecewise integral projective* transformations consists of local homeomorphisms of $\mathbf{R}^n$ for which there is a subdivision of the domain into n-simplices, such that on each one, the homeomorphism is induced by the action of $\mathrm{PGL}(n+1, \mathbf{Z})$ on $\mathbf{R}^n \subset \mathbf{RP}^n$.

Since PIP takes points in $\mathbf{R}^n$ with rational coordinates to points with rational coordinates, it is not transitive.

Problem 3.2.18 (piecewise linear and piecewise integral projective structures). Show that every piecewise linear manifold can be given a piecewise integral projective structure, unique up to piecewise integral projective equivalence.

The unusual aspect to this structure is that the group of piecewise integral projective homeomorphisms of a manifold is only countable. Richard Thompson has shown that the group of piecewise integral projective transformations of the circle is finitely presented and simple. It is not known whether the piecewise integral projective homeomorphism group of a surface is finitely presented.

Reference for Richard Thompson? He uses different language

It turns out that the Teichmüller space of a surface (section 3.8) can be given a sphere at infinity, analogous to that for hyperbolic space, such that the transformations induced by homeomorphisms of the surface are piecewise integral projective.

Section "Smoothings" differentiable manifold Section "diffeomorphism classes of polyhedra" Theorem "linearizing diffeomorphisms" Convention "continuous dependence on diffeomorphism"

3.3. Smoothings

We now have a pretty good theoretical understanding of which gluings of polyhedra give topological three-manifolds, but there still remains a significant issue. Does a space obtained by gluing polyhedra have also the structure of a differentiable manifold? Here is a warm-up exercise:

Exercis 3.3.1. (a) Show that any two n-gons are diffeomorphic. (To be a diffeomorphism, a map should have a maximum rank derivative at every point, including the vertices.)

(b) Show that two combinatorially equivalent convex polyhedra in $\mathbf{R}^3$ having no more than four edges at any vertex are diffeomorphic.

(c) Show that a typical convex icosahedron in the combinatorial class of the regular icosahedron is not diffeomorphic to the regular icosahedron.

This exercise illustrates the need to be careful.

To construct a differentiable structure on a manifold obtained by gluing convex polyhedra, it is a mistake to keep track of the differentiable structure of the entire polyhedra; a better approach is to insist only on retaining their piecewise-differentiable structure.

The construction or proof of uniqueness of a smooth structure is done by induction on dimension. The main step in the induction procedure investigates the link of a simplex. For this reason, the crucial results relate to the structure on a sphere, and our initial objective is to understand better the group of diffeomorphisms of the sphere.

Let $\mathrm{Diff}(S^n)$ be the group of smooth diffeomorphisms of the n-sphere. We topologize $\mathrm{Diff}(S^n)$ with the C^1-topology. That is to say, a diffeomorphism is near the identity if both its values and its derivatives are uniformly near the identity. The orthogonal group $O(n+1)$ is contained in $\mathrm{Diff}(S^n)$. In low dimensions, the inclusion is a homotopy equivalence, but this is false in general.

Silvio, Please check font for Diff. Also, there are two macros, namely diff and Diff. Should these be the same? David

We need references and more facts here. David

Theorem 3.3.2 (linearizing diffeomorphisms). *The inclusion* $O(n+1) \subset \mathrm{Diff}(S^n)$ *is a deformation retract for* $n = 1$ *and* $n = 2$.

This theorem is due to Smale [Sma59].

Proof of 3.3.2:

Convention 3.3.3 (continuous dependence on diffeomorphism). Let f be a diffeomorphism of S^1 or S^2. We will perform various operations on f to change it to an orthogonal transformation. One can check as one goes that each operation is straightforward enough so that if it is carried out simultaneously on all diffeomorphisms, then the effect is continuous as a function of f, so that the desired deformation retraction is indeed constructed.

Silvio. This is not a convention. One needs a macro to enable a reference to a paragraph like this? David

In order to prepare for the non-trivial part of the proof, we will first make the derivative of a diffeomorphism orthogonal at one point, and then make the diffeomorphism equal to that orthogonal transformation in a neighborhood of

::flag
::Gram–Schmidt process
::Gram–Schmidt homotopy
GL$(n, ()\mathbf{R})$
$O(n)$
Lemma "nice patch"
% continuous dependence on diffeomorphism
% parabolic exercise needed

the point. These results are true for all values of n. The problem will then reduce to showing that the group of diffeomorphisms of a disk which are fixed on the boundary is contractible, which is false for general n.

First we do some standard linear algebra. Let $e_0, \dots, e_n$ be the standard orthonormal basis for $\mathbf{R}^{n+1}$. Any basis $x_0, \dots, x_n$ gives rise to a *flag* of half-spaces $X_0, \dots, X_n$, where X_i has dimension $i+1$ and is the unique halfspace containing X_{i-1} and x_i. If we have two bases giving rise to the same flag, then there is a canonical homotopy from one to the other, moving each basis vector linearly, and the flag is fixed by the homotopy. Given a flag, there is a unique orthonormal basis giving rise to the flag. This assignment of an orthonormal basis to an arbitrary basis is known as the *Gram–Schmidt process* and we will call the canonical homotopy of an arbitrary basis to an orthonormal basis the *Gram–Schmidt homotopy*. Note that if the first i vectors of the basis are orthonormal, then they stay fixed during the Gram-Schmidt homotopy.

Exercise 3.3.4. Show that GL$(n, ()\mathbf{R})$ deformation retracts to $O(n)$.

Lemma 3.3.5 (nice patch). *Let D be a round disk in S^n with center $e_0 \in S^n$, and let $G'_n \subset$ Diff(S^n) be the space of diffeomorphisms f of S^n, whose restriction to D is equal to the restriction of some element of $O(n+1)$ and such that $f(-e_0) = -f(e_0)$. Then* Diff(S^n) *deformation retracts to G'_n*

Proof of 3.3.5: We will work as though there is only one diffeomorphism f to worry about (see 3.3.3).

First we show that we can make the derivative orthogonal rather than linear at e_0. Let $x_0 = f(e_0)$ and let $x_1, \dots, x_n$ be the image of $e_1, \dots, e_n$ under the derivative of f at e_0. Let X be the $(n+1) \times (n+1)$ matrix with columns the x_i, and let X_t be the Gram–Schmidt homotopy, with $X_0 = X$ and X_1 orthogonal. The first column of X_t is x_0. Let $\rho\colon \mathbf{R}^{n+1} \to S^n$ be radial projection. Consider the isotopy $\rho X_t X^{-1} f\colon S^n \to S^n$. The isotopy keeps the image of e_0 at x_0 the image of $-e_0$ at $-f(e_0)$ and, at time t, the derivative at e_0 is X_t. It follows that we may restrict our attention to diffeomorphisms with X orthogonal.

exercise about parabolics needed: David

By ??, if we are given three points A, B and C on S^n, such that A is distinct from B and C, there is a unique parabolic element g with fixed point A such that $g(B) = C$. The map g depends continuously on A, B and C. Moreover the derivative of g at the fixed point is the identity.

We apply this to our situation with $A = f(e_0)$, $B = f(-e_0)$ and $C = -f(e_0)$ to show that we can restrict attention to f such that $f(-e_0) = -f(e_0)$ and with derivative at e_0 orthogonal. Let $X \in O(n+1)$ be the unique orthogonal map given by f and the derivative of f at e_0. Then $f^{-1}X\colon S^n \to S^n$ fixes e_0 and $-e_0$ and has derivative equal to the identity at e_0.

Conjugating with stereographic projection from $-e_0$, we get a smooth map $h\colon \mathbf{R}^n \to \mathbf{R}^n$ fixing the origin and with derivative the identity. We use a bump function ϕ, with support in a small disk centered at the origin, to create an isotopy moving from h to a map which is equal to the identity near the origin

% averaging
% nice patch
Exercise "averaging"
% nice patch
% nice patch
% logarithm

and is equal to h outside the unit disk. There are some technical points to check here (see exercise 3.3.6).

Finally, we expand the small disk radially until it is equal in size to the disk D in the statement of lemma 3.3.5. $\boxed{3.3.5}$

Exercise 3.3.6 (averaging). Given a bump function ϕ with support in the unit disk in $\mathbf{R}^n$, let $\phi_\varepsilon(x) = \phi(x/\varepsilon)$. Let h be a diffeomorphism of $\mathbf{R}^n$ keeping the origin fixed with derivative the identity.

(a) Show that $t\phi_\varepsilon \mathrm{id} + (1 - t\phi_\varepsilon)h$ gives an isotopy of h for ε small enough, depending on h.

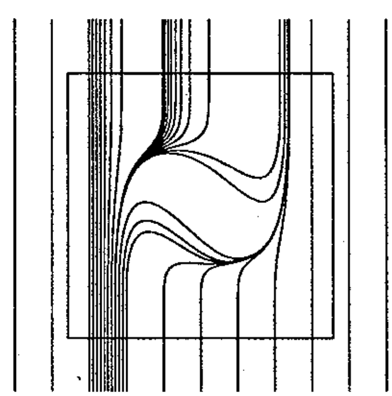
(b) Explain why it is probably not an isotopy when ε is too large.

(c) Show that ε can be chosen in a manner which depends only on the restriction of h to the unit disk, and in a manner which depends continuously on this restriction. (This only makes sense if we have a topology on the space of maps of the closed unit disk into $\mathbf{R}^n$—we use the C^1-topology.)

(d) The isotopy created in the proof of lemma 3.3.5 depends continuously on the restriction of h to the unit disk and therefore continuously on f.

We have shown in lemma 3.3.5 that we may assume that the diffeomorphisms we are considering are orthogonal in a disk D. In the case of S^1, we can now change the diffeomorphism by a linear isotopy on the complementary interval (see exercise 3.3.7).

It follows that we need only show that the group G of diffeomorphisms of S^2, which are equal to the identity on D, is contractible. This means that we only need to prove that the group of smooth diffeomorphisms of the plane, which are the identity outside the unit square, is contractible.



VectorField

Figure 3.2. Vector field on the square. Trajectories of a smooth vector field which is constant and vertical outside the unit square.

Now that we have set the stage, we come to the important part of the proof. Let $h\colon \mathbf{R}^2 \to \mathbf{R}^2$ be fixed outside the unit square. and let V be the image of the unit vector field $(0,1)$ under h. We normalize V to get a unit length vector field. We have a map $\mathbf{R}^2 \to S^1$ which gives the direction of V.

% closed orbits
Exercise "logarithm"
Problem "closed orbits"
% classification of surfaces
% Poincaré index theorem
Problem "reparametrization"
% linearizing diffeomorphisms
% linearizing diffeomorphisms
% linearizing diffeomorphisms
Corollary "extending diffeomorphisms"
% linearizing diffeomorphisms

Outside the unit square the map is constant. This map is homotopic by a smooth homotopy to a point without changing it on the complement of the unit square, as we can see by lifting to the covering $\mathbf{R} \to S^1$. Let V_t be the unit vector field at time t. V_t depends continuously on the pair (t, h), where $0 \leq t \leq 1$, and h is a diffeomorphism of $\mathbf{R}^2$ which is fixed outside the unit square, and, for fixed $V_t(x)$ depends smoothly on the pair (x, t), where $x \in \mathbf{R}^2$.

Now consider the trajectories of V_t for some fixed value of t. We claim that the orbit of any point in the square starts from the bottom edge of the square and ends on the top edge. For otherwise the orbit accumulates in the square at some point p. We then obtain a closed orbit by adjusting the vector field in a neighborhood of p. This is impossible (see problem 3.3.8)

We now reparametrize the trajectories so that it takes unit time to go from the bottom of the square to the top along any trajectory. Integrating the vector fields, we get the required isotopy. $\boxed{3.3.2}$

Exercise 3.3.7 (logarithm). Make sense of the assumption made in the proof above that an orthogonal map of S^1 can be regarded as linear on an interval contained in the circle.

Problem 3.3.8 (closed orbits). Prove that a never zero vector field in the plane has no closed orbits. (Hints: Use problem 1.3.12 and proposition 1.3.10. Move the vector field to make it point inwards along the closed orbit.) The answer to this question is a part of the famous Poincaré-Bendixson theorem.

Problem 3.3.9 (reparametrization). We give the space of smooth vector fields on the square the C^1-topology. Here are some details that should be checked in the proof of theorem 3.3.2.

(a) Prove that, with an appropriate choice of homotopy, V_t varies continously in the C^1-topology as a function of the given diffeomorphism and of t.

(b) Prove the reparametrization described in the last paragraph of the proof of theorem 3.3.2 can be done in a way that varies continously with the vector field in the compact open topology. (Hint: use the inverse function theorem on a single vector field. Use facts about approximate solutions to differential equations and continuous dependence of inverse functions on parameters for the continuity.)

I am fairly confident this hint is right, but I haven't checked in painful detail. David

Exercise 3.3.10. Why doesn't the proof of theorem 3.3.2 go through in higher dimensions?

1

Corollary 3.3.11 (extending diffeomorphisms). *Let $n = 1$, 2 or 3. Let D^n be the unit disk, with boundary S^{n-1}. Then the homomorphism* $\mathrm{Diff}(D^n) \to \mathrm{Diff}(S^{n-1})$, *given by restriction, is a homotopy equivalence. In particular any diffeomorphism of S^{n-1} can be extended to D^n.*

Proof of 3.3.11: To construct the inverse map to the homomorphism, we start with a diffeomorphism and use Smale's theorem (3.3.2) to construct an isotopy

% extending diffeomorphisms
Challenge "extending diffeomorphisms to a simplex"
::triangulation
::C^r-triangulation
Theorem "smoothing"
S^7

to an orthogonal map. Let B be a small ball with center the origin of D^n. The isotopy is used to construct a diffeomorphism of the complement of B to itself, which restricts to the given diffeomorphism on S^{n-1} and it an orthogonal map on the boundary of B. We extend the orthogonal map to B radially from the origin. [3.3.11]

Exercise 3.3.12. Fill in the details of the proof of 3.3.11. Show that all the constructions depend continously on the data. Show how to carry out the construction so that the final map is smooth in a neighborhood of the boundary of B. Explain how a careless interpretation of the words of the proof would lead to a map which is continuous but not differentiable on the boundary of B.

Challenge 3.3.13 (extending diffeomorphisms to a simplex). Amongst other things this exercise gives a very nice and very explicit way of approximating the boundary of a simplex. exercise??? David

(a) Let σ be a linear n-dimensional simplex in $\mathbf{R}^n$ and let $\lambda_0, \ldots, \lambda_n$ be the barycentric coordinate functions for σ. Prove that for fixed small positive ε $\lambda_0 \ldots \lambda_n = \varepsilon$ is the equation of an $(n-1)$-manifold, one of whose components S_ε inside σ is diffeomorphic to S^{n-1}.
(b) Show that S_ε is smooth and that it encloses a convex set.
(c) As ε tends to zero, S_ε tends to the boundary of σ in the Hausdorff topology
(d) Let f be a diffeomorphism of a neighborhood U in σ of the boundary of σ into σ. For ε sufficiently small, fS_ε is a sphere bounding a set which is star-shaped from the barycenter of σ.
(e) Let W be the complement in σ of a small round ball centered at the barycenter. The restriction of f to a smaller neighborhood of the boundary can be extended to a diffeomorphism of W with itslf.
(f) If $n \leq 3$, the restriction of f can be extended to a diffeomorphism of σ with itself.

A *triangulation* of a topological space X is a simplicial complex K and a homeomorphism $\phi: K \to X$. Let M be a smooth manifold. A triangulation of M is said to be a C^r*-triangulation* if the restriction of the homeomorphism to each closed simplex is a C^r-diffeomorphism.

Theorem 3.3.14 (smoothing). *Let M be a piecewise linear manifold of dimension one, two or three. Then M has a smooth structure and a C^∞-triangulation, such that the triangulation has the given piecewise linear structure. Any two such structures are diffeomorphic to each other.*

We emphasize that these conclusions are false in higher dimensions. There are piecewise linear manifolds that cannot be made smooth, and there are also piecewise linear manifolds which have essentially distinct smoothing. The famous first examples of these were due to Milnor [Mil56], who proved that there are differentiable manifolds homeomorphic to S^7 but not diffeomorphic to S^7. In dimensions one, two and three, the concepts of topological, piecewise linear and differential manifolds are essentially equivalent, but we do not prove

% linearizing diffeomorphisms
% triangulating piecewise linear manifolds

the equivalence of the topological case with the others. (See Moise [Moi77], [Bin59] and other references for a discussion of the problem.)

Freedman and Donaldson should be asked to give an authoritative statement on the current situation in dimension 4. David

Proof of 3.3.14: This proof consists largely of not very interesting technical details, which could be carried out in different ways. Whatever one does, the crux of the matter is the application of Smale's theorem (3.3.2).

It would be nice to have a historical discussion, references to the Hauptvermutung statement and Sullivan's solution, references to discussions of topological three-manifold (Moise), references to multiple differentiable structures on spheres ... but

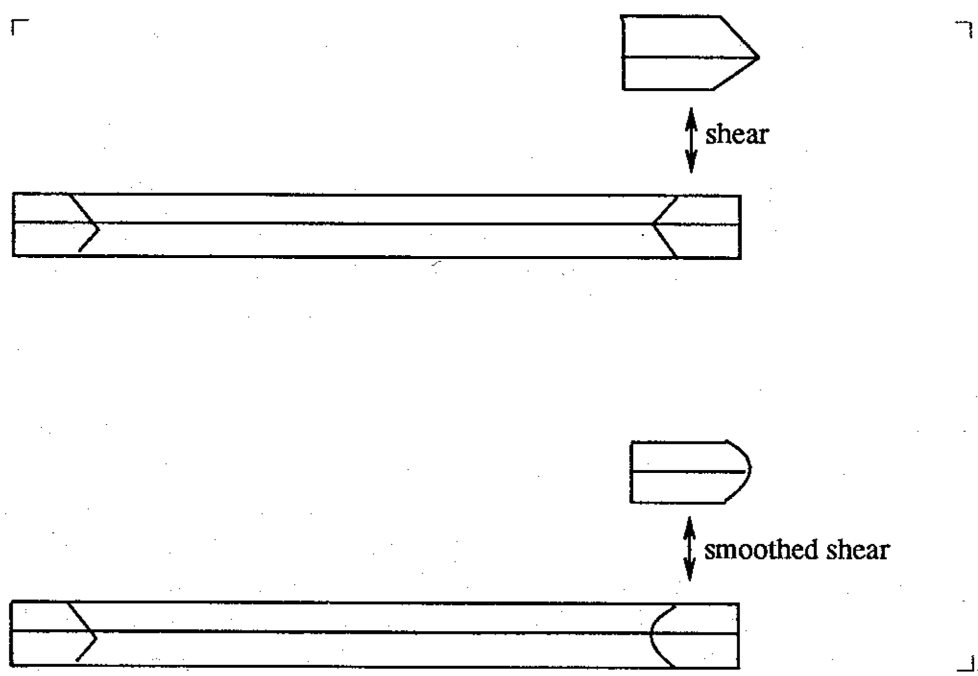
We prove existence first. Let n be the dimension of our manifold. We may assume that our manifold is triangulated, that for each simplex σ there is a piecewise linear coordinate chart which is defined on a neighborhood of the closed star of σ and that each coordinate chart maps each simplex linearly (see ??).

The differentiable coordinate charts in the neighborhood of each vertex are precisely the same as the piecewise linear coordinate charts. Now we extend the differentiable structure to a neighborhood of the edges. Let e be an edge, and let ϕ be a piecewise linear coordinate chart defined on the small open neighborhood N of the interior of e to $\mathbf{R}^n$ which is linear on the intersection of N with each simplex.

1manifold

Figure 3.3. charts for one dimensional manifold. Charts for a one-manifold which are linear on each simplex give an affine structure.

When $n = 1$, the transition functions are already smooth (in fact affine), and so the existence of a differentiable structure is proved.



Shear

Figure 3.4. smoothing a shear. In a neighborhood of an edge, the transition function of a triangulated PL manifold is a shear.

% extending diffeomorphisms
% 2smoothing

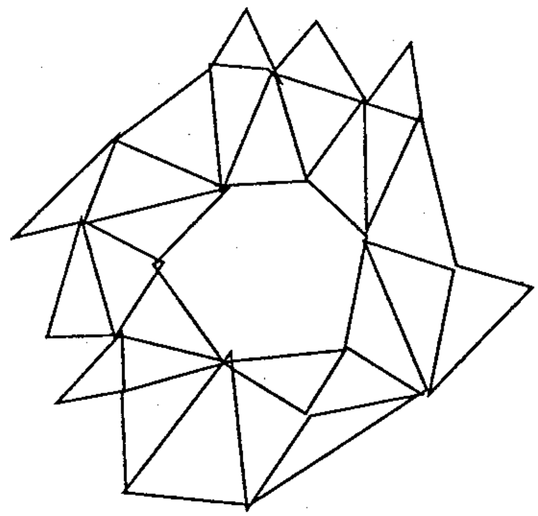
When $n = 2$, N consists of two closed strips S' and S'', whose intersection is e. Consider the portion of N on which two possibly different differentiable structures have been defined. This portion is the disjoint union of two connected pieces N_1 and N_2. The transition function is linear on $S' \cap N_1$ and on $S'' \cap N_1$. We can therefore deform ϕ so that the transition function becomes smooth. It is only necessary to perform the deformation in an extremely small neighborhood of e. Away from e, ϕ is unaltered.

On each triangle, we now have a differentiable structure in a neighborhood of the edges. Moreover, in an annulus going round the boundary of the triangle, a little away from the edge, the original linear chart on the triangle gives the differentiable structure. Therefore the differentiable structure extends. This completes the proof of the existence of a differentiable structure on a piecewise linear surface.

Now let $n = 3$. We first have to extend the differentiable structure already found in the neighborhood of each vertex over the sets N defined above. The picture of N is a bar with central axis e and flanges radiating out from e. At the two ends of N, we have pieces N_1 and N_2 on which two distinct differentiable structures have been defined. We restrict our attention to N_1. N_1 consists of a finite number of wedges and each wedge has two flanges appearing in its boundary. We will change ϕ, the coordinate chart for N, so that the transition function becomes smooth. We will express this purely in terms of changing the transition function itself. Since we are fixing the coordinate charts for the neighborhoods of the vertices, this forces a change in ϕ on N_1. We will omit the easy details associated to the concomitant change in ϕ which is necessary near N_1.

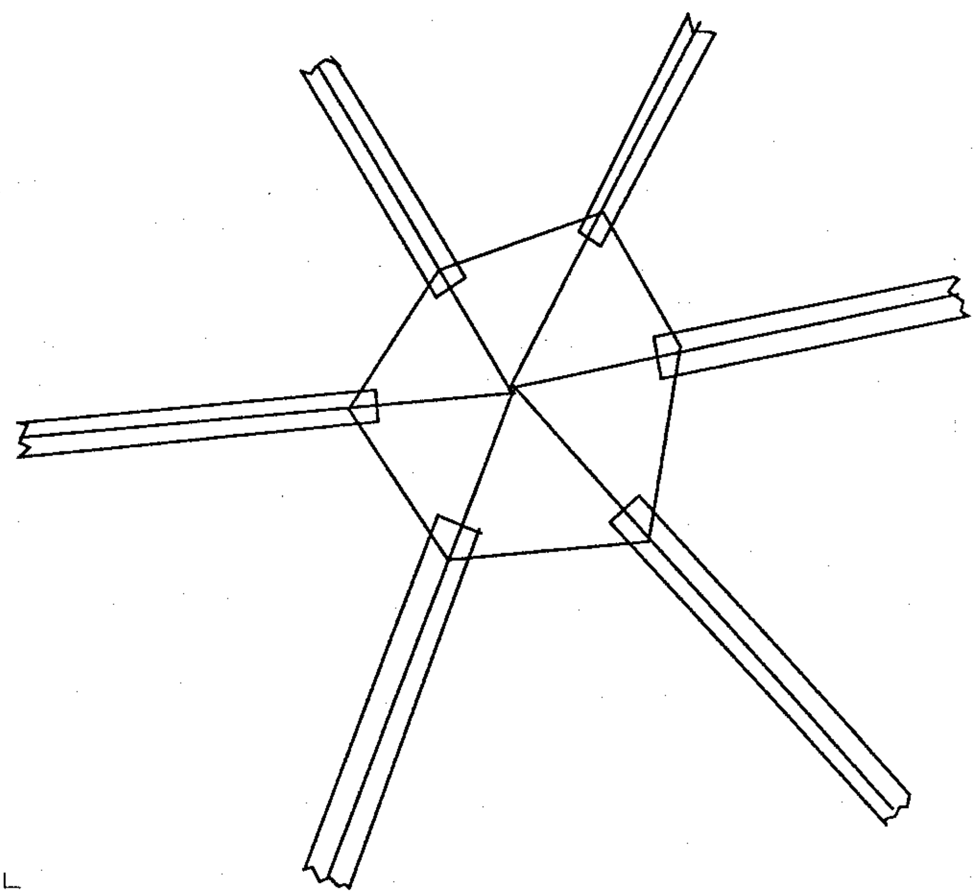
The initial transition function is linear on the intersection of N_1 with any simplex. We regard it as being the identity on e. The first step is to make the transition function for N_1 send each two-dimensional disk normal to e into itself. The circles which bound these disks are triangulated by their intersection with the triangulation of M. We adjust the transition function so that these circles are mapped diffeomorphically and so that vertices are fixed. The diffeomorphism can be extended to the disks bounded by the circles, using corollary 3.3.11 (see figure 3.7). The differentiable structure has now been extended to a neighborhood of e, consistently with the given differentiable structure on each closed simplex.

Now we extend the differentiable structure to a neighborhood of a 2-simplex σ. Let N be an open neighborhood of the interior of σ. We have a chart $\phi\colon N \to \mathbf{R}^3$ which is linear on the intersection of N with any simplex. The boundary of σ has an annular neighborhood in σ where the differentiable structure has already been defined. Let $N_1 \subset N$ be a neighborhood of the annulus in M where the differentiable structure has already been defined. There are therefore two differentiable structures on N_1 and the transition function is smooth except on $\sigma \cap N_1$, on which it is smooth in the tangential direction. We straighten out the normals to σ to make the transition function smooth on N_1.



This stupid fig program erased the middle of the picture above. Please reinsert with a decent drawing program.

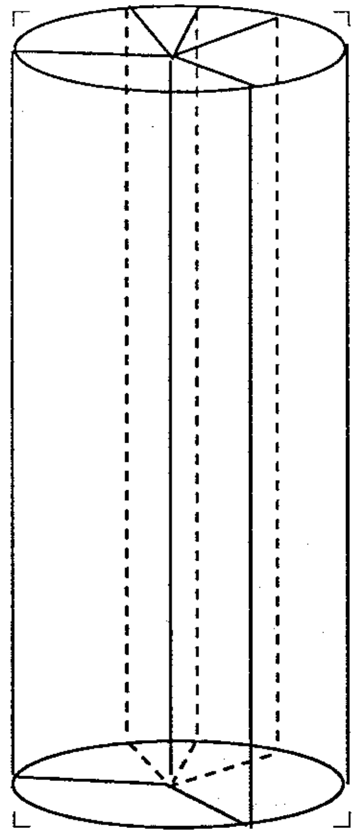
Neighborhood of vertex scaled up.



VertexNeighborhood

Figure 3.5. neighborhood of a vertex. The first picture shows the star of a vertex in a triangulated surface, and some of the extra structure we are using.

Lemma "dimension 1 uniqueness"
Lemma "dimension 2 uniqueness"



3dEdgeNhd

Figure 3.6. flanges. The neighborhood of an edge in dimension three.

We now have a differentiable structure in a neighborhood of the 2-skeleton, which agrees with the given differentiable structure on each closed simplex. Therefore we can extend the structure to each 3-simplex by using a linear chart on the interior of the 3-simplex. This completes the proof of the existence of a smooth structure on a 3-dimensional piecewise linear manifold.

The next task is to prove the uniqueness of the smooth structure. We recall that we are assuming that there is a triangulation of M and two differentiable structures, such that the differentiable structure on each closed simplex is given by a linear embedding in $\mathbf{R}^n$.

Lemma 3.3.15 (dimension 1 uniqueness). *Let M be a one-manifold and let $\phi: K \to M$ be a triangulation of M. Suppose we have two differentiable structures on M, such that ϕ is a C^∞-triangulation for each of them. Then the two structures are diffeomorphic by a diffeomorphism which sends each simplex to itself.*

Proof of 3.3.15: We construct a diffeomorphism near the vertices and extend it. $\boxed{3.3.15}$

Lemma 3.3.16 (dimension 2 uniqueness). *Let M be a two-manifold and let $\phi: K \to M$ be a triangulation of M. Suppose we have two differentiable structures on M, such that ϕ is a C^∞-triangulation for each of them. Let N*

% 2smoothing
% extending diffeomorphisms
Lemma "uniqueness dimension 3"
% dimension 2 uniqueness
% extending diffeomorphisms
% xxx
% extending diffeomorphisms

be a neighborhood of the vertices. Then the two structures are diffeomorphic by a diffeomorphism which sends each vertex to itself, and, for each simplex σ*, sends* $\sigma \backslash N$ *to* σ*.*

Proof of 3.3.16: We construct a diffeomorphism h near the vertices which sends each vertex to itself and, near each vertex, but not too near, sends each point on an edge e into e (see figure 3.7). It is easy to extend h to a neighborhood of the 1-skeleton in such a way that any point which is not too near the 0-skeleton and which lies in an edge e is sent into e.

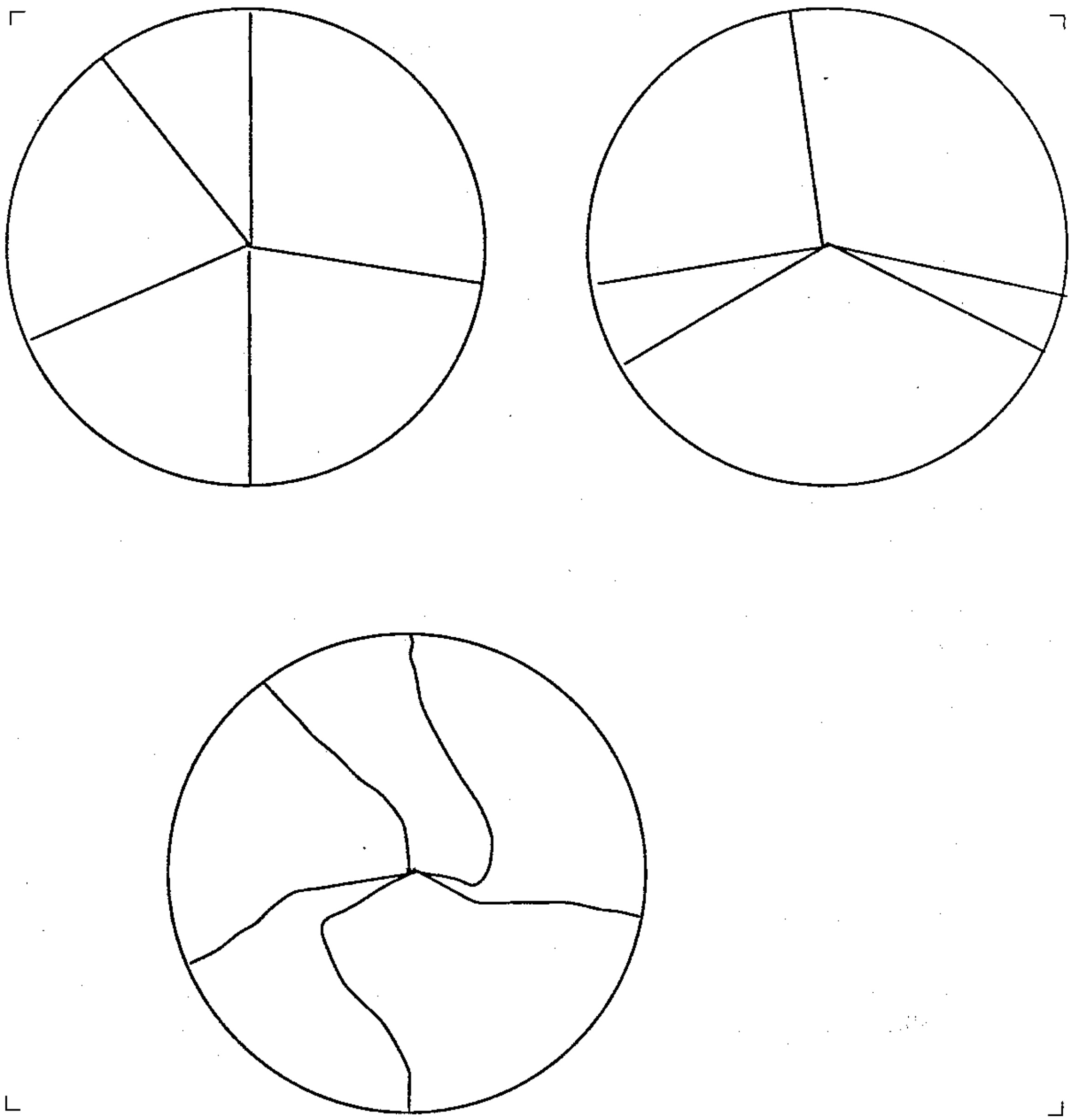
Now we want to extend h to each triangle, having already defined h on a neighborhood of the boundary. There is enough control of the diffeomorphism near the vertices (but not too near) and near the edges, to enable us to construct a smooth simple closed curve, which is star-shaped from the barycenter, whose image under h is also star-shaped from the barycenter. We can then extend h to the whole triangle, using corollary 3.3.11 in dimension one. 3.3.16

Lemma 3.3.17 (uniqueness dimension 3). *Let* M *be a three-manifold and let* $\phi: K \to M$ *be a triangulation of* M*. Suppose we have two differentiable structures on* M*, such that* ϕ *is a* C^∞*-triangulation for each of them. Let* N_0 *be an open neighborhood of the vertices and let* N_1 *be a neighborhood of the 1-skeleton. Then the two structures are diffeomorphic by a diffeomorphism which sends each vertex to itself, for each edge* e *sends* $e \backslash N_0$ *to* e*, and for each triangle* t *sends* $t \backslash N_1$ *to* t*.*

Proof of 3.3.17: Let v be a vertex and let S' and S'' be small smooth 2-spheres surrounding v in the two differentiable structures. Then S' and S'' give two smooth structures which are equivalent as PL manifolds. By lemma 3.3.16, there is a diffeomorphism from S' to S'' which sends vertices of S' to vertices of S'', most of each edge of S' to an edge of S'' and most of each triangle of S' to a triangle of S''. By corollary 3.3.11, h can be extended to map the three-dimensional ball bounded by S' to the ball bounded by S''. We may assume that h sends the vertex at the center of the ball to itself.

Now we extend h to a neighborhood of the edges. Such a neighborhood is a solid cylinder with central axis an edge e and with flanges corresponding to the various triangles which have e as an edge. The boundary of the solid cylinder consists of two disks, lying on spheres surrounding vertices of M, and a cylindrical surface S. The intersection of S with the triangles of M consists of a finite set of arcs (see figure ??) going from one end of the cylinder to the other.

We extend h to a diffeomorphism of S with the following properties. If D is a disk normal to e in the one structure, then it is mapped to a disk normal to e in the other structure, through the same point of e. If I is an arc of intersection of a two-simplex σ with S, then it is mapped into σ. Using corollary 3.3.11, we extend the diffeomorphism to D, sending the edge e to itself by the identity. We have now defined h in a neighborhood of the one-skeleton.



2smoothing

Figure 3.7. smoothing a piecewise linear homeomorphism. The first and second circles enclose neighborhoods of a vertex. The third circle shows the image of the second neighborhood in the first under a diffeomorphism

% dimension 2 uniqueness
% Alexander filling
% extending diffeomorphisms
diffeomorphic embeddings of the unit ball
diffeomorphism of the unit disk
Riemann mapping theorem
Section "["
% smoothing algorithms

Next we extend h to the two-dimensional simplices, as in lemma 3.3.16. Then we extend h to a neighborhood of each triangle, by defining it on short curves normal to the triangles.

Finally, we want to extend h to each tetrahedron, having defined it on a neighborhood of the boundary of the tetrahedron. Let v be the barycenter of the tetrahedronand let B be a small round ball centered at v. We have sufficient control so that we can construct a smooth 2-sphere near the boundary, star-shaped from v, sent to a smooth 2-sphere near the boundary, also star-shaped from v. Since the spheres are star-shaped, they bound balls. An alternative to making the spheres star-shaped it to quote Alexander's theorem (??). By the corollary to Smale's theorem (3.3.11), we can extend h to the tetrahedron. $\boxed{3.3.17}$

This completes the proof of the theorem. $\boxed{3.3.14}$

Problem 3.3.18. (a) Suppose a smooth manifold M of dimension m is embedded smoothly as a closed subset of a smooth manifold N of dimension $m+1$ and that M and N are orientable. Suppose this is done twice, with the same M, but with two different manifolds N, namely N_1 and N_2. Prove that there is a neighborhood of M in N_1 which is diffeomorphic to a neighborhood of M in N_2, such that the diffeomorphism is the identity on M. For the purpose of this chapter, this only needs to be proved for $m = 0$, 1 and 2.

(b) Prove that any diffeomorphic embeddings of the unit ball B^n in $\mathbf{R}^n$ is isotopic among diffeomorphic embeddings to the linear map having the same derivative at the origin.

(c) Prove that any diffeomorphism of the unit disk D^2 to itself that fixes three points on its boundary is isotopic among such diffeomorphisms to the identity. [Hint: use the Riemann mapping theorem creatively.]

Exercis 3.3.19. canonical smoothing of a three-manifold] From theorem 3.3.14 we know that a three-manifold has a differentiable structure canonical up to diffeomorphism. Can you find a canonical definition for a differentiable structure, given a gluing pattern? In other words, can you find rules for the construction of a differentiable structure so that no arbitrary choices need to be made in the course of the construction?

If such a definition exists, it can be expressed as a collection of open coordinate charts in $\mathbf{E}^3$, together with pasting diffeomorphisms identifying an open subset of one with an open subset of another, such that there is some "formula" for the pasting diffeomorphism in terms of the combinatorics of the polyhedra and their gluing pattern.

Section "Geometric structures on manifolds"
::(G, X)-manifolds
Example "Euclidean manifolds"
::Euclidean
::flat
% Polygons and surfaces
% The totality of surfaces
% The Teichmüller space of a surface
% Euclidean manifolds and crystallographic groups
% euclideanmanifold
% Three-dimensional Euclidean manifolds
% euclideanmanifold
% gluing is manifold iff $\chi = 0$

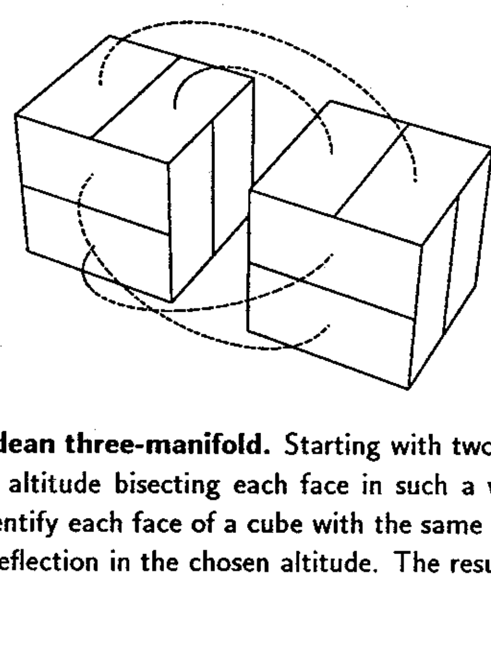
3.4. Geometric structures on manifolds

It is sometimes convenient to permit $\mathcal{G}$ to be a pseudogroup acting on an arbitrary manifold, although as long as $\mathcal{G}$ is transitive, this does not give any new types of manifolds. Given a group G acting on a manifold X, we can consider the pseudogroup $\mathcal{G}$ of restrictions of elements of G to open sets in X; we will generally talk about *(G, X)-manifolds* instead of $\mathcal{G}$-manifolds. Many important pseudogroups are of this form:

Example 3.4.1 (Euclidean manifolds). If G is the group of isometries of Euclidean space $\mathbf{E}^n$, a $(G, \mathbf{E}^n)$-manifold is called a *Euclidean*, or *flat*, manifold; the structure of these manifolds is what we discussed informally in section 1.1. As we saw in section 1.3, the torus and the Klein bottle are the only compact two-dimensional manifolds that can be given Euclidean structures, but there are many such structures. We'll return to this question in section 3.8.

It is an altogether non-trivial result, due to Bieberbach [Cha86], that, in any dimension, there are only finitely many compact Euclidean manifolds up to homeomorphism, and that any such manifold can be finitely covered by a torus of the same dimension. For more details, see section 4.3.

Figure 3.8 shows a gluing construction for a three-dimensional example, which we'll call G_6 in section 4.4, where three-dimensional Euclidean manifolds are classified.



euclideanmanifold

Figure 3.8. A Euclidean three-manifold. Starting with two identical cubes, each marked with an altitude bisecting each face in such a way that no two altitudes intersect, identify each face of a cube with the same face of the other cube, by means of a reflection in the chosen altitude. The result is a Euclidean manifold.

Exercise 3.4.2. Prove the last assertion in the caption of figure 3.8.

(a) Show that the identification space is a manifold. (This will follow trivially from proposition 3.2.10.)

(b) Find a group of isometries of $\mathbf{E}^3$ whose quotient is our manifold. The existence of this group shows that the manifold has a Euclidean structure.

Example "elliptic manifolds"
::spherical
::elliptic
% Poincaré dodecahedral space
% lens spaces
Example "hyperbolic manifolds"
::hyperbolic manifold
% Hyperbolic surfaces
% Seifert–Weber dodecahedral space
% G-manifold
% Poincaré dodecahedral space
% Seifert–Weber dodecahedral space
% lens spaces
% euclideanmanifold
Example "the figure-eight knot complement"
% a knottier example
% idealtetrahedron
% idealtetrahedron
% a knottier example

(c) Find a subgroup of finite index isomorphic to $\mathbf{Z}^3$; by looking at the quotient of $\mathbf{E}^3$ by this subgroup, we see the manifold can be finitely covered by a torus, verifying Bieberbach's theorem.

Example 3.4.3 (elliptic manifolds). If G is the orthogonal group $O(n+1)$ acting on the sphere S^n, a (G, S^n)-manifold is called *spherical*, or *elliptic*. The Poincaré dodecahedral space (example 1.4.4) and lens spaces (example 1.4.6) are spherical manifolds.

Example 3.4.4 (hyperbolic manifolds). If G is the group of isometries of hyperbolic space $\mathbf{H}^n$, a $(G, \mathbf{H}^n)$-manifold is a *hyperbolic manifold*. We discussed hyperbolic surfaces in section 1.2 and a three-dimensional example, the Seifert–Weber dodecahedral space, in example 1.4.5.

In each of the three preceding examples, showing that a manifold has the specified geometric structure amounts to showing that each point has a neighborhood isometric to the appropriate ball. (You should justify this in the light of definition 3.1.2.) This is certainly the case if the manifold is a quotient of $\mathbf{E}^3$, S^3 or $\mathbf{H}^3$ by a group of isometries.

Alternatively, if the manifold is defined by gluing pieces of $\mathbf{E}^3$, S^3 or $\mathbf{H}^3$ with geodesic boundary, the condition can be verified by checking that the dihedral angles add up to 360° around each edge, and that corners fit together exactly. We checked the edge condition (but not the corner condition!) in examples 1.4.4, 1.4.5 and 1.4.6. Likewise, we can see that the manifold of figure 3.8 is Euclidean simply by observing that the edges of the cubes are identified in groups of four, with dihedral angles of 90°, and the corners are identified in groups of eight octants.

We use the preceding remark to give a hyperbolic structure to several open manifolds:

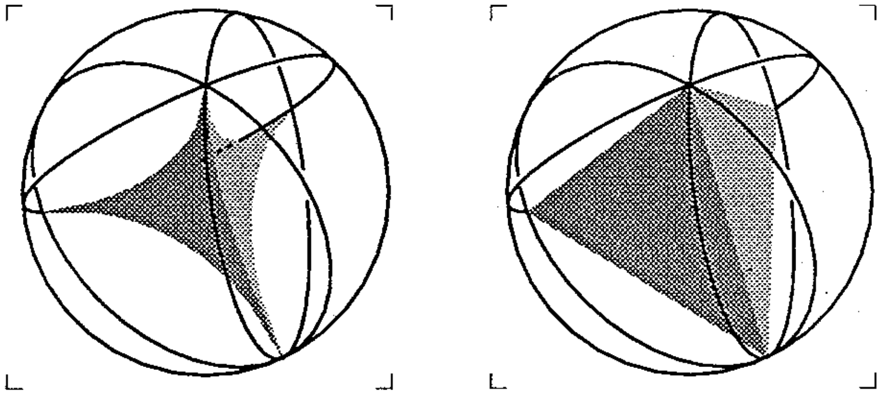
Example 3.4.5 (the figure-eight knot complement). We saw in example 1.4.8 how a certain gluing of two tetrahedra (minus their vertices) yields a space homeomorphic to the complement of a figure-eight knot in S^3. We now give this space a hyperbolic structure.

Let the two tetrahedra be regular ideal tetrahedra in hyperbolic space, that is, regular tetrahedra whose vertices are at infinity. Combinatorially, a regular ideal tetrahedron is a simplex with its vertices deleted; geometrically, it can be modeled on a regular Euclidean tetrahedron inscribed in the unit sphere, interpreted in the projective model. The dihedral angles of this polyhedron are 60°, as can be seen from figure 3.9.

Exercise 3.4.6. Justify the caption of figure 3.9, by showing that the angle between two planes in hyperbolic space is the same as the angle of their bounding circles on the sphere at infinity, seen in the Poincaré ball model.

Now we glue the faces of the two tetrahedra using hyperbolic isometries, following the combinatorial pattern of example 1.4.8. As we discussed there, edges are identified six at a time, so the dihedral angles around each edge add

% the figure-eight knot complement
Daryl Cooper
Colin Adams
Jeff Weeks
% whitelink
::Whitehead link
% whitelink



idealtetrahedron

Figure 3.9. A regular ideal tetrahedron. The dihedral angles of a polyhedron can be read off from the angles between the circles at infinity determined by the faces in the Poincaré model (a); for an ideal polyhedron, the circles are the same as in the projective model, and in this case they equal 60°, by symmetry (b).

up to 360°. We don't have to check the vertex condition, because there are no vertices.

Apparently, this manifold was first constructed by Gieseking in 1912 [Mag74], without any relation to knots. Riley [Ril82] independently constructed this hyperbolic manifold and proved that it was homeomorphic to the figure-eight knot complement by an indirect method, involving fundamental groups. Troels Jorgensen constructed the same manifold from an entirely different point of view, as a punctured torus bundle over the circle. In addition, this example is closely connected to a family of groups studied by Bianchi [Bia92] in connection with number theory: the fundamental group of the figure-eight knot complement is a subgroup of index 12 in $PSL(2, \mathbf{Z}[\omega])$, where $\omega = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ is a primitive sixth root of unity. reference to Jorgensen

The technique of example 3.4.5 may seem very special, but it actually applies to many different knot and link complements. One divides the complement into a union of ideal polyhedra, then attempts to realize these polyhedra as ideal hyperbolic polyhedra and glue them together to form a hyperbolic manifold. Several people have implemented this procedure on computers, including Daryl Cooper, Colin Adams, and Jeff Weeks.

Problem 3.4.7. Figure 3.10 shows the *Whitehead link*,

/u/gt3m/3mbook/pictures/chap3/4/whitelink.ps not found

whitelink

Figure 3.10. The Whitehead link and its spanning two complex

The Whitehead link may be spanned by a two-complex which cuts the complement into an octahedron, with vertices deleted (see figure 3.10). The one-cells are the three arrows, and the attaching maps for the two-cells are

% whitelinkoct
% borromspan
% Borromgludiag
% borromspan
Example "affine manifolds"
::affine manifold

indicated by the dotted lines. The three-cell is an octahedron (with vertices deleted), and the faces are identified as in figure 3.11

/u/gt3m/3mbook/pictures/chap3/4/whitelinkoct.ps not found

whitelinkoct

Figure 3.11. One vertex has been removed so the polyhedron can be flattened in the plane.

A hyperbolic structure may be obtained from a Euclidean regular octahedron inscribed in the unit sphere. Interpreted as lying in the projective model for hyperbolic space, this octahedron is an ideal octahedron with all dihedral angles $90°$ (again, this can be seen by extending the faces to circles on the sphere at infinity). Gluing it in the indicated pattern, again using hyperbolic isometries between the faces gives a hyperbolic structure for the complement of the Whitehead link.

/u/gt3m/3mbook/pictures/chap3/4/idealoct.ps not found

idealoct

Figure 3.12. Ideal octahedron with $90°$ dihedral angles.

/u/gt3m/3mbook/pictures/chap3/4/borromspan.ps not found

borromspan

Figure 3.13. The Borromean rings and their spanning two-complex

A slightly more intricate example of the above technique is given by considering the complement of the Borromean rings. The Borromean rings are spanned by a two-complex which cuts the complement into two ideal octahedra (figure 3.13). The two octahedra are glued as in figure 3.14.

To see Figure 3.13, the octahedron on the left should be imagined below the Borromean rings, the octahedron on the right above (so that they have opposite orientations). Each face of the upper octahedron is glued to the face directly below it with a $120°$ rotation; the sense of the rotation is alternately clockwise and counter-clockwise, alternating in direction like gears. It is possible to visualize this decomposition reasonably well – make sure to delete the vertices so the corners can open up.

To give this manifold a hyperbolic structure we can use two octahedra with $90°$ dihedral angles like the one in the previous example, since four faces are glued to each edge in the resulting complex.

[Tak85] showed that every three-manifold with non-empty boundary can be obtained by gluing together simplices, then removing regular neighborhoods of their vertices. Of course, not every way of gluing simplices is related to a geometric construction.

Will we discuss Ford domains and Delaunay triangulations later, so that there is a forward reference?

Example 3.4.8 (affine manifolds). If G is the group of affine transformations of $\mathbf{R}^n$, an $(G, \mathbf{R}^n)$-manifold is called an *affine manifold.*

% foliatetorus
% Manifolds obtained from the figure eight knot
% Flexibility and rigidity of geometric structures
Example "complex manifolds"
::complex manifold
::uniformization theorem
†Riemann mapping theorem

[/u/gt3m/3mbook/pictures/chap3/4/Borromgludiag.ps not found]

Borromgludiag

Figure 3.14. The gluing diagram for the complement of the Borromean rings

As an example, consider again the homothety of figure 3.1. As the quotient of $\mathbf{R}^2 \setminus \{0\}$ by a group of affine transformations (the group generated by this homothety), the torus has an affine structure (why?).

Here is another method, due to John Smillie, for constructing affine structures on T^2 from any quadrilateral Q in the Euclidean plane. Identify the opposite edges of Q by the orientation-preserving similarities that carry one to the other. Since similarities preserve angles, the sum of the angles about the vertex in the resulting complex is 2π, so a neighborhood of the vertex has an affine structure (why?). We will see in chapters **??** and **??** how such structures on T^2 are intimately connected with questions concerning Dehn surgery in three-manifolds. reference

Milnor [Mil58] showed that the only closed two-dimensional affine manifolds are tori and Klein bottles. An important question about affine manifolds is whether in general a compact affine manifold has Euler number zero.

Are all affine structures on tori obtained in this way?

reference; ask David Fried if question is still open

A Euclidean structure on a manifold automatically gives an affine structure. Bieberbach [Cha86] proved that closed Euclidean manifolds with the same fundamental group are equivalent as affine manifolds.

Example 3.4.9 (complex manifolds). When n is even, $\mathbf{R}^n$ can be identified with $\mathbf{C}^{n/2}$. Let Hol be the pseudogroup of local biholomorphic maps of $\mathbf{C}^{n/2}$, that is, holomorphic maps that have holomorphic local inverses. (It turns out that this is always the case for holomorphic local homeomorphisms.) A Hol-manifold is called a *complex manifold* of (complex) dimension $n/2$.

When $n = 2$, a map is holomorphic if and only if it is conformal and preserves orientation. Therefore an orientation-preserving isometry of the Poincaré disk model for $\mathbf{H}^2$ is biholomorphic, and every orientable hyperbolic surface inherits the structure of a complex manifold.

Stereographic projection from the unit sphere to $\mathbf{C}$ is a conformal map. A collection of maps obtained by rotating the sphere and then mapping by stereographic projection to $\mathbf{C}$ constitutes an atlas for a complex structure on S^2 (provided they don't all omit the same point); this is the complex structure of $\mathbf{CP}^1$, the Riemann sphere.

Similarly, orientation-preserving isometries of $\mathbf{E}^2$ are holomorphic. It follows that a complex structure is inherited from a hyperbolic, Euclidean or elliptic structure, so all orientable surfaces have the structure of complex manifolds.

This easy observation has a converse: every complex structure on a closed surface comes from a hyperbolic, Euclidean, or elliptic structure. The converse is a celebrated result known as the *uniformization theorem.* It is closely related to the †Riemann mapping theorem. The uniformization theorem was

Poincaré
Klein

the subject of much attention (and contention) by Poincaré, Klein, and others in the latter part of the nineteenth century.

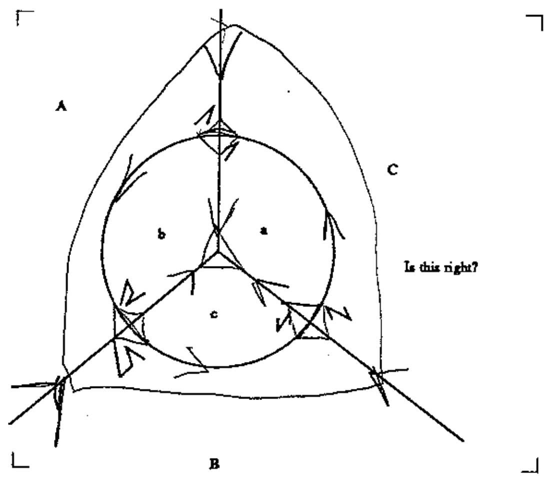
Bill: this seems unsatisfactory. Where did the contention come from? Silvio

Section "A hyperbolic manifold with geodesic boundary"
A HYPERBOLIC MANIFOLD WITH GEODESIC BOUNDARY
% geodgluing
% gluing is manifold iff $\chi = 0$
% wormholes

3.5. A hyperbolic manifold with geodesic boundary

Here is another manifold which is obtained from a gluing of two tetrahedra. First glue the two tetrahedra along one face; then glue the remaining faces according to figure 3.15 to obtain a simplicial complex K. Note that the orientations of the edges determine how A is to be glued to A', etc.

Task: put in at least crude drawings here for reference in text. I've changed stunted back to truncated. Does this occur elsewhere? The originals must be in a file drawer in Princeton.



geodgluing

Figure 3.15. geodesic boundary gluing diagram. Two tetrahedra can be glued together to form a polyhedron which looks like this, laid out in the plane. Glue lower case faces to upper case faces in this diagram.

This is a confusing picture. -yair

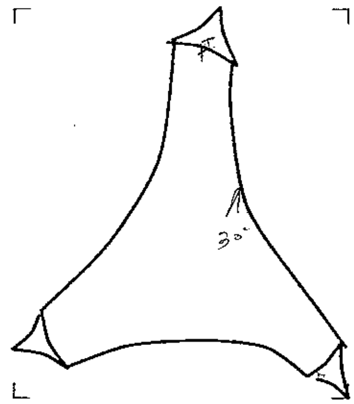
The complex K obtained from this gluing diagram has only one vertex, one edge, three faces, and two tetrahedra, so its Euler number is -1. Therefore, by Lemma 3.2.10, the link of the vertex has Euler number -2 so it is a surface of genus two. Let M be the manifold obtained by removing an open neighborhood of the vertex, so ∂M is a surface of genus two. We will construct a hyperbolic structure for M so that the boundary is geodesic.

To accomplish this, consider now a one-parameter family of regular tetrahedra in the projective model for hyperbolic space centered at the origin in Euclidean space, beginning with the tetrahedron whose vertices are on the sphere at infinity, and expanding until the edges are all tangent to the sphere at infinity. The dihedral angles vary from 60° to 0°, so somewhere in between, there is a tetrahedron with 30° dihedral angles. Truncate this simplex along each plane $v^\perp$, where v is a vertex (outside the unit ball), to obtain a *truncated tetrahedron.* (Note that the planes $v^\perp$ are disjoint, because the lines joining the vertices intersect $\mathbf{H}^3$.)

Two copies glued together give a hyperbolic structure for M with totally geodesic boundary, because $v^\perp$ is perpendicular to the faces containing v. A *closed hyperbolic three-manifold* can be obtained by doubling this example: i.e., taking two copies of M and gluing them together by the "identity" map on the boundary

The manifold M turns out to be a submanifold of S^3: it is the complement of the handlebody shown in figure 3.18. The manifold may be imagined as

% geodgluing
% HANDLB
% HANDLB
% HANDLB
% HANDLB
% HANDLB
% HANDLB
% HANDLB



30truncate

Figure 3.16. truncated 30-90 tetrahedron. A truncated hyperbolic tetrahedron with all edge angles 90° or 30°.

created from an apple by three worms, tunneling into the apple from three locations spaced around its surface. The worms start burrowing in, but then each worm turns toward its counterclockwise neighbor. However, the tunnels fail to connect, passing beneath the neighboring worm's tunnel. The three worms then turn up and toward the center, passing above their neighbor's tunnel, and finally they all unite in the core of the apple.

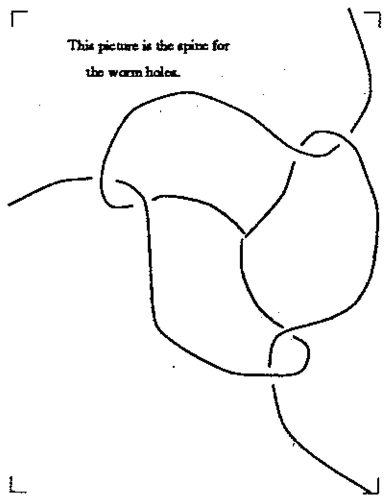
In order to see the homeomorphism, we begin with the diagram of figure 3.15. Cut out neighborhoods of the vertices to give the polyhedron which will be glued to yield the manifold shown in Figure 3.17(a). Also temporarily remove a regular neighborhood of each of the nine edges separating a pair of lettered faces (see Figure 3.17(b)). These nine wedges of cheese are glued to each other to form a copy of $D^2 \times I$, i.e., a two-handle. We will glue the two-handle on at the end. Now deform the picture by an isotopy, so the pairs of lettered faces are in position for easy gluing (as in Figure 3.17(c)).

Now glue the paired faces together, creating a genus-three handlebody with an annulus on its boundary where a two-handle is to be attached (as in Figure 3.17(d)). It is drawn as a ball with three worm holes. Deform the picture so the annulus becomes an equatorial belt and the worm holes are longer (Figure 3.17(e)) and simplify (Figure 3.17(f)). Attach the two-handle to obtain a manifold M with boundary contained in S^3, whose complement is a genus two handlebody made from two balls, joined by three one-handles. When the balls are appropriately arranged so that one is small, and one is centered at ∞, Figure 3.17(a) results.

HANDLB

HANDLB

Figure 3.17. HANDLB. A George Francis style figure



wormholes

Figure 3.18. ball with wormholes. If three wormholes are removed from a ball, intertwining and meeting at the center as illustrated, the result is homeomorphic to the the three-manifold obtained from the gluing diagram of two tetrahedra of figure 3.15, minus small neighborhoods about the vertices.

Section "The developing map"

3.6. The developing map

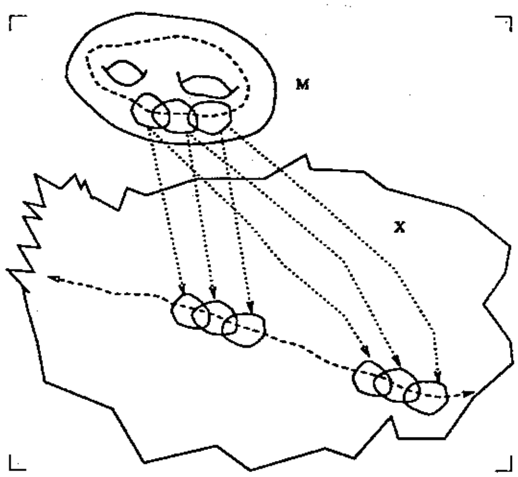
Several times, beginning in chapter 1, we have used the intuitive idea of 'unrolling' a geometric manifold. For instance, a Euclidean torus can be unrolled in the plane, to give a tiling pattern in the plane. In this section we will give a proper definition for 'unrolling' (developing), and generalize it to its natural context.

Let X be any (connected) real analytic manifold, and G a group of real analytic diffeomorphisms acting transitively on X. An element of G is then completely determined by its restriction to any open subset of X.

Suppose that M is any (G,X)-manifold. Let $U_1, U_2, \ldots$ be coordinate charts for M, with maps $\phi_i : U_i \mapsto X$ and transition functions γ_{ij} satisfying

$$\gamma_{ij} \circ \phi_i = \phi_j.$$

In general the γ_{ij}'s are local G-diffeomorphisms of X defined on $\phi_i(U_i \cap U_j)$ so they are determined by locally constant maps, also denoted γ_{ij}, of $U_i \cap U_j$ into G.



develop

Figure 3.19. multivalued development. The developing map is defined for any (G,X)-manifold M when G acts analytically. It can be thought of either as a multi-valued map from M to X or as a map from the universal cover of M to X. It is determined by analytic continuation, which is global.

Consider now an analytic continuation of ϕ_1 along a path α in M beginning in U_1. It is easy to see, inductively, that on a component of $\alpha \cap U_i$, the analytic continuation of ϕ_1 along α is of the form $\gamma \circ \phi_i$, where $\gamma \in G$. *Hence, ϕ_1 can be analytically continued along every path in M.* The result of analytic continuation on nearby paths with the same endpoints is identical. It follows that there is a global analytic continuation of ϕ_1 defined on the universal cover of M. (Use the definition of the universal cover as a space of homotopy classes of paths on M.) This map,

$$D : \widetilde{M} \to X$$

::developing map
Exercise "developing map is unique"
% developing map is unique
Proposition "existence of holonomy map"
% Euclidean manifolds
% affine manifolds
Proposition "(G, X)-manifold is quotient of X"

is called the *developing map* (the formal name for unrolling). D is a local (G, X)-diffeomorphism (i.e., it is an immersion inducing the (G, X)-structure on $\widetilde{M}$).

Exercise 3.6.1 (developing map is unique). The developing map D is unique up to post-composition (composition on the left) with elements of G.

Although G acts transitively on X in the cases of primary interest, this condition is not necessary for the definition of D. For example, if G is the trivial group and X is closed, then closed (G, X)-manifolds are precisely the finite-sheeted covers of X, and D is the covering projection.

From the uniqueness property (exercise 3.6.1) of D, we have in particular that

Proposition 3.6.2 (existence of holonomy map). *Fixing a developing map D, for any covering transformation T_α of $\widetilde{M}$ over M, there is some (unique) element $g_\alpha \in G$ such that*

$$D \circ T_\alpha = g_\alpha \circ D.$$

Since $D \circ T_\alpha \circ T_\beta = g_\alpha \circ D \circ T_\beta = g_\alpha \circ g_\beta \circ D$ it follows that the correspondence

$$H : \pi_1(M) \longrightarrow G$$

$$\alpha \mapsto g_\alpha$$

is a homomorphism, called the *holonomy* of M. The image of H is called the *holonomy group* of M. Because D is only determined up to left-composition with elements of G, the holonomy H is determined up to conjugacy in G.

Exercise 3.6.3. What are the developing maps and holonomies of

(a) a Euclidean torus, as in example 3.4.1,

(b) an affine torus, as in example 3.4.8.

In general, the holonomy of M need not determine the (G, X)-structure on M, but there is an important special case in which it does.

Definition 3.6.4. M is a *complete* (G, X)-manifold if $D : \widetilde{M} \to X$ is a covering map. In particular, if X is simply-connected, this means D is a homeomorphism.

It follows that

Proposition 3.6.5 ((G, X)-manifold is quotient of X). *If X is simply connected, then any complete (G, X)-manifold M may be reconstructed from the image $\Gamma = H(\pi_1(M))$ of the holonomy, as the quotient space $M \approx X/\Gamma$.*

% (G, X)-manifold is quotient of X
% {
Proposition "closed (G, X)-manifolds are complete"
::isotropy group
::stabilizer
Lemma "existence of invariant metric"

Remark: Since Γ acts on X from the left, it would be more correct to write the quotient as $\Gamma\backslash X$. However, for our purposes the distinction is so seldom important that for aesthetic reasons we will continue to write X/Γ.

Because of the significant relation between Γ and M expressed in proposition 3.6.5, it is often worthwhile replacing a non-simply-connected model space X by its universal cover $\widetilde{X}$. There is a covering group $\widetilde{G}$ acting on $\widetilde{X}$, defined to be the group of homeomorphisms of $\widetilde{X}$ which are lifts of elements of G. Then $\widetilde{G}$ can be described in the form of an extension

$$1 \to \pi_1(X) \to \widetilde{G} \to G \to 1.$$

For more about the structure of $\pi_1(X)$ and $\widetilde{G}$, see section 3.3.19analysis of stabilizers]. There is a one-to-one correspondence between (G, X)-structures and $(\widetilde{G}, \widetilde{X})$-structures, but the holonomy for a $(\widetilde{G}, \widetilde{X})$-structure contains more of the information.

Here is a useful sufficient condition for completeness.

Proposition 3.6.6 (closed (G, X)-manifolds are complete). *Let G be a Lie group acting analytically and transitively on a manifold X, such that for any $x \in X$, the isotropy group G_x of x is compact. Then every closed (G, X)-manifold M is complete.*

The *isotropy group* or *stabilizer*, G_x, is defined as $\{g \in G \mid gx = x\}$. Let $T_x(X)$ be the tangent space to X at x. There is an analytic homomorphism of G_x to the group of automorphisms of $T_x(X)$. The proof of the proposition utilizes the following fact, which is important in its own right:

Lemma 3.6.7 (existence of invariant metric). *If G acts transitively, X admits a G-invariant Riemannian metric if and only if the image of G_x in $GL(T_x(X))$ has compact closure. Any G-invariant Riemannian metric is analytic.*

Transitivity implies that the given condition at one point x is equivalent to the same condition everywhere.

Proof of 3.6.7: In one direction this is clear: If G preserves a metric then G_x maps to a subgroup of $O(n)$, which is compact. Now fix some x and assume that the image of G_x has compact closure H_x. Let Q be any positive definite form on the tangent space $T_x(X)$. Since H_x is compact there is a finite Haar measure on it, i.e. a finite measure invariant under the action of H_x on itself (see [MZ55].) Average the set of transforms g^*Q, $g \in H_x$, using this measure, to obtain a quadratic form on $T_x(X)$ which is invariant under H_x. Define a Riemannian metric $(ds^2)_y = g^*Q$ on X, where $g \in G$ is any element taking y to x. This definition is independent of the choice of g, and the resulting Riemannian metric is invariant under G. [3.6.7]

Such a metric then pieces together to give a Riemannian metric on any (G, X)-manifold, which is invariant under any (G, X)-map. We can now prove the proposition.

% closed (G,X)-manifolds are complete
% closed (G,X)-manifolds are complete
Exercise "analysis of stabilizers"

Proof of closed (G,X)-manifolds are complete: If M is any closed (G,X)-manifold, then there is some $\varepsilon > 0$ such that the ε-ball in the inherited Riemannian metric on M is always contractible and convex, i.e. any two points in it are joined by a unique geodesic inside the ball. (We use the compactness of M to guarantee that we can choose ε uniformly for all of M.) We may also choose ε so that all ε-balls in X are contractible and convex, since G is a transitive group of isometries. Then for any $y \in \widetilde{M}$, $B_\varepsilon(y)$ is mapped homeomorphically by D, for if $D(y) = D(y')$ for $y' \neq y$ in the ball the geodesic connecting y to y' is mapped to a self-intersecting geodesic, which can be seen to be a contradiction to the convexity of ε-balls in X. Then D is also an isometry between $B_\varepsilon(y)$ and $B_\varepsilon(D(y))$, by definition. If x is any point in X, consider the inverse image of a neighborhood, $D^{-1}(B_{\varepsilon/2}(x))$. For any y in the inverse image, the ε-ball around y maps isometrically and thus must properly contain a homeomorphic copy of $B_{\varepsilon/2}(x)$. The entire inverse image must then be a disjoint union of such homeomorphic copies. Therefore D evenly covers X, so it is a covering projection, and M is complete.

closed (G,X)-manifolds are complete

Corollary 3.6.8. *Every closed elliptic three-manifold M has a finite fundamental group. The universal covering space of M is S^3, and in particular, if M is simply-connected it is homeomorphic to S^3.*

Proof of 3.6.8: proposition 3.6.6 says the universal covering of M is S^3. Since S^3 is compact, $\pi_1(M)$ is finite. 3.6.8

A topological space X whose universal covering space is contractible is called an Eilenberg-MacLane space, and denoted $K(G,1)$ where G is $\pi_1(X)$ (and 1 means this X satisfies the condition that $\pi_k(X)$ is trivial for $k \neq 1$. There are analogous definitions for $K(G,n)$ for arbitrary n). A space which is a $K(G,1)$ is determined by G up to homotopy equivalence. From proposition 3.6.6 we see that closed hyperbolic manifolds and closed Euclidean manifolds are $K(G,1)$'s. This means that their fundamental groups are extremely important.

Exercise 3.6.9 (analysis of stabilizers). Let X be a manifold and G a group of homeomorphisms acting transitively on X. Let $x_0 \in X$ be any point.

This exercise needs work. More explicit about fundamental group of group abelian. More examples: G/Gamma has G action on other side, not abelian fundamental group. (negative example). What are some good positive examples, besides torus? nilgeometry mod Z? compact example?

(a) If the stabilizer G_{x_0} is path-connected, then $\pi_1(X)$ is abelian.
[*Hint:* If α and $\beta : S^1 \to G$ are loops in G based at 1, then the map $\alpha * \beta : S^1 \times S^1 \to G$ of the torus into G (where $*$ is multiplication in G) shows that α and β commute. Now consider the map from G to X given by $g \mapsto gx_0$].

(b) Let H_{x_0} be the group of components of G_{x_0} (in other words, G_{x_0} modulo its connected component of the identity). Construct a homomorphism $\pi_1(X) \to H_{x_0}$.

(c) Show that the kernel of this homorphism is central.

foobar

foobar

Figure 3.20. foobar. The developing map of an affine torus constructed from a quadrilateral (see example 3.4.8). The torus is plainly not complete.

Section "Discrete groups"
Definition "discrete group action"
::discrete
Definition "group action with discrete orbits"
::has discrete orbits
Definition "wandering group action"
::wanders
Definition "properly discontinuous group action"
::properly discontinuously
::proper
Exercise "definitions are succesively stronger"
% discrete group action
% properly discontinuous group action
::freely
Proposition "quotient by a wandering group is a manifold"
% nonhausdorff

3.7. Discrete groups

When G is a group of analytic diffeomorphisms of a simply-connected manifold X, then complete (G, X)-manifolds (up to isomorphism) are in one-to-one correspondence with certain subgroups of G (up to conjugacy by elements of G). There are certain traditional fallacies concerning the characterization of the groups which are holonomy groups for complete (G, X)-manifolds, so it is worth going through the definitions carefully.

Definition 3.7.1 (discrete group action). When G is a topological group, then a *discrete* subgroup $\Gamma \subset G$ is any subgroup on which the induced topology is the discrete topology. In other words, there is a neighborhood U of 1 such that $\Gamma \cap U = 1$.

Definition 3.7.2 (group action with discrete orbits). Let Γ be a group acting on a space X. Then Γ *has discrete orbits* if for every $x \in X$ there is a neighborhood U of x such that the set of $\gamma \in \Gamma$ for which $\gamma x \in U$ is finite.

Definition 3.7.3 (wandering group action). Let Γ be a group of homeomorphisms of X. Then Γ *wanders* if for every $x \in X$ there is a neighborhood U of x such that the set of $\gamma \in \Gamma$ for which $\gamma U \cap U \neq \phi$ is finite.

Definition 3.7.4 (properly discontinuous group action). Let Γ be a group of homeomorphisms of a locally compact space X. Then Γ acts *properly discontinuously* if for every compact subset K of X there is only a finite set of elements $\gamma \in \Gamma$ such that γK intersects K, i.e., for every compact K the map $\Gamma \times K \to X$ is a proper map. (A *proper* map f is a map such that the inverse image of any compact set is compact.)

Exercise 3.7.5 (definitions are succesively stronger). When the group of homeomorphisms of X is given the compact-open topology, then for subgroups Γ of this group the properties 3.7.1-3.7.4 are successively stronger.

Recall that a group Γ acts *freely* on a space X if no point of X is fixed by any non-trivial element of Γ.

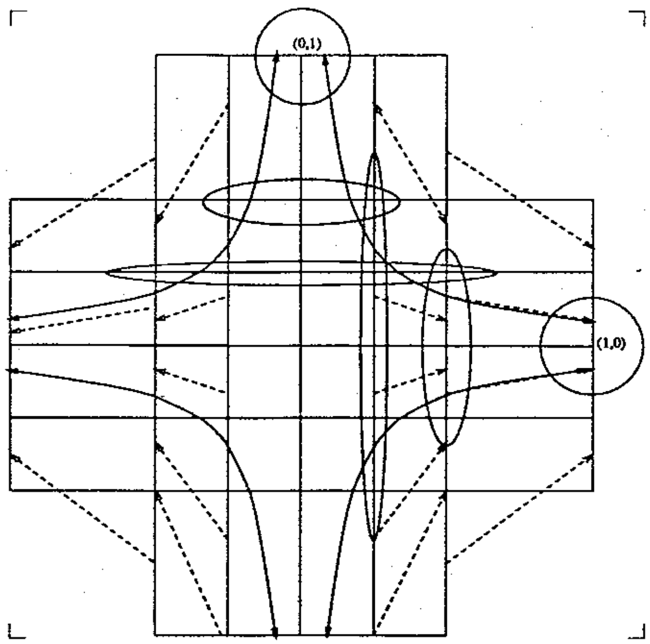
Proposition 3.7.6 (quotient by a wandering group is a manifold). *Let Γ be a group acting on a manifold X. If Γ wanders and acts freely, then the quotient space X/Γ is a manifold and the map $X \to X/\Gamma$ is a covering map.*

Proof of 3.7.6: If Γ wanders and the action is free, one easily finds a neighborhood U of any point x such that the translates γU are all disjoint. This shows that the image of x in the quotient topology has a neighborhood homeomorphic to U, and that U is evenly covered. 3.7.6

Warning: The quotient space X/Γ need not be Hausdorff. For example, consider the action of $\mathbf{Z}$ on $\mathbf{R}^2 - 0$ as a group of linear diffeomorphisms (see figure 3.21) generated by

$$(x, y) \mapsto (2x, y/2).$$

% quotient by a wandering group is a manifold
% nonhausdorff
% discrete group action
% properly discontinuous group action



nonhausdorff

Figure 3.21. Action with nonhausdorff quotient. The linear transformation $(x, y) \mapsto (2x, y/2)$, acting on the plane minus the origin, is a wandering action so its quotient space is a manifold: for each point, a sufficiently small neighborhood is disjoint from all its translates. However, the quotient manifold is not Hausdorff. In fact, every neighborhood of $(0, 1)$ in the quotient space intersects every neighborhood of $(1, 0)$.

Then each point in $\mathbf{R}^2 - 0$ has a neighborhood whose translates are disjoint, so by proposition 3.7.6 the quotient is a manifold M and the map $\mathbf{R}^2 - 0 \to M$ is a covering map. However, every neighborhood of the image of $(1, 0)$ in M intersects every neighborhood of the image of $(0, 1)$ in M, so M is not Hausdorff. Its fundamental group is $\mathbf{Z}$.

There is remarkable theorem of [Bro12b], the plane translation theorem, which asserts (among other things) that *every* free action of $\mathbf{Z}$ on $\mathbf{R}^2$ is wandering. There are, in fact, many ways that $\mathbf{Z}$ can act on $\mathbf{R}^2$. For instance, when the negative y-axis is removed from the example of figure 3.21, the remainder of the plane is homeomorphic to the plane. In particular, there are non-homeomorphic non-Hausdorff two-manifolds with fundamental group $\mathbf{Z}$, and universal covering space homeomorphic to $\mathbf{R}^2$.

Although we customarily suppose that manifolds are Hausdorff, without explicit mention of the fact, non-Hausdorff manifolds do sometimes arise naturally, and they can often be useful.

Exercise 3.7.7. Find examples of groups of homeomorphisms of manifolds to demonstrate that none of the properties 3.7.1-3.7.4 are equivalent.

[Hint: An action of $\mathbf{Z}$ with discrete orbits can be constructed on $(S^1 \times \mathbf{R} \times [-1, 1]) - \{\pi\} \times \mathbf{R} \times \{0\}$ by choosing a family of diffeomorphisms of the cylinders $S^{-1} \times \mathbf{R} \times \{t\}$ whose orbits spiral with slope t.]

% quotient by a wandering group is a manifold
% definitions are succesively stronger
Proposition "proper image of properly discontinuous group is properly discontinuous"
Exercise "discrete subgroups of Lie groups are properly discontinuous"
% proper image of properly discontinuous group is properly discontinuous
% discrete subgroups of Lie groups are properly discontinuous
% discrete group action
% properly discontinuous group action

Proposition 3.7.8. *Let Γ be a group acting on a (Hausdorff) manifold X. The quotient space X/Γ is a Hausdorff manifold with $X \to X/\Gamma$ a covering map if and only if Γ acts freely and properly discontinuously.*

Proof of 3.7.8: If Γ acts freely and properly discontinuously, then by proposition 3.7.6 and exercise 3.7.5 we know that X/Γ is a manifold and $p : X \to X/\Gamma$ is a covering map. Suppose x and y are points in X on distinct orbits. Let K be a union of two disjoint compact neighborhoods of x and y which contain no translates by Γ of x or y. Then $K - \bigcup_{\gamma\neq 1} \gamma K$ is still a neighborhood of x union a neighborhood of y, and these neighborhoods project to disjoint neighborhoods in X/Γ, so X/Γ is Hausdorff.

For the converse, suppose that X/Γ is Hausdorff. Consider any compact set $K \subset X$. Suppose there were an infinite sequence of equations $\gamma_i x_i = y_i$ with $x_i, y_i \in K$ and distinct γ_i. Then if (x, y) is any accumulation point of $\{(x_i, y_i)\}$, $p(x)$ and $p(y)$ do not have disjoint neighborhoods. These images $p(x)$ and $p(y)$ would have to be distinct, for otherwise, if $\gamma x = y$, there would be an infinite number of transformations $\gamma^{-1}\gamma_i$ of any neighborhood U of x which intersect it, and p would not be a covering map. Therefore X/Γ is not Hausdorff, a contradiction. [3.7.8]

Here is a criterion which is often convenient for checking proper discontinuity:

Proposition 3.7.9 (proper image of properly discontinuous group is properly discontinuous). *Suppose X and Y are spaces on which a group Γ acts and $f : X \to Y$ is a proper, surjective map which respects the group action. If the action on X is properly discontinuous, then the action on Y is also properly discontinuous.*

Proof of 3.7.9: If $K \subset Y$ is any compact set then $f^{-1}(K) \subset X$ is compact and $K \cap \gamma K \neq \phi \Leftrightarrow f^{-1}(K) \cap \gamma f^{-1}(K) \neq \phi$. [3.7.9]

Exercise 3.7.10 (discrete subgroups of Lie groups are properly discontinuous). A discrete subgroup Γ of a Lie group G acts properly discontinuously on G by left-translation. [Hint: Find a sufficiently small neighborhood V of 1 so that whenever gV intersects hV, and γgV intersects hV, then $\gamma = 1$. Cover any compact set K by a finite collection $\{h_i V\}$.]

Corollary 3.7.11. *Suppose G is a Lie group and X is a manifold on which G acts transitively with compact stabilizers G_x. Then any discrete subgroup of G acts properly discontinuously on X.*

Proof of 3.7.11: The map $G \to X = G/G_x$ is proper. Apply proposition 3.7.9 and exercise 3.7.10. [3.7.11]

Thus, in the cases of most interest to us, properties 3.7.1-3.7.4 are in fact equivalent.

When a group Γ acts properly discontinuously on a space X in such a way that the quotient space X/Γ is compact, then Γ is called a *cocompact* group. If the quotient space X/Γ has finite volume, then Γ is a *co-finite-volume* group. Often in the literature the terminology "lattice" is used for a co-finite-volume group, and "uniform lattice" for a cocompact group. We will avoid this terminology, since it seems to be confusing to the uninitiated.

What about the problem of writing actions on the left or on the right? This issue should be faced head on. It looks ugly to write it as is logical, on the left, so I'm tempted just to explain our choice somewhere, and leave the symbols as is. But does this get into difficulty where we really need the distinction?

We added a remark in 3.5. Is it enough? DY

3.8. The Teichmüller space of a surface

Section "The Teichmüller space of a surface"
% The totality of surfaces
::Teichmüller space
::Teichmüller modular group
::Dehn twists

We have seen, in section 1.3, that every closed surface M_g of genus $g > 1$ has a hyperbolic structure. Actually, there is a lot of freedom in the construction, with the consequence that surfaces of genus $g > 1$ have not just one hyperbolic structure, but many hyperbolic structures. Consider, for example, a surface of genus two, obtained from gluing opposite sides of an octagon. What shapes of octagon will do?

The length of each side of the octagon has to equal the length of the opposite side. In addition, the sum of the angles has to be 2π, or equivalently, the area must be 2π. A crude dimensional analysis suggests that there are many solutions to these conditions: an octagon is determined by 8 vertices in $\mathbf{H}^2$ giving 16 degrees of freedom, but the isometry group of $\mathbf{H}^2$ has dimension 3, leaving 13 degrees of freedom for the shape of an octagon. The constraint that the sides match up in pairs should leave 9 degrees of freedom, and the constraint that the angles sum to 2π should leave an 8-dimensional family of octagons which can be glued together to give a hyperbolic surface of genus 2. Under the gluing, all 8 vertices become identified and correspond to a point on the surface. This base point itself has 2 degrees of freedom. Consequently there appears to be a 6-dimensional family of distinct hyperbolic structures on a surface of genus 2.

The main issue, in defining the space of hyperbolic structures on a surface, is what equivalence relation to use. There are two good choices, giving two different spaces.

The set of all hyperbolic structures on an oriented surface forms a topological space called *Teichmüller space* $\mathcal{T}_g$. Alternately, one may think of the collection of all surfaces each with its own hyperbolic structure. Then $\mathcal{T}_g$ may be defined as the set of pairs (N, f) with $f : M \to N$ an orientation preserving diffeomorphism, with the equivalence $(N, f) \equiv (N_1, f_1)$ if and only if $f \circ f_1^{-1}$ is homotopic to an isometry (here (M, id) is taken as the origin). A Hausdorff topology is introduced by declaring that two points (N, f), (N_1, f_1) are close if $f \circ f_1^{-1}$ is homotopic to a diffeomorphism which nearly preserves the hyperbolic metric — that is, every non-zero tangent vector X on one surface is taken to a vector Y on the other such that the lengths of X and Y have a ratio near 1. That is, if the image of each infinitessimal circle, which is an ellipse, has the ratio of its major to minor axes uniformly close to 1.

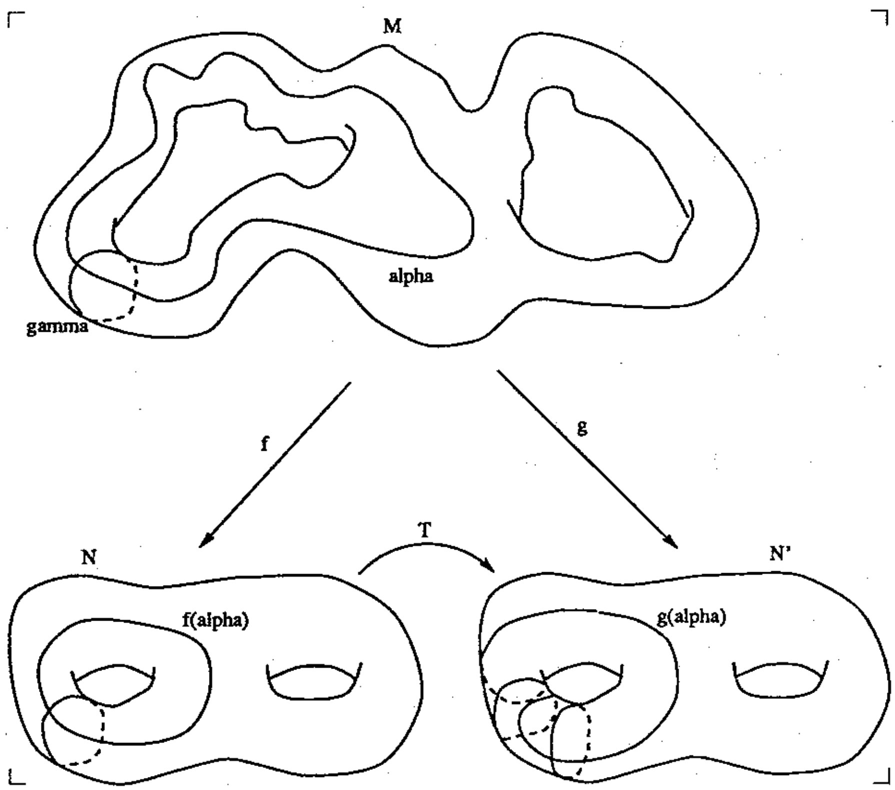
An orientation preserving automorphism $h : M \to M$ induces the automorphism $h^* : (N, f) \to (N, f \circ h)$ of $\mathcal{T}_g$. This action depends only on the homotopy class of h. The group of all such automorphisms of $\mathcal{T}_g$ acts discretely and is called the *Teichmüller modular group* Γ_g. Note that h^* fixes the point (N, f) if and only if $f \circ h \circ f^{-1}$ is homotopic to an isometry. Any surface with a symmetry lies in the fixed set of a finite subgroup $\neq \{id\}$ of Γ_g.

The modular group is generated by *Dehn twists*. A Dehn twist about a simple loop γ on an oriented surface M is defined as follows. Let A be an annular neighborhood of γ. The orientation of M induces an orientation of A.

::moduli space
::Riemann moduli space

Let τ denote the automorphism of A which fixes one boundary component and rotates the other by 2π in the positive direction. Extend τ to the whole surface by setting it equal to the identity outside A. Then τ, or any automorphism homotopic to τ, is called a Dehn twist about γ. A twist acting on $\mathcal{T}_g$ has no fixed points.

The quotient space, $\mathcal{T}_g/\Gamma_g$ (which is an orbifold) is called *moduli space* or *Riemann moduli space.* In other words, in Teichmüller space, we pay attention not just to what metric a surface is wearing, but also how it is worn. This is in contrast to moduli space where all surfaces wearing the same metric are equivalent. (The importance of the distinction will be clear to anybody who at some time has inadvertently slipped a sweater on backwards, or who after putting a pajama suit on an infant has found one leg to be twisted).



twistexample

Figure 3.22. Two isometric surfaces that define different points in Teichmüller space. The surfaces N and N' are isometric, but the map T for which $T \circ f = g$ is not homotopic to the isometry. It differs from the isometry by two twists around the curve γ. Note how the homotopy class α in R is realized in N and N' as geodesics of different lengths.

Note that in two dimensions, the notions of *homotopy* and *isotopy* can be used interchangeably ([Bae28, Eps66].) Two homeomorphisms (or two embeddings) are homotopic if there is a flow from one to the other; they are isotopic, if the flow can be taken through homeomorphims (resp., embeddings).

% twistexample
Exercise "holonomy defines structure"
% Teichmüller space of a surface
Proposition "embedding of geodesics determined by homotopy classes"

For another illustration of the difference between Teichmüller and moduli space, consider any element of the Teichmüller space of M and any curve α. There is a well-defined quantity $length(\alpha)$, defined to be the infimum of lengths of curves in the homotopy class of α. No such quantity is defined in the moduli space: there, we could only define what it means to say that *some* homotopy class of curves on the surface has length l since there is no *fixed* identification with homotopy classes of curves on M (see figure 3.22).

A surface with a complex structure (roughly, the assignment of a rule to measure angles) is necessarily orientable and is called a *Riemann surface.* As a consequence the uniformization theorem of complex analysis, a hyperbolic structure is associated with, for example, every closed Riemann surface of genus exceeding one, or every surface that results from removing at least 3 points from a Riemann surface. Conversely, it follows from the definition that every orientable surface with a hyperbolic structure has an associated complex structure. Orientation preserving isometries between hyperbolic structures correspond to conformal mappings between the associated complex structures. In this book we will work only with hyperbolic structures, which leads to a real analytic structure of $\mathcal{T}_g$.

There is an extensive literature on Teichmüller spaces beginning with Teichmüller, who introduced a metric. In the theory, points of Teichmüller space are often referred to as *marked Riemann surfaces*, because the points of $\mathcal{T}_g$ are specified by equivalence classes of pairs (*Riemann surface, marking*), where marking means designating a standard set of generators for the fundamental group.

Exercise 3.8.1 (holonomy defines structure). Show that two structures are equivalent in Teichmüller space if and only if their holonomy maps are conjugate by an element of $\mathrm{isom}(\mathbf{H}^2)$.

As we shall prove (theorem 3.8.8), Teichmüller space is homeomorphic to $\mathbf{R}^{6g-6}$. (The first proof of this is due to [Fricke, Klein 1889]; see [Bers, 1960] for a good modern reference.)

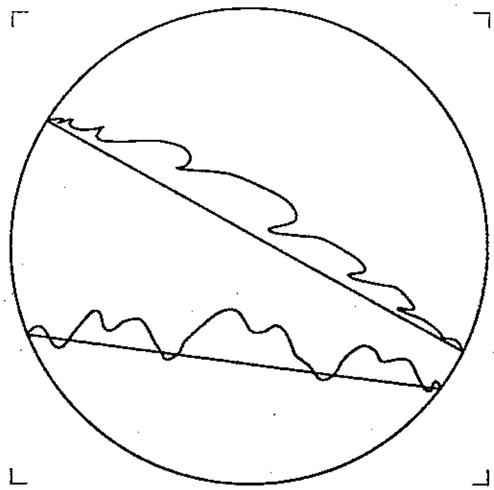
A *free homotopy class* of maps $f : X \to Y$ means a homotopy class without reference to basepoints.

Exercise 3.8.2. Free homotopy classes of curves $f : S^1 \to Y$ in a connected space Y are in one-to-one correspondence with conjugacy classes in $\pi_1(Y)$.

We will say that two curves $\alpha, \beta : S^1 \to X$ are homotopically distinct if α is not freely homotopic either to β or to $\beta \circ \gamma$, where $\gamma : S^1 \to S^1$ reverses orientation.

Proposition 3.8.3 (embedding of geodesics determined by homotopy classes). *Let M be a (complete) oriented hyperbolic surface and $\{\gamma'_i\}$ a collection of mutually disjoint simple loops which are homotopically distinct and non trivial. Then each γ'_i is freely homotopic to a unique geodesic γ_i and the collection $\{\gamma_i\}$ is likewise of mutually disjoint simple loops.*

Proof of 3.8.3: We will carry out the proof for the case that the geodesic γ is homotopic to a simple loop γ'; the general case of multiple loops follows by a similar argument. Consider the picture in the universal cover $\mathbf{H}^2$. Let $\tilde{\gamma}'$ be a preimage (lift) of γ'. There is a uniquely determined cyclic subgroup G of the group of cover transformations that preserves $\tilde{\gamma}'$. The simple arc $\tilde{\gamma}'$ has two distinct end points p, q on S^1_∞ which are the common fixed points of G. Any cover transformation not in G maps $\tilde{\gamma}'$ to a disjoint arc.



geodembed

Figure 3.23. geodesics are embedded. In an oriented hyperbolic surface, for any collection of embedded, homotopically distinct curves, the corresponding system of closed geodesics is also embedded. The condition is equivalent to the condition that in the universal covering space, the endpoints of paths which cover the curves do not separate each other.

The hyperbolic line $\tilde{\gamma}$ between p, q projects to a geodesic γ freely homotopic to γ'. (To see this, consider the intermediate projection to the open cylinder $\mathbf{H}^2/G$.) Because there is only one line between p, q, γ is unique in its free homotopy class.

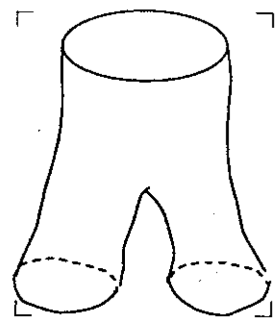
Two distinct hyperbolic lines in $\mathbf{H}^2$ are either disjoint or they cross once, separating the respective end points. Because no two lifts in the orbit of $\tilde{\gamma}$ separate their respective end points, the same is true for the orbit of $\tilde{\gamma}$; consequently γ must be a simple loop.

3.8.3

In a non-oriented two-manifold, any simple curve with a non-oriented regular neighborhood (i.e., with a Möbius band neighborhood) is represented by a simple geodesic. The boundary of its regular neighborhood is represented by a geodesic running twice around the original geodesic. This is the only situation where the non-oriented case is different.

Let's analyze an arbitrary hyperbolic structure on a closed oriented surface M_g of genus g. First, pick any maximal collection of non-trivial, homotopically distinct, disjoint simple curves Γ' on M_g. Such a collection of curves cuts M_g

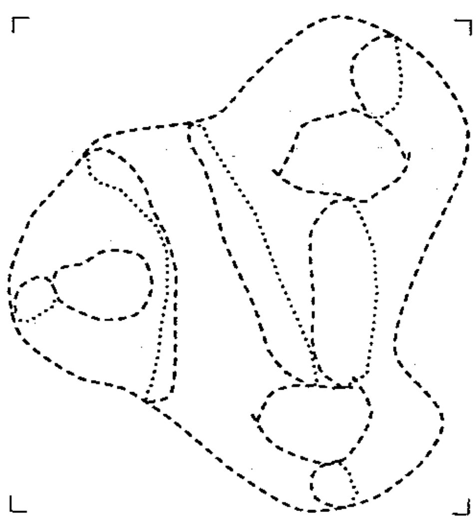
% pants
Proposition
"Teichmüller space
of the pair of
pants"
% pantsew



pants

Figure 3.24. A pair of pants. This surface, the sphere minus three disks, or a pair of pants, serves as a basic building block from which all closed hyperbolic surfaces can be built.

into subsurfaces with boundary which are each necessarily diffeomorphic to a pair of pants, P (figure 3.24), the only surface which cannot be cut further except by curves that bound disks or are homotopic to its boundary. There are $|\chi(M_g)| = 2g - 2$ copies of P obtained by cutting along Γ', so there are $\frac{3}{2}|\chi(M_g)| = 3g - 3$ components of Γ'.



pantsdecompose

Figure 3.25. A pants decomposition of a surface. Any oriented surface can be decomposed into pairs of pants sewen together along boundary components.

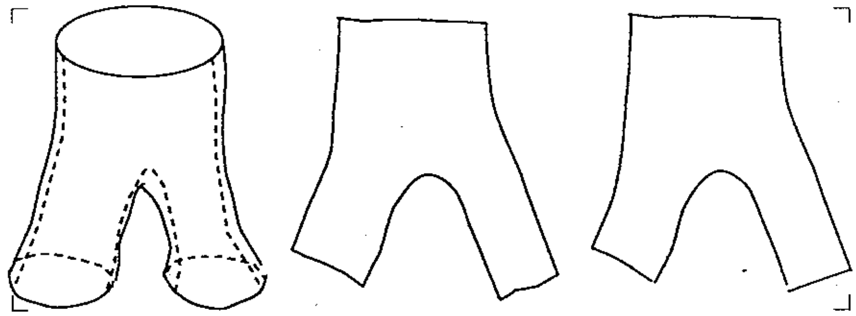
The collection Γ' is represented uniquely by a collection Γ of disjoint simple geodesics in a hyperbolic structure on M_g. Cut M_g along Γ to obtain $2g - 2$ hyperbolic structures with geodesic boundary on P.

Proposition 3.8.4 (Teichmüller space of the pair of pants). *Hyperbolic structures on P with geodesic boundary (up to isometries homotopic to the identity) are in one-to-one correspondence with a choice of three lengths for the three components of ∂P.*

should there also be a picture in the hyperbolic plane?am

Proof of 3.8.4: P can be cut along three seams S_1, S_2 and S_3 into two hexagons, H and H' as shown in figure 3.26. These seams are uniquely represented in

% geodesics on hyperbolic surfaces with geodesic boundary
% Some computations in hyperbolic space
% existence of right hexagons with any three lengths
Exercise "existence of right hexagons with any three lengths"
% buildhex



pantsew

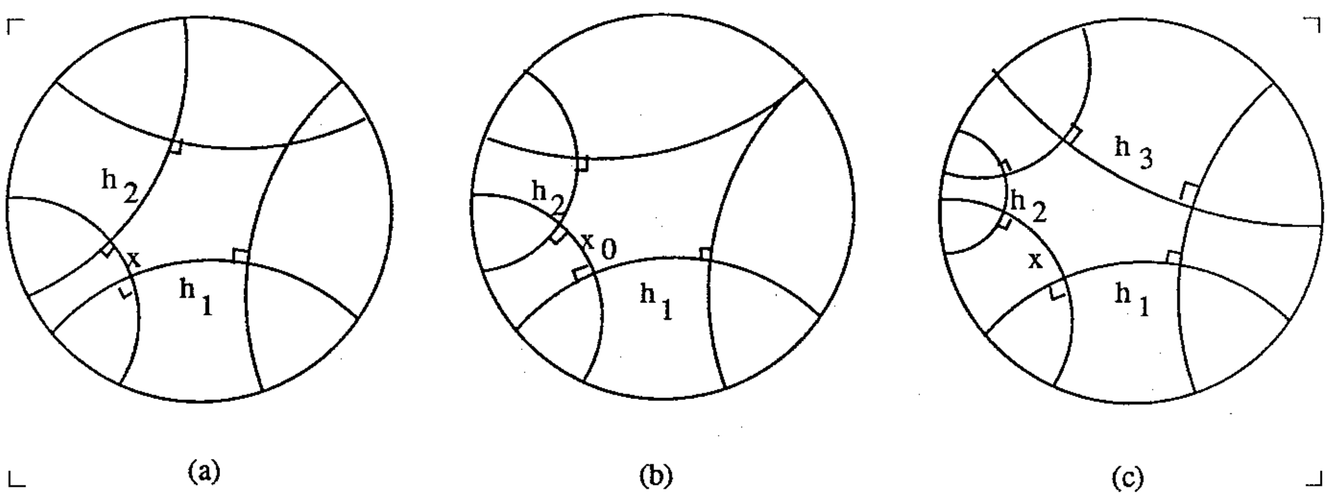
Figure 3.26. Pants sewn from two hexagons. A pair of pants can be formed by sewing together two right-angled hexagons along three of the sides.

the hyperbolic structure as geodesic arcs orthogonal to ∂P (cf. exercise 3.8.7). Cutting along these arcs, we obtain two right-angled hyperbolic hexagons H and H'. The hyperbolic structure on H or H' is uniquely determined by the lengths of S_1, S_2 and S_3, in view of the hexagon cosine law (see section 2.4). Therefore H is isometric to H', so the lengths of h_i and h_i' are equal, and equal to half the length l_i. The lengths l_1, l_2 and l_3 determine H and H', hence the hyperbolic structure on P. Any choice of three lengths corresponds to some hyperbolic structure, by exercise 3.8.5. 3.8.4

Exercise 3.8.5 (existence of right hexagons with any three lengths). Show that any three positive real numbers h_1, h_2 and h_3 occur as the lengths of alternate sides of an all-right hyperbolic hexagon.

[Hint: *Method 1:* Construct a quadratic form of type (2,1) , using basis $S_1^{\perp}, S_2^{\perp}, S_3^{\perp}$, by writing down all the inner products. Use the quadratic form model to obtain the desired hexagon.

Method 2: Given h_1 and h_2, consider the family of figures determined by x in figure 3.27. Show (geometrically) that picture (c) occurs for all $x > x_0$, and that h_3 is a strictly monotone function of $x > x_0$].



buildhex

Figure 3.27. Constructing an all-right hexagon

Exercise "covering pants"
% Teichmüller space of the pair of pants
Exercise "geodesics on hyperbolic surfaces with geodesic boundary"
% existence of closed geodesics on a hyperbolic manifold
% embedding of geodesics determined by homotopy class
Theorem "Teichmüller space of a surface"

Exercise 3.8.6 (covering pants). From two right angled hexagons as in the proof of proposition 3.8.4 show how to find a set of generators of a fuchsian group G which is the universal covering group of the corresponding P. The group G is uniquely determined by P up to conjugation within $PSL(2, \mathbf{C})$.

Exercise 3.8.7 (geodesics on hyperbolic surfaces with geodesic boundary). Prove the analogs of ?? and ?? for a hyperbolic manifold M with geodesic boundary, where free homotopy classes of arcs from boundary to boundary are allowed, as well as closed curves. [Hint: Consider the double of M, which is the union of two copies of M glued along the boundary by the identity map. Represent the double of an arc by a closed curve, and show it must be orthogonal to ∂M, because it is invariant by the symmetry exchanging the two halves of the double].

Theorem 3.8.8 (Teichmüller space of a surface). *Teichmüller space $\mathcal{T}(M_g)$ is homeomorphic to $\mathbf{R}^{6g-6}$.*

Proof of 3.8.8: For each component of Γ', we have a free parameter, its length, in $\mathbf{R}_+ \approx \mathbf{R}$. A choice of all these lengths determines $2g-2$ hyperbolic structures on P which can be glued together to form a hyperbolic structure on M_g. This gluing can be done in more than one way: for each component of Γ' we can vary the gluing map by twisting one side with respect to the other side by an arbitrary signed hyperbolic distance.

To describe the gluing process, we will examine below the particular case of two pants P, P_1 to be glued along a common geodesic boundary loop σ. The case of gluing two edges on the same pants is quite similar. Successively applying one technique or the other to the growing surface element, the collection of pants is assembled into the full surface. It must be kept in mind that in joining adjacent surface elements (or two edges on the same element) it is not enough simply to specify an identification between points on the opposite sides of the edge. We must indicate how the fundamental groups of the elements, based at origins O, O_1, say, are to be fused to form the fundamental group of the joined element. This can be done by introducing an arc crossing the common edge σ between O, O_1 together with a multiple of 2π to indicate how many times the arc should circle σ; two choices differ by a multiple of the Dehn twist about σ. We will use a gluing parameter x, where $-\infty < x < \infty$. The pure gluing depends on the quantity $X = e^{2\pi i x/L}$, where L is the length of σ.

We will work in $\mathbf{H}^2$ in the upper half plane model. Orient the common boundary geodesic σ so that P lies to its left. We may assume that the positive imaginary axis σ^* in $\mathbf{H}^2$ lies over σ and that its left side lies over P, its right lies over P_1. Denote by G the fuchsian covering group of P that acts to the left of σ^* and G_1 that over P_1 that acts to the right. Fix a point O^* on σ^* as origin. Let x denote distance along σ^* measured from O^* with positive or negative sign depending on direction.

Consider the Möbius transformation

Bill: are you happy with this explicit formulation? am

$$T_x : z \rightarrow (e^x)z.$$

Each point z on σ^* is moved signed distance x along σ^*. Now given an x, consider the fuchsian group

$$G_x = < G, T_x G_1 T_x^{-1} >$$

which is generated by G and $T_x G_1 T_x^{-1}$. This is a universal covering group for the four-holed sphere P_2 resulting from gluing P and P_1 along σ with twist parameter x. Two values of x which differ by an integral multiple of L give rise to the same group G_x. However no conjugacy will induce the automorphism of G_x that comes from the automorphism of $\pi_1(P_2)$ induced by a non-zero power of a Dehn twist about σ.

figure of developing image of adjacent pairs of pants is left out. Instead, make figure of joining the fuchsian groups???am

This parametrization gives a continuous map from Teichmüller space to $\mathbf{R}^{6g-6} \approx \mathbf{R}_+^{3g-3} \times \mathbf{R}^{3g-3}$ where the first $3g-3$ coordinates are the lengths of geodesics in Γ' and the second $3g-3$ coordinates are their associated twist parameters. Every set of parameters determines a hyperbolic surface up to homotopy, so this map is a homeomorphism. [3.8.8]

Exercise 3.8.9. (a) Let M be a closed hyperbolic manifold. The space of maps of S^1 into M representing a given free homotopy class retracts into the space of uniformly parametrized geodesics in that homotopy class. (Hint: consider the appropriate covering space.)

(b) If there is a homotopy $h_t : M \to M$ such that $h_0 = \mathrm{id}$ and h_1 is an isometry between two hyperbolic structures on M as constructed above, then there is a homotopy h' with $h'_0 = h_0$, $h'_1 = h_1$ and h'_t carrying each curve of Γ' to a reparametrization of itself.

(c) Therefore, the difference of the twist parameters on the two sides of a curve of Γ' is an invariant of the element of Teichmüller space.

Section "Completeness of non-compact manifolds"
COMPLETENESS
% Hyperbolic surfaces obtained from ideal triangles
% The developing map
Section "COMP CRITS"
% The developing map

3.9. Completeness of non-compact manifolds

Even in the nicest cases, it can be subtle to determine whether or not non-compact (G, X)-manifold is complete. For example, consider the thrice-punctured sphere, which is obtained by gluing together two triangles minus vertices in this pattern:

[/u/gt3m/3mbook/pictures/chap3/9/whatisthis.ps not found]

whatisthis

Figure 3.28. 1.5

A hyperbolic structure can be obtained by gluing together two ideal triangles (with all vertices on the circle at infinity).

Each side of an ideal triangle is isometric to $\mathbf{R}$, so there is freedom in the choice of the gluing maps: the identification between paired sides may be modified by an arbitrary translation. Therefore, we have not just one, but a family of hyperbolic structures on the thrice-punctured sphere, parametrized by $\mathbf{R}^3$. (They need not be all distinct, and in fact they are not). We will see in section 3.10 that exactly one of these is complete!

To prepare for understanding such structures, we will collect some useful conditions for completeness of a (G, X)-structure, where G is a group of real-analytic diffeomorphisms acting on a manifold X, as in section 3.6. For convenience, we fix some natural metric on X, thereby imparting a metric to any (G, X)-manifold.

This proposition helps justify the use of the term "complete."

Propositio 3.9.1. *Let G be a transitive group of real analytic diffeomorphisms of X with compact stabilizers G_x. Let M be a (G, X)-manifold. Then the following conditions are equivalent:*

(a) M is a complete (G, X)-manifold. (See section 3.6.)

(b) For some $\varepsilon > 0$, every closed ε-ball in M is compact.

(c) For every $a > 0$, every closed a-ball in M is compact.

(d) There is some family $\{S_t\}_{t\in\mathbf{R}_+}$ of compact sets which exhaust M such that S_{t+a} contains the neighborhood of radius a about S_t.

(e) M is complete as a metric space.

Proof of 3.9.1: We will check that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b). Let $x_0 \in \widetilde{X}$ be any point. There is some $\varepsilon > 0$ such that the closed ε-ball about x_0 is compact. This ε works for all $x \in \widetilde{X}$, since $\widetilde{G}$ acts transitively on $\widetilde{X}$. Any ε-ball in M is the image of an ε-ball in $\widetilde{X}$, so it is also compact.

(b) $\Rightarrow$ (c). Suppose that every ball of radius a is compact for some $a \geq \varepsilon$. Then such a ball is covered by a finite number of $\varepsilon/2$-balls. A ball of radius $a + \varepsilon/2$ is therefore covered by a finite number of ε-balls, so it is compact. By induction closed balls of every radius are compact.

(c) $\Rightarrow$ (d). Let S_t be the ball of radius t about x_0.

Section "3.9.3" % FIG-8

(d) $\Rightarrow$ (e). If $\{x_i\}$ is any Cauchy sequence, then $\{x_i\} \subset S_T$ for some sufficiently high T, hence it converges.

(e) $\Rightarrow$ (a). Suppose M is metrically complete. We will show that the developing map $D : \widetilde{M} \to X$ is a covering map by proving that any path α_t in X can be lifted to $\widetilde{M}$ (since local homeomorphisms with the path-lifting property are covering maps). First we need to see that $\widetilde{M}$ is metrically complete. In fact, the projection to M of any Cauchy sequence in $\widetilde{M}$ has a limit point $x \in M$. Since x has a compact neighborhood which is evenly covered in $\widetilde{M}$, and whose components are definitely separated in the metric of $\widetilde{M}$, the Cauchy sequence converges also in $\widetilde{M}$.

Consider now any path α_t in X. If it has a lifting $\widetilde{\alpha}_t$ for t in a closed neighborhood $[0, t_0]$, then it has a lifting for $t \in [0, t_0 + \varepsilon)$ (for some $\varepsilon > 0$). On the other hand, if it has a lifting for t in a half-open interval $[0, t_0)$, the lifting extends to $[0, t_0]$ by the completeness of $\widetilde{M}$. Hence, M is complete as a (G, X)-manifold. **3.9.1**

Note: For hyperbolic manifolds obtained by gluing polyhedra, condition (d) seems to be the most useful for showing a structure is complete, and condition (e) for showing a structure is not complete.

Propositio 3.9.2. *The hyperbolic structures for the figure-eight knot, Whitehead link and Borromean rings complements (Section ??) are complete.*

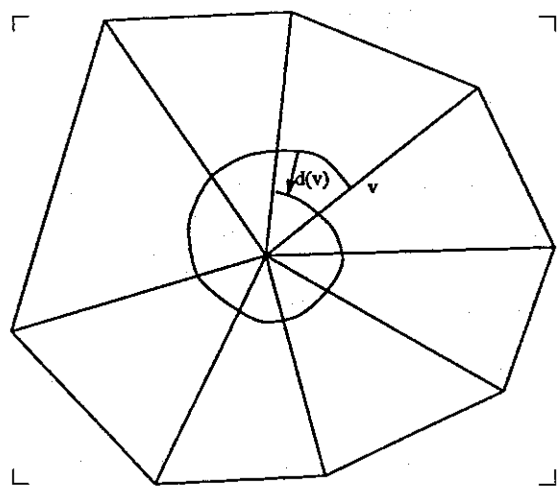
Proof of 3.9.2: In any of the three cases, let S_0 be the union of centers of the polyhedra to be glued. The balls of radius t about S_0, intersected with the polyhedra, are matched nicely by the gluing maps, so they yield sets S_t meeting criterion (d) . **3.9.2**

Section "Hyperbolic surfaces obtained from ideal triangles"
GLUINGTRIANGS
Section "DV=0 IS COMPL"
% COMP CRITS

3.10. Hyperbolic surfaces obtained from ideal triangles

Consider an oriented surface S obtained by gluing ideal triangles in some pattern.

Let K be the complex obtained by including the ideal vertices. Associated with each ideal vertex v of K, there is an invariant $d(v)$, defined as follows. Let h be a horocycle in one of the ideal triangles, centered about a vertex which is glued to v and "near" this vertex. Extend h as a horocycle in S counterclockwise about v. It meets each successive ideal triangle as a horocycle orthogonal to two of the sides, until finally it re-enters the original triangle as a horocycle h' concentric with h, at a distance $\pm d(v)$ from h. The sign is chosen to be positive if and only if the horoball bounded by h in the ideal triangle contains that bounded by h'. (Why doesn't $d(v)$ depend on the initial choice of horocycle h?)



extendhoro

Figure 3.29. Extending a horocycle. In a hyperbolic surface obtained by gluing ideal triangles together, the invariant $d(v)$ is defined for each of its ideal vertices by extending a horocycle around v and measuring how far its position has changed upon return to the starting line.

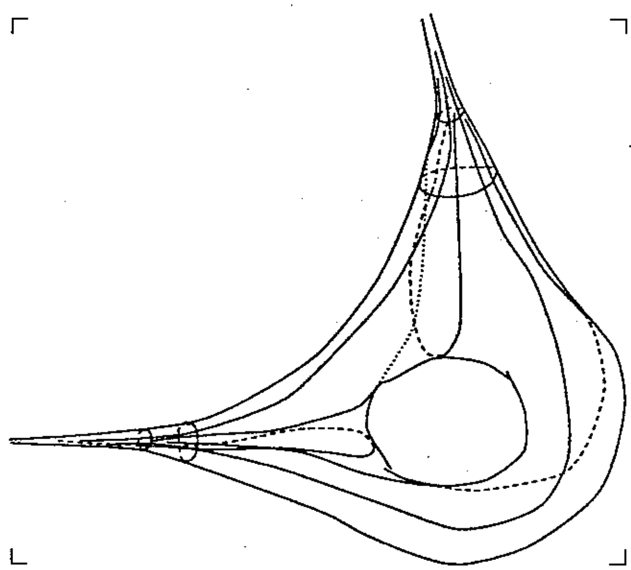
Propositio 3.10.1. *The surface S is complete if and only if $d(v) = 0$ for all vertices v.*

Proof of 3.10.1: Suppose, for instance, that some invariant $d(v) > 0$. Continuing h further around v; the length of each successive circuit around v is reduced by a constant factor < 1, so the total length of h after an infinite number of circuits is bounded. A sequence of points evenly spaced along h is a non-convergent Cauchy sequence, so S flunks condition 3.9.1(e).

If all invariants $d(v) = 0$, on the other hand, one can remove neighborhoods of each vertex in K defined by fitting together pieces of horoballs to obtain a compact subsurface S_0. Let S_t be the surface obtained by removing smaller horoball neighborhoods bounded by horocycles a distance of t from the original

% COMP CRITS
% incompletesurf

ones. The surfaces S_t satisfy the hypotheses of 3.9.1(d), hence S is complete. $\boxed{3.10.1}$



completesurf

Figure 3.30. A complete hyperbolic surface. When all the invariants $d(v)$ are zero, then the ideal triangles glue together to form a complete hyperbolic surface, since the horocycles match up around each ideal vertex. Note that each end of S looks like a pseudosphere.

For any hyperbolic manifold M, let $\overline{M}$ denote the metric completion.

Propositio 3.10.2. *Let S be a hyperbolic surface obtained by gluing together ideal hyperbolic triangles, (leaving no free edges). Then $\overline{S}$ is a hyperbolic surface with geodesic boundary. It has one boundary component of length $|d(v)|$ for each invariant $d \neq 0$.*

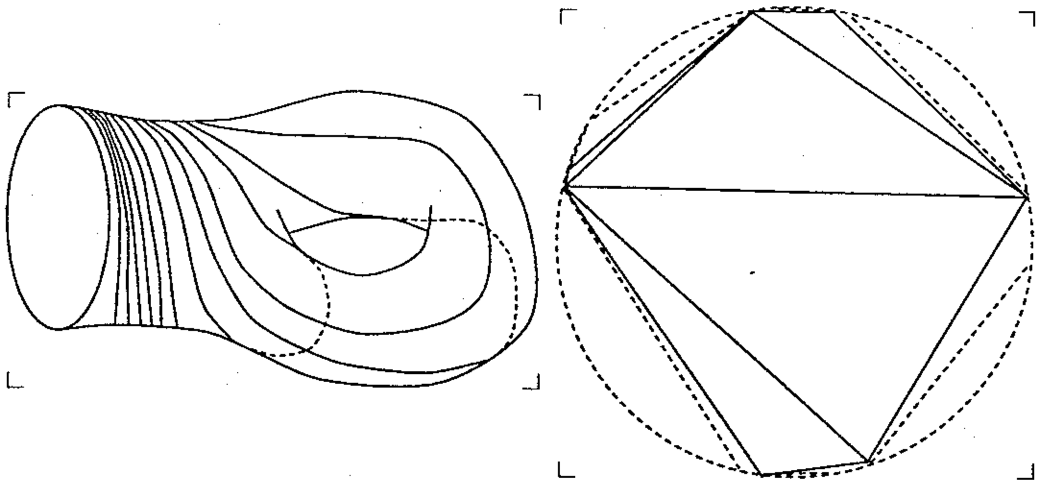
Proof of 3.10.2: Each horocycle, extended as above, which "spirals out" toward a missing vertex v of S has an endpoint in the metric completion. Distinct horocycles are determined by a parameter t representing the position of intersection with some edge heading toward v, well-defined modulo $d(v)$. Distinct horocycles are uniformly spaced, so they have distinct endpoints. After adjoining these endpoints, one readily verifies that every Cauchy sequence converges. Hence $\overline{S}$ is a surface with boundary.

Each boundary component of $\overline{S}$ arises as a limit of geodesics which are orthogonal to the horocycles around the corresponding missing vertex, and thus ∂S is an union of geodesics. $\boxed{3.10.2}$

Figure 3.31 is a sketch of $\overline{S}$ and of the developing map of S, in the projective disk model.

In the case that S is a thrice punctured sphere, note that the conditions $d(v) = 0$ for the three ideal vertices give 3 linearly independent linear conditions in 3 variables, so they uniquely determine the three gluing maps of edges

Section "Teichmüller spaces by gluing"
% The Teichmüller space of a surface



incompletesurf

Figure 3.31. The development of an incomplete surface. When some of the invariants $d(v)$ are not zero, then the surface obtained by gluing ideal triangles is the interior of a complete hyperbolic surface with geodesic boundary. The boundary components have length $|d(v)|$. In the developing map the edges of the triangles accumulate on the boundaries, and on the completed surface they spiral around each boundary component.

of two ideal triangles. This reflects the fact that the Teichmüller space of a thrice-punctured sphere is a single point.

Exercis 3.10.3. (a) Consider a complete hyperbolic surface with at least one cusp. Show that it can be decomposed into ideal triangles by cutting in any combinatorial pattern which works topologically.

(b) Show that for any complete hyperbolic surface M^2 with geodesic boundary, the interior of M^2 can be decomposed into ideal triangles in any combinatorial pattern which works topologically and with any direction of spiraling specified around each boundary component of M^2. (We assume here, of course, that M^2 is not closed.)

(c) Describe the space of hyperbolic structures with geodesic boundary on M^2, (as above) up to isometries homotopic to the identity. Check the answer against the method of Section 3.8.

Section "Hyperbolic manifolds obtained by gluing ideal polyhedra"
GLU ID POLYHED
Section "SIM-COMPLET"

% SIMCOMPLET
% DV=0 IS COMPL
% DV=0 IS COMPL
Section "COMP IFF EUC"

3.11. Hyperbolic manifolds obtained by gluing ideal polyhedra

Consider now the more general case of a hyperbolic manifold M obtained by gluing pairs of faces of a finite collection of polyhedra, with some vertices at infinity and leaving no free faces. Let $\widehat{M}$ be the simplicial complex obtained by including the ideal vertices.

Let $\mathbf{Sim}^n$ denote the group of similarities of Euclidean n-space $\mathbf{E}^n$. An element of $\mathbf{Sim}^n$ can be composed from a rotation ($\in O(n)$), an expansion or contraction by a scalar factor, and a translation. A $(\mathbf{Sim}, \mathbf{E}^n)$-structure is called a *similarity* structure.

Propositio 3.11.1. *The link $l(v)$ of an ideal vertex v of $\widehat{M}$ has a canonical similarity structure. M is complete if and only if for each ideal vertex v, the similarity structure on $l(v)$ is complete.*

Proof of 3.11.1: To obtain the similarity structure on $l(v)$, take a horospherical cross-section of the set of rays heading toward an ideal vertex of one of the polyhedra. The horosphere is modeled on Euclidean space, and projection from one horosphere to a concentric one, along their common "radii", is a Euclidean similarity. These similarity structures piece together to give a similarity structure on $l(v)$.

We will show that the two conditions of Proposition 3.11.1 are equivalent, by relating them to the way horospheres fit together around the ideal vertex. Begin with one horospherical section, far out near an ideal vertex of one of the polyhedra, and begin extending it, around the link of v in various directions. Whenever this extension process comes back to the original face of the polyhedron, if it does not match up exactly, then M is not complete, since one may construct a non-convergent Cauchy sequence as in Proposition 3.10.1. If the horospheres always match up, then M is complete, again by the same reasoning as in Proposition 3.10.1.

When the horospheres match up, a horospherical cross-section of $l(v)$ refines the similarity structure to a Euclidean structure. Since Euclidean structures on compact manifolds are complete, the similarity structure is complete.

It remains to show that the similarity structure is not complete when the horospheres fail to match up. We need the following lemma.

Lemm 3.11.2. *A similarity structure on a closed manifold N is complete*

$\Leftrightarrow$ *it comes from a Euclidean structure*

$\Leftrightarrow$ *its holonomy group consists of Euclidean isometries.*

Proof of 3.11.2: The universal cover $\widetilde{N}$ always has a Euclidean structure (induced by the developing map to $\mathbf{E}^n$). If the holonomy group acts isometrically, this induces a Euclidean structure on N. This shows that the last two conditions are equivalent.

If the holonomy is not Euclidean, there is some loop with contracting holonomy. The lift of this loop in $\widetilde{N}$ develops in $\mathbf{E}^n$ in a geometric progression,

% SIMCOMPLET
% COMP IFF EUC
Section "similarity torus"
% The totality of surfaces

converging at one end to a point in $\mathbf{E}^n$. This point has no neighborhood which is evenly covered, so N is not complete. [3.11.2]

Continuation of proof of 3.11.1: When some horospherical cross-section does not match up, there is an element of the similarity structure of the link of some ideal vertex which contracts. Thus, applying lemma 3.11.2, the similarity structure is not complete. [3.11.1]

Even when M is not complete, the fact that there is a similarity structure on the link imposes topological constraints on it. In particular, when $n = 3$ the links must all be tori, as follows from this exercise:

Exercis 3.11.3. Show that no closed oriented surface other than the torus can be given a similarity structure. [Hint: compare with the proof of the similar statement for Euclidean structures in 1.3]

The completion $\overline{M}$ gives a different picture in each of the cases $n = 2, n = 3$ or $n > 3$. We shall be mainly concerned with the case $n = 3$, which is the most interesting. We will study it in detail in Chapter 7, showing by example that we can deduce non-trivial topological information by studying the completions.

Warning: The similarity structure on $l(v)$ for a hyperbolic manifold M obtained by gluing ideal polyhedra definitely depends not just on M but on the decomposition of M into ideal polyhedra, when M is not complete.

Section "The eight model geometries"
8 GEOMS

3.12. The eight model geometries

What is a geometry? Up till now, we have discussed three kinds of three-dimensional geometry: hyperbolic, Euclidean and spherical. These three geometries have in common the property of being as homogeneous as possible: not only do they have self-isometries which move any point to any other point, but these three spaces also have isometries which can take any orthonormal frame in the tangent space at any point to any other orthonormal frame at any other point. If we remove this latter condition, there are many more possibilities: there are 3-dimensional spaces which are still homogeneous, but they are (so to speak) warped, so that there are certain directions in the space which are geometrically distinguished from other directions.

An enumeration of additional 3-dimensional geometries depends on what spaces we wish to consider and what structures we use to define and to distinguish the spaces. For instance, do we think of a geometry as a space equipped with such notions as lines and planes, do we think of a geometry as a space equipped with a notion of congruence , or do we think of a geometry as a space equipped with either a metric or a Riemannian metric? There are deficiencies in any of these approaches.

The problem with using structures such as lines and planes is that they are not general enough. The five new geometries which we will describe here do not have any good notion of a plane – there are no flat (totally geodesic) surfaces in these geometries passing through certain tangent planes. Besides, even in Euclidean geometry, information about geometric shapes is not determined by incidence properties of lines and planes.

The problem with using the notion of congruence to define the structure of a geometry is that the number of different geometries proliferates more than we would really like. For instance, we could consider Euclidean space with the group of translations as the group of congruences, or Euclidean space with horizontal translations together with vertical screw motions (where the amount of rotation is proportional to the vertical motion), etc., and get a number of different geometries. Sometimes it is interesting to distinguish these different structures, but for the broad picture these variations should all be considered under the category of Euclidean geometry.

Finally, the problem with using a notion of distance to determine the geometry is, first, that by simply rescaling something like $\mathbf{H}^3$ or S^3, we get different metric spaces. Moreover, even if we consider metric spaces which are isometric up to a constant scaling factor to be equivalent, many of the spaces have a whole family of homogeneous metrics (sometimes with varying degrees of homogeneity) which are not even equivalent up to scaling. The three-sphere, for instance, has an interesting family of homogeneous metrics obtained from the usual metric by picking a family of Hopf circles and contracting or expanding the lengths of these circles, while keeping the metric constant in the orthogonal directions.

::model geometry
% existence of
invariant metric

The best way to think of a geometry, really, is to keep in mind these different points of view all at the same time. If we allow changes of the group of congruences which do not change the metric and changes of the metric which do not change the group of congruences as inessential changes in the geometry, and if we also group geometries together when the sets of compact manifolds modeled on them are identical, we end up with a reasonable enumeration of geometries.

For logical purposes, we must pick only one definition. We choose to represent a geometry as a space equipped with a group of congruences, that is, a (G, X) space.

Definitio 3.12.1. A *model geometry* (G,X) is a manifold X together with a Lie group G of diffeomorphisms of X, subject to the following conditions.

a) The space X is connected and simply-connected.

b) G acts transitively on X, and the stabilizers G_x of points $x \in X$ are compact.

c) G is not contained in any larger group of diffeomorphisms of X with compact stabilizers of points.

d) There exists at least one compact manifold modeled on (G,X).

Remarks. Condition a) selects one representative from each class of locally equivalent geometries with different fundamental groups, so that they serve as models for identical classes of 3-manifolds. Condition b) means that the space possesses a homogeneous Riemannian metric invariant by G (see lemma 3.6.7). Condition c) says that no Riemannian metric which is invariant by G is also invariant by any larger group. In particular, it selects at most one geometry for each isometry class of metric spaces. Another reason for condition c) is that by enlarging the structure group G, we do not decrease the set of manifolds with that structure. Condition d) is not phrased in an intrinsic way, but it is useful because it eliminates a whole continuous family of 3-dimensional geometries which do not serve as models for any compact manifolds – this family consists of 3-dimensional solvable Lie groups (or extensions of them by small compact groups) which act on themselves by left multiplication. These geometries are not unimodular, i.e., they have geometrically invariant vector fields which have negative divergence (decrease volume). Clearly, such a geometry cannot be a model space for anything compact or even of finite volume. We chose the phrasing of d), rather than the condition that the geometry be unimodular, because in higher dimensions unimodularity is not sufficient to guarantee that there are compact manifolds modeled on a geometry.

Now we will give a brief description of the other five 3-dimensional model geometries, along with simple examples of manifolds modeled on each. We will not prove that this list of eight model geometries is correct, even though it is possible to develop the list in an elementary way in terms of the local geometric pictures which are consistent with the possible choices of stabilizers of points G_x (for a discussion, see [Sco83]).

Section "classification of $S^2 \times \mathbf{E}^1$-manifolds"

Before actually enumerating the new geometries, however, we should point out that all five are obtained, in one way or another, by combining one of the two-dimensional geometries with the geometry of the line. Put another way, each of the five has a canonical fibration either over a 2-dimensional geometry or a 1-dimensional geometry (or both), with each fiber being a geometry of complementary dimension. The complete geometric description of the fibration involves some measure of how it is twisted. Four of the new geometries have fibrations over a 2-dimensional geometry, with 1-dimensional fibers. Missing from the list are twisted fibrations over the 2-sphere – such fibrations are swallowed up by the geometry of the 3-sphere, because of condition c), as well as untwisted fibrations over the Euclidean plane, which are swallowed up by 3-dimensional Euclidean geometry. Three of the new geometries have fibrations over the line, with fiber the 2-sphere ($S^2 \times \mathbf{E}^1$), the hyperbolic plane ($\mathbf{H}^2 \times \mathbf{E}^1$), and the Euclidean plane (solvegeometry). (Actually, in the last case, the fiber should more properly be thought of as the more restrictive geometry of the group $\mathbf{R}^2$ acting on itself by left multiplication.) Furthermore, solvegeometry in fact has a somewhat weird fibration over the hyperbolic plane – see section ??discussed with fig eight knot?? – it is just that this fibration is not as canonical nor is it as topologically meaningful as the fibrations of the other geometries.

This weird fibration is never discussed in the current version of the book–Dick

3.12.2 The 2-Sphere Cross the Line

The underlying space for this geometry is $S^2 \times \mathbf{E}^1$, and the group G is $O(3) \times \text{isom}(\mathbf{E}^1)$. One family of examples of (G, X) manifolds is the quotient of $S^2 \times \mathbf{E}^1$ by the discrete group generated by the transformation

$$(\theta, t) \mapsto (A(\theta), t + c)$$

where $\theta \in S^2, t \in \mathbf{E}^1$, A is an arbitrary orthogonal transformation and c is an arbitrary positive real number. The manifold is diffeomorphic to $S^2 \times S^1$ when $\det(A) > 0$ and to a related unorientable manifold otherwise.

There is one more family of examples, with group $\mathbf{C}_2 * \mathbf{C}_2$ generated by

$$(\theta, t) \mapsto (-\theta, -t)$$

and

$$(\theta, t) \mapsto (-\theta, a - t)$$

where $a > 0$. The quotient three-manifold is orientable, and it can be described as $S^2 \times I$, with each end identified to itself by the antipodal map. In the three-manifold the two ends become copies of $\mathbf{RP}^2$, with orientable regular neighborhoods wich are therefore one-sided and have boundary S^2.

These are the only possible examples of compact manifolds modeled on $S^2 \times \mathbf{E}^1$:

Exercis 3.12.3. (a) Any discrete subgroup of isometries of $S^2 \times \mathbf{E}^1$ acts discretely (but not necessarily effectively) on $\mathbf{E}^1$.

(b) A discrete group of isometries of $\mathbf{E}^1$ is isomorphic to $\mathbf{Z}$ or to $\mathbf{C}_2 * \mathbf{C}_2$.

(c) The only closed three-manifolds modeled on $S^2 \times \mathbf{E}^1$ are the examples given above.

3.12.4 The Hyperbolic Plane Cross the Line

The underlying space for this geometry is $\mathbf{H}^2 \times \mathbf{E}^1$, and the group is the group of isometries of the hyperbolic plane cross the group of isometries of the Euclidean line. The product of any hyperbolic surface with a circle gives an example of a $(\mathrm{isom}(\mathbf{H}^2) \times \mathrm{isom}(\mathbf{E}^1), \mathbf{H}^2 \times \mathbf{E}^1)$-manifold. More generally, if S is any hyperbolic surface and $\phi : S \to S$ is any isometry of it, then the mapping torus of ϕ,

$$M_\phi = (S \times I)/\{(x,0) \sim (\phi(x),1)\}$$

has an $\mathbf{H}^2 \times \mathbf{E}^1$ structure.

3.12.5 The Geometry of Twisted Bundles over the Hyperbolic Plane.

This geometry is needed for modeling most circle bundles over surfaces of negative Euler number The simply-connected underlying space for this geometry can be constructed by taking the universal covering space of the tangent circle bundle $TS(\mathbf{H}^2)$ of the hyperbolic plane, that is, the space of non-zero tangent vectors up to multiplication by positive real numbers. The group I of isometries of the hyperbolic plane acts, via derivatives, on the tangent circle bundle $TS(\mathbf{H}^2)$, which is homeomorphic to an open solid torus. It is a curious fact that even orientation-reversing isometries of the $\mathbf{H}^2$ plane preserve orientation in their action on $TS(\mathbf{H}^2)$, since they reverse the direction of the circles as well as the orientation of the base. $TS(\mathbf{H}^2)$ can also be thought of as $\mathrm{PSL}(2,()\mathbf{R})$, the group of orientation-preserving isometries of $\mathbf{H}^2$, since such an isometry is determined by where it takes a fixed element of the tangent circle bundle.

Exercis 3.12.6. How do the various kinds of isometries of the hyperbolic plane act on its tangent circle bundle, visualized as a solid torus? Describe, in particular, how a rotation of (say) $2\pi/3$ in the hyperbolic plane acts, and how a reflection through a line acts.

Let $\widetilde{I}$ consist of all homeomorphisms of the universal cover $W = \widetilde{TS}(\mathbf{H}^2)$ which commute with covering transformations and whose quotient action on $TS(\mathbf{H}^2)$ comes from the action of an element of I. Let J be the group generated by $\widetilde{I}$, together with vertical translations (which preserve each fiber, in the fibration of W over $\mathbf{H}^2$). Note that vertical translations commute with those elements of $\widetilde{I}$ that preserve orientation in the hyperbolic plane, and that a translation by any multiple of 2π is again in $\widetilde{I}$. The entire group J preserves the orientation of W, so every (J, W)-manifold can be given a canonical orientation if we choose an orientation for W.

% klein237 Section "SL2R METRIC"

There is a nice construction for some interesting examples of 3-manifolds modeled on (J, W), which begins with an arbitrary discrete group Γ of orientation-preserving isometries of the hyperbolic plane – for instance, the orientable $(2,3,7)$ triangle group (the orientation-preserving subgroup of the group generated by reflections in the sides of a triangle with angles $\pi/2$, $\pi/3$, and $\pi/7$. See Figure 2.18). Even though there are sometimes elements of such groups which fix points on the hyperbolic plane, they do not fix any points on the tangent sphere bundle of the hyperbolic plane, since their derivatives are not the identity. The quotient of $TS(\mathbf{H}^2)$ by Γ is a 3-manifold with a (J, W)-structure.

Other examples can be constructed from these by fiddling with the amounts of vertical translation in their holonomy groups ($\subset J$). For instance, the holonomy group for the manifold constructed from the $(2,3,7)$ group above is an infinite cyclic central extension of the $(2,3,7)$ triangle group,

$$\mathbf{Z} \to \widetilde{\Gamma} \to \text{(2,3,7)-group}.$$

If this group is enlarged by adjoining translations by multiples of $2\pi/11$, the group is still discrete and still acts without fixed points.

Exercis 3.12.7. For which infinite cyclic central extensions of the $(2,3,7)$ triangle group can you construct isomorphic discrete subgroups of J? Which of these act without fixed points? Suppose the base group is the fundamental group of a surface of genus two, instead of the $(2,3,7)$ group – can you think of any additional construction? [Hint: what is the abelianization of $\pi_1(TS(\text{surface}))$?]

Instead of visualizing $W = \widetilde{\mathrm{PSL}(2,()\mathbf{R})}$ as something constructed from little vectors on the hyperbolic plane, it is better to think of it in three dimensions. One image is the universal covering of the solid torus, $D^2 \times \mathbf{R}$. This description as a product is unnatural, however; it is really better to try to visualize the internal geometry of W from the point of view of someone living inside it. To form a picture of the internal geometry, begin by imagining what it is like to live inside $\mathbf{H}^2 \times \mathbf{E}^1$. The horizontal planes are like hyperbolic planes, while any plane which passes through the vertical direction is a Euclidean plane. Now modify this picture by introducing a consistent warp of space. There is still an infinitesimal definition of vertical and horizontal, but the warping eliminates any global notion of horizontal levels, since the new picture is not geometrically a product. By consistent warping, one can create the effect that a horizontal path which turns at a constant speed to the left never closes up, but instead, spirals upward – it returns to the initial vertical fiber successively at higher and higher points. A horizontal path which turns at a constant speed to the right spirals downward.

In the geometry of W, the foliation by vertical lines is natural – it is invariant by the full group of congruences J. The distinction between it and the geometry of $\mathbf{H}^2 \times \mathbf{E}^1$ is that in W, there is no horizontal foliation.

NILGEOM
% SEMIDIR DEF

Exercis 3.12.8. Discuss the three-dimensional picture of W in more precise terms. Suppose that you move along a path in W which is infinitesimally horizontal, and arrive back at the vertical line where you started. Express your change in altitude in terms of the signed area enclosed by your shadow in $\mathbf{H}^2$. Give a formula for the metric on W in terms of upper half-plane coordinates cross $\mathbf{R}$ (check your answer against [Sco83]).

3.12.9 Nilgeometry.

This geometry is needed for modeling twisted circle bundles over surfaces of zero Euler number It has many similarities to the previous case. To give an algebraic description, we can begin with the Heisenberg group, which is the group of upper-triangular 3 by 3 matrices

$$N = \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$$

This group is the only 3-dimensional nilpotent but non-abelian connected and simply-connected Lie group. The Heisenberg group acts on itself by left translation to give a 3-dimensional geometry, but this group of transformations is not a maximal group with compact stabilizers. In order to visualize the Heisenberg group and also see how it can be enlarged, let us use slightly different coordinates (a,b,d), where $d = c - \frac{1}{2}ab$. The rule for multiplication becomes

$$(a,b,d)\cdot(a',b',d') = (a+a', b+b', d+d'+\frac{1}{2}(ab'-ba')).$$

Fortunately, it is not too hard to see how left multiplication by an element $n = (a,b,d)$ acts on $\mathbf{R}^3$: in fact, n acts as an affine transformation of $\mathbf{R}^3$ which takes vertical lines (parallel to the d-axis) to vertical lines, and projects to the (a,b)-plane to act as a pure translation. The transformation may be constructed by first translating the origin to the point (a,b,d) in $\mathbf{R}^3$, and then shearing $\mathbf{R}^3$ by a linear transformation which is the identity if $(a,b) = (0,0)$, and otherwise fixes the vertical plane which contains the d-axis as well as the point (a,b,d), moving a general point (x,y,z) in the d-direction a distance equal to the signed area of the triangle $\Delta((0,0),(a,b),(x,y))$.

Using the (a,b,d)-coordinates, we see that there is a group of automorphisms of N which is isomorphic to $O(2)$, acting as a group of orientation-preserving isometries of $\mathbf{R}^3$ by first performing its usual action on the (a,b)-plane, and then correcting the orientation if necessary by a reflection through this plane. As the group G of congruences of nilgeometry, we take the semi-direct product corresponding to this action of $O(2)$,

$$G = N \ltimes O(2).$$

(See exercise 3.12.10 for a definition of semi-direct products).

Section "SEMIDIR DEF"

As with the preceding geometry, nilgeometry has a natural 1-dimensional vertical foliation, but no natural transverse 2-dimensional foliation.

As an example, we can construct a 3-manifold with a nilgeometry structure from the subgroup of the Heisenberg group consisting of matrices with integer entries. This subgroup is discrete, and has compact quotient.

junglejim

junglejim

Figure 3.32. junglejim. A homogeneous jungle gym in nilgeometry. This is a picture of the subgroup Γ of the Heisenberg group consisting of matrices with integral entries. The group is transferred to (a, b, d)-coordinates, where the orbit of the origin under the group forms a lattice in $\mathbf{R}^3$, although it is not the integer lattice. Lines are drawn to show the graph of the group: three generators are chosen (the generators are represented by matrices such that (a, b, c) is one of the three coordinate vectors) and pairs of lattice points are connected by a line whenever one is obtained from the other by multiplication on the right by one of the generators. Since multiplication on the right commutes with multiplication on the left, the picture is invariant by the discrete group Γ acting by left multiplication, so from it one can see how Γ acts.

Exercis 3.12.10. Let G be a group with a subgroup H and a normal subgroup N. G is the *semi-direct product* of N and H ($G = N \ltimes H$) if $G = NH$ and $N \cap H = \{1\}$.

a) There is a natural homomorphism $\phi : H \to Aut(N)$ given by conjugation.

b) This action of H on N determines the group: given any two groups N and H, and a homomorphism $\phi : H \to Aut(N)$, show how to construct the (unique) semi-direct product of N and H consistent with this action.

3.12.11 Solvegeometry.

This geometry is the least symmetric of all, having the stabilizer of any point a finite group. It is based on a three-dimensional solvable Lie group $S = \mathbf{R}^2 \ltimes \mathbf{R}$ where the action of $\mathbf{R}$ by conjugation on $\mathbf{R}^2$ is that $t \in \mathbf{R}$ sends $(x, y) \in \mathbf{R}^2$ to $(e^t x, e^{-t} y)$. If we use coordinates $(x, y, t) = (x, y, 0) \cdot (0, 0, t)$, then the formula for multiplication is

$$\begin{aligned}(x, y, t) \cdot (x', y', t') &= (x, y, 0) \cdot (0, 0, t) \cdot (x', y', 0) \cdot (0, 0, -t) \cdot (0, 0, t' + t) \\ &= (x + e^t x', y + e^{-t} y', t + t').\end{aligned}$$

% MAP TOR
Section "MAP TOR"

Thus, multiplication on the left by (x, y, t) is an affine transformation of $\mathbf{R}^3$, which first distorts in the x and y directions according to t, then translates by (x, y, t). The action of $(\mathrm{C}_2)^2$ on $\mathbf{R}^3$, in which the three non-trivial elements act as 1/2-rotations about the three orthogonal lines $x = y = 0$, $x - y = t = 0$ and $x + y = t = 0$, preserves the group structure. Our space is S, and the group is $S \ltimes (\mathrm{C}_2)^2$.

If $\phi : T^2 \to T^2$ is any linear map of the torus to itself – that is, it comes from a linear map of $\mathbf{R}^2$ to itself which preserves the integer lattice – then the mapping torus M_ϕ always has either a Euclidean structure, a nil-geometry structure or a solve-geometry structure. (Exercise 3.12.13). Compare [Poi95].

Exercis 3.12.13. Let $\phi : T^2 \to T^2$ be a linear map of the torus to itself, and M_ϕ be the mapping torus. Let λ_1 and λ_2 be the roots of the characteristic polynomial of ϕ, so $\lambda_1 \cdot \lambda_2 = \det \phi = \pm 1$ and $\lambda_1 + \lambda_2 = \mathrm{tr}(\phi)$.

(a) If λ_1 and λ_2 are not real, then they are roots of unity, and ϕ has finite order. T^2 has a Euclidean structure for which ϕ is an isometry, and M_ϕ has a Euclidean structure.

(b) If λ_1 and λ_2 are real and distinct, then ϕ has two eigenspaces. The torus has a Euclidean metric for which these eigenspaces are orthogonal, and M_ϕ has a solve-geometry structure.

(c) If ϕ does not fit into case (a) or (b), then $\lambda_1 = \lambda_2 = \pm 1$, and then either $\phi^2 = 1$ and M_ϕ has a Euclidean structure, or else ϕ is conjugate to the form $\pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ $[n \neq 0]$ and M_ϕ has a nil-geometry structure.

(d) Find examples of linear maps ϕ_1 and ϕ_2 which have the same characteristic polynomials with real distinct roots such that M_{ϕ_1} and M_{ϕ_2} are not homeomorphic.

Chapter "The structure of discrete groups"

Chapter 4

The structure of discrete groups

There are often strong consequences for the topology of a manifold M which can be expressed as the quotient space of a homogeneous space (G, X) by a discrete group $\Gamma \subset G$. These consequences arise from geometric and algebraic restrictions on discrete groups. In some cases, the information is strong enough to enable a complete classification of closed (G, X)-manifolds.

In this chapter, we will investigate the structure of discrete subgroups of automorphisms of a homogeneous space, with an emphasis on the three-dimensional model geometries.

Section "The thick-thin decomposition" ::injectivity radius Proposition "thick compact" % Groups generated by small elements

4.1. The thick-thin decomposition

Let M^n be a complete hyperbolic manifold, possibly with infinite volume. For any point $x \in M$, consider the set of all its lifts to the universal cover $\widetilde{M^n} = \mathbf{H}^n$. They form a regular array, like atoms in a crystal. Because the group of covering transformations is discrete, there is a minimum distance d between any two lifts of x. This minimal distance d is also the shortest possible length of a loop based at x which represents a non-trivial element of $\pi_1(M)$. A ball of radius $r = d/2$ about x is embedded, since all its lifts are disjoint, but no larger ball can be embedded.

This number r is called the *injectivity radius* of M at x, written $r = i_r(x)$. (The terminology arises from the more general setting of a complete Riemannian manifold M. There is a canonical map, the exponential map, from the tangent space of M at x to M, defined by extending geodesics. The injectivity radius is the radius of the largest open ball on which the exponential map is injective.)

We can decompose M into two pieces, $M = M_{thick(\varepsilon)} \bigcup M_{thin(\varepsilon)}$, where

$$M_{thick(\varepsilon)} = \{x \in M : i_r(x) \geq \varepsilon\}$$

and

$$M_{thin(\varepsilon)} = closure(\{x \in M : i_r(x) < \varepsilon\}).$$

When the particular value of ε is not important, we simply write M_{thick} and M_{thin}.

We shall see that $M_{thin(\varepsilon)}$ has a very standard form, provided ε is chosen small enough. In particular $M_{thin(\varepsilon)}$ and $M_{thick(\varepsilon)}$ are sub-manifolds with boundary, $M_{thick(\varepsilon)}$ does not fit into any standard mold, but on the other hand,

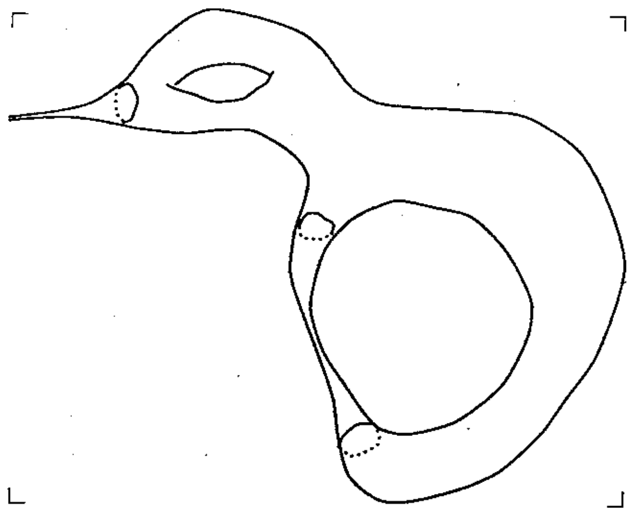
Proposition 4.1.1 (thick compact). *If M^n is a complete hyperbolic manifold of finite volume, then M_{thick} is compact.*

Proof of 4.1.1: The ball of radius ε in M about a point in M_{thick} has the same volume V as a ball in $\mathbf{H}^n$ of the same radius. Therefore, there is an upper bound $\mathrm{vol}(M)/V$ for the number of points which can be put in M_{thick} such that the distance between any two is at least 2ε, since their balls would be disjoint. If S is any maximal set of points in M_{thick} with this property, then the closed neighborhood of radius 2ε about S contains M_{thick}. This neighborhood is compact; since M_{thick} is a closed subset, it is also compact. 4.1.1

(complete)

Let us turn now to M_{thin}. If $x \in M_{thin}$, then for any point $\widetilde{x} \in \mathbf{H}^n$ lying above x there are elements $\gamma \in \pi_1(M)$ which move $\widetilde{x}$ only a short distance. The key to understanding M_{thin} is to analyze discrete groups generated by "small" elements. We will state now the main conclusion in this direction for hyperbolic space, but we will give the proof in a somewhat more general context in the next section (4.2).

Proposition "thin abelian subgroup"
% thin abelian subgroup
% distance function is convex



thickthinsurf

Figure 4.1. The thick-thin decomposition of a surface. The thin part of a hyperbolic surface of finite area consists of neighborhoods of short geodesics and cusps which are isometric to pseudospheres.

Proposition 4.1.2 (thin abelian subgroup). *In every dimension n there is an $\varepsilon > 0$ and an integer m such that if B is a ball of radius ε in $\mathbf{H}^n$ and if $\pi_1(M)$ is any discrete subgroup of* $\text{isom}(\mathbf{H}^n)$ *generated by elements γ such that $\gamma(B)$ intersects B, then $\pi_1(M)$ has a normal abelian subgroup of index at most m.*

One reason the integer m is necessary is that there are nonabelian discrete groups which fix a point. For instance, in dimension $n = 2$, the group generated by reflections in two lines meeting at angle π/k is discrete. It is a dihedral group, and has a normal abelian subgroup (rotations) of index two. For $n = 2$, m can be taken as 2.

In dimension 3, the most complicated example of a discrete group which fixes a point is the group of symmetries of an icosahedron. It has order 120. The largest normal abelian subgroup has order 2 (generated by the antipodal map of an icosahedron). The minimal value for m in dimension 3 is 60.

Proposition 4.1.2 enables us to analyze the structure of $M_{thin(\varepsilon)}$, where M is a complete hyperbolic manifold. Choose an ε sufficiently small for the proposition.

For any set $S \subset \pi_1(M)$, let $T(S)$ denote the set of points in $\mathbf{H}^n$ moved a distance less than ε by some non-trivial element of S. In particular, $T(\pi_1(M))$ is the interior of $\widetilde{M}_{thin(\varepsilon)}$, the subset of $\widetilde{M}$ lying above $M_{thin(\varepsilon)}$.

For any particular isometry γ, the translation distance $d_\gamma(x) = d(x, \gamma(x))$ is a convex function on $\mathbf{H}^n$ (2.5.8). Therefore the set $T(\{\gamma\})$ is convex.

Two sets $T(\{\beta\})$ and $T(\{\gamma\})$ can have non-empty intersection only if the group $G(\beta, \gamma)$ generated by β and γ contains a normal abelian subgroup $A(\beta, \gamma)$ with bounded index.

::glide-reflection
::star-shaped

Since the group $\pi_1(M)$ consists entirely of hyperbolic and parabolic transformations, there are two possibilities. The first possibility is that there exists a hyperbolic element $\phi \in A(\beta, \gamma)$. If so, then ϕ has a unique axis l. The group $A(\beta, \gamma)$ therefore must take l to itself, so it acts as a group of translations on l. Therefore $A(\beta, \gamma)$ is isomorphic to the infinite cyclic group, $\mathbf{Z}$. Since $A(\beta, \gamma)$ is a normal subgroup of $G(\beta, \gamma)$, the latter group also takes l to itself, so $G(\beta, \gamma)$ as well is infinite cyclic. In particular, β and γ commute.

In fact, if B is the subgroup of $\pi_1(M)$ stabilizing l, then B too must be infinite cyclic, so the component of $T(\pi_1(M))$ containing $T(\{\gamma\})$ is the same as $T(B)$. (It is often not true, however, that $T(B)$ is identical with $T(\{\alpha\})$, where α generates B.)

In dimension two, for any hyperbolic element ϕ the set $T(\{\phi\})$ is a neighborhood of some radius r about its axis, since ϕ commutes with reflection through its axis. It follows that in dimension two, $T(B)$ is also a neighborhood of some radius r about l.

In dimension three, if ϕ is a hyperbolic element which preserves orientation, it is composed of a translation along l followed by a rotation about l. This commutes with rotation about l, so $T(\{\phi\})$ in this case is a cylindrical neighborhood of some radius r about l. If M is orientable, then $T(B)$ is a union of cylindrical neighborhoods, so it is also a cylindrical neighborhood.

If ϕ is a hyperbolic orientation-reversing transformation of $\mathbf{H}^3$, it is a *glide-reflection*, composed of a translation along l followed by a reflection through some plane P through l. Then $T(\{\phi\})$ will have an eccentric cross-section, extending further in the plane of P than in the plane through l perpendicular to P. In this case, a cyclic generator α for B must also be a glide reflection, and $T(B) = T(\{\alpha\})$.

In higher dimensions, $T(B)$ need not be convex. Nonetheless, it is a union of convex sets containing l, so it is star-shaped with respect to any point on the axis l. (A subset X of $\mathbf{H}^n$ is *star-shaped* with respect to a point $x \in X$ if the geodesic between x and any other point of X is contained entirely within X.)

The other possibility is that $A(\beta, \gamma)$ consists entirely of parabolic elements. A parabolic element fixes a unique point $p \in S^{n-1}_\infty$, so $A(\beta, \gamma)$ and $G(\beta, \gamma)$ must fix p. We may arrange coordinates so that $p = \infty$ in the upper half-space model.

Let B be the subgroup of $\pi_1(M)$ fixing p. It must consist entirely of parabolic elements: if it contained a hyperbolic element, then the conjugates of a parabolic element by this hyperbolic element would have arbitrarily short translation distances.

Then the component of $T(\pi_1(M))$ containing $T(\{\gamma\})$ is $\bigcup_{\phi \in B} T(\{\phi\})$. Along any vertical ray in upper half-space, the translation distance of any parabolic element ϕ fixing ∞ decreases exponentially. Therefore, $T(\{\phi\})$ intersects such a ray in a half-line. This means that $T(B)$ is star-shaped in a

Problem "shapes of thin sets"
::cusps
Proposition "thick compact implies finite volume"
% thick compact

generalized sense with respect to p (although p, being on S_∞, is not in $T(B)$ at all).

In the two- or three-dimensional orientable case, each $\beta \in B$ must look like a pure translation in the upper half-space model, so $T(\{\beta\})$ and $T(B)$ are open horoballs.

Problem 4.1.3 (shapes of thin sets). (a) Give an example of an orientation-preserving hyperbolic transformation α of $\mathbf{H}^3$ such that $T(\{\alpha\})$ is a proper subset of $T(\{\alpha^{100}\})$.

(b) Give an example of a non-convex component of $T(A)$ for a cyclic subgroup $A(\beta,\gamma)$ of isometries of $\mathbf{H}^4$.

From the picture of the components of $T(\pi_1(M))$, the picture of M_{thin} follows easily.

Some components of M_{thin} may be compact. These are diffeomorphic to $(D^{n-1} \times \mathbf{R})/\mathbf{Z}$, so the quotient is $D^{n-1} \times S^1$ if the action of $\mathbf{Z}$ preserves orientation, or a non-orientable disk bundle over S^1 otherwise.

Other components of M_{thin} may be non-compact. These all have the form (up to diffeomorphism) of $N^{n-1} \times [0,\infty)$, where N^{n-1} is an arbitrary Euclidean $(n-1)$-manifold. These non-compact components of M_{thin} are neighborhoods of *cusps* of M. Cusps are not points of M, but ideal points, which can be defined as (parabolic fixed points in S_∞^{n-1} of elements of $\pi_1(M)$)/(Action of $\pi_1(M)$).

Should more be said here, or is it obvious enough - Dick

Proposition 4.1.4 (thick compact implies finite volume). *A complete hyperbolic manifold M has finite volume iff M_{thick} is compact.*

Proof of 4.1.4: We have already seen that M_{thick} is compact if M has finite volume (4.1.1). For the converse, we need to show that when M_{thick} is compact, then each of the finitely many components of M_{thin} has finite volume. This is trivial for compact components. For a non-compact quotient, C, consider its universal cover $\widetilde{C}$ in upper half-space, $x_n > 0$, arranged so the parabolic fixed point is the point at ∞. The volume of C is computed as the volume of $\widetilde{C}$ lying above some compact region in $X_1, ..., X_{n-1}$ plane, i.e. a fundamental region for the action of B on $\widetilde{C}$. This cross-section of $\widetilde{C}$ lies above some height $x_n = h > 0$ and since the hyperbolic volume element decreases as $(1/X_n)^n$, the volume is clearly finite. 4.1.4

Section "Groups generated by small elements"
::Hausdorff topology
::Hausdorff distance
Proposition "closed subgroups Hausdorff closed"

4.2. Groups generated by small elements

This section has been totally rewritten. Not all the previous results, nor the previous exercises, have been put in. Be on the lookout for results that are needed. The nilpotent subgroup theorem has been strengthened, so it should be more powerful.

To get an overall image of discrete groups Γ of a Lie group G, you need to understand not only the well-behaved examples but the extreme examples as well, when Γ is almost-indiscrete, in the sense that there are many elements very close to the identity.

For this, the *Hausdorff topology* is quite helpful. The *Hausdorff distance* between two closed subsets A and B of a compact metric space X is the greatest distance a point in either set is from any point in the other set, or in symbols,

$$d(A,B) = \max\left(\max_{a\in A}(\min_{b\in B} d(a,b)), \max_{b\in B}(\min_{a\in A} d(a,b))\right).$$

The Hausdorff metric makes the set of closed subsets of X into a compact metric space $\mathcal{H}(X)$. The topology of $\mathcal{H}(X)$ depends only on the topology of X, not on the metric on X.

For subsets of a noncompact metric space, the appropriate definition is not as clear, and in fact there are two good definitions. Most obvious is to use exactly the same definition as for compact spaces. Unfortunately, with this topology, the space is not compact (and not even paracompact).

For our purposes, the best choice is to define the Hausdorff topology in terms of the intersections with compact sets. That is, if X is a locally compact, complete metric space and $A \subset X$ is a closed subset, we define neighborhoods $\mathcal{N}_{K,\varepsilon}(A)$ for compact $K \subset X$ and $\varepsilon > 0$ to consist of sets B such that $d(A \cap K, B \cap K) < \varepsilon$. With this topology $\mathcal{H}(X)$ is a compact topological space.

Should we have a reference for this fact?-Dick

Let's apply this to the case X is a Lie group G: the Hausdorff topology induces a topology on the set of closed subgroups $H \subset G$. Here is an elementary fact:

Proposition 4.2.1 (closed subgroups Hausdorff closed). *The set of closed subgroups of a Lie group G form a closed, hence compact, set in $\mathcal{H}(G)$.*

Proof of 4.2.1: Let $A \subset G$ be a closed subset which is a limit point of closed subgroups of G. Let $a, b \in A$. We need to verify that $a^{-1} \in A$ and that $ab \in A$. Choose a compact set $K \subset G$ whose interior contains neighborhoods of a, b, ab and a^{-1}. Then $\mathcal{N}_{K,\varepsilon}(A)$ contains subgroups with elements within ε of a and b, hence elements near to a^{-1} and to ab. Taking the limit as $\varepsilon \to 0$, the proposition follows. 4.2.1

For any closed subgroup $H \subset G$, the component H_0 of H which contains 1 is a Lie subgroup: it is a connected submanifold, and it is determined by its tangent space at 1. The entire subgroup H is a union of a discrete set of cosets of H_0; in the special case that $H_0 = 1$, H is a discrete group.

If G admits discrete groups with elements arbitrarily close to 1, then the cyclic subgroups generated by these elements form a dotted line going a definite distance away from 1. If this happens, the set of discrete subgroups do not

Theorem "nilpotent limits" ::Lie algebra

form a closed subspace of $\mathcal{H}(G)$: the Hausdorff limit of a sequence of discrete groups, as elements get closer and closer to 1 has a nontrivial component of the identity.

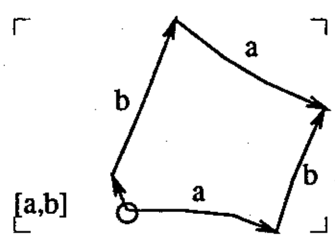
If we can understand what are the Lie subgroups $H_0 \subset G$ which can occur as the identity components of such limits, we will have a good tool for understanding subgroups of G with small elements. The key result is the following:

Theorem 4.2.2 (nilpotent limits). (a) *For any Lie group G, there is a neighborhood U of 1 in G such that any discrete subgroup of G which is generated by its intersection with U is nilpotent.*

(b) *Furthermore, any discrete subgroup of G generated by its intersection with U is a cocompact subgroup of a connected, closed, nilpotent subgroup of G.*

(c) *The closure of the set of discrete subgroups of a Lie group G consists only of discrete subgroups together with certain closed subgroups whose identity component is nilpotent.*

Proof of 4.2.2: The basic idea is that for elements close to the identity in a Lie



smallnil

Figure 4.2. commutator of small elements. In any Lie group, the commutator $[a,b] = aba^{-1}b^{-1}$ of two elements near 1 is even closer. Its distance from 1 is estimated by the product of the distances of a and b from 1, by Taylor's theorem.

group, multiplication is approximated by vector addition in its tangent space (its *Lie algebra*). Since addition of vectors is commutative, multiplication of small elements is almost commutative: it is commutative up to second order. In symbols, there is an $\varepsilon > 0$ and a constant C such that for any $a, b \in G$ such that $d(1,a) < \varepsilon$ and $d(1,b) < \varepsilon$, then

$$d(1, [aba^{-1}b^{-1}]) < Cd(1,a)d(1,b).$$

In particular, $[a,b]$ is considerably closer to 1 than either a or b. This inequality follows from Taylor's theorem applied to the map of $G \times G \to G$ defined by the commutator $[a,b]$, taking into account that $[1,b] = [a,1] = 1$.

Consider any discrete subgroup Γ which is generated by its intersection with the ε neighborhood of 1, where ε is sufficiently small that $d(1,[a,b]) < 1/2d(1,a)$ for any a and b within ε of 1. Then any nested commutator

$$[g_1, [g_2, [g_3, \ldots, [g_{k-1}, g_k] \ldots]]]$$

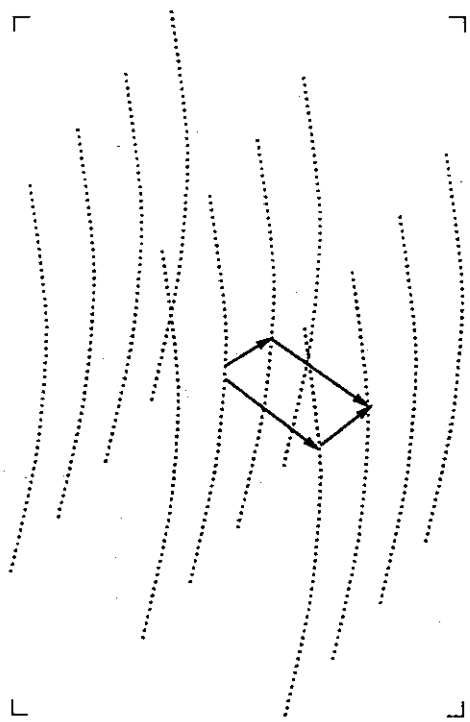
::left-invariant
exp(V)

of elements g_i within ε of 1 is within distance $2^{-k}\varepsilon$ of 1. Since there is a minimum distance of nontrivial elements of Γ from 1, there is some k such that all such nested commutators of length k equal 1.

It easily follows that Γ is nilpotent: define subgroups

$$\Gamma = \Gamma_0 \supset \Gamma_1 \supset \ldots \supset \Gamma_k = 1$$

to be the subgroups generated by length i nested commutators of elements of Γ in the ε-neighborhood of 1. This is a central series for Γ, which proves that it is nilpotent.



smallgroup

Figure 4.3. group generated by small elements. A discrete subgroup of a Lie group generated by small elements must look something like this. It is something like a lattice in $\mathbf{R}^n$, but multiplication might not be quite commutative: the commutators of elements in spaced out directions might be nontrivial in directions where the dots are spaced much closer together.

To prove the second statement, we need to use the exponential map and the logarithm in a Lie group G. The Lie algebra of G is the tangent space $\mathcal{G}$ to G at the point 1. An element V in $\mathcal{G}$ can be transported by the action of G on itself by multiplication on the left to give a *left-invariant* vector field L_V on G. The flow of L_V is the action on G on the right by a 1-parameter subgroup of G. The image of 1 by the time one map of the flow of L_V is called $\exp(V)$—the time one map of the flow acts by multiplication on the right by $\exp(V)$. The derivative of the map $V \to \exp(V)$ is the identity at 1. It follows, in particular that the exponential map is injective from a neighborhood of 0

log
Ad
::Grassmanian
% slow recovery

in $\mathcal{G}$ to a neighborhood of 1 in G. In such a neighborhood, we can define log to be the inverse function of exp.

The usual exponential function is the exponential map for the multiplicative group of positive real numbers.

It is easy to check that for any integer n, just as with the usual exponential function, $\exp(nV) = \exp(V)^n$. However, $\exp(V+W)$ does not in general equal $\exp(V)\exp(W)$, since multiplication in G might not be commutative.

Let U be a small neighborhood of 1 in G. Let us study a discrete group Γ generated by its intersection with U. We may assume that the logarithm is well-defined on U and that the left-invariant vector fields L_V are nearly parallel in U, as viewed in some local coordinate system. So that we have a clear picture, we can take U to be the image of a ball of radius ε in $\mathcal{G}$. We will construct a connected nilpotent subgroup containing Γ by building an increasing family of nilpotent subgroups

$$\Gamma_0 = N_0 \subset N_1 \subset \ldots \subset N_k,$$

containing Γ, of increasing dimension, until we obtain a connected subgroup N_k.

The first step is the easiest. Let a_1 be a nontrivial element of Γ which is closest to 1, and let V_0 be its logarithm. The flow from V_0 is a 1-parameter subgroup of G containing a_0. We know that a_1 is central in Γ. It follows that the nth root $a_1^{\frac{1}{n}} = \exp((1/n)\log(V_0))$ is also centralized by Γ, because of the uniqueness of nth roots within U. Passing to the limit as $n \to \infty$, it follows that the 1-parameter subgroup M_1 with tangent V_0 is centralized by Γ. Because the powers of a_1 are uniformly spaced in M_1, it is a closed subgroup. Thus $N_1 = M_1 \cdot \Gamma$ is a closed nilpotent subgroup of G on which *Gamma* acts cocompactly.

The induction step is similar. We assume, by induction, that we have a sequence $N_0 \subset \ldots \subset N_i$ of closed nilpotent groups containing Γ, each of the form $N_i = M_i \cdot \Gamma$, where M_i is connected and closed, of dimension i. We also assume that $[N_i, M_i] \subset M_{i-1}$ and that Γ is a cocompact subgroup of N_i.

We take a_i to be a smallest element of N_i which is not in M_i. The commutator of any element of $N_i \cap U$ with a_i is smaller than a_i, so it must be in M_i. In particular, a_i normalizes M_i. We want to show that the 1-parameter group $G(a_i)$ containing a_i also normalizes M_i. To see this, consider its action by conjugacy on the group G. The derivative of this action at 1 is an action (called *Ad*) on $\mathcal{G}$. This induces an action on the space of d_i-dimensional subspaces of $\mathcal{G}$, where d_i is the dimension of M_i. (In general, the space of k-dimensional subspaces of an an n-dimensional vector space is called the *Grassmanian* manifold $G(n,k)$.) Exercise 4.2.3 asserts that for any smooth flow on any compact manifold, there is some $t_0 > 0$ such that if a point is not fixed by the flow, it cannot be fixed by the time t map of the flow for any time less than t_0. It follows that, if U has been chosen small enough, $G(a_i)$ normalizes M_i. Therefore, $M_{i+1} = M_i \cdot G(a_i)$ is a closed, connected subgroup of G.

I think the fact that M_{i+1} is closed depends on the cocompactness of the action, the only argument I could come up with used

Exercise "slow recovery"
% nilpotent limits
Corollary "discrete orthogonal almost abelian"
% nilpotent limits
% invariant metric exists
::bi-invariant metric
% nilpotent limits

Applying the uniqueness of roots within U as before, it follows that Γ normalizes M_{i+1}, since it normalizes $a_i \cdot M_i$. Therefore, $N_{i+1} = \Gamma \cdot M_{i+1}$ is a closed subgroup of G.

To verify the other inductive hypotheses, apply similar arguments to the action by conjugacy of N_{i+1} on the M_j, and to the quotient action on M_{i+1}/M_j.

The proof that $[N_i, M_i] \subset M_{i-1}$ needs considerable expansion, the cocompactness of the action seems clear-Dick

The third statement of the theorem is easy to deduce from the second. 4.2.2

Exercise 4.2.3 (slow recovery). Let M be a closed manifold, and X a C^2-smooth vector field on M. Let ϕ_t be the flow coming from X. Show that there is a real number $t_0 > 0$ such that for any $p \in M$, if ϕ_t does not fix p for every t, then no $1 < t < t_0$ fixes p.

[Hint: Consider the picture when you blow up a local coordinate patch so that εX has unit length at p.]

The proof actually uses a continuous version of this exercise -Dick

In particular cases, theorem 4.2.2 can be made much more specific.

Corollary 4.2.4 (discrete orthogonal almost abelian). *For any dimension n, there is an m such that any discrete subgroup of $O(n)$ contains an abelian subgroup with index at most m.*

Proof of 4.2.4: Let U be a neighborhood of 1 guaranteed by theorem 4.2.2. For any discrete (or equivalently, finite) subgroup Γ of $O(n)$, the subgroup Γ_U generated by its intersection with U has bounded index. Specifically, we can find a symmetric neighborhood $W = W^{-1}$ of 1 such that $W^2 \subset U$. If c_i are coset representatives for Γ_U, then $c_i W$ and $c_j W$ are disjoint for $i \neq j$. Therefore the index $i(\Gamma, \Gamma_U)$ is less than the volume of $O(n)$ divided by the volume of W.

Since $O(n)$ is compact, it admits a metric which is invariant by the action of $O(n)$ on itself by both left and right multiplication (??). In general, a metric of this sort on a Lie group is called a *bi-invariant metric.*

According to 4.2.2, Γ_U is contained in a connected nilpotent subgroup N of $O(n)$. The bi-invariant metric restricts to a bi-invariant metric on N. Let Z be the center of N. The action Ad of N on its Lie algebra must fix the tangent space z to Z, since this action is the derivative of the action by conjugacy. Conversely, any tangent vector which is fixed by Ad generates a 1-parameter subgroup which is central, and therefore the vector is in z.

The perpendicular subspace p to z is also invariant, since the metric is invariant by Ad. Notice that p does not contain any invariant vectors.

If $Z \neq N$, the quotient group Z/N is nilpotent and has positive dimension, then Ad similarly has a nontrivial fixed subspace on its Lie algebra. But p projects isomorphically to the Lie algebra of Z/N, which contradicts the fact that p has no invariant vectors. Therefore, $Z = N$, so N is actually abelian. 4.2.4

Exercise "Isometries of $\mathbf{E}^n$"
Lemma "maximal plane of translation"
% Isometries of $\mathbf{E}^n$
Corollary "discrete Euclidean almost abelian"
% nilpotent limits
% discrete orthogonal almost abelian
% Isometries of $\mathbf{E}^n$

We will derive a similar corollary for Euclidean isometry groups, but first let's recall some basic facts about Euclidean isometries.

Exercise 4.2.5 (Isometries of $\mathbf{E}^n$). (a) Every isometry ϕ can be written as $V \mapsto A(V) + V_0$, where $A \in O(n)$ and V_0 is some vector.

(b) ~~If V_0 is not in the 1-eigenspace of A, then ϕ has a unique fixed point.~~ ϕ has a unique fixed pt. $\Leftrightarrow$ A has no 1-eigenspace

(c) If G is a subgroup of the group of isometries of E^n then then the subgroup $G(T)$ consisting of all the elements of G which are pure translations is normal in G.

Lemma 4.2.6 (maximal plane of translation). *Let ϕ be an isometry of $\mathbf{E}^n$. There is an unique maximal plane Euclidean subspace $\mathbf{E}(\phi)$ which ϕ maps to itself by a translation.*

Proof of 4.2.6: Let ϕ have the form $V \mapsto A(V)+V_0$. Let U be the 1-eigenspace of A. If U is trivial ϕ fixes a point and we are done. If not, then ϕ takes planes in $\mathbf{E}^n$ parallel to U to planes parallel to U, so it has an induced action as an isometry ϕ/U of $\mathbf{E}^n/U$. In the induced action, the 1-eigenspace is trivial.

It follows, exercise 4.2.5, that ϕ/U has a unique fixed point. In terms of ϕ, this means that there is a unique maximal plane Euclidean subspace $\mathbf{E}(\phi)$ which ϕ maps to itself by a translation. 4.2.6

Corollary 4.2.7 (discrete Euclidean almost abelian). *For each dimension n, there is an integer m such that a discrete subgroup of isometries of $\mathbf{E}^n$ has an abelian subgroup of index at most m.*

Proof of 4.2.7: Let U be some neighborhood of 1 in isom($\mathbf{E}^n$) as guaranteed by theorem 4.2.2. We may assume that this neighborhood has the form $U_1 \times U_2$, where U_1 is a neigborhood of 1 in $O(n-1)$ and U_2 is a neighborhood of 1 in the group of translations. If we conjugate isom($\mathbf{E}^n$) by contractions, we see that U_2 can be made as large as we please!

The subgroup of Γ generated by $U_1 \times \lambda U_2$ is contained in a connected, closed, nilpotent subgroup, for every λ. As $\lambda \to \infty$, these subgroups must eventually stabilize, because they can only increase by increasing their dimension. Therefore, the subgroup $\Gamma(U_1)$ generated by elements of Γ with derivative in U_1 is nilpotent, and is contained in a connected, closed, nilpotent subgroup N of isom($\mathbf{E}^n$). It follows, as in the proof of corollary 4.2.4, that the index of Γ_{U_1} in Γ is bounded.

We claim that N is abelian.

Let Z be the center of N. Suppose there is an element $\phi \in Z$ which is not a translation, *i.e.*, $\mathbf{E}(\phi)$ is a proper subspace of $\mathbf{E}^n$, of dimension $k < n$. Then all of N must preserve $\mathbf{E}(\phi)$, so that $N \subset O(n-k) \times \text{isom}\,\mathbf{E}^k$. Since N is nilpotent, its projections to the two factors are also nilpotent, hence, by induction, almost abelian, so N itself is almost abelian.

Otherwise, the center Z must consist of translations.

Let T be the normal subgroup of N consisting of translations (see exercise 4.2.5). Let $\mathbf{E}(T)$ be the plane spanned by the translational components of the elements of T. $E(T)$ is preserved by N. The quotient N/T is also nilpotent

% discrete orthogonal almost abelian
Problem "abelian subgroups translate"
Proposition "short motion almost nilpotent"
% nilpotent limits

and acts as a group of isometries of $\mathbf{E}^n/\mathbf{E}(T)$. Therefore N may be naturally thought of as a subgroup of $O(n-k) \times \mathbf{E}^k$ where k is the dimension of $\mathbf{E}(T)$. The result then follows by applying corollary 4.2.4 and induction. 4.2.7

Problem 4.2.8 (abelian subgroups translate). For every abelian subgroup A of isom $\mathbf{E}^n$, show that there is a unique maximal Euclidean subspace $\mathbf{E}(A)$ on which A acts by translations.

Consider a group G acting on a space X with compact stabilizers of points. We need to extend our analysis to the case of groups generated by elements which move some point point $x \in X$ a short distance in X, but where the group elements may not be close to the identity.

need argument to show G acts invariant for some Riemannian metric. (3.6.7)

Proposition 4.2.9 (short motion almost nilpotent). *Let G be a Lie group acting on a manifold X so that stabilizers of points are compact. Let $x \in X$ be any point. There exists an integer m and an $\varepsilon > 0$ such that any discrete subgroup Γ of G generated by elements which move x a distance less than ε has a normal nilpotent subgroup of finite index no more than m.*

Furthermore, Γ is contained in a closed subgroup of G with no more than m components with a nilpotent identity component.

split up into orbits of G. What if ∞ #?

Proof of 4.2.9: To clarify our picture, we may assume without loss of generality that $X = G/G_x$, or equivalently, that G acts transitively. In that case, G fibers over X with fiber G_x.

Do we need to explain this? I'm not sure its worth since all the cases we will be applying it to have this property-Dick

Let U be a neighborhood of 1 as guaranteed by 4.2.2, Then the subgroup of any discrete group Γ generated by its intersection with U is nilpotent. The main complication in this proof arises from the possibility that this subgroup may not be normal.

(a) Let $V = V^{-1}$ be any symmetric neighborhood of G_x whose closure is compact. Then V contains all elements of G which move x a distance less than ε, for some ε. Inductively define a sequence of neighborhoods $U_0 = U \supset U_1 \supset \ldots$ by setting

$$1 \in U_{i+1} = \bigcap_{g \in V} gU_ig^{-1} \subset 1U_i1^{-1} = U_i$$

Any conjugate of an element of U_{i+1} by an element of V is in U_i.

(b) Let d be the dimension of G. Since V has a compact closure, there is some integer m such that among any m elements $g_1, \ldots, g_m$ of V, there is at least one pair (g_i, g_j) which differ by an element $g_i^{-1}g_j \in U_d$.

For any subgroup $\Gamma \subset G$ and any subset $Y \subset G$, we use the notation Γ_Y to denote the subgroup generated by $\Gamma \cap Y$.

Choose a symmetric neighborhood $W = W^{-1}$ of G_x such that if $\{b_1, \ldots, b_m\}$ are any m elements contained in W then their product $b_1b_2\cdots b_m$ is in V. Consider any discrete group Γ generated by its intersection with W, that is, $\Gamma = \Gamma_W$. Each subgroup Γ_{U_i} is nilpotent, and is a cocompact subgroup of a connected nilpotent subgroup $N_i \subset G$. The subgroups $N_i \subset N_{i-1}$ can increase only by increasing dimension, so there must be some index $1 \leq i \leq d$

Corollary "small discrete hyperbolic almost abelian"
% short motion almost nilpotent
% discrete orthogonal almost abelian
% discrete Euclidean almost abelian
% discrete orthogonal almost abelian

such that $N_i = N_{i-1}$. Since any element $g \in W \cap \Gamma$ must conjugate Γ_{U_i} into $\Gamma_{U_{i-1}}$, it follows that g normalizes N_i. So $H = \Gamma \cap N_i$ is a normal, nilpotent subgroup of Γ.

We claim that the quotient group Γ/H has order at most m. Let $a_1, \ldots, a_n$ be a set of generators for H, each of which lies in W. If Γ/H has order greater than M there would be a collection of $m+1$ words $\{w_i\}$, in the generators $\{a_i\}$, of word length not exceeding m, which map to distinct elements of Γ/H. But each of these elements w_i is contained in V, so that some pair of them would have to differ by an element of $\Gamma_{U_d} \subset H$, contrary to hypothesis. $\boxed{4.2.9}$

Corollary 4.2.10 (small discrete hyperbolic almost abelian). *For every dimension n there is an integer m and a distance $\varepsilon > 0$ such that any discrete subgroup Γ of* isom($\mathbf{H}^n$) *generated by elements which move some point x a small distance ε contains an abelian subgroup with index at most m.*

Proof of 4.2.10: In light of theorem 4.2.9 it will suffice to show that every connected closed nilpotent subgroup N of isom($\mathbf{H}^n$) is abelian. If there is a hyperbolic element ϕ in the center of N, then the entire group must preserve the axis of ϕ, N is a subgroup of $O(n-1) \times \mathbf{R}$, and the conclusion follows from 4.2.4. If there is a parabolic element ϕ in the center of N, then the entire group must preserve the fixed point of ϕ on S^{n-1}_∞, so it is a subgroup of the group of Euclidean similarities. However, it must actually be a subgroup of the group of Euclidean isometries, since ϕ does not commute with a similarity that actually expands or contracts. The result follows from 4.2.7.

In the remaining case the center of N contains an elliptic element ϕ. Then N must preserve the fixed hyperbolic subspace of dimension $0 \le d < n$ fixed by ϕ, and it is a subgroup of $\mathbf{H}^d \times O(n-d)$. The result follows by induction, using 4.2.4. $\boxed{4.2.10}$

Section "Euclidean manifolds and crystallographic groups"
EUCLID MFDS
% discrete Euclidean almost abelian
Theorem "Bieberbach"
Corollary "classification of Euclidean manifolds"
% Bieberbach
% discrete Euclidean almost abelian

4.3. Euclidean manifolds and crystallographic groups

Does someone have the list of references here?

We have seen (4.2.7) that every discrete group of isometries of $\mathbf{E}^n$ contains an abelian subgroup A with finite index. There is a beautiful, more complete, classical theory primarily due to [Bie11, Bie12]. See also [Zas48].

There seems to be an argument for also including Frobenius and Burckhardt in this citation.-Dick

Theorem 4.3.1 (Bieberbach). (a) *A group G is isomorphic to a discrete group of isometries of $\mathbf{E}^m$ (for some m) iff G contains a free abelian subgroup with finite index.*

(b) *G is isomorphic to a discrete cocompact group of isometries of $\mathbf{E}^m$ iff G contains a subgroup A isomorphic to $\mathbf{Z}^m$ with finite index such that A is its own centralizer. In that case, A is also normal, and it is the unique maximal free abelian subgroup of finite index. A is the subgroup of all translations.*

(c) *If G_1 and G_2 are discrete cocompact groups of isometries of $\mathbf{E}^{m_1}$ and $\mathbf{E}^{m_2}$ which are isomorphic as groups, then $m_1 = m_2$ and there is an affine isomorphism $a : \mathbf{E}^{m_1} \to \mathbf{E}^{m_2}$ taking G_1 to G_2.*

(d) *For any given dimension m, there is only a finite collection of cocompact discrete groups of isometries of $\mathbf{E}^m$, up to affine equivalence.*

Cocompact discrete groups of isometries of $\mathbf{E}^m$ are called *crystallographic groups*, or sometimes *Bieberbach groups.*

Charlap attributes part b to Auslander and Kuranishi-Dick

Corollary 4.3.2 (classification of Euclidean manifolds). *Diffeomorphism classes of closed Euclidean m-manifolds are in one-to-one correspondence with torsion-free groups containing a subgroup with finite index isomorphic to $\mathbf{Z}^m$. The correspondence is*

$$M \longleftrightarrow \pi_1(M).$$

Proof of classification of Euclidean manifolds: (assuming 4.3.1) If G is torsion-free and contains a subgroup A_1 with finite index which is free abelian of rank m, then any maximal abelian subgroup S containing A_1 satisfies (b). Therefore, we need only verify that a discrete group G of Euclidean isometries acts freely if and only if G is torsion-free. Clearly any element which fixes a point must have finite order if G is discrete.

Conversely, if F is any finite subgroup of isometries, then F has at least one fixed point: for any point y in $\mathbf{E}^n$, the center of mass of its orbit Fy is fixed by F. classification of Euclidean manifolds

Proof of Bieberbach: The forward direction of part (a) follows from 4.2.7. There is a straightforward construction to prove the backward direction of part (a), which produces an action of G on a fairly high-dimensional Euclidean space. Let A be a free abelian subgroup of rank n and index p. Choose a faithful representation ρ of A as a discrete group of translations of $\mathbf{E}^n$. There is an associated action of G on p disjoint copies of $\mathbf{E}^n$, one for each coset γA. One way to define this action is to let

$$X = G \times \mathbf{E}^n/(ga, x) \sim (g, \rho(a)x)$$

Proposition "dimension n invariant subspace"
% abelian subgroups translate
% Bieberbach
% discrete Euclidean almost abelian
% abelian subgroups translate
% dimension n invariant subspace

for $a \in A$. Then X consists of one copy of $\mathbf{E}^n$ for each coset gA, and G acts on it naturally. (Some topologists may find it helpful to think of this as the $\mathbf{E}^n$ bundle over G/A associated to the principal bundle G with fiber A acting on the right). G therefore acts faithfully as a discrete group of isometries on the product of the components of X, which is $\mathbf{E}^{np}$. ($\mathbf{E}^{np}$ is the space of sections of the $\mathbf{E}^n$ bundle, above. This construction of an action of a big group from an action of a subgroup is known as the induced representation).

To prove parts (b) and (c), we need only the following fact.

Proposition 4.3.3 (dimension n invariant subspace). *Let G be any discrete group of isometries of $\mathbf{E}^m$. Let n be the rank of any normal free abelian subgroup of finite index. Then there is an $\mathbf{E}^n \subset \mathbf{E}^m$ invariant by G.*

Proof of dimension n invariant subspace: By problem 4.2.8, there is some unique maximal plane E_A invariant by any abelian subgroup A such that A acts by translations on E_A. There is a foliation of E_A by parallel n-dimensional planes, each invariant by A, where n is the rank of the image of A in $Isom(E_A)$. If A is already free abelian, n is the rank of A, because a free rank $n+1$ action by translations on $\mathbf{E}^n$ cannot be discrete. If A is normal in G, then G preserves the leaves of the foliation, so it acts on the space of leaves, which is another Euclidean space. But the action factors through G/A, a finite group, so there is a fixed point, i.e., a fixed leaf. [dimension n invariant subspace]

Continuation of proof of 4.3.1 To prove (b) in the forward direction consider any cocompact group G of isometries of $\mathbf{E}^m$. By 4.2.7, there is an abelian subgroup $B \subset G$ of finite index (hence still cocompact). By problem 4.2.8, B acts by translations on some plane E_B. Since B is cocompact, $E_B = \mathbf{E}^m$ and B contains m linearly independent translations. Let $A \subset G$ be the subgroup of G consisting of all translations. Then A is isomorphic to $\mathbf{Z}^m$. Any element γ which commutes with all elements of A must have a trivial rotational part, so it is a translation and therefore in A.

To prove (b) in the reverse direction, suppose G is a group and A a free abelian subgroup of finite index which is its own centralizer. Then, by part (a), G acts effectively as a discrete group on some Euclidean space $\mathbf{E}^n$. Since A has finite index, the intersection of all the finitely many subgroups of G conjugate to A is a normal subgroup B of finite index.

Applying 4.3.3, we find that there is an $\mathbf{E}^m \subset \mathbf{E}^n$ invariant by G on which B acts by translations, where m is the rank of B. Since $B \subset A$ and A is abelian, the rotational parts of elements $a \in A$ must act trivially on $\mathbf{E}^m$ so A also acts on $\mathbf{E}^m$ by translations. The action is effective since for $a \in A$, some power a^k is an element of B (non-trivial because A is free). Let g be an element in $G - A$, if g acts trivially on $\mathbf{E}^m$

Since A is its own centralizer, the rotational parts of elements $g \in G - A$ are non-trivial, so the action of G is effective.

wrong see page 7

The proof that elements of $G - A$ act effectively needs work-Dick

For part (c), if ρ_1 and ρ_2 are two faithful representations of a group G as discrete cocompact groups of isometries of $\mathbf{E}^{m_1}$ and $\mathbf{E}^{m_2}$, we know from (b)

% dimension n invariant subspace
% discrete Euclidean almost abelian
Theorem "finitely many automorphisms of abelian group"
% thin cubic lattice
% finitely many automorphisms of abelian group

that $m_1 = m_2 = rank(A)$. Consider the diagonal action (ρ_1, ρ_2) on the product $\mathbf{E}^{m_1} \times \mathbf{E}^{m_2}$. From 4.3.3, there is a copy of $\mathbf{E}^{m_1} \subset \mathbf{E}^{m_1} \times \mathbf{E}^{m_2}$ invariant by the action. This plane must project surjectively to each factor, since it is invariant, so it is the graph of an affine isomorphism from $\mathbf{E}^{m_1}$ to $\mathbf{E}^{m_2}$ conjugating ρ_1 to ρ_2.

Part (d) may seem plausible now, because from (b) and 4.2.7 we have an upper bound to the order of G/A (for G a cocompact group in any fixed dimension). The details, however, take a little care.

First, consider the representation of G/A in $aut(A) = \mathrm{GL}(n, \mathbf{Z})$.

Theorem 4.3.4 (finitely many automorphisms of abelian group). *There are only finitely many finite subgroups of* $\mathrm{GL}(n, \mathbf{Z})$*, up to conjugacy in* $\mathrm{GL}(n, \mathbf{Z})$.

Should homology be mentioned below, when a cocycle is considered? the fact that a finite group has a classifying space finite in each dimension? Other, more 'mathematical', approaches to this?

Proof of finitely many automorphisms of abelian group: Given a finite subgroup $F \subset \mathrm{GL}(n, \mathbf{Z})$, we will describe a procedure to find a basis for the integer lattice L so that elements of F are represented by matrices with entries whose absolute values are bounded by a constant C_n depending only on n.

First, choose a positive quadratic form on $\mathbf{R}^n$ which is invariant by F (by averaging the standard one over F, for instance). Adjust the scale so that the minimal length of a lattice point (other than 0) is 1. Let $V \subset \mathbf{R}^n$ be the vector subspace spanned by lattice elements of length 1. Adjust the scale in the orthogonal complement $V^\perp$ until the minimal length of a lattice point in $\mathbf{R}^n - V$ is 1. F still preserves the metric because V and $V^\perp$ are invariant under F. Continue until you obtain a quadratic form such that the minimum non-zero length of a lattice element is 1, and the lattice points of length 1 span $\mathbf{R}^n$ as a vector space.

There may not be any free basis for L consisting of elements of length 1. (See problem 4.3.7.) However, we can find a free basis for L consisting of elements with length $\leq (n+1)/2$. To do this, choose a set $a_1, \ldots, a_n$ of linearly independent lattice elements of length 1 . Suppose inductively that we have found our desired basis $b_1, \ldots, b_k$ for the intersection of the lattice with the subspace W_k spanned by $a_1, \ldots, a_k$, $k < n$. Let l be a lattice element in $W_{k+1} - W_k$ such that $l + W_k$ has minimum distance to W_k. This distance does not exceed 1 (since a_{k+1} has ~~that~~ distance ≤ 1).

Now add multiples of $a_1, \ldots, a_k$ to make the orthogonal projection of l to W_k reasonably small: The parallelepiped P spanned by $a_1, \ldots, a_k$ has diameter less than k, and with appropriate choice of $m_1, ..., m_k$, the projection of $l' = l + \sum_{i=1}^k m_i a_i$ can be placed in a copy of P centered at the origin, which is contained in the ball of radius $k/2$. Therefore, the element l' of $L \cap (l + W)$ has length not greater than $\sqrt{(k/2)^2 + 1} \leq ((k+2)/2)$. Adjoin l' to the basis, completing the inductive step.

To complete the proof of theorem 4.3.4, we will now show that matrices representing elements F with respect to our basis $b_1, ..., b_n$ have bounded entries. It suffices to show that the coordinates of any vector V of length

% Bieberbach
% SEMIDIR DEF
% finitely many automorphisms of abelian group
% finitely many automorphisms of abelian group
% semidirect crystallographic supergroup
Problem "estimating lattice automorphisms"
% finitely many automorphisms of abelian group

$\leq (n+1)/2$ are bounded. The i^{th} coordinate of V is the ratio of the volume of the parallelepiped P spanned by $b_1, ..., b_{i-1}, V, b_{i+1}, ..., b_n$ to the volume of the parallelepiped Q spanned by $b_1, ..., b_n$. The volume of P is at most $[(n+1)/2]^n$, while the volume of Q is greater than the volume of a ball in $\mathbf{E}^n$ of radius $1/2$ (since the injectivity radius of $\mathbf{R}^n/L$ is the constant $1/2$). Hence, the coordinates are bounded. finitely many automorphisms of abelian group

Continuation of the proof of 4.3.1(d). We have proved so far that (fixing the dimension n) there are only finitely many choices for G/A together with its action on A by conjugation. This gives us the behaviour of cosets of A, but not of individual elements within the cosets.

Enough data to determine the structure of G can be defined by choosing representatives $c(g) \in G$ for each coset $g \in G/A$, and for each pair g, h specifying the multiplication rule in G,

$$c(g) \cdot c(h) = c(gh) \cdot \alpha(g, h),$$

where $\alpha(g, h)$ is in A. The choice of $\alpha(g, h)$ along with the action of G/A on A gives the abstract isomorphism type of G, which by part (c) classifies G up to affine isomorphism. Note that if the $\alpha(g, h)$ are all trivial then G is the semi-direct product $A \ltimes G/A$ (see 3.12.10). Any choice of $\alpha(g, h)$ satisfying certain consistency conditions will produce a group, but many of them may be isomorphic. To find a bound for the number of groups, we need only show that, given a group G, we can make appropriate choices of $c(g)$ and $c(h)$ that will bound the elements $\alpha(g, h)$. Use the quadratic form used for the proof of 4.3.4 to impart a new Euclidean structure to $\mathbf{R}^n$ which is invariant by G. Choose any point $x_0 \in \mathbf{E}^n$, and for each element $g \in G/A$ choose $c(g)$ to be a representative of the coset g which moves x_0 a minimal distance.

Kurosh was given as the reference. Is this best? What references to extensions of groups?

Let $a_1, \ldots, a_n$ be a set of linearly independent lattice elements of length 1, as in the proof of theorem 4.3.4, and let P be the parallelogram they span. Then P is a fundamental domain for the subgroup generated by $a_1, \ldots, a_n$ and has diameter less than n. Therefore one can always alter $c(g)$ by composing with elements of this subgroup to assure that $c(g)$ moves x_0 a distance less than n. Then $\alpha(g, h) = c(gh)^{-1}c(g)c(h)$ moves x_0 a distance less than $3n$, so the number of choices is bounded. Bieberbach

Another way to show that there are only a finite number of groups G with a given action of G/A on A is to show they are subgroups with bounded index in the semi-direct product $A \ltimes G/A$. See exercise 4.3.8.

Problem 4.3.5 (estimating lattice automorphisms). (a) Any finite subgroup F of $SL(n, \mathbf{Z})$ injects into $SL(n, \mathbf{C}_m)$, for $m \geq 3$. [Hint: In the metric used for the proof of 4.3.4, the unit ball in L injects into $\mathbf{C}_m^n$, for $m \geq 3$.]

(b) The kernel of $F \to SL(n, \mathbf{C}_2)$ has order a power of 2.

(c) The order of $SL(n, \mathbf{C}_p)$ is $p^{-1}(p^n - 1)(p^n - p) \cdots (p^n - p^{n-1})$, for p a prime. By considering the *gcd* of the order of $SL(n, \mathbf{C}_p)$, what restrictions can you find for

Problem "isometric lattices"
Problem "thin cubic lattice"
Exercise "semidirect crystallographic supergroup"
% Bieberbach

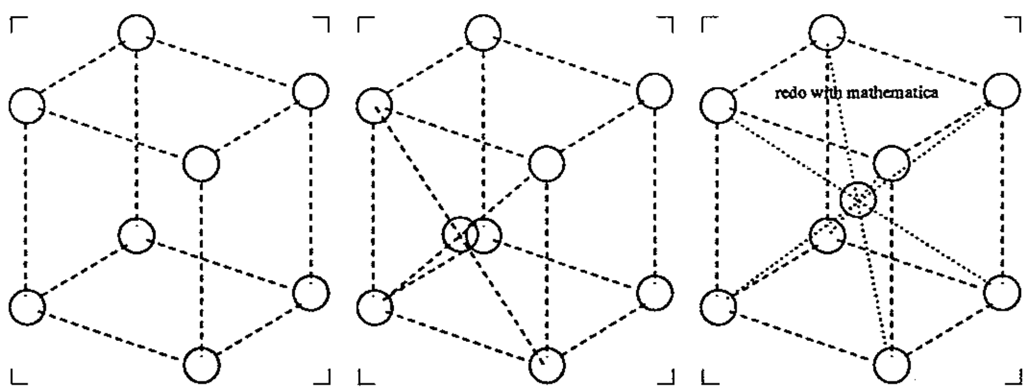
finite subgroups of SL(n, Z), for small n? Do you get any more information by considering $\mathrm{SL}(n, \mathbf{C}_m)$ for general m?

Why has the question changed to one about GL(n, Z) in part (c) of 4.3.5?-Dick

Problem 4.3.6 (isometric lattices). Show that the group of symmetries of an octahedron can be embedded in $\mathrm{GL}(3,\mathbf{Z})$ in exactly three different ways, conjugate in $\mathrm{GL}(3,\mathbf{R})$ to the standard action but not conjugate in $\mathrm{GL}(3,\mathbf{Z})$.

[Hint: to construct lattices invariant by the octahedral group, begin with the ordinary cubic lattice and adjoin halves of certain lattice elements. To show that these lattices are the only types, use the observation that if L is a lattice invariant by a finite subgroup $F \subset 0(n)$, then for any element V of L with minimal length and any $f \in F$, the angle between V and $f(V)$ must be at least 60^o if $f(V) \neq V$.]

Why are the three actions not conjugate in $\mathrm{GL}(n,\mathbf{Z})$?



isometric

Figure 4.4. the three isometric lattices. The three lattices which have the largest possible symmetry, that of the octahedron. In crystallography, these are called generically the *isometric lattices*, and specifically the *cubic lattice*, the *face-centered cubic lattice*, and the *body-centered cubic lattice*.

Problem 4.3.7 (thin cubic lattice). In dimension $n \geq 5$, the standard cubic lattice $\mathbf{Z}^n \subset \mathbf{R}^n$ is contained in a lattice which is not generated by its intersection with the unit ball.

Exercise 4.3.8 (semidirect crystallographic supergroup). Prove that every crystallographic group G whose maximal abelian subgroup of finite index is A is a subgroup of the semi-direct product $A \ltimes G/A$, with bounded finite index.

Let n be the order of G/A.

(a) G acts as a group of affine maps of A^n, using the induced representation described in 4.3.1.

(b) Let $B \subset A^n$ be the orbit of the action of A which contains the origin $(1,\ldots,1)$. There is an induced affine action of G on A^n/B, which factors through G/A.

(c) Embed A^n in a larger free abelian group C of the same rank by adjoining n^{th} roots to all elements. G acts on C. Let D be the maximal subgroup containing B which has the same rank as B. Then the action of G on C/D has a fixed point, so there is a coset E of D which is invariant by G.

(d) The action of G on E is effective. The group $G \cdot E$ consisting of transformations in G followed by translations in E is isomorphic to the semi-direct product $A \ltimes G/A$.

(e) The above shows that G with n^{th} roots adjoined for all translations in G is a semi-direct product. The index is n^m, where m is the rank of A. Show that the minimal index with which G can be embedded in a semi-direct product is at most n. (Note: the action of G/A on the lattice, in this minimal embedding, might not be isomorphic to the action of G/A on A).

Section "Three-dimensional Euclidean manifolds"
EUC 3-MFDS
% maximal subgroups of SL(3, Z)
% Bravais lattices
Lemma "Euclidean splitting"

4.4. Three-dimensional Euclidean manifolds

The geometric theory of 3-dimensional crystallography was mainly developed during the course of the 19th century.

There are 32 subgroups of $O(3)$ that occur as subgroups of a crystallographic group which stabilize a point in $\mathbf{E}^3$. They are known in crystallography as the point groups, and were first enumerated by [Hes30].

There are 14 distinct crystallographic groups which are the full groups of symmetries of a lattice (as an array of points) in $\mathbf{R}^3$, up to affine equivalence. These were enumerated by [Bra49], and the classes of lattices, according to their symmetry, are known as Bravais lattices. (See problem 4.4.13 and challenge 4.4.14.)

There are 65 orientation-preserving crystallographic groups, which were classified by [Soh79]. Finally, three people independently derived the full list of 230 crystallographic groups [vF91], [Sch91], and [Bar94],

From a list of crystallographic groups, one can of course derive a list of Euclidean three-manifolds by crossing out the groups which do not act freely ([Now34]). This involves much more work than necessary, though, and [HW35] gave a direct classification of Euclidean three-manifolds.

With such a large number of crystallographic groups, is there a conceptual way we can sort out the fundamental groups of three-manifolds?

Suppose that Γ is the fundamental group of a closed Euclidean 3-manifold.

The first step is to understand better how the finite subgroup $F \subset O(3)$ which consists of the derivatives (or rotational parts) of elements of Γ relates to the lattice of translations in Γ. For each subgroup $H \subset F$, there is a subgroup Γ_H consisting of those elements of Γ whose derivative lies in H. Of course, $\Gamma_{\{1\}}$ is $\mathbf{Z}^3$, and its quotient is T^3.

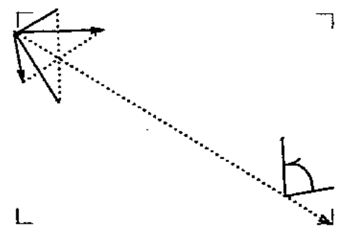
The next cases are when H is a cyclic group C_k, generated by the derivative $d\alpha \in O(3)$ of some element $\alpha \subset \Gamma$.

Since α has no fixed points, $d\alpha$ must have 1 as an eigenvalue. Either the 1-eigenspace is 1-dimensional, and $d\alpha$ is a rotation of order k about a line, or it is 2-dimensional and $d\alpha$ is reflection through a plane. In either case, it determines an orthogonal splitting of $\mathbf{R}^3$ as a two-dimensional space V plus a one-dimensional space W. A first observation is that this splitting is commensurate with the lattice $\Gamma_{\{1\}}$ of translations in Γ:

Lemma 4.4.1 (Euclidean splitting). *If the group Γ_{C_k} preserves orientation, then the group of translations $\Gamma_{\{1\}}$ is generated by its intersection with V and its intersection with W.*

If Γ_{C_k} contains orientation-reversing elements, then $k = 2$, $d\alpha$ is a reflection, and the subgroup of $\Gamma_{\{1\}}$ generated by its intersections with V and W has index at most 2.

Proof of 4.4.1: Since the group Γ_{C_k} preserves the factorization of $\mathbf{E}^3$ into



euclidsplits

Figure 4.5. Euclidean manifolds are almost products. When a cocompact Euclidean group Γ has an element α whose derivative is a rotation about some axis, then for any translation $T \in \Gamma$, the composition of conjugates of T by powers of α translates along the axis of α.

$\mathbf{E}^2 \times \mathbf{E}^1$, it is contained in the product:

$$\Gamma_{C_k} \subset \mathrm{isom}(\mathbf{E}^2) \times \mathrm{isom}(\mathbf{E}^1).$$

First we'll do the case that $d\alpha$ is a rotation, so α is a screw motion in the direction of $\mathbf{E}^1$. Consider an arbitrary element $\gamma \in \Gamma_{C_k}$, and its projection $p(\gamma)$ to $\mathrm{isom}(\mathbf{E}^1)$. We claim that $p(\gamma)^k$ is in the image of Γ_{C_k}. First, if γ is not a translation, then it is a screw motion in the direction of $\mathbf{E}^1$. In that case, its kth power is a translation, necessarily in the same direction. Otherwise, γ is a translation. It is an easy exercise to check that in this case the product of the conjugates of γ by all powers of α, i.e. $\gamma \cdot (\alpha\gamma\alpha^{-1}) \cdot \ldots \cdot (\alpha^{k-1}\gamma\alpha^{1-k})$, lies in Γ_{C_k}. Hence, $p(\gamma)^k$ lies in the image of Γ_{C_k}.

It follows that the image of Γ_{C_k} in $\mathrm{isom}(\mathbf{E}^1)$ is discrete, since it has index at most k in the discrete subgroup $\Gamma_{C_k} \cap \mathrm{isom}(\mathbf{E}^1)$.

The kernel of the homomorphism p from Γ_{C_k} to $\mathrm{isom}(\mathbf{E}^1)$ consists entirely of translations, for otherwise there would be elements with fixed points. Therefore, the translations of Γ_{C_k} along $\mathbf{E}^2$ together with those along $\mathbf{E}^1$ have index at most k in all of Γ_{C_k}. Since the group of all translations also has index k, these two subgroups must be identical. The first assertion of the lemma is established.

Now we take up the case that Γ_{C_k} contains orientation-reversing elements. Then $k = 2$, and the elements which reverse orientation must have derivative equal to reflection in $\mathbf{E}^2$. In this case, consider the projection $q : \Gamma_{C_k} \to \mathrm{isom}(\mathbf{E}^2)$. The image consists of translations, by hypothesis. We claim that for any element $\gamma \in \Gamma_{C_2}$, the square of its image $q(\gamma)^2$ is in the intersection $\Gamma_{C_2} \cap \mathrm{isom}(\mathbf{E}^2)$. The reasoning is exactly the same as before: if the derivative of γ is not trivial, then γ^2 is in the intersection. Otherwise, γ is a translation, and the composition of γ with its conjugate by an orientation-reversing element lies in the intersection.

We conclude that the image $q(\Gamma_{C_2})$ has index at most 4 in the intersection $\Gamma_{C_2} \cap \mathrm{isom}(\mathbf{E}^2)$. (Both groups are isomorphic to $\mathbf{Z}^2$; the image of the map $g \to g^2$ of $\mathbf{Z}^2$ to itself has index 4.)

Problem "splitting crystallographic groups"
Theorem "Euclidean covered by torus cross E"
% other product coverings
% Euclidean splitting
% axis of composition

The kernel of q consists entirely of translations, for if any element of the kernel had a non-trivial derivative it would have a fixed point. We can conclude that the subgroup of Γ_{C_2} generated by translations along $\mathbf{E}^2$ together with translations along $\mathbf{E}^1$ has index at most 4 in Γ_{C_2}, so it has index 1 or 2 in its subgroup $\Gamma_{\{1\}}$ of translations. 4.4.1

Problem 4.4.2 (splitting crystallographic groups). Is the conclusion of the lemma false for a general crystallographic group in $\mathbf{E}^3$? If so, how would the conclusion need to be weakened to make it valid in this more general context?

The stepping stone to an understanding of the classification of Euclidean three-manifolds is the following:

Theorem 4.4.3 (Euclidean covered by torus cross E). *Every closed Euclidean three-manifold M can be expressed as a quotient of $T^2 \times \mathbf{E}^1$ by the action of a discrete group G, where T^2 is a Euclidean torus.*

From this theorem, there is a fairly routine procedure for recovering the list of Euclidean three-manifolds. The discrete group G is a subgroup of $\operatorname{isom}(T^2) \times \operatorname{isom}(\mathbf{E}^1)$. If G does not project faithfully to $\operatorname{isom}(\mathbf{E}^1)$, we can factor $T^2 \times \mathbf{E}^1$ by the kernel to re-express M as a quotient either of $T^2 \times \mathbf{E}^1$ or $K^2 \times \mathbf{E}^1$, where K^2 is a Klein bottle, by a group which is represented faithfully by its projection to $\operatorname{isom}(\mathbf{E}^1)$.

The group of deck transformations is $\mathbf{Z}$ or $\mathbf{Z}_2 * \mathbf{Z}_2$ (the only discrete cocompact subgroups of $\operatorname{isom}(\mathbf{E}^1)$), and it is not hard to list all possible cases, and then check for redundancy. (Compare problem 4.4.11)

First we will prove the theorem, as this will give us additional useful information.

Proof of 4.4.3: Let Γ be the group of deck transformations of $\mathbf{E}^3$ over M; let $F \subset O(3)$ be the group of rotational parts of elements of Γ and let A be the group of translations in Γ.

If $F = \{1\}$, Γ is a group of translations and the result is trivial.

If F is a cyclic group generated by some element α, then from lemma 4.4.1, there is a splitting $\mathbf{E}^3 = \mathbf{E}^2 \times \mathbf{E}^1$ such that the intersection H of $\Gamma_{\{1\}}$ with $\operatorname{isom}(\mathbf{E}^2)$ is isomorphic to $\mathbf{Z}^2$. This subgroup is normal, and the quotient $\mathbf{E}^3/H$, acted on by Γ/H, fulfills the requirements for the theorem.

In the remaining cases, F has order greater than 2, so its orientation-preserving subgroup F_0 is nontrivial. For each nontrivial element $\alpha \in F_0$, there is a splitting $\mathbf{E}^3 = \mathbf{E}^2 \times \mathbf{E}^1$, with $\mathbf{E}^1$ parallel to the axis of α, such that the subgroup $\Gamma_{\{1\}}$ of translations is generated by its elements that translate along one of the two factors.

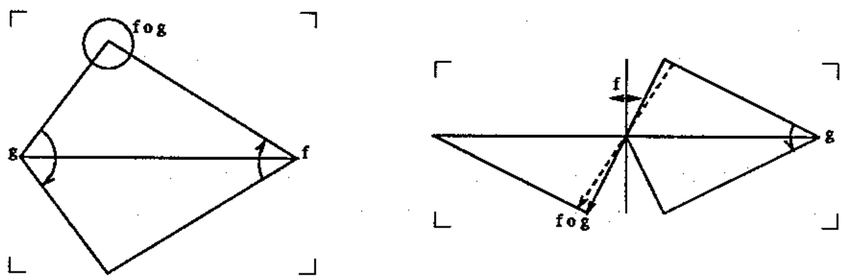
If there are elements $\alpha, \beta \in F_0$ whose axes do not coincide, then the axes must be orthogonal, since the first nonzero lattice point on either must project to a lattice point on the other. Then $\alpha^2 = \beta^2 = 1$, otherwise the axis of $\alpha\beta$ is not orthogonal to the axes of α and of β. (See problem 4.4.4.) It follows that α takes the axis of β to itself, so that it commutes with β, and the group F_0 is

% axis of composition
Problem "axis of composition"
% composition

$C_2{}^2$. If M is orientable, we are done: each of the three splittings, determined by the axis of α, of β, or of $\alpha\beta$, is invariant by Γ. Take H to be the group generated by translations in the plane perpendicular to the axis of α, say, and $\mathbf{E}^3/H$ satisfies the requirements.

What remains is the case that F is not orientation-preserving and not cyclic. Then the fixed plane of any reflection must contain every axis, since the composition of a reflection with a rotation whose axis is not contained in the plane of reflection does not have 1 as an eigenvalue, hence an isometry with this derivative has a fixed point. (Problem 4.4.4.) It follows that there can be only one such axis, since the composition of rotations about two different axes has an axis not in the same plane. The splitting defined by this axis enables us to finish the construction. 4.4.3

Problem 4.4.4 (axis of composition). (a) Let f and g be rotations of $\mathbf{E}^2, \mathbf{H}^2$ or the elliptic plane $\mathbf{RP}^2$, with fixed points x and y. Give a description of $f \circ g$, including a construction for its fixed point (which may be at ∞ or beyond in the projective embeddings of $\mathbf{E}^2$ or $\mathbf{H}^2$). [Hint: see figure 4.6]



composition

Figure 4.6. axis of composition. This diagram gives hints for problem 4.4.4. The diagram on the left indicates a construction for the composition of two rotations, and on the right for the composition of a rotation and a reflection.

(b) Verify that an orientation-preserving group of isometries of S^2 has all elliptic axes mutually orthogonal if and only if it is a cyclic group or $\mathbf{C}_2 \oplus \mathbf{C}_2$ acting by 180° rotations about three mutually orthogonal transformations.

(c) Describe the composition of the reflection in a line in $\mathbf{E}^2$ with a rotation about a point not on that line, and show that it has no fixed points. Generalize this to S^2 and to $\mathbf{E}^2$.

(d) Describe the composition of reflection in a plane in $\mathbf{E}^3$ with a rotation about an axis intersecting the plane in a point.

(e) Generalize the construction of Part a) to the case that f and g are arbitrary orientation-preserving isometries of $\mathbf{E}^3, \mathbf{H}^3$ or $\mathbf{P}^3$.

Now we can enumerate the Euclidean three-manifolds. We will follow the notation of [Wol67], denoting the six orientable manifolds $G_1, ..., G_6$, and the four non-orientable manifolds $B_1, ..., B_4$. Three of the orientable manifolds,

% Euclidean covered by torus cross E
% Euclidean splitting
% finite subgroups of SL(2, Z)
% sqtorus
% hextorus
% axis of composition
% Euclidean manifolds
% Euclidean covered by torus cross E
% Euclidean splitting
Section "F Z2"
% Euclidean splitting

G_3, G_4, and G_5, have two different oriented forms, distinguished by the handedness of the screw motion with shortest translation distance. The orientations with left-handed screws can be denoted G', etc.

In order to avoid repetitions, we can enumerate the manifolds in order of the group F of rotational parts of their deck transformations.

Case 4.4.5 The group F is trivial.

The manifold G_1 is the three-torus T^3.

Bill says that he's going to change the medthod of exposition in the case by case analysis-Dick

Has mapping torus been defined somewhere?

Case 4.4.6 The group F is cyclic and preserves orientation.

From propositions 4.4.3 and 4.4.1, it follows that the manifold is a mapping torus M_ϕ, where ϕ is a linear transformation in $SL(2, \mathbf{Z})$ which acts as an isometry of some Euclidean torus which rotates by angle $l\pi/k$, with $(l, k) = 1$.

By challenge 4.4.10, every non-trivial finite subgroup of $SL(2, \mathbf{Z})$ is a subgroup of $\mathbf{C}_4$ (acting by rotations of the square torus, 1.2) or $\mathbf{C}_6$ (acting by rotations of the hexagonal torus, 1.3). Then $k = 2, 3, 4$ or 6, and the manifolds are these:

F	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$	$\mathbf{C}_6$
Angle of rotation	180°	120°, 240°	90°, 270°	60°, 300°
Type of two-torus	arbitrary	hexagonal	square	hexagonal
Name of three-manifold	G_2	G_3, G_3'	G_4, G_4'	G_5, G_5'

Case 4.4.7 The group F preserves orientation, but it is not cyclic.

Then, (from problem 4.4.4(b)), F is $\mathbf{C}_2 \times \mathbf{C}_2$ acting as rotations about three orthogonal axes. We have seen one such manifold, G_6. (See example 3.4.1). To prove that this is the only possible example, we use 4.4.3 to conclude that our manifold M must be of the form $T^2 \times \mathbf{E}^1/(\mathbf{C}_2 * \mathbf{C}_2)$, where the rotational part of the action of $\mathbf{C}_2 * \mathbf{C}_2$ on T^2 is generated by reflections through orthogonal axes. The two actions of $\mathbf{C}_2$ on T^2 must have no fixed points, and an application of proposition 4.4.1 (in dimension 2 instead of 3) shows that the torus must be a rectangular torus, and the reflections must be parallel to the sides. This proves there is only one such manifold.

Case 4.4.8 The group F is $\mathbf{C}_2$, acting by reflection through a plane.

We may assume the fixed plane of the reflection is the x-y plane. Every element of the group G which reverses orientation leaves invariant some plane $z =$ constant, so we may normalize so the x-y plane is such a plane, for some $g \in G$. Let $A \subset G$ be the translations in G, and $A_0 \subset A$ be those translations which leave the x-y plane invariant. By Proposition 4.4.1, A_0 has either index 1 or index 2 in the image of A projected orthogonally to the x-y plane. We can again normalize so that an infinite cyclic subgroup of A_0 preserves the x-axis, and that a minimal translation along the x-axis has an orientation-reversing square-root, which translates half the distance and reflects through the x-y plane.

We can think of the manifold in two ways: as $T^2 \times \mathbf{E}^1$ modulo the action of $\mathbf{C}_2 * \mathbf{C}_2$ (acting in the z direction), or as a Klein bottle $K^2 \times \mathbf{E}^1$ modulo

% F Z2
% F Z2
Section "F¿2"
% Euclidean splitting

an action of $\mathbf{Z}$, where the fundamental group of the Klein bottle is the group that stabilizes the x-z plane. It is clear that this is a normal subgroup.

Case 4.4.8 (a): The projected image of A is A_0. This is equivalent to the translational part of any element with a non-trivial rotational component lying in the x-z plane, because otherwise by composition one can obtain a pure translation whose projection does not lie in A_0.

In this case, the manifold is B_1, the product of the Klein bottle with S^1. The alternative description is $T^2 \times \mathbf{E}^1$ modulo $\mathbf{C}_2 * \mathbf{C}_2$, with each of the $(\mathbf{C}_2)$'s acting in precisely the same way as translations of the torus.

Case 4.4.8 (b): The projected image of A contains A_0 with index 2. Thus A contains a pure translation with x-y component not contained in A_0, and whose z component is half the minimal translation fixing the z-axis. The manifold B_2 is $T^2 \times \mathbf{E}^1/(\mathbf{C}_2 * \mathbf{C}_2)$ where the two $(\mathbf{C}_2)'s$ are acting on T^2 in two different ways as order-two translations. The alternative description of B_2 is the mapping torus of an isometry $\phi : K^2 \to K^2$, where ϕ interchanges the two parallel simple geodesics on K^2 which have unoriented neighborhoods (diffeomorphic to Moebius bands). ϕ is the projection of the translation mentioned above.

Case 4.4.9 The group F contains a reflection, and has order greater than 2.

We will see that F must then be $\mathbf{C}_2 \oplus \mathbf{C}_2$, generated by reflections in two orthogonal planes. First, recall that any rotation axis in F must pass through any fixed plane of any reflection in F. The composition of any two rotations with distinct axes has an axis which does not lie in the plane spanned by them. Therefore, there is a unique axis for rotations in F which we can arrange to be the z-axis.

G acts on the z-coordinate by translations. Let t be the minimal non-zero translation in the z direction of a pure translation in G. Then there is a homomorphism from F to $\mathbf{R}/\langle t\rangle$ taking $\gamma \in F$ to the z translational part (mod t) of any element of G whose rotational part is γ. Let $G_0 \subset G$ be the subgroup which acts trivially on z. If G_0 consists entirely of translations, this map is injective, so F is cyclic. This contradicts our assumption, so G_0 is the fundamental group of the Klein bottle.

By proposition 4.4.1, applied once to G_0 and once to G, we can normalize so the group $A \subset G$ of translations is generated by one translation along each of the three coordinate axes, and G_0 is generated by $A_0 = A \cap G_0$ together with a translation along the x-axis composed with reflection in the x-z plane. The group F must normalize G_0, which proves that F is $\mathbf{C}_2 \oplus \mathbf{C}_2$ generated by reflection in the x-z and y-z planes.

Our manifold is the mapping torus of an isometry ϕ of a Euclidean Klein bottle to itself. The gluing map ϕ has various lifts to an isometry of the Euclidean plane which normalizes G_0; some lifts are reflections composed with a translation while others are 180° rotations composed with a translation. There are two cases for ϕ, depending on whether the lifts have a translational

% F¿2
% F¿2
Challenge "finite subgroups of SL(2, Z)"
% finitely many automorphisms of abelian group
Problem "other product coverings"
Problem "abelian structure"
% Euclidean splitting
Problem "maximal subgroups of SL(3, Z)"
% abelian structure

component in the y direction which is not a translational component of an element of G_0.

Case 4.4.9(a): The manifold B_3 is obtained when the gluing map ϕ preserves each of the two simple geodesics with non-oriented neighborhoods.

Case 4.4.9(b): The manifold B_4 is obtained when ϕ interchanges the two simple geodesics with non-oriented neighborhoods.

Someone should check these tables and problems. I have only done it cursorily. -Bill

Challenge 4.4.10 (finite subgroups of SL(2, Z)). (a) Every finite subgroup of SL(2, Z) is a conjugate to a subgroup contained in $\mathbf{C}_6$ acting by rotations of the equilateral triangular lattice, or $\mathbf{C}_4$ acting as rotations of the square lattice. [Hint: see the proof of 4.3.4]

(b) * SL(2, Z) is isomorphic to $\mathbf{C}_4 *_{\mathbf{C}_2} \mathbf{C}_6$. [Hint: Let SL(2, Z) act on the upper half-plane in the standard way, as fractional linear transformations. Find a fundamental domain for the action.]

(c) Show that the Teichmüller space of Euclidean structures on the torus of area 1, $\mathcal{T}(T^2)$, is parametrized by the upper half-plane, and the action of SL(2, Z) on it as automorphisms of $\pi_1(T^2)$ corresponds to the standard action used in part b) above.

Problem 4.4.11 (other product coverings). In what other ways can the ten Euclidean three-manifolds be described as $M^2 \times \mathbf{E}^1/G$ where M^2 is a Euclidean two-manifold and G is $\mathbf{C}_2 * \mathbf{C}_2$?

Problem 4.4.12 (abelian structure). Modify proposition 4.4.1, so it applies to arbitrary cocompact groups acting in Euclidean n-space. (The conclusion will be vacuous unless some non-trivial element of F has a non-trivial subspace.)

Problem 4.4.13 (maximal subgroups of SL(3, Z)). Show that there are four maximal finite subgroups F of SL(3, Z) up to conjugacy, by finding the maximally symmetric lattices A as follows:

(a) If A is invariant by a rotation of order m about an axis V then from problem 4.4.12, $(A \cap V) \oplus (A \cap V^{\perp}) \subset mA \subset A$, so $m = 2, 3, 4$ or 6.

(b) The only lattices invariant by a cyclic group of rotations of order 6 are the products of a triangular lattice in $\mathbf{E}^2$ with a lattice in $\mathbf{E}^1$. Any such lattice has maximal symmetry D_{12}. [Hint: consider (a) for ϕ^2 and ϕ^3, where ϕ generates the group.]

(c) If F contains a cyclic group of order 3, then either it is subsumed under b) or the orthogonal image A_1 of A in $V^{\perp}$ contains $A \cap V^{\perp}$ with index 3. The three cosets must look like this:
Such a lattice automatically has symmetry D_3, generated by a 180° revolution about two lines. This is not a maximal finite subgroup, since by adjusting the scale in the V direction one can obtain the cubic, the face-centered cubic or the body-centered cubic lattice.

(d) If F contains a group of rotations of order 4, then A is either $(A \cap V) \oplus (A \cap V^{\perp})$ or contains this with index 2. In the latter case, A is obtained from the body-centered cubic lattice. By adjusting the vertical scale, one can obtain either the face-centered cubic or the body-centered cubic lattice.

Challenge "Bravais lattices"
Problem "two-dimensional groups, three-dimensional manifolds"
% Orbifolds and Seifert fiber spaces
% Fiber bundles
Challenge "intransigent groups"

blankfig

blankfig

Figure 4.7. blankfig

(e) $\mathbf{C}_2$ has two embeddings in $SL(3, \mathbf{Z})$ up to conjugacy, subsumed in the embeddings of $\mathbf{C}_4$.

(f) If F does not contain a normal cyclic subgroup, then it contains the group of symmetries of a tetrahedron (which can be inscribed in a cube). There are three orthogonal order 2 axes corresponding to the three common perpendiculars to opposite edges of a tetrahedron, so A is either a cubic, a face-centered cubic or a body-centered cubic lattice.

(g) Since $-I$ is central in $GL(3, \mathbf{Z})$, maximal finite subgroups of $GL(3, \mathbf{Z})$ are in one-to-one correspondence with maximal finite-subgroups of $SL(3, \mathbf{Z})$.

Challenge 4.4.14 (Bravais lattices). Enumerate the fourteen Bravais lattice classes. Here is a table which can serve as a guide:

Name for group of Bravais lattices	Subgroups of $O(3)$ normalizing the lattice	Number of lattice classes in group
Triclinic	1	1
Monoclinic	$\mathbf{C}_2$	2
Orthorhombic	$\mathbf{C}_2 \oplus \mathbf{C}_2$	4
Tetragonal	D_8	2
Hexagonal	D_6 or D_{12}	2
Isometric	Octahedral group	3

Problem 4.4.15. Show that there are two crystallographic groups G_1 and G_2 (not uniquely determined) in dimension 3 such that every other three-dimensional crystallographic group is isomorphic to a subgroup of G_1 or G_2.

Should we give them a hint to look at 4.4.13 and 4.3.8 or does this make it too easy?-Dick

Problem 4.4.16 (two-dimensional groups, three-dimensional manifolds). Show that every closed orientable three-manifold arises as the tangent circle bundle of a two-dimensional Euclidean orbifold See Chapter 5, especially section 5.7, for the appropriate definitions. Which Euclidean 3-manifold occurs twice?

Challenge 4.4.17 (intransigent groups). Show that there are interesting discrete subgroups of $\text{isom}(\mathbf{E}^2) \times \text{isom}(\mathbf{E}^2)$ whose projections to the factors are not discrete.

(a) Consider the group G with presentation $< \alpha, \beta, \gamma \mid \alpha^2 = \beta^2 = \gamma^2 = (\alpha\beta)^5 = (\beta\gamma)^5 = (\gamma\alpha)^5 = 1 >$. The groups generated by reflections in the sides of either of the two golden isosceles triangles, with angles $\pi/5, 2\pi/5$ and $2\pi/5$ or $3\pi/5, \pi/5$ and $\pi/5$, are homomorphic images of G. If ρ_1 and ρ_2 are the two

% intransigent groups

homomorphisms, show that the image of the diagonal homomorphism $\rho_1 \times \rho_2 : G \to \mathrm{isom}(\mathbf{E}^2) \times \mathrm{isom}(\mathbf{E}^2)$ is a crystallographic group in $\mathbf{E}^4$, yet the images of ρ_1 and ρ_2 are not discrete. [Hint: pass to a subgroup of index 2 of G, and consider how $\rho_1 \times \rho_2$ acts on the ring $\mathbf{Z}$ [5th roots of unity], which maps into $\mathbf{C}^2$ as a lattice.]

(b) Show that $\rho_1 \times \rho_2$ is not faithful by comparing it to an action ρ on the hyperbolic plane. Construct a "picture" of the crystallographic group of (a) by showing that $\mathbf{H}^2/\rho(G_0)$ is a surface M of genus 2, where G_0 is the kernel of a map from G to D_5. Then G acts on the maximal abelian cover $\widetilde{M}$, and its image there has a subgroup of index 10 isomorphic to $\mathbf{Z}^4$, the group of deck transformations of $\widetilde{M}$ over M. Map $\widetilde{M}$ equivariantly into $\mathbf{E}^4$ (with polyhedral image), using maps of its triangles to copies of the two golden triangles. Can you decide whether this is an embedding?

(c) Generalize this construction for any prime p to produce crystallographic groups in $\mathbf{E}^{p-1}$ which are generated by three elements of order 2 and contain elements of order p. Show that this is the minimum possible dimension for a crystallographic group having elements of order p.

Problem 4.4.18. Translate challenge 4.4.17 into a complex analytic form. In other words, make use of the conformal equivalence between any two triangles rather than the affine equivalence. The surface M of genus 2 is a 5-fold regular branched covering of the Riemann sphere $\mathbf{CP}^1$, so it can be put in the form $\{(x, y) \in \mathbf{CP}^1 \times \mathbf{CP}^1 : x^2 + y^5 = 1\}$. The two maps of the universal abelian cover $\tilde{M}$ of M to $\mathbf{C}$ have differentials which can be taken as $y^{-3}dx$ and $y^{-4}dx$. (Derive this from the qualitative form of the branching.) What happens when 5 is replaced by an arbitrary prime p?

Section "Elliptic three-manifolds"
% The geometry of the three-sphere
% free order two elements
Exercise "free order two elements"

4.5. Elliptic three-manifolds

Bill has not finished looking at this section yet-Dick

The classification of elliptic three-manifolds was given in outline by [Hop26] and in detail by [ST30, ST32] The analysis is made much easier by the fact that the group of isometries of S^3 is almost a product. (See section 2.7, the geometry of the three-sphere).

We understand manifolds in this section to be closed.

Proposition 4.5.1. *Every elliptic three-manifold is orientable.*

Proof of 4.5.1: An orientation-reversing element of $O(4)$ must have an odd dimensional (-1)-eigenspace, since it preserves orientation on its $(+1)$-eigenspaces and its two-dimensional irreducible subspaces. Therefore, its $(+1)$-eigenspace is non-empty, so it has a fixed point. $\boxed{4.5.1}$

Remark. Readers more familiar with topology will recognize the topological explanation for this fact: one of the first corollaries of the Lefschetz fixed-point theorem is that every orientation-reversing continuous map of an odd-dimensional sphere to itself has a fixed point.

Proposition 4.5.2. *If the holonomy group of an elliptic three-manifold M does not contain $-I$, then $-I$ acts on M as a covering transformation (i.e., without fixed points).*

Proof of 4.5.2: For $-I$ to have a fixed point in its action on M means that some element $\gamma \in \pi_1(M)$ takes some point $x \in S^3$ to its antipodal point $-x$. Then $\gamma^2(x) = x$, so $\gamma^2 = I$, and since γ acts without fixed points, $\gamma = -I$ (see exercise 4.5.3). $\boxed{4.5.2}$

Exercise 4.5.3 (free order two elements). Show that

(a) the only order two element of $O(3)$ which acts freely on S^2 is the antipodal map $-I$, and

(b) the only order two element of $O(n)$ which acts freely on S^n is the antipodal map $-I$.

In view of these two propositions, the classification of elliptic three-manifolds can be broken into two stages.

First, we study three-manifolds whose holonomy groups contain $-I$. These are the quotients of elliptic three-space, $\mathbf{RP}^3 = \mathrm{SO}(3)$, by a group of orientation-preserving isometries. The advantage is that the orientation-preserving isometry group of $\mathbf{RP}^3$ is a product, $\mathrm{SO}(3) \times \mathrm{SO}(3)$, rather than almost a product as for S^3.

The full list can then be obtained by adjoining certain double covers of manifolds in the first list, those double covers whose holonomy does not contain $-I$. This second step is easy:

Proposition 4.5.4. *A three-manifold $\mathbf{RP}^3/\Gamma$ has a double cover by a three-manifold whose holonomy does not contain $-I$ if and only if Γ has odd order. Such a double covering is unique.*

% Frobenius map Proposition "diagonal groups are fiber product"

Proof of 4.5.4: If Γ has even order, it contains an element γ of order 2. If $\widetilde{\gamma}$ is any isometry of S^3 covering γ, then $\widetilde{\gamma}^2 = -I$ (since if $\widetilde{\gamma}^2 = I$ then $\widetilde{\gamma} = -I$ so γ itself would have to be trivial. In that case, M does not have a covering of the desired form.

If Γ has odd order, consider the holonomy group $\widetilde{\Gamma}$ for M, which has a two-to-one homomorphism to Γ. Think of left-multiplication by an element in $\widetilde{\Gamma}$ as a permutation of $\widetilde{\Gamma}$. Multiplication by $-I$ is an odd permutation. Therefore the subgroup $\Gamma_0 \subset \widetilde{\Gamma}$, consisting of even permutations, maps isomorphically to Γ, and $\widetilde{\Gamma} = \Gamma \times \mathbf{C}_2$ (see problem 4.5.12). The desired double cover is S^3/Γ_0.

Every subgroup $\Gamma_1 \subset \widetilde{\Gamma} = \Gamma \times \mathbf{C}_2$ which projects isomorphically to Γ is the graph of a homomorphism from Γ to $\mathbf{C}_2$. The only homomorphism is trivial, so $\Gamma_1 = \Gamma_0$. 4.5.4

There remains the classification of finite subgroups $H \subset \mathrm{SO}(3) \times \mathrm{SO}(3)$ which act freely on $\mathbf{RP}^3$. We have just used one construction for subgroups of a product $\Gamma_1 \times \Gamma_2$: the graph of a homomorphism $\Gamma_1 \to \Gamma_2$ is a subgroup of $\Gamma_1 \times \Gamma_2$, and this construction covers all possible subgroups which project isomorphically to Γ_1. There is also a symmetric construction, the graph of a homomorphism $\Gamma_2 \to \Gamma_1$. These constructions are special cases of a nice construction for all subgroups of a product. Given a pair of groups Γ_1 and Γ_2, and a pair of homomorphisms h_1, h_2 to a third group B, the *fiber product* $G(h_1, h_2)$ of h_1 and h_2 is the subgroup of $\Gamma_1 \times \Gamma_2$

$$G(h_1, h_2) = \{(\gamma_1, \gamma_2) : \ h_1(\gamma_1) = h_2(\gamma_2), \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2\}$$

$$\begin{array}{ccc} & G(h_1,h_2) & \\ \swarrow & & \searrow \\ \Gamma_1 & & \Gamma_2 \\ \searrow & & \swarrow \\ & B & \end{array}$$

The special case when h_1 or h_2 is an isomorphism reduces to the graph of a homorphism.

Proposition 4.5.5 (diagonal groups are fiber product). *Every subgroup $H \subset \Gamma_1 \times \Gamma_2$ which projects surjectively to both factors is the fiber product of a pair of surjective homomorphisms of Γ_1 and Γ_2 to some group B.*

Proof of 4.5.5: Let H_1 and H_2 be the intersections of H with the two factors, $\Gamma_1 \times \{1\}$ and $\{1\} \times \Gamma_2$.

We claim that H_1 is normal in $\Gamma_1 \times \{1\}$. To check this, consider an element $(h_1, 1) \in H_1$, and conjugate it by an arbitrary element $(\gamma_1, 1)$. By hypothesis, there is some $\gamma_2 \in \Gamma_2$ such that $(\gamma_1, \gamma_2) \in H$. It follows that $(\gamma_1, \gamma_2) \cdot (h_1, 1) \cdot (\gamma_1^{-1}, \gamma_2^{-1}) = (\gamma_1 h_1 \gamma_1^{-1}, 1) \in H_1$. Therefore, H_1 is normal in $\Gamma_1 \times \{1\}$. Similarly, H_2 is normal in $\{1\} \times \Gamma_2$.

Now we can take the quotient of the entire picture by the normal subgroup $H_1 \times H_2$. We obtain a subgroup $B = H/(H_1 \times H_2)$ of the product $(\Gamma_1/H_1) \times (\Gamma_2/H_2)$ which intersects each factor of the product trivially. In this quotient

% diagonal groups are fiber product
% The geometric classification of 2-dimensional orbifolds
Problem "grundgy classification of subgroups of SO(3)"
% axis of composition

picture, B is the graph of an isomorphism between Γ_1/H_1 and Γ_2/H_2. We can reconstruct H as the fiber product of the maps of Γ_1 and Γ_2 to B. 4.5.5

Using 4.5.5, finite groups of isometries of $\mathbf{RP}^3$ can be enumerated fairly readily in terms of subgroups of SO(3). These subgroups will be classified conceptually in section 5.5, but it is worth working them out by a straightforward albeit grundgy approach:

Problem 4.5.6 (grundgy classification of subgroups of SO(3)). Classify all finite subgroups of SO(3).

(a) For any spherical triangle with angles α, β and γ there are elements A, B and C of SO(3) such that $ABC = 1$ which rotate by angles 2α, 2β and 2γ. (Compare problem 4.4.4).

(b) Conversely, if $A, B, C \in \mathrm{SO}(3)$ satisfy $ABC = 1$, and if the three axes are not coplanar, they can be described as above using any of eight spherical triangles.

(c) Three positive real numbers α, β and γ are the angles of a nondegenerate spherical triangle if and only if $\alpha + \beta + \gamma > \pi$, and they satisfy the triangle inequalities $\alpha + \beta > \gamma$ *etc.*

(d) In a spherical triangle, the shortest side is opposite the smallest angle.

(e) If A and B generate a finite group F, and if they have axes closer than any other pair of elements of F, then at least one has order 2 and the other has order 2 or 3.

(f) If A and B as above have order 2 and generate a finite subgroup, the group is a dihedral group D_{2n} of order $2n$.

(g) If A and B as above have order 2 and 3, their product has order 3, 4 or 5, and they generate the tetrahedral group (order 12), the octahedral group (order 24), or the icosahedral group (order 60).

(h) Every finite subgroup of SO(3) is generated by 1 or 2 elements.

(i) The finite subgroups of SO(3) are the cyclic groups, the dihedral groups, the tetrahedral group, the octahedral group, and the icosahedral group.

The other piece of information necessary is a good criterion for a group to act freely on $\mathbf{RP}^3$.

Proposition 4.5.7. *A subgroup $H \subset \mathrm{SO}(3) \times \mathrm{SO}(3)$ acts freely on $\mathbf{RP}^3$ if and only if there is no element $(h_1, h_2) \in H$ (except $(1,1)$) such that h_1 is conjugate in* SO(3) *to h_2.*

Remark: Every element of SO(3) is a rotation by some angle α about an axis, and two elements are conjugate if they rotate by the same angle. By reversing the direction of an axis, a rotation by α is also conjugate to a rotation by $2\pi - \alpha$.

Proof of 4.5.7: The action of $\mathrm{SO}(3) \times \mathrm{SO}(3)$ on $\mathbf{RP}^3 = \mathrm{SO}(3)$ is by multiplication on the right and left,

$$(h_1, h_2)(x) = h_1 x h_2^{-1}.$$

For x to be a fixed point means $x = h_1 x h_2^{-1}$ or $x h_2 x^{-1} = h_1$. 4.5.7

Corollary "low order commitment"
Exercise "Sylow subgroups"
Corollary "one factor is cyclic"
% grundgy classification of subgroups of SO(3)
% polyhedra and permutations
% one factor is cyclic

Corollary 4.5.8 (low order commitment). *Let $H \subset \mathrm{SO}(3) \times \mathrm{SO}(3)$ be a group which acts freely on $\mathbf{RP}^3$, and let Γ_1 and Γ_2 be its two projections. Then for $p = 2$ and for $p = 3$, all order p elements in one of the Γ_i are in H, and no order p elements of the other are in H.*

Proof of 4.5.8: All elements of order $p = 2$ or $p = 3$ are conjugate in SO(3). Therefore, H_1 and H_2 cannot both have elements of order p—at least one of them, say H_2, does not.

Consider any element $\alpha \in \Gamma_1$ of order p. We will show that $(\alpha, 1) \in H$.

By definition, there is some β in Γ_2 such that $(\alpha, \beta) \in H$. Then its pth power $(1, \beta^p) \in H_2$ has order m not a multiple of p (but possibly $m = 1$). Hence (α^m, β^m) has order p, so β^m must equal 1. Therefore $(\alpha^m, 1) \in H$, and since m is prime to p, also $(\alpha, 1) \in H$. **4.5.8**

Exercise 4.5.9 (Sylow subgroups). (a) Construct an isometric action of $\mathbf{C}_5$ on $\mathbf{RP}^3$ such that H_1 and H_2 are both trivial.

(b) Construct examples of groups of isometries acting freely on $\mathbf{RP}^3$ whose orders are powers of 2 and powers of 3 for which both Γ_1 and Γ_2 are not trivial.

Corollary 4.5.10 (one factor is cyclic). *(Seifert) With notation as above, either Γ_1 or Γ_2 must be trivial or cyclic. There is some one-parameter subgroup ($\approx S^1$) of one of the two factors of* $\mathrm{SO}(3) \times \mathrm{SO}(3)$ *which commutes with the action of H.*

Proof of 4.5.10: If one of the Γ_i, say, Γ_2 is cyclic (or trivial), then it is contained in a 1-parameter subgroup P of SO(3), isomorphic to S^1, so that multiplication on the right by P commutes with the action of H. Thus all we need to show is that one of the Γ_i is cyclic.

By 4.5.6, the finite subgroups of SO(3) are the cyclic groups, the dihedral groups, and the three groups of rigid motions of regular polyhedra: the tetrahedral, octahedral and icosahedral groups. All but the cyclic groups contain elements of order 2, and all others except the tetrahedral group are generated by elements of order 2.

If H contains no elements of order 2, then both Γ_i are cyclic, and we are done.

Otherwise, we may suppose for concreteness that H_1 contains elements of order 2. We need to show that if Γ_1 is not cyclic then Γ_2 is cyclic.

The only choice for Γ_1, with a proper normal subgroup containing its order two elements, is the tetrahedral group. The subgroup generated by order two elements is the subgroup taking each pair of opposite edges to itself. It has the structure of $\mathbf{C}_2 \oplus \mathbf{C}_2$. It has index three—see problem 4.5.17.

Therefore, the quotient group $\Gamma_1/H_1 = \Gamma_2/H_2$ is $\mathbf{C}_3$. Since H_2 contains no elements of order 2, neither does Γ_2, so it is cyclic. **4.5.10**

The significance of 4.5.10 is that any elliptic three-manifold $M = \mathbf{RP}^3/H$ (or its double cover) is foliated by circles, coming from the family of Hopf circles which are orbits of any S^1 action on $\mathbf{RP}^3$ which is normalized by H.

% The geometry of the three-sphere
% one factor is cyclic
% two-dimensional description of elliptic three-manifolds
% Orbifolds and Seifert fiber spaces
Theorem "classification of elliptic manifolds"

This is not quite a fiber bundle, because sometimes there are circles where the circles in a neighborhood wind around several times before closing. A foliation like this is called a *Seifert fibering* of M.

We can use these fiberings to better understand the topology and geometry of elliptic manifolds. In the correspondence (discussed in section 2.7) between $\mathbf{RP}^3$ and the tangent circle bundle of S^2, the tangent circles form a family of Hopf circles. The action of SO(3) on the right is the action of isometries of SO(3) by their derivatives on the tangent sphere bundle of S^2.

The action on the left is a little harder to visualize in this picture: to make it well-defined, we must first pick a base point (tangent vector) X_0, corresponding to the identity element of the group. Given an element γ of SO(3), γ acting on the left sends any tangent vector Y to the vector $Y\gamma$ which has the same geometrical relation to Y as X_0 does to $\gamma X_0 = X_0\gamma$. Thus, γ acts on the left by issuing a set of instructions to a tangent vector such as "turn 10° to the left in the circle fiber, proceed 100 kilometers in S^2 along the resulting direction, turn 45° to the right and halt."

Using this description we see that we can arrange the action of $O(2) \subset$ SO(3) so that it acts on the right as the group of diffeomorphisms of the tangent sphere bundle of S^2 consisting of simultaneous rotations of the fibers, possibly composed with the derivative of the antipodal map. This corresponds to regarding $O(2) \subset$ SO(3) as the actions which preserve the standard Hopf circles.

By 4.5.10, any finite group Γ acting freely on $\mathbf{RP}^3$ is contained in SO(2) × SO(3). Using the action as described above, we can think of the quotient space reasonably well in terms of a two-dimensional picture (Challenge 4.5.14). We will discuss the general theory of Seifert fiber spaces in chapter 5.

Theorem 4.5.11 (classification of elliptic manifolds). *The elliptic three-manifolds M are of one of the following four types:*

(a) *If $\pi_1 M$ is abelian, then it is cyclic and M is a lens space*

$$L_{p,q} = S^3/\mathbf{C}_p$$

(p and q relatively prime), where a generator of $\mathbf{C}_p$ (p not necessarily a prime) acts, using complex coordinates (z_1, z_2), as

$$(z_1, z_2) \mapsto (\zeta z_1, \zeta^q z_2)$$

with ζ the primitive p^{th} root of unity $e^{2\pi i/p}$.

(b) *$\pi_1(M)$ has the form $\widetilde{H}_1 \times H_2$, where $\widetilde{H}_1$ is a binary dihedral group, the binary tetrahedral group, the binary octahedral group or the binary icosahedral group, and H_2 is a cyclic group (possibly $\{1\}$) with order relatively prime to the order of $\widetilde{H}_1$. In this case M has the form $\mathbf{RP}^3/(H_1 \times H_2)$, where $H_1 = \widetilde{H}_1/\{\pm I\} \subset$ SO(3) and $H_2 \subset$ SO(3).*

(c) *$\pi_1(M)$ has the form $\widetilde{H}$ where H is a subgroup of index 3 of $T \times \mathbf{C}_{3m}$ where m is odd and T is the tetrahedral group. M has the form $\mathbf{RP}^3/H$*

% low order commitment
% one factor is cyclic
Problem "Frobenius map"
% classification of elliptic manifolds

(d) $\pi_1(M)$ *has the form* $\widetilde{H}$ *where* H *is a subgroup of index* 2 *in* $\mathbf{C}_{2n} \times D_{2m}$ *where* n *is odd and* m *and* n *are relatively prime. Again,* M *has the form* $\mathbf{RP}^3/H$.

Note: If Γ is a subgroup of SO(3), then the group $\widetilde{\Gamma} \subset S^3$ consisting of elements which project to Γ is called the *binary* Γ – e.g., if Γ is the icosahedral group, $\widetilde{\Gamma}$ is the binary icosahedral group. As we have seen, $\widetilde{\Gamma} = \Gamma \times \mathbf{C}_2$ iff Γ has odd order.

Proof of 4.5.11: Consider first the case $M = \mathbf{RP}^3/H$. If H is abelian, then Γ_1 and Γ_2 are both either cyclic or $\mathbf{C}_2 \oplus \mathbf{C}_2$. But if $\pi_1(M)$ is abelian, $\mathbf{C}_2 \oplus \mathbf{C}_2$ cannot occur, since its binary counterpart $\widetilde{\Gamma}$ is the quaternion group $\{\pm 1, \pm i, \pm j, \pm k\}$ which is non-abelian. It is the homomorphic image of $\pi_1(M)$ induced by the map $SO(4) \to S^3$. Therefore, Γ_1 and Γ_2 are cyclic, and H must itself be cyclic – otherwise for some prime p, the group of elements of order p would have rank 2, which would mean $\mathbf{C}_p \subset H_1$ and $\mathbf{C}_p \subset H_2$, contradicting the fact that H acts freely. Since H is cyclic, H has the structure $H = \mathbf{C}_{2^k} \oplus \mathbf{C}_m$ for some k and some odd m, and $\widetilde{H} = \widetilde{\mathbf{Z}}_{2^k} \oplus \mathbf{C}_m$. But $\widetilde{\mathbf{Z}}_{2^k} = \mathbf{C}_{2^{k+1}}$ since it has a unique element $(-I)$ of order 2.

If $\pi_1(M)$ is abelian but $M \neq \mathbf{RP}^3/H$, then $\pi_1(M/\pm I)$ is cyclic, so $\pi_1(M) \subset \pi_1(M/\pm I)$ is also cyclic.

The description given in (a) for an arbitrary free action of a cyclic group on S^3 is derived from elementary linear algebra.

If $M = \mathbf{RP}^3/H$, where H is not cyclic, then we can assume (from 4.5.8) that H_1 contains all elements of order 2 in Γ_1 and that $H_2 = \mathbf{C}_m$ where m is odd. So either $H_1 = \Gamma_1$, or Γ_1 is cyclic or the tetrahedral group. $H_1 = \Gamma_1$ implies also $H_2 = \Gamma_2$ and $\pi_1(M) = \widetilde{H_1 \oplus H_2} = \widetilde{H}_1 \oplus H_2$. If Γ_1 is the tetrahedral group, we get case (c), as in the proof of 4.5.10. If Γ_1 is cyclic, Γ_2 must be dihedral, since H is not abelian and the groups of the regular polyhedra are not cyclic extensions of cyclic groups. Γ_2/H_2 must be $\mathbf{C}_2$, since $\mathbf{C}_m$ is the only normal cyclic subgroup of odd order of a dihedral group D_{2m}. Thus, $\Gamma_2 = D_{2m}$, $\Gamma_1 = \mathbf{C}_{2n}$, and H is a subgroup of index 2 of $\Gamma_1 \oplus \Gamma_2$. 4.5.11

Problem 4.5.12 (Frobenius map). Show that for any group $\widetilde{\Gamma}$ which is a central extension $A \to \widetilde{\Gamma} \to \Gamma$, where Γ is finite, the "Frobenius map" $\widetilde{\Gamma} \to A$ which sends any element to its $|\Gamma|^{th}$ power is a homomorphism. In particular, if $|\Gamma|$ is relatively prime to the order of any element in A, then $\widetilde{\Gamma}$ is canonically isomorphic to the product $\Gamma \times A$.

[Hint: choose a product structure for $\widetilde{\Gamma}$ as a set, but respecting the action of A. Reinterpret the map $\gamma \to \gamma^{|\Gamma|}$ as the total amount γ "rotates" the cosets of A].

Exercise 4.5.13. Give the correspondences between these various descriptions of lens spaces:

(a) $M = \mathbf{RP}^3/H$, where $H \subset$ (cyclic group $\times$ cyclic group), or the double cover of a manifold of this form.

(b) The description given in Theorem 4.5.11. What repetitions are there in the list?

Challenge "two-dimensional description of elliptic three-manifolds"

Problem "polyhedra and permutations"

(c) Take a lens, $L = (D^2 \times I)/$ (each $\theta \times I$, for $\theta \in \partial D^2$), and glue the top to the bottom by a rotation of order p.

(d) Glue together two copies of the solid tori $D^2 \times S^1$ by any affine homeomorphism ϕ between their boundaries such that $\phi(\partial D^2)$ is not parallel to ∂D^2.

Challenge 4.5.14 (two-dimensional description of elliptic three-manifolds). Give a two-dimensional description of the elliptic three-manifolds other than lens spaces, expressed in terms of the bundle of regular p-gons in the tangent space of S^2. Which lens spaces can be expressed analogously? What are the two ways of describing M in the case $\pi_1(M) =$ (cyclic group) $\times$ (binary dihedral group)?

Exercise 4.5.15. Elliptic three-manifolds $\longleftrightarrow$ subgroups of $U(2)$ which act freely on $\mathbf{C}^2 - \{0\}$.

Problem 4.5.16. Describe $S^3/$ (binary tetrahedral group) and $S^3/$ (binary octahedral group) geometrically by describing a fundamental polyhedron in S^3 and how it is glued together.

Problem 4.5.17 (polyhedra and permutations). Give a geometric description of the isomorphisms

tetrahedral group $\approx A_4$

octahedral group $\approx S_4$

icosahedral group $\approx A_5$.

[Hint: consider the action on various geometrical figures inside a regular polyhedron].

Describe the structure of the full group of isometries of a regular polyhedra (allowing reversals of orientation). [Hint: $-I$ is central in $O(3)$].

Give a description of the automorphism of the icosahedral group $\approx A_5 \subset S_5$ which does not come from an isometry, by imagining a dodecahedron made of pentagrams.

Section "Discrete groups of isometries of the five fibered geometries"
FIBERED GROUPS
% classification of $S^2 \times \mathbf{E}^1$-manifolds
Section "DISC OR ABEL"
Section "COFIN TWISTHYP"

4.6. Discrete groups of isometries of the five fibered geometries

Bill has not looked at this section at all yet-Dick

In the preceding two sections, we have seen that Euclidean three-manifolds and elliptic three-manifolds have a close relation to discrete groups acting in one and two dimensions. It is less surprising that there are similar descriptions for manifolds modeled on any of the five geometries which have a natural fibering over either two-dimensional or one-dimensional geometries.

For $S^2 \times \mathbf{E}^1$, this was easy (section 3.12.3). For the remaining four geometries – $\mathbf{H}^2 \times \mathbf{E}^1$, $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$, nilgeometry and solvgeometry – the description in terms of lower dimensional geometries still requires analysis.

Propositio 4.6.1. *A discrete group, Γ, of isometries of $\mathbf{H}^2 \times \mathbf{E}^1$ or $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ has image in the group of isometries of $\mathbf{H}^2$ which is either discrete or contains an abelian subgroup with finite index.*

Proof of 4.6.1: Let X denote either of the two geometries, and G_0 the component of its group of automorphisms which contains 1 . First, assume that $\Gamma \subset G_0$. The center $Z_0 = Z(G_0)$ is the kernel of the map $p : G_0 \to Isom(\mathbf{H}^2)$. Let U be any minipotent neighborhood of 1 in G. Z_0U is also minipotent since $[z_1u_1, z_2u_2] = [u_1, u_2]$. Therefore the subgroup H of $p(\Gamma)$ generated by $p(\Gamma) \cap p(U) = p(\Gamma \cap Z_0U)$ is nilpotent. We will assume that $p(\Gamma)$ is not discrete, and prove that it contains an abelian subgroup with finite index. The non-discreteness of $p(\Gamma)$ implies that H is not empty, and one readily sees that it is contained in a one-parameter subgroup A which is either elliptic, hyperbolic, or parabolic.

minipotent is no onger defined in 4.2

If $\gamma \in \Gamma$ is any element, and if $h = p(h')$ is a sufficiently small element, then the commutator $[\gamma, h']$ is close to the identity (since one can write $h' = zh_0$, for some h_0 near the identity). Since Γ is discrete, h may be chosen small enough to guarantee that γ commutes with h', so $p(\gamma)$ centralizes A. The centralizer of A is A, so $p(\Gamma)$ is abelian. (In the more general case that Γ is not contained in the identity component G, $p(\Gamma)$ would only have an abelian subgroup of finite index). 4.6.1

Corollar 4.6.2. *A discrete group Γ of isometries of $\mathbf{H}^2 \times \mathbf{E}^1$ or $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ is a co-finite-volume group iff it has an infinite subgroup $H \subset \Gamma$ which acts trivially on $\mathbf{H}^2$, and the group Γ/H is a discrete co-finite-area group of isometries of $\mathbf{H}^2$.*

Proof of 4.6.2: If $p(\Gamma)$ is abelian, it acts preserving a line, point or horocycle in H^2, so Γ preserves and acts discretely on the preimage in X (where X is $\mathbf{H}^2 \times \mathbf{E}^1$ or $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$). Thus X/Γ has infinite volume.

Since X/Γ has finite volume, $p(\Gamma)$ is a discrete group of motions of $\mathbf{H}^2$. Let H be the kernel of p. The volume of X/Γ is $\mathrm{area}(\mathbf{H}^2/p(\Gamma)) \cdot \mathrm{length}(\mathbf{E}^1/H)$, and the corollary follows. 4.6.2

% NILGEOM
% NILGEOM
% DISC AND GRPS
Section "COFIN NIL"
Section "DISC SOL"
% DISC AND GRPS

Propositio 4.6.3. *A discrete group* Γ *of isometries of nilgeometry (3.12.9) projects to a group of isometries of* $\mathbf{E}^2$ *which is either discrete, or preserves some line or some point in* $\mathbf{E}^2$.

Proof of 4.6.3: Let $p : G \to Isom(\mathbf{E}^2)$ be the projection of G to its action on $\mathbf{E}^2$ (in the (a,b,d) coordinates of 3.12.9). Let us first suppose that Γ is contained in the identity component G_0 of G, the group of isometries of nilgeometry. Then the commutator subgroup of Γ acts by translations on $\mathbf{E}^2$, since the commutator of any two orientable isometries of $\mathbf{E}^2$ is a translation. If γ_1 and γ_2 are elements of Γ, then $[\gamma_1, \gamma_2]$ is a vertical translation t of nilgeometry by a distance equal to the area of the parallelogram spanned by $p(\gamma_1)(0) - 0$ and $p(\gamma_2)(0) - 0$ to $\mathbf{E}^2$.

Suppose that there is some such parallelogram with non-zero area. Let T be the group of all vertical translations in Γ. T is central in G_0, and the group G_0/T acts on $\mathbf{E}^2$ with compact stabilizers. Therefore, by ??, the group Γ/T of isometries of $\mathbf{E}^2$ is discrete.

If on the other hand all translations of $\mathbf{E}^2$ in $p([\Gamma, \Gamma])$ are linearly dependent, then either $p([\Gamma, \Gamma])$ acts trivially on $\mathbf{E}^2$ – in which case either some element of $p(\Gamma)$ has non-trivial rotation and thus $p(\Gamma)$ fixes a point or $p(\Gamma)$ consists entirely of pure translations and is thus discrete – or $p([\Gamma, \Gamma])$ acts by translations along some line. Then the rotational parts of the action of $p(\Gamma)$ can only be the identity or 180° rotations, and one sees that $p(\Gamma)$ must preserve some line in $\mathbf{E}^2$.

The general case, when $\Gamma \not\subset G_0$, follows readily. 4.6.3

Corollar 4.6.4. *A co-finite-volume group* Γ *of isometries of nilgeometry is cocompact, and its image in* $Isom(\mathbf{E}^2)$ *is discrete.*

Finally, consider three-dimensional solve-geometry. Denote the space by X, its group by G and the identity component of G by G_0. Recall the structure of G_0 as an extension $\mathbf{R}^2 \to G_0 \to \mathbf{R}$.

Propositio 4.6.5. *If* $\Gamma_0 \subset G_0$ *is any discrete subgroup, then* $\Gamma_0/(\Gamma_0 \cap \mathbf{R}^2)$ *acts discretely on* $\mathbf{R}$. *Any discrete subgroup* Γ *of* G *acts as a discrete group on the line* $\mathbf{R}$, *and the kernel of the action on the line is a discrete group of isometries of the Euclidean plane.*

Proof of 4.6.5: Any element γ in Γ_0 which is not in $\mathbf{R}^2$ acts as an affine map in $\mathbf{R}^3$ which translates some unique vertical line (since γ acts on $\mathbf{R}^2$ without any eigenvalue of 1).

If Γ_0 is abelian, it is either contained in $\mathbf{R}^2$, or all of Γ_0 must act as translations on a single vertical line. The proposition follows in either case.

If Γ_0 is not abelian, its commutator subgroup is contained in $\mathbf{R}^2$. If $t \in \Gamma_0 \cap \mathbf{R}^2$ and $\gamma \in \Gamma_0 - \mathbf{R}^2$, then t and $\gamma t \gamma^{-1}$ are linearly independent translations, so $\Gamma_0 \cap \mathbf{R}^2$ is cocompact. The action of Γ_0 on $\mathbf{R}$ factors through the action on $X/(\Gamma_0 \cap \mathbf{R}^2)$, so by ?? the image of Γ_0 in $\mathbf{R}$ is discrete.

The statements concerning Γ follow readily. 4.6.5

% intransigent groups
% Some computations in hyperbolic space
% discrete Euclidean almost abelian
% COFIN NIL

Exercis 4.6.6. (This is standard material in the theory of arithmetic groups.) Show that there are interesting discrete groups of isometries of $\mathbf{H}^2 \times \mathbf{H}^2$ which project to non-discrete groups on both factors. Do challenge 4.4.17 first.

(a) Now consider the group G with presentation

$$< \alpha, \beta, \gamma \mid \alpha^2 = \beta^2 = \gamma^2 = (\alpha\beta)^7 = (\beta\gamma)^7 = (\gamma\dot{\alpha})^7 = 1 > .$$

Consider three triangles: two in the hyperbolic plane with angles $\pi/7, 2\pi/7$ and $2\pi/7$, and $3\pi/7, \pi/7$ and $\pi/7$, the third on S^2 with angles $2\pi/7, 3\pi/7$ and $3\pi/7$. This gives three homomorphisms of G into isometries of some geometry. Show that the diagonal homomorphism, into $Isom(\mathbf{H}^2 \times \mathbf{H}^2 \times S^2)$, has discrete image, and consequently the projection to $Isom(\mathbf{H}^2 \times \mathbf{H}^2)$ is also discrete.

[Hint: write down matrices for the quadratic form descriptions of $\mathbf{H}^2$ and S^2 based on these triangles, as in Section 2.4. Note that all coefficients are algebraic integers. Describe a lattice in $\mathbf{R}^9$ invariant by the diagonal action of G, and conclude that G is discrete.

(b) Prove that the image in $Isom(\mathbf{H}^2)$ is indiscrete by showing that the two images are isomorphic, and that there is a subgroup which is elliptic in one image but hyperbolic in the other.

Exercise Show a closed Nil-manifold is Haken.

(c) Show that these homomorphisms of G are not faithful.

We now have enough information to be able to find algebraic properties of cocompact groups in any of the eight geometries which identify the geometry. First, a convenient piece of terminology. If X is a property of certain groups, then a group is *almost* X if it contains a subgroup of finite index which is X. For instance, discrete groups of Euclidean motions are almost abelian (4.2.7); discrete cocompact groups of automorphisms of nilgeometry are almost nilpotent (from 4.6.4), etc.

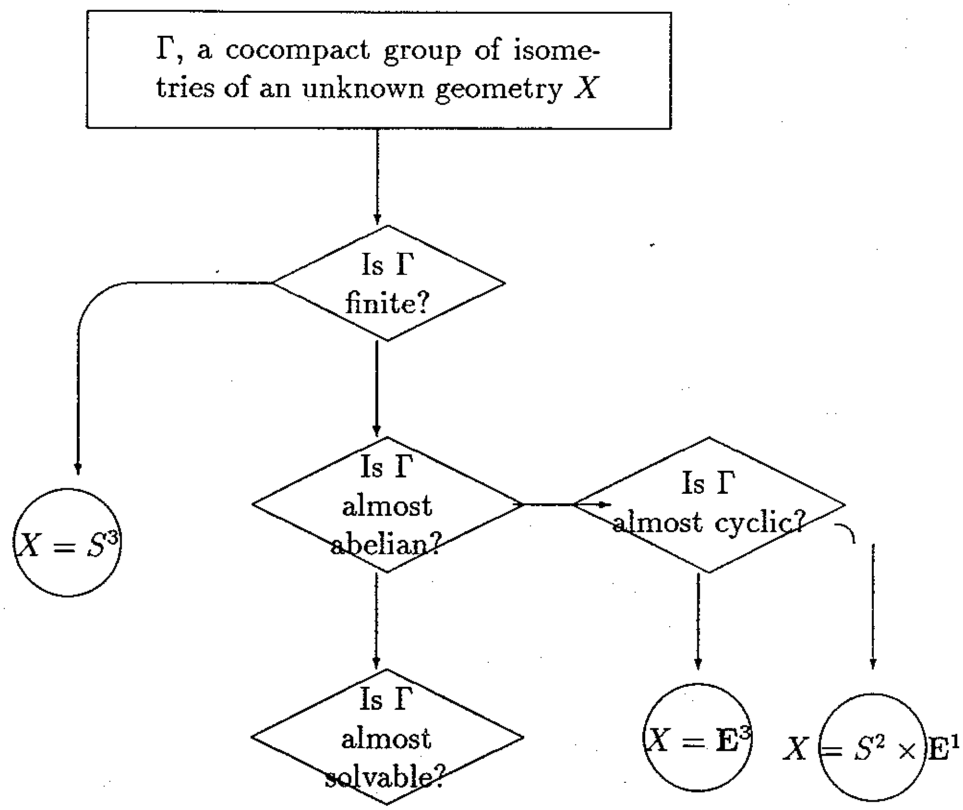
Someone needs to fin[d] a 1982 copy of chapter 4 and insert this flowchart and th[e] one later in the section-Dick

Theore 4.6.7. *A cocompact group Γ of automorphisms of any one of the eight basic three-dimensional geometries is not isomorphic to a cocompact group in any of the others.*

The following flowchart reconstructs the geometry.

% classification of $S^2 \times E^1$-manifolds
% discrete Euclidean almost abelian
% COFIN TWISTHYP

% COFIN NIL
% DISC SOL



Proof of 4.6.7: We already have most of the information, so we only need to consider a few extra points. From section 3.12.3, corollary 4.2.7, 4.6.2, 4.6.4 and 4.6.5, it is clear that a group always takes a yes branch when the flowchart

% DISC OR ABEL
% NOLIFT
Section "NOLIFT"
% SL2R METRIC
Section "NONCOM COFIN"
% NONCOM COFIN
Section "COCOM NIL"

says it should. An infinite cyclic normal subgroup in the case of $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ on $\mathbf{H}^2 \times \mathbf{E}^1$, is $\Gamma \cap Z_0$, where Z_0 is the center of the identity component of the group. The only problem is to make sure that no group takes a yes branch when the flowchart says it shouldn't.

Cocompact groups of isometries of $\mathbf{H}^3$ are prevented from taking any yes branch by the fact that they have no normal abelian subgroups, by a familiar line of reasoning. (To prove that no cocompact subgroup of $Isom(\mathbf{H}^3)$ is almost solvable, we further observe that any maximal abelian subgroup is only normalized by itself). Groups of isometries of $\mathbf{H}^2 \times \mathbf{E}^1$ or $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ are separated from the earlier cases similarly. (See the proof of 4.6.1.) To separate them from each other, note that any normal infinite cyclic subgroup $A \subset \Gamma$ is a subgroup of $\Gamma \cap Z_0$). The centralizer of A is $\Gamma \cap G_0$. If A maps injectively to the abelianization of $\Gamma \cap G_0$, then there is some homomorphism of $\Gamma \cap G_0$ to $\mathbf{Z}$ which is injective on A. Let H be the kernel. Since the action of H on $\mathbf{H}^2$ is the same as the action of HA, which has finite index in Γ, it follows that H is cocompact. A cocompact subgroup of $\mathrm{PSL}(2,\mathbf{R})$ does not lift to $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ (Exercise 4.6.8), so the geometry must be $\mathbf{H}^2 \times \mathbf{E}^1$.

It is elementary to check the other cases. **4.6.7**

Exercis 4.6.8. A cocompact subgroup of $\mathrm{PSL}(2,\mathbf{R})$ does not lift to $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$. [Hint: see exercise 3.12.8.]

The non-cocompact but cofinite volume case is somewhat different.

Theore 4.6.9. *Non-cocompact cofinite volume groups of automorphisms exist only for* $\mathbf{H}^3, \mathbf{H}^2 \times \mathbf{E}^1$ *and* $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ *(out of the eight basic geometries). Any such group of automorphisms of* $\mathbf{H}^2 \times \mathbf{E}^1$ *acts also as a cofinite volume group of automorphisms of* $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$*, and vice versa. Otherwise, groups are distinguished by this flowchart:*

WE NEED A FLOWCHART HERE

The separation of the case of $\mathbf{H}^3$ from the other two cases is exactly the same as for cocompact groups. We will not prove the rest of 4.6.9.

Here are purely algebraic characterizations of cocompact discrete subgroups of nilgeometry and solvegeometry, analogous to a Bieberbach theorem.

Theore 4.6.10. *A group* Γ *has an action as a discrete cocompact group of automorphisms of nilgeometry with finite kernel if and only if* Γ *contains a subgroup of finite index isomorphic to the group*

$$H = < a, b \mid [a,[a,b]] = [b,[a,b]] = 1 > .$$

The action is effective if and only if the centralizer of H *is infinite cyclic* ($\Leftrightarrow$ *torsion free).*

Section "COCOM SOL"
% COCOM NIL
% finitely many automorphisms of abelian group
% NILGEOM

Theore 4.6.11. *A group Γ has an action as a discrete cocompact group of automorphisms of solvegeometry with finite kernel if and only if Γ contains a subgroup H of finite index which is an extension of the form*

$$\mathbf{Z}^2 \to H \leftrightarrows \mathbf{Z}$$

where the action of $\mathbf{Z}$ on $\mathbf{Z}^2$ is generated by a hyperbolic element of $\mathrm{SL}(2,\mathrm{Z})$, *i.e., an element with $|trace| > 2$.*

The action of Γ is effective if and only if the centralizer of H is trivial.

startproof[4.6.10] If Γ is a discrete cocompact group of isometries of nilgeometry, then Γ acts on the plane as a cocompact discrete group of isometries. Let a and b be two elements of Γ which act on the plane as linearly independent translations. The subgroup they generate is isomorphic to H. Since the commutator $[a,b]$ is a non-zero vertical translation, $\mathbf{E}^3/H$ must be compact, so H is of finite index.

Conversely, suppose Γ is a group which contains H with finite index. Let H_0 be a subgroup with finite index in H which is normal in Γ (take for example the intersection of H's conjugates). Then H_0 has various actions as a discrete cocompact group of automorphisms of nilgeometry. The first step is to choose an action such that conjugation by an element $\gamma \in \Gamma$ induces an automorphism of H_0 which comes from an isometry of nilgeometry. Toward this end arrange and choose an action of H_0 in nilgeometry so that conjugation of its induced action on $\mathbf{E}^2$ by an element $\gamma \in \Gamma$ comes from an isometry of $\mathbf{E}^2$. This can be done by choosing a metric on $\mathbf{E}^2$ invariant by the conjugation action, as in 4.3.4. The action on the plane does not determine the action in space: in fact, for any linear map $f : \mathbf{R}^2 \to \mathbf{R}$, the linear map L_f of $\mathbf{R}^3$ into itself which preserves the z-axis and sends (X,z) to $(X, z+f(X))$ (where $X \in \mathbf{R}^2$), induces an automorphism of the Heisenberg group (using the (a,b,d) coordinates of 3.12.9). Two actions of H_0 as isometries of nilgeometries which give rise to the same action in the plane differ by some map L_f – note in particular, that the action of $[H_0, H_0]$ is always the same, being determined by the areas of various parallelograms in the plane.

A finite group of automorphisms of H_0 which respects the action of H_0 on $\mathbf{E}^2$ acts by affine transformations on the set of choices of actions of H_0 in nilgeometry. This finite group has a fixed point – the center of mass of any orbit – so the actions of H_0 can be chosen so the finite group comes from isometries of nilgeometry.

Finally, to realize Γ we consider the induced representation of Γ as a group of affine transformations of $N^{(\Gamma/H_0)}$. This space is constructed by considering

$$(\Gamma/H_0) \times N = (\Gamma \times N)/(gh, x) \sim (g, p(h)x)$$

where N is nilgeometry and p is the representation of H_0 as a group of isometries of nilgeometry, and taking the cartesian product of the copies of N. There is a foliation of $N^{(\Gamma/H_0)}$ by copies of N, parallel to a kind of diagonal embedding $x \to ((g_1, x), ..., (g_n, x))$ where $i = 1, ..., n$ is an index for the cosets of H_0

% dimension n invariant subspace
% COCOM NIL
% COCOM SOL
% DISC SOL
% COCOM SOL
% MAP TOR
% COCOM SOL
% COCOM SOL

in Γ and each g_i is a choice of an isometry of nilgeometry which realizes the automorphism of H_0 defined by the i-th coset. This foliation is invariant by the action of Γ, and each leaf is invariant by H_0. As in the proof of proposition 4.3.3, there must be some leaf invariant by Γ. The action of Γ on this leaf gives the desired action on N.

The condition for the action of Γ to be effective is easily checked. An element γ cannot possibly act trivially on N unless it centralizes H (or H_0). The group which centralizes H_0 in the full group of isometries of N is $\mathbf{R}$ acting as pure vertical translations), so the centralizer of H acts effectively if and only if it is infinite cyclic. finishproof4.6.10

startproof[4.6.11] If Γ is a discrete cocompact group of automorphisms of solvegeometry, then $\Gamma \cap G_0$ has the form indicated for H, where G_0 is the identity component of G (using 4.6.5).

If Γ is a group having a subgroup H of finite index of the form described in 4.6.11, then we can construct a discrete cocompact action of H on X (by Exercise 3.12.13(b)). We may as well assume that H is normal in Γ. The next claim is that every possible automorphism of H is induced by conjugation by some element of G which normalizes the action of H on X. Indeed, any automorphism must preserve the $\mathbf{Z}^2$ subgroup, and it must preserve or interchange the two eigenspaces in $\mathbf{R}^2 = \mathbf{Z}^2 \otimes \mathbf{R}$. The automorphism restricted to $\mathbf{Z}^2$ can therefore be realized by an element of G of finite order, followed, if necessary, by some translation in the (vertical) t-direction to adjust the relative scaling factors of the eigenspaces. The full group H has one remaining generator, which is a translation along some line. After the automorphism, it becomes a translation along a new line, in the same or opposite sense depending on whether the eigenspaces in $\mathbf{R}^2$ are preserved or interchanged. A horizontal translation now adjusts this final generator, without changing the action of the others.

Note that two different elements of G can never induce the same automorphism of H. Therefore the construction above gives a group homomorphism $Aut(H) \to G$, which extends the given embedding of H in G. The composition $\Gamma \to Aut(H) \to G$ completes the proof of 4.6.11. finishproof4.6.11

Chapter "Orbifolds and Seifert fiber spaces"
ORBIFOLDS
::orbifolds
::V-manifold

Chapter 5

Orbifolds and Seifert fiber spaces

There are two images which even though logically equivalent, are quite different in appearance: the first image is that of a group Γ acting freely and properly discontinuously on a simply-connected space X (such as the hyperbolic plane); the second image is its quotient manifold X/Γ.

In this chapter we will enlarge our vocabulary, so that we can discuss in a similar way the two images associated with groups which act properly discontinuously but do not necessarily act freely — such a group is discrete, but some of its elements may have fixed points. The quotient spaces of these groups (equipped with enough additional structure to describe the way they act) we will call *orbifolds*. (The name *V-manifold*, has also been used for a related concept.)

There are several reasons to study groups which do not act freely. In the first place, there are many beautiful and simple examples, which are often easier to come by than examples of groups acting freely.

Second, groups which do not act freely arise in connection with the study of symmetries of manifolds, for instance, the theory of symmetries of knots.

Third, as we have seen in the preceding chapter, groups acting freely in three dimensions are sometimes built out of groups which do not act freely in one and two dimensions. One and two dimensional orbifolds give a more rational base for the theory of Seifert fiber spaces.

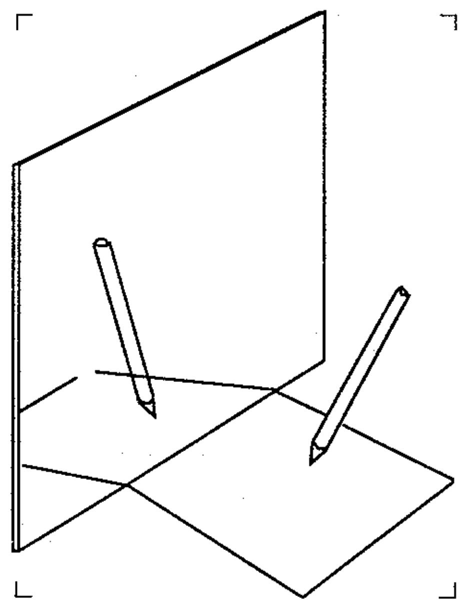
Finally, orbifolds provide a useful tool for constructing hyperbolic structures on three-manifolds: see [Thurston,] and [Thurston,]

Section "Some examples of orbit spaces"
Example "ordinary mirror"
% mirror
Example "barber shop"
Example "cone"
% picone
Example "billiard table"

5.1. Some examples of orbit spaces

Let us begin with a few examples of quotient spaces or orbit spaces in order to get a taste of their geometric flavor.

Example 5.1.1 (ordinary mirror). Consider the action of the cyclic group $\mathbf{C}_2$ of order 2 on $\mathbf{R}^3$ by reflection in the y-z plane. The quotient space is the half-space $x \geq 0$. Physically, one may imagine a mirror placed on the y-z wall of this half-space. The scene as viewed by a person in this half-space is like all of $\mathbf{R}^3$, with scenery invariant by the $\mathbf{C}_2$ symmetry (figure 5.1).



mirror

Figure 5.1. Symmetry of a mirror. An ordinary mirror creates a scene with 2-fold symmetry.

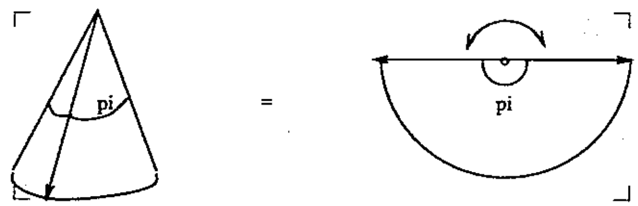
Example 5.1.2 (barber shop). Consider the group G generated by reflections in the planes $x = 0$ and $x = 1$ in $\mathbf{R}^3$. G is the infinite dihedral group $\mathbf{D}_\infty = \mathbf{C}_2 * \mathbf{C}_2$. The quotient space is the slab $0 \leq x \leq 1$. A physical model is formed by two mirrors on parallel walls, as commonly seen in barber shops.

Example 5.1.3 (cone). The cyclic group C_k acts in the plane as the group of rotations of order k about the origin. The region between two rays at a $2\pi/k$ angle is a fundamental domain. The two sides of the angle are identified by the group action, forming a cone.

The most common cones, for instance, many ice cream cones, are formed when $k = 2$ from an angle of π, or a half-plane. Such a cone, when rolled around on table rolls around its point, returning to its initial orientation when it reaches the original position. (figure 5.2).

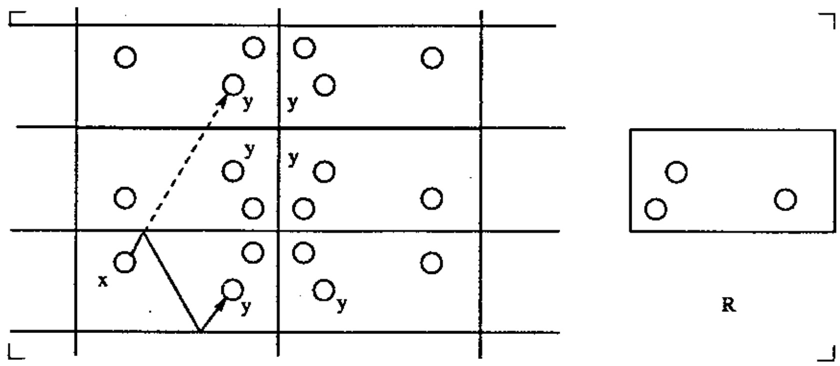
Example 5.1.4 (billiard table). Let G be the group of isometries of the Euclidean plane generated by reflections in the four sides of a rectangle R. G

Example "rectangular pillow"



picone

Figure 5.2. A cone with cone angle π.



billiard

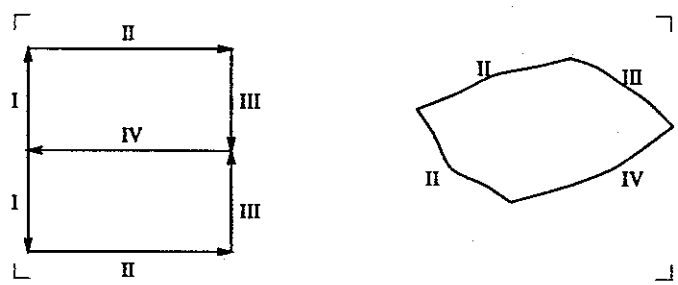
Figure 5.3. A billiard table. (Notation: (*2222).) A billiard table and multiple reflections.

is isomorphic to $\mathbf{D}_\infty \times \mathbf{D}_\infty$, and the quotient space is R. A physical model is a billiard table. A collection of balls on a billiard table gives rise to an infinite collection of balls on $\mathbf{R}^2$, invariant by G.

To make the physical model work more accurately, at least for a single ball, think of the mathematical billiard table as a smaller rectangle inside the physical billiard table, with margins of one ball radius. Then the centers of balls can take positions exactly delimited by this rectangle. If the ball has no spin so that its angle of incidence equals to angle of reflection, its trajectory on the billiard table is the image of a straight line in the plane, folded up by $\mathbf{D}_\infty \times \mathbf{D}_\infty$.

the labelling on figure 5.4 is wrong

Example 5.1.5 (rectangular pillow). (Notation: (2222)). Let H be the



pillow

Figure 5.4. The orientation preserving billiard group

subgroup of index 2 which preserves orientation in the group G of the preceding example. If the images of the billiard table are alternately colored black and

% pillow
::cone points
Problem "tetrahedron as orbifold"
Example "cone axis"
::cone axis
Example "borromean orbifold"
% borrocube
% borroident

white in a checkerboard pattern, this is the subgroup which takes black to black and white to white. The quotient space of $\mathbf{E}^2$ by H is a rectangular pillow.

Two adjacent rectangles form a fundamental domain for the group action. The quotient space is obtained by identifying the edges of the two rectangles by reflection in the common edge (figure 5.4). Topologically, this quotient space is a sphere, with four distinguished points or singular points, which come from points in $\mathbf{R}^2$ with nontrivial isotropy ($\mathbf{C}_2$). Geometrically, it is like a pillow. The local intrinsic geometry of the pillow near each of its points is the same as the geometry of a cone of cone angle π. These are called order 2 *cone points.* Quotient spaces of group actions often have cone-type singularities.

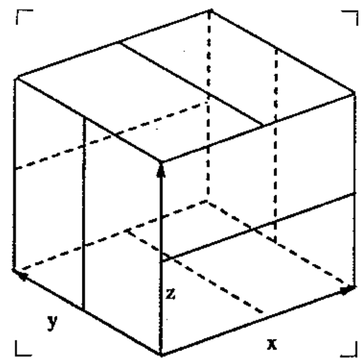
Problem 5.1.6 (tetrahedron as orbifold). Consider any tetrahedron T in $\mathbf{E}^3$ which is made from four congruent triangles. Show that T is isometric to the quotient space of $\mathbf{E}^2$ by a discrete group. Construct a physical model, and watch what happens if you roll it carefully on a table. Wrap a ribbon or string around the tetrahedron so that it follows a geodesic. It should never cross itself. Why not?

Example 5.1.7 (cone axis). The example of a cone generalizes to three dimensions in a straightforward way. The cyclic group $\mathbf{C}_k$ acts in $\mathbf{E}^3$ as the group of rotations of order k about a line. A fundamental domain is a wedge between two planes which meet at angle $2\pi/k$ along the line, and one face of the wedge is identified to the other via the action. The quotient space is homeomorphic to $\mathbf{R}^3$. Geometrically, it inherits a Euclidean metric in the complement of the cone axis. The geometry along the axis is that of a product of an order k cone with the line $\mathbf{E}^1$. A singular curve like this is called an order k *cone axis.*

Can you visualize what it would look like to live inside this quotient space? The appearance would be the identical to that of a pattern with k-fold rotational symmetry about a line in $\mathbf{E}^3$. Suppose, for example, that $k = 2$. You would see an image of yourself directly through the cone axis. The axis itself would not be visible. If you threw a ball toward your image, it would whip around the axis, just like a light ray, and come back to you.

Example 5.1.8 (borromean orbifold). (Notation: $(2_0{}^+2_12_1{}')$.) A crystallographic group. Here is one more 3-dimensional example to illustrate the geometry of quotient spaces.

Consider three families of lines parallel to the three axes in $\mathbf{E}^3$ interlaced evenly so they do not intersect: in coordinates, $(t, n, m + 1/2)$, $(m + 1/2, t, n)$ and $(n, m + 1/2, t)$ where n and m are integers and t is a real parameter. The lines intersect each cube in the unit lattice as in figure 5.5. Let G be the group generated by 180° rotations about these lines. The unit cube can be taken as fundamental domain for this group action. We may construct the quotient space by making all identifications coming from elements of G which take faces of the cube to faces. These identifications fold each side shut, like a book. In doing this, we keep track of the axes, which form the singular locus of the final result. As you can see by studying figure 5.6, the quotient space is S^3 with the

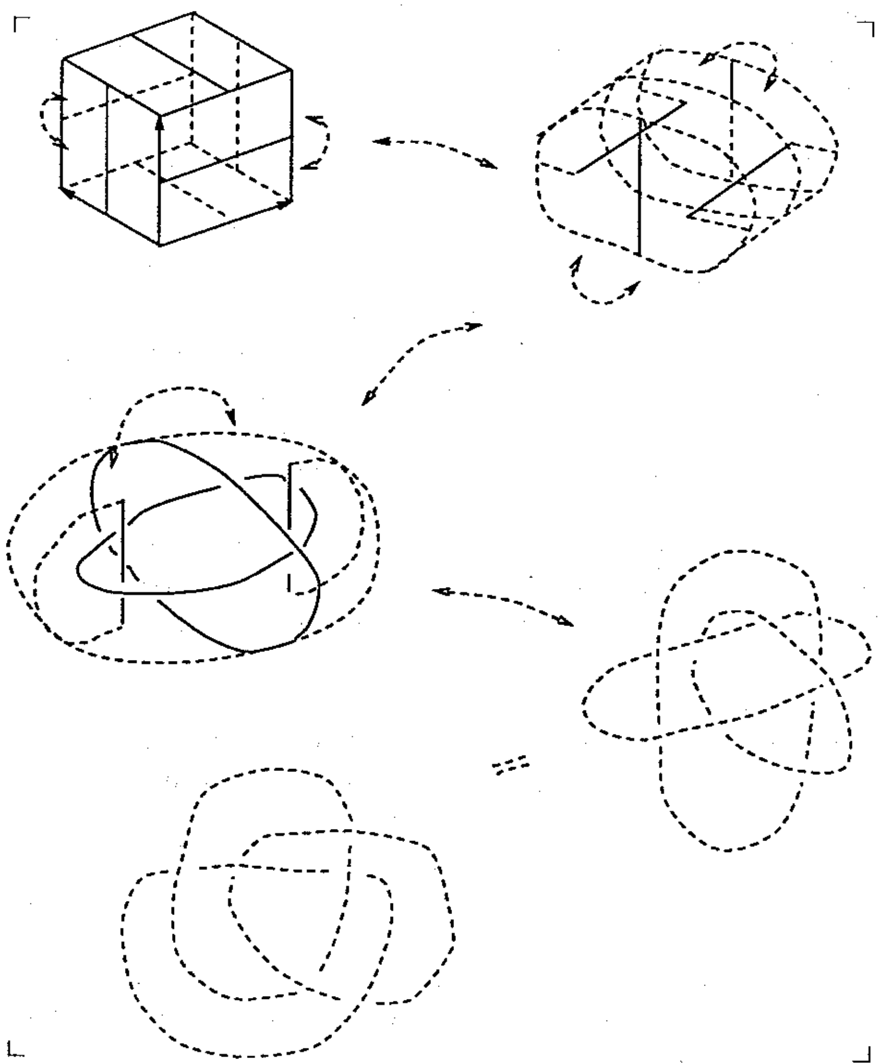


borrocube

Figure 5.5. A group generated by rotations in R^3. The quotient space of a group generated by 180° rotations about 3 non-intersecting families of parallel lines. A unit cube is a fundamental domain for the group action.

singular locus consisting of three circles in the form of the Borromean rings. Metrically, this means that S^3 has a Euclidean metric in the complement of these rings. Each ring is an order 2 cone axis.

In these examples, it turned out not to be very hard to construct the quotient space from the group action. We need to develop the connection between group actions and quotient spaces in the opposite direction as well: given a picture of the quotient space, to be able to say, aha–that picture arises from a certain group acting on Euclidean space (or hyperbolic space, or ...). To do this requires knowledge about the singular locus (that is, points coming from elliptic axes, etc.), with appropriate information about how the group acts above this locus.



borroident

Figure 5.6. Identifications of a cube yielding the Borromean rings. These pictures show how to fold up the sides of the fundamental domain according to the group action, one pair of opposite sides at a time, finally yielding the 3-sphere with three singular axes in the form of a link which is called the Borromean rings.

Section "Basic definitions for orbifolds"
::orbifold
% Some examples of orbit spaces

5.2. Basic definitions for orbifolds

An *orbifold* is a space locally modeled on $\mathbf{R}^n$ modulo faithful finite group actions. Before filling in the formal details of what this means, we will try to elucidate it by examples and an intuitive disucssion.

The idea is that an orbifold is pieced together from the structure that is locally visible in the quotient space of the action of a discrete group on a manifold. In the section 5.1, we gave a number of examples of quotient spaces of groups acting by isometries. In these instances, the quotient space, or quotient orbifold, has a *geometric* structure along with its *topological* structure. For such examples, the geometry, without any additional information about group actions, is sufficient to capture the essence of the example. For instance, the geometry near an order k cone point is enough to deduce the value of k.

For a topological orbifold, not equipped with a geometric structure metric, one has to take more care. For instance, as topological objects, all cones $\mathbf{E}^2/\mathbf{C}_k$ are equivalent. In order to capture the nature of a cone as a quotient space of the plane, an additional piece of information besides the topology is necessary: the additional information is the group $\mathbf{C}_k$, its action on $\mathbf{E}^2$, and a homeomorphism of a neighborhood of the origin in the quotient space to a neighborhood of the cone point. Really, there is only one way to do this, up to homeomorphism, once the location of the cone point and the integer k are specified, so the description is often abbreviated by labeling the point by the integer k.

For any finite collection $\{(p_1, n_1), (p_2, n_2), \ldots, (p_k, n_k)\}$ of points p_i in the plane with an integer n_i greater 1 attached to each point, there is an orbifold $O((p_1, n_1), (p_2, n_2), \ldots, (p_k, n_k))$: it is locally modeled on $\mathbf{R}^2$ in every neighborhood not intersecting the points. In a neighborhood of the point p_i, the orbifold is modeled on $\mathbf{E}^2/\mathbf{C}_{n_i}$. If we identify $\mathbf{E}^2$ with the complex numbers $\mathbf{C}$, the modeling of the neighborhood of p_i on $\mathbf{C}/\mathbf{C}_{n_i}$ is conveniently expressed by the formula $z \mapsto$ $(z - p_i)e^{2\pi i/n_i} + p_i$ ← completely wrong! This is not a periodic map

An orbifold structure should be interpreted as describing, or potentially describing, the quotient space of a discrete group acting on some manifold. One thing you can 'do' with an orbifold is try to find a manifold, together with a discrete group acting on it, whose orbit space is the given orbifold. Sometimes this is possible, sometimes not. For the orbifold $O((p_1, n_1), (p_2, n_2), \ldots, (p_k, n_k))$, this would amount to constructing a manifold which 'unwraps' n_i times about each p_i.

There may be many different manifolds with discrete groups acting on them, whose quotient space is the same orbifold.

Consider, for example, the orbifold $O((1, 2), (-1, 2))$. This (and other similar examples) work out nicely when considered as complex orbifolds. One way to 'unwrap' about the order two cone points at 1 and -1 is by taking the graph of the ambiguous function $\sqrt{(z-1)(z+1)}$, that is,

$$M^2 = \left\{(z, w) \in \mathbf{C}^2 | w^2 = (z-1)(z+1)\right\}.$$

Problem "cyclic branched cover of orbifold"

This set M^2 is a surface (or in the lingo of complex analysis and algebraic geometry, a curve) which maps by a mostly $2-1$ map to the first coordinate. The map $(w,z) \mapsto (-w,z)$ generates a $\mathbf{C}_2$ action on the surface, with quotient orbifold $O((1,2),(-1,2))$. Topologically, the surface is a cylinder. You can see this by making a branch cut along the interval $[-1,1]$. The two branches of $\sqrt{}$ are interchanged every time you cross this arc. When you cut the plane along the arc, you get a half-infinite cylinder, and when you glue the two copies (for the two branches) together, you get a doubly-infinite cylinder.

The surface M^2 can be analytically parametrized by $\mathbf{C} - 0$: a formula is $1/2(t - t^{-1}, t + t^{-1})$. The group $\mathbf{C}_2$ of symmetries, in these coordinates, is generated by inversion, $t \mapsto t^{-1}$. The two fixed points are ± 1. The function $g(t) = (t - t^{-1})/2$ gives the map to the quotient orbifold.

The punctured complex plane can easily be expressed as a quotient space of the plane modulo a group of translations, using the function exp. If we choose the translations to be tranlations by $2Pi$, so that $t = \exp(2Piis)$, then the composed map of $\mathbf{C}$ to $O((1,2),(-1,2))$ is cos. Somehow, the cosine function is implicit in the orbifold $O((1,2),(-1,2))$: the cosine function obeys the laws

$$\cos(s) = \cos(-s)$$

and

$$\cos(s) = \cos(s + 2\pi),$$

in other words, it is invariant by an action of the infinite dihedral group on $\mathbf{C}$, generated alternately by order 2 rotations about the origin and about π. The cosine function itself is the quotient map to the quotient orbifold!

This gives two different ways to 'unwrap' the orbifold $O((1,2),(-1,2))$. There are an infinite number of intermediate examples, using the various finite dihedral groups acting on the punctured plane, generated by inversion together with multiplication by some group of roots of unity. The maps to the quotient orbifold are $g_k(t) = 1/2(t^k - t^{-k}, t^k + t^{-k})$.

In general, there is at least one easy way to 'unwrap' the orbifold $O((p_1,n_1),\ldots,(p_k,n_k))$. Let N be the least common divisor of the n_i, and use the function

$$\left(\prod_i (z - p_i)^{N/n_k}\right)^{1/N}.$$

(We use this notation, rather than cancelling the exponents, to emphasize that there are N branches of the function above a generic point in $\mathbf{C}$.)

Problem 5.2.1 (cyclic branched cover of orbifold). What is the genus of this surface? How many punctures does it have?

We can also add the point at infinity in $\mathbf{C}$, to turn it into a sphere $\hat{complexes}$ (the Riemann sphere), and try to understand the orbifolds $\hat{O}((p_1,n_1),\ldots(p_k,n_k))$ defined in a similar way. It is a good idea to use ∞ as one of the points p_k, as a reminder not to forget it. The analysis

::coordinate patch
::coordinate chart

becomes trickier. Some cases can be derived from the Schwartz-Christoffel method for conformal mappings of the upper half-plane to a polygon. For instance, consider $\hat{O}((0,3),(1,3),(\infty,3))$. The graph of the multi-valued function $z \mapsto (z(z-1))^{1/3}$ is one way to 'unwrap' this orbifold. (Notice that the three sheets are permuted cyclically when one goes around a large circle, that is, a small circle about ∞.) The surface can be triangulated using the images of the three segments of the real axis. This triangulation has 3 vertices, 9 edges and 6 triangles: its Euler number is 0, and it is a torus. ('Curves' which are toruses are called elliptic curves.) The orbifold is the quotient of the torus by a group of order 3.

Topologically, we can express the torus as the quotient of the plane by a group of translations. Combining with the order three rotations of the torus, we obtain a topological description of our orbifold as the quotient of the plane by a group generated by order 3 rotations about the three corners of an equilateral triangle.

We can derive the complex analytic formula using the Schwarz-Christoffel formula for the Riemann mapping to a polygon. More explicitly, integrate the equation

$$dw = (z)^{-2/3}(1-z)^{-2/3}dz.$$

In this formula, the argument of the coefficient of dz is constant in each of the three intervals $(-\infty,0)$, $(0,1)$, and $(1,\infty)$. Therefore, each of the three segments maps to a straight line. These lines are parallel to the sides of an equilateral triangle. Since dw never vanishes in the upper half plane, the integral defines the the requisite mapping. The map can be extended across each of the three intervals of the real line by the Schwarz reflection principle. When this is continued recursively, one obtains an analytic description of $\hat{O}((0,3),(1,3),(\infty,3))$ as the quotient of the plane by the $(3,3,3)$ group.

Here is a more formal expansion of the definition for an orbifold: an orbifold Q consists of an underlying space X_Q, together with some additional structure. The structure can be described by covering X_Q with a collection of connected open sets U_i which (for convenience in the definition) is closed under taking components of finite intersections. Each U_i is called a *coordinate patch.* To each U_i is associated a finite group Γ_i, a faithful action of Γ_i on some connected open subset $\widetilde{U}_i$ of $\mathbf{R}^n$, and a homeomorphism, called a *coordinate chart* $\phi_i : U_i \to \widetilde{U}_i/\Gamma_i$. Whenever $U_i \subset U_j$, there is to be an injective homomorphism

$$f_{ij} : \Gamma_i \to \Gamma_j$$

and an embedding

$$\widetilde{\phi}_{ij} : \widetilde{U}_i \to \widetilde{U}_j$$

equivariant with respect to f_{ij} (that is, for $\gamma \in \Gamma_i$, $\widetilde{\phi}_{ij}(\gamma x) = f_{ij}(\gamma)\widetilde{\phi}_{ij}(x)$), such that the diagram below commutes:

Definition "geometric orbifolds"
Example "manifolds are orbifolds"
Example "mirror of manifold"
::mirror
::silvered
% orbifold with boundary
Example "manifold with cone axes"

$$
\begin{array}{ccc}
\widetilde{U}_i & \xrightarrow{\widetilde{\phi}_{ij}} & \widetilde{U}_j \\
\downarrow & & \downarrow \\
\widetilde{U}_i/\Gamma_i & \xrightarrow{\phi_{ij}=\widetilde{\phi}_{ij}/\Gamma_i} & \widetilde{U}_j/f_{ij}(\Gamma_i) \\
\uparrow & & \downarrow \\
\Big| \phi_i & & \widetilde{U}_j/\Gamma_j \\
 & & \uparrow \phi_j \\
U_i & \subset & U_j
\end{array}
$$

We regard $\widetilde{\phi}_{ij}$ as being defined only up to composition with elements of Γ_j, and f_{ij} as being defined up to conjugation by elements of Γ_j. It is not generally true that $\widetilde{\phi}_{ik} = \widetilde{\phi}_{jk} \circ \widetilde{\phi}_{ij}$ when $U_i \subset U_j \subset U_k$, but there should exist an element $\gamma \in \Gamma_k$ such that $\gamma\widetilde{\phi}_{ik} = \widetilde{\phi}_{jk} \circ \widetilde{\phi}_{ij}$ and $\gamma \cdot f_{ik}(g) \cdot \gamma^{-1} = f_{jk} \circ f_{ij}(g)$, for each $g \in \Gamma_i$.

Of course, the covering $\{U_i\}$ is not an intrinsic part of the structure of an orbifold; two coverings of a space X give rise to the same orbifold structure if there is a larger cover, containing each of the given covers, and still satisfying the definitions.

The notion of a geometric structure can be carried over to orbifolds in a straightforward way:

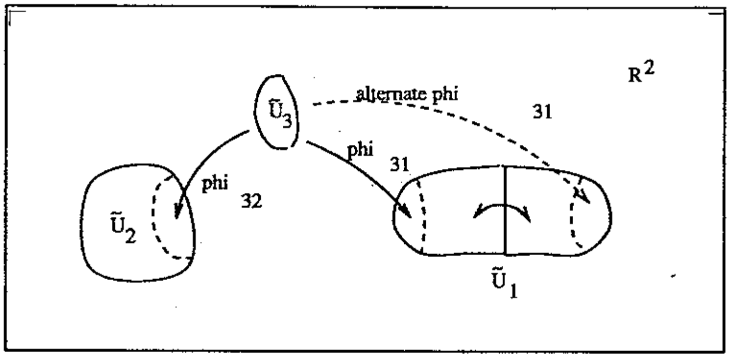
Definition 5.2.2 (geometric orbifolds). A Γ-orbifold, where Γ is a pseudogroup of local homeomorphisms of $\mathbf{R}^n$, means that in the definition all maps and group actions are contained in the the pseudogroup Γ. As with manifolds, if G is a group acting on a space X, a (G, X)-orbifold is the same as a Γ orbifold, where Γ is the pseudogroup of restrictions of elements of G to some coordinate patch in X.

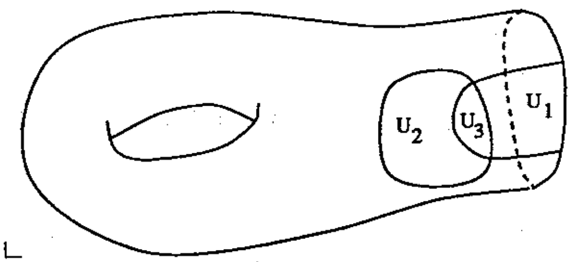
Example 5.2.3 (manifolds are orbifolds). A closed manifold is an orbifold, where each group Γ_i is the trivial group, so that $\widetilde{U}_i = U_i$.

Example 5.2.4 (mirror of manifold). A manifold M with boundary can be given an orbifold structure mM in which its boundary becomes a *mirror.* Any point on the boundary has a neighborhood modeled on $\mathbf{R}^n/\mathbf{C}_2$, where $\mathbf{C}_2$ acts by reflection in a hyperplane. We will say that a boundary component of a manifold which has been given an orbifold structure in this way is *silvered.* One can imagine living inside a manifold whose walls are glass; some of the walls may be transparent, while others may be silvered so that they have become mirrors. The unsilvered walls are still boundary components (see definition 5.2.8.)

Example 5.2.5 (manifold with cone axes). If M is a differentiable n-manifold, and $N \subset M$ is any codimension two submanifold, then a local

Proposition "quotients of manifolds are orbifolds"





orbichart

Figure 5.7. A manifold with boundary underlies a mirrored orbifold. A manifold M with boundary is the underlying space for an orbifold mM (without boundary), by making the boundary into a mirror. The figure shows how a covering works near the boundary of M.

model for the pair (M, N) is $(\mathbf{R}^n, \mathbf{R}^{n-2})$. For any integer $k \geq 1$, this pair is homeomorphic to $\mathbf{R}^n/\mathbf{C}_k$, where $\mathbf{C}_k$ acts by rotations about $\mathbf{R}^{n-2}$. Therefore there is an orbifold structure on M which makes N into an order k cone axis. This structure can be made differentiable in a natural way.

This orbifold is 'M with order k N', symbolized $M(kN)$.

Proposition 5.2.6 (quotients of manifolds are orbifolds). *The quotient space of a manifold M by a group Γ which acts properly discontinuously on M is an orbifold.*

Proof of 5.2.6: For any point $x \in M/\Gamma$, choose $\widetilde{x} \in M$ projecting to x. Let I_x be the stabilizer of $\widetilde{x}$. (I_x depends, of course, on the particular choice of $\widetilde{x}$.) There is a neighborhood $\widetilde{U}_x$ of $\widetilde{x}$ invariant by I_x and disjoint from its translates by elements of Γ not in I_x. The projection induces a homeomorphism from $\widetilde{U}_x/I_x$ to U_x. To obtain a suitable covering of M/Γ, augment some covering $\{U_x\}$ by adjoining components of finite intersections. If V is a component of $U_{x_1} \cap \ldots \cap U_{x_k}$, then there is a corresponding component $\widetilde{V}$ of some set of translates $\gamma_1\widetilde{U}_{x_1} \cap \ldots \cap \gamma_k\widetilde{U}_{x_k}$. The group $\gamma_1 I_{x_1}\gamma_1^{-1} \cap \ldots \cap \gamma_k I_{x_k}\gamma_k^{-1}$ acts on $\gamma_1\widetilde{U}_{x_1} \cap \ldots \cap \gamma_k\widetilde{U}_{x_k}$ and on its set of components. Let H be the stabilizer of $\widetilde{V}$. Then V, $\widetilde{V}$ and H give the required local structure. 5.2.6

% manifold with cone axes
::singular locus
Proposition "singular category"
% differentiable orbifolds locally orthogonal
Definition "orbifold with boundary"
::orbifold with boundary
Definition "suborbifold"
::suborbifold
::suborbifolds with boundary
::proper suborbifold

The orbifold mM arises in this way, for instance: it is obtained as the quotient space of the C_2 action on the double dM of M which interchanges the two halves. The construction of example of 5.2.5 often produces an orbifold which is a quotient of a manifold by a group action, but not always.

Henceforth, we shall use the terminology M/Γ to mean M/Γ as an orbifold.

Note that each point x in an orbifold Q is associated with a group Γ_x, well-defined up to isomorphism. In a local coordinate system $U = \widetilde{U}/\Gamma$, Γ_x is the stabilizer of any point in $\widetilde{U}$ mapping to x. (Alternatively, Γ_x may be defined as the smallest group associated with the various different coordinate patches containing x.) The set $\Sigma_Q = \{x \mid \Gamma_x \neq \{1\}\}$ is the *singular locus* of Q. We shall say that Q is a manifold when $\Sigma_Q = \phi$.

Warning: It happens often that the underlying space X_Q is a topological manifold, especially in dimensions 2 and 3, although Q is not a manifold. Do not confuse properties of Q with properties of X_Q.

Proposition 5.2.7 (singular category). *The singular locus of an orbifold is a closed set with empty interior.*

Proof of 5.2.7: For any coordinate patch $U = \widetilde{U}/\Gamma$, $\Sigma_Q \cap U$ is the image of the union of the fixed point sets in U of elements of Γ. Hence $\Sigma_q \cap U$ is closed, so Σ_Q is closed.

The fact that Σ_Q has empty interior is a consequence of the theorem of [New31] that a non-trivial homeomorphism of a manifold which has finite order cannot fix any open set. This theorem is deep in the topological case, but it is elementary in the differentiable case that mainly concerns us. (See 5.4.1 for the discussion in the differentiable case.) 5.2.7

Here are two more definitions which are completely parallel to the definitions for manifolds.

Definition 5.2.8 (orbifold with boundary). An *orbifold with boundary* means a space locally modeled on $\mathbf{R}^n$ modulo finite groups and $\mathbf{R}^n_+$ modulo finite groups.

Warning: Do not confuse the boundary of an orbifold with the boundary of its underlying space (when this makes sense.) The actual boundary of the orbifold depends not just on the picture, but on the data of the local groups as well.

Definition 5.2.9 (suborbifold). A d-dimensional *suborbifold* Q_1 of an orbifold Q_2 means a subspace $X_{Q_1} \subset X_{Q_2}$ locally modeled on $\mathbf{R}^d$ modulo the induced actions of the local groups of Q_2 on invariant d-dimensional subspaces.

We will also have occasion to talk of *suborbifolds with boundary*, and we have a variety of different possibilities. We take for granted that the various actions involved in the local structure are induced in the obvious way by the largest group associated to a point. By a *proper suborbifold* Q_1 of Q_2 we mean that the boundary of Q_1 is a suborbifold of the boundary of Q_2, and no other point of Q_1 is contained in the boundary of Q_2. If an interior point of Q_1 lies in the boundary of Q_2, then the whole of Q_1 must lie in the boundary of Q_2.

Example "(2,3,6) prism"
Example "Borromean suborbifolds"
% borromean orbifold
Problem "suborbifolds and the singular locus"
% differentiable orbifolds locally orthogonal
Section "V-manifold explanation"

The orbifold mI, the mirrored unit interval, is non-orientable and one-dimensional. An endpoint is a zero-dimensional manifold—the group of order two is changed to a group of order one, to satisfy the requirement of the definition that only faithful actions are used. The quotient of S^2 by $\mathbf{C}_k$ acting by rotations has two cone points. A geodesic joining the two cone points gives a suborbifold if and only if $k = 2$.

Example 5.2.10 ((2,3,6) prism). Consider a rectangular prism in euclidean three space, with triangular base, having angles $\pi/2$, $\pi/3$ and $\pi/6$. Suppose all the faces are mirrors and all edges are corner reflectors. Then the triangular faces are suborbifolds and the edges perpendicular to the trianglular faces are suborbifolds. No other faces or edges are suborbifolds. To see this phenomenon physically, build a large box of prismatic shape, make one two triangular mirrors (reflecting inwards) with red glass. Step inside the box and replace the red triangular lid. You see an infinite sequence of parallel red floors and ceilings. If an edge of a red triangle on the floor is painted black, you see on the floor a pattern of black edges which is not a line. A vertical edge does however give rise to a line.

Example 5.2.11 (Borromean suborbifolds). In example 5.1.8 of the three-sphere with order 2 Borromean rings, consider a disk in S^3 whose boundary is one of the three rings and which intersects the other rings transversely in two points. This is a suborbifold. The boundary of the disk is locally modeled on a plane modulo a reflection in a line in the plane. At the transverse intersections with cone axes, it is locally modeled on a plane modulo rotations about a perpendicular line.

Problem 5.2.12 (suborbifolds and the singular locus). Let Q_1 be a suborbifold of Q_2 and suppose that a non-singular point of Q_1 is a singular point of Q_2. Prove that Q_1 is contained in the singular locus of Q_2. (The topological version of this uses Newman's theorem [New31]. The differentiable version can be proved using only proposition 5.4.1.)

trivial?

5.2.13 Note on history and terminology

The concept of an orbifold has been implicitly used by mathematicians for a long time – by Poincaré, for instance – but the absence of a formalized definition has often created awkwardness in language and communication, and a consequent distortion in the development of mathematics.

A formal definition for a V-manifold was introduced by [Sat56, Sat57], although in his definition he requires Σ_Q to have codimension at least 2, and the definition was based on the Riemannian geometry rather than modelling modulo groups. There has been a fair amount of literature treating V-manifolds.

The term "orbifold" was coined during my course at Princeton in 1976-77, when I was not aware of the earlier definition of V-manifolds. The word was obtained by a democratic process after earlier terminology proved unsuccessful. The word "orbifold" seems to suggest the correct concept to people more quickly than "V-manifold", so I have decided to stick with it. The "V" of

"V-manifold" is meant to suggest the cone-like shape near its singular locus. Unfortunately, people first encountering the word tend to assume a V-manifold is a kind of manifold. "Orbifold" suggests the orbit space of a group action on a manifold. Unfortunately, it tends to look like nothing when abbreviated (for symbolic use), so I have found it expedient to add a tail to $O \to Q$ to give a cue that it is not 0.

Section "Covering orbifolds and the fundamental group"
Definition "covering orbifold"
::covering orbifold
% rectangular pillow
% billiard table
% paperdolls

5.3. Covering orbifolds and the fundamental group

When M is a simply-connected manifold and Γ is a properly discontinuous group, we would like to be able to reconstruct Γ as the "fundamental group" of M/Γ, and M as the "universal covering" of M/Γ, just as in the case that Γ acts freely. In this section we will show how to make the reconstruction.

Definition 5.3.1 (covering orbifold). A *covering orbifold* of an orbifold Q is an orbifold $\widetilde{Q}$, with a projection $p : X_{\widetilde{Q}} \to X_Q$ between the underlying spaces, such that

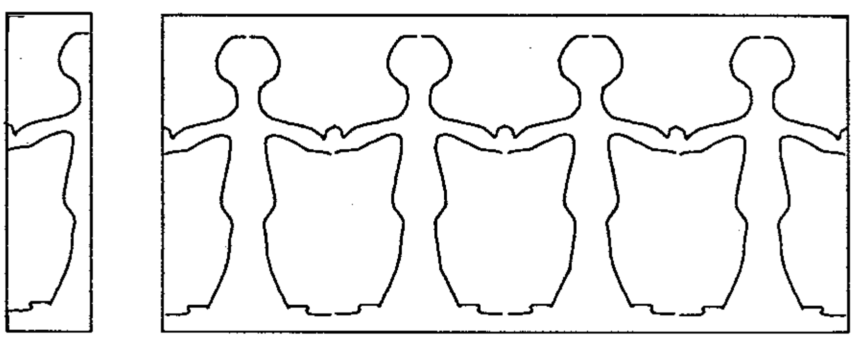
i) p is a *local covering* that is, each point $x \in X_{\widetilde{Q}}$ in the domain has a neighborhood $U = \widetilde{U}/\Gamma$ (where $\widetilde{U}$ is an open subset of $\mathbf{R}^n$) such that p restricted to U is isomorphic to a map $\widetilde{U}/\Gamma \to \widetilde{U}/\Gamma'$ $(\Gamma \subset \Gamma')$

and

ii) p is an *even covering*, that is, each point $x' \in Q$ in the range has a neighborhood $V = \widetilde{V}/\Gamma$ for which each component U_i of $p^{-1}(V)$ is isomorphic to $\widetilde{V}/\Gamma_i$, where $\Gamma_i \subset \Gamma$ is some subgroup. The isomorphism must respect the projections.

Note that the underlying space $X_{\widetilde{Q}}$ is not generally a covering space of X_Q.

As a basic example, when Γ is a group acting properly discontinuously on a manifold M, then M is a covering orbifold of M/Γ. In fact, for any subgroup $\Gamma' \subset \Gamma$, M/Γ' is a covering orbifold of M/Γ. Thus, the rectangular pillow (5.1.5) is a two-fold covering space of the billiard table (5.1.4)

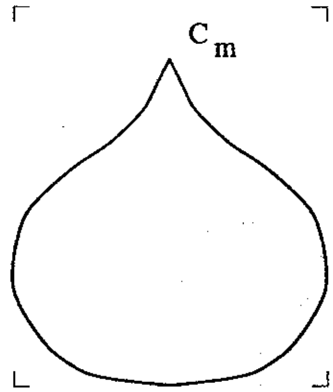


paperdolls

Figure 5.8. Covering spaces of a mirrored strip. The construction of covering spaces of a mirored strip by larger mirrored strips is often taught to five year olds by folding paper dolls. (Notation: A mirrored strip is also called $({}^*\infty\infty)$.)

Here is another explicit example to illustrate the notion of covering orbifold. Let S be the infinite strip $0 \le x \le 1$ in $\mathbf{R}^2$; consider the orbifold mS (named $({}^*\infty\infty)$). This orbifold is a covering space of itself with any number of sheets, by folding back and forth. An eight-sheeted covering is depicted in figure 5.8.

Definition "good and bad"
::good
::bad
% teardrop
% 235orb
Proposition "universal cover of orbifold"
% universal cover of orbifold



teardrop

Figure 5.9. The teardrop is a bad orbifold. It has no covering space other than itself. (Notation: (m))

Definition 5.3.2 (good and bad). An orbifold is *good* if it has some covering orbifold which is a manifold. Otherwise it is *bad.*

The teardrop (figure 5.9) is an example of a bad orbifold. The underlying space for a teardrop is S^2. Its singular locus consists of a single point, whose neighborhood is modeled on $\mathbf{R}^2/\mathbf{C}_n$, where $\mathbf{C}_n$ acts by rotations. By comparing possible coverings of the upper half with possible coverings of the lower half, you may easily see that the teardrop has no non-trivial connected covering.

Similarly, you may verify that an orbifold $Q = (\_m)$ with underlying space $X_Q = S^2$ having two cone points of orders n and m is bad unless $m = n$.

Two-dimensional orbifolds with three or more cone points, as we shall see, are always good. For example, the orbifold with underlying space S^2 having three cone points of orders 2, 3 and 5 is isomorphic to S^2 modulo the group of orientation preserving symmetries of a dodecahedron (figure 5.10)

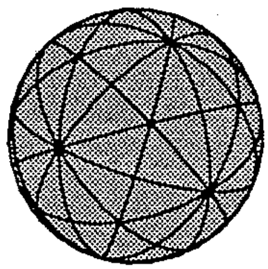
Proposition 5.3.3 (universal cover of orbifold). *An orbifold Q has a universal cover. In other words, if $x \in X_Q - \Sigma_Q$ is a base point for Q then the universal covering $p : \widetilde{Q} \to Q$ is a connected covering with base point $\widetilde{x}$ (with $p(\widetilde{x}) = x$), such that for any other covering $p' : \widetilde{P} \to Q$ with base point $\widetilde{x}'$ (and $p(\widetilde{x}') = x$), there is a unique lifting $q : \widetilde{Q} \to \widetilde{P}$ of p to a covering map of $\widetilde{P}$.*

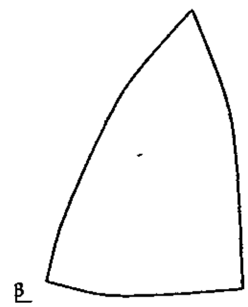
This proof has been refurbished – wpt

The universal covering orbifold $\widetilde{Q}$ corresponds to what is sometimes called the universal branched covering of X_Q subject to certain branching conditions along Σ_Q. The universal covering is a manifold iff Q is good. In that case, it is the same as the universal covering space of any manifold which is a covering orbifold of Q. One fairly elementary method to prove 5.3.3 goes in outline as follows:

First define the orientation covering Q_1 of Q, a two-fold covering orbifold which is oriented. The singular locus Σ_{Q_1} now has codimension at least 2. The universal covering of Q is obtained from a certain covering space of $Q_1 - \Sigma_{Q_1}$

% fiberprod





235orb

Figure 5.10. The 235 orbifold. The orbifold (235) with underlying space S^2 and cone points of orders 2, 3 and 5 is the quotient space of S^2 by the group of orientation-preserving isometries of the dodecahedron.

by "filling it back in" above Σ_{Q_1}. The desired covering space of $Q_1 - \Sigma_{Q_1}$ has fundamental group the subgroup generated by all loops which are freely homotopic to the form α/Γ, in some local coordinate system $\widetilde{U}/\Gamma$, where α is a loop in $\widetilde{U}$.

We will give a different proof, because it seems more illuminating.

Proof of universal cover of orbifold: The key step is to be able, given a collection of coverings $Q_i \to Q$, to find another covering $P \to Q$ which is a common covering of the Q_i.

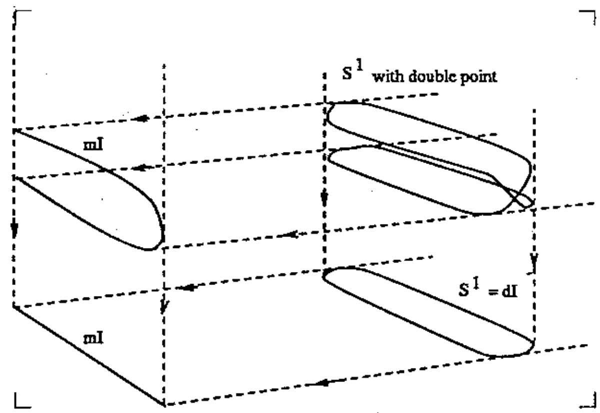
For the construction of the universal covering space of a topological space, the fiber product works toward this effect. Recall that the fiber product of a collection of maps $f_i : X_i \to X$ of topological spaces is the subset of the cartesian product consisting of all tuples in $\prod_i X_i$ that map to the same point in X: $\{(x_i)_{i\in I} \in \prod_{i\in I} X_i \mid \forall i, \in I,\ f_i(x_i) = f_j(x_j)\}$.

In the case of orbifolds, it does not quite work to take the fiber product of the underlying spaces. The difficulty is best illustrated by example. Two covering maps $S^1 = dI \to mI$ and $mI \to mI$ are shown in figure 5.11, along with the fiber product of the underlying spaces. There is an extraneous double point, which we must eliminate in the right definition of fiber product of covering maps of orbifolds.

First let's see what happens in local coordinates. Let $U \cong \widetilde{U}/\Gamma$ be a coordinate patch. We may suppose that U is sufficiently small so that in every covering of Q, $p^{-1}(U)$ consists of components of the form $\widetilde{U}/\Gamma'$, where $\Gamma' \subset \Gamma$.

Let $Q_i \xrightarrow{p_i} Q$ be a collection of covering orbifolds, and consider components V_j of $p_i^{-1}(U)$. Each V_j has the form $\widetilde{U}/\Gamma_j$, for some group $\Gamma_j \subset \Gamma$, where j ranges over the set J of all components for all i.

The esistence of such a sufficintly small U is not trivial in the topological case-Dick



fiberprod

Figure 5.11. The orbifold fiber product is not the topological fiber product. There are often extraneous intersections in the topological fiber product. In this case, the orbifold fiber product of degree two coverings of $mI \to mI$ and $S^1 \to mI$ is a degree four covering $S^1 \to mI$, while the topological fiber product of the maps between underlying spaces is a figure eight. The orbifold fiber product takes into account the local groups to separate these intersections into different sheets, making it an orbifold.

If x is a regular point in U, then preimages of x in V_j are parametrized by the cosets $\Gamma_j\gamma \in \Gamma/\Gamma_j$. (Once we pick some $\widetilde{x} \in \widetilde{U}$ which maps to x, then the preimage in $\widetilde{U}$ is $\Gamma\widetilde{x}$ and the preimage in V_j is $\Gamma\widetilde{x}/\Gamma_j \cong \Gamma/\Gamma_j$.) Then the part of the fiber product mapping to x is parametrized by the product

$$\prod_j (\Gamma/\Gamma_j) = \prod_j \Gamma / \prod_j \Gamma_j.$$

This expression has the disadvantage that it depends on the choice of an $\widetilde{x}$ lying above x. To get a uniform formula, we use all possible choices for $\widetilde{x}$ and divide by the action of Γ which permutes these choices. The result is

$$V = (\prod_{j\in J} \Gamma \times \widetilde{U})/(\Gamma \times \prod_{j \in J} \Gamma_j)$$

where the action of $\Gamma \times \prod_j \Gamma_j$ is

$$(\gamma, (\gamma_j)_{j\in J}) \colon ((\alpha_j)_{j\in J}, u) \mapsto ((\gamma_j \alpha_j \gamma^{-1})_{j \in J}, \gamma u).$$

In these formulas, when the index set is infinite, the products are unrestricted products equipped with the discrete topology.

The covering maps $V \to \widetilde{U}/\Gamma_j$ are defined by sending $((\alpha_i)_{i\in I}, \widetilde{u})$ to $\alpha_j \widetilde{u}$. Note that this respects the equivalence relations. If $u \in U$ is any regular point this construction agrees with the usual fiber product description, so it is the same as the topological construction.

It is clear that this construction gives a covering orbifold of U.

Definition "fundamental group"

::fundamental group
% no universal cover
% no fundamental group

These local constructions piece together to give a kind of global fiber product. (The piecing together automatically works because the local construction has the universal property that any common covering of the $(\widetilde{U}/\Gamma_i)$ has a unique factorization as a covering of V). Note that V has many components, reflecting the fact that no choice of base points was made.

To form a universal covering of an orbifold Q, take a set I of coverings representing all possible isomorphism classes of covering, form the fiber product of all these coverings, and take the component of the base point.

universal cover of orbifold

The universal covering $\widetilde{Q}$ of an orbifold Q is automatically a regular covering: for any preimage $\widetilde{x}'$ of the base point $\widetilde{x}$ there is a deck transformation taking $\widetilde{x}$ to $\widetilde{x}'$.

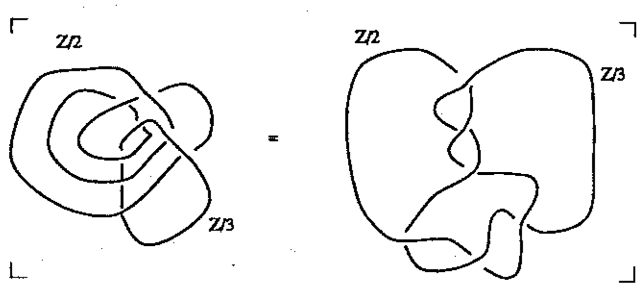
Definition 5.3.4 (fundamental group). The *fundamental group* $\pi_1(Q)$ of an orbifold Q is the group of deck transformations of the universal covering $\widetilde{Q}$.

Perhaps something ought to be said about the group of deck transformations of the fiber product-Dick

Although this is the most natural definition for the fundamental group of an orbifold, it is not usually the best way to compute the fundamental group. As with manifolds, the fundamental group can also be described in terms of closed loops in the orbifold through a base point, or in terms of a cell division of the orbifold. Loops are trickier than in manifolds, however, because of ambiguities that can arise from the singular locus. A closed curve in the underlying space X_Q determines an element of the fundamental group only if it is equipped with enough additional information to determine continuations of liftings to arbitrary covering space. To describe in detail how this works must wait until we have analyzed the nature of the singular locus. That this analysis must be special rather than general is really a consequence of the phenomena of exercises 5.3.7 and 5.3.8.

Covering spaces of orbifolds can be very useful for understanding examples.

The labelling is wrong in figure 5.12. dbae



borromcov

Figure 5.12. The quotient of the Borromean rings by a threefold symmetry. The three-sphere with order 2 Borromean rings is a 3-fold covering space of the 3-sphere with order 2 A and order 3 B, where A and B are circles linked as illustrated. The latter orbifold has a 2-fold covering which is S^3 with order 3 figure eight knot.

% borromean orbifold
% borromcov
% borromean orbifold
% order 3 figure eight
Challenge "order 3 figure eight"
% borromcov
% borromean orbifold
% borromcov
% no universal cover
% no fundamental group
% simplicial orbifolds
Problem "fiber product of bands"
Problem "no universal cover"
Problem "no fundamental group"
Problem "simplicial orbifolds"
% singular category

For example, the Borromean ring example (5.1.8) has a 3-fold axis of symmetry, so it is a 3-fold covering space as shown in figure 5.12. A two-fold covering space formed from the latter orbifold by unwrapping around the C_2 axis which is the image of the rings is the figure eight knot labeled C_3. These two orbifolds obtained from our original example 5.1.8 are also quotient spaces of Euclidean crystallographic groups (challenge 5.3.5). The figure eight knot example is hard to see directly, at least compared to the Borromean ring example.

Challenge 5.3.5 (order 3 figure eight). Show that the two orbifolds derived in figure 5.12 are quotient spaces of Euclidean crystallographic groups, by using example 5.1.8, finding an appropriate supergroup of index three and then a subgroup of index two in the supergroup. Study how these groups act on a fundamental domain, so you can relate them to figure 5.12.

It is tempting to try to generalize the theory of orbifolds to encompass more complicated topology such as simplicial complexes modulo properly discontinuous groups. This does not work well. In particular, problem 5.3.7 and problem 5.3.8 show that the local description of the quotient space is not sufficient to construct either a universal covering space or a fundamental group, without adding an undesirable condition to the definition (exercise 5.3.9).

Problem 5.3.6 (fiber product of bands). Let C be a cylinder and M be a Moebius band. Both mC and mM are double coverings of mC, by reflecting in the median circle. What is the orbifold fiber product of these coverings? What is the smallest regular covering which factors through both of these coverings?

Problem 5.3.7 (no universal cover). Consider the complex A obtained by gluing a disk to the median circle of C, and the complex B obtained by gluing a disk to the median circle of M. There are $\mathbf{C}_2$ actions on A and B, such that the quotient spaces, as modeled on simplicial complexes modulo finite groups, are isomorphic. Therefore, the quotient does not have a universal covering, using definitions parallel to the definitions for orbifolds.

Problem 5.3.8 (no fundamental group). There is a $(\mathbf{C}_2)^2$ action and a $(\mathbf{C}_4)$ action on the letter **H** whose quotient spaces are isomorphic, as spaces locally modeled on simplicial complexes modulo finite groups.

Problem 5.3.9 (simplicial orbifolds). What additional hypothesis is necessary to make it possible to reconstruct a universal covering and a fundamental group from a space locally modeled on simplicial complexes modulo finite groups? [Hint: consider proposition 5.2.7.]

Section "Geometric structures on orbifolds and the developing map"
Proposition "differentiable orbifolds locally orthogonal"
Proposition "local picture of 2-orbifolds"
% dihedral
% classification of 2-orbifolds
% The developing map

5.4. Geometric structures on orbifolds and the developing map

The singular locus of a topological orbifold can be wild, that is, it can have a complicated topological structure even locally. A discussion of this is outside our scope, so from now on we will focus on differentiable orbifolds.

The singular locus of a differentiable orbifold may be understood as follows. Let $U = \widetilde{U}/\Gamma$ be any local coordinate system. There is a Riemannian metric on $\widetilde{U}$ invariant by Γ: such a metric may be obtained by beginning with any Riemannian metric on $\widetilde{U}$, and averaging by the finite group Γ. For any point $\widetilde{x} \in \widetilde{U}$ consider the exponential map, which gives a diffeomorphism from the ε ball in the tangent space at $\widetilde{x}$ to a small neighborhood of $\widetilde{x}$. Since the exponential map commutes with the action of the isotropy group of $\widetilde{x}$, it gives an isomorphism between a neighborhood of the image of $\widetilde{x}$ in Q, ind a neighborhood of the origin in the orbifold $\mathbf{R}^n/\Gamma$, where Γ is a finite subgroup of the orthogonal group $O(n)$.

Proposition 5.4.1 (differentiable orbifolds locally orthogonal). *A differentiable n-orbifold is locally modeled on $\mathbf{R}^n$ modulo finite subgroups of the orthogonal group. The local coordinate systems are glued together by diffeomorphisms.*

Proposition 5.4.2 (local picture of 2-orbifolds). *The singular locus of a two-dimensional orbifold has these types of local models:*

i) *The mirror, $\mathbf{R}^2/\mathbf{C}_2$, where $\mathbf{C}_2$ acts by reflection in the y-axis.*

ii) *Cone points of order n: $\mathbf{R}^2/\mathbf{C}_n$, with $\mathbf{C}_n$ acting by rotations.*

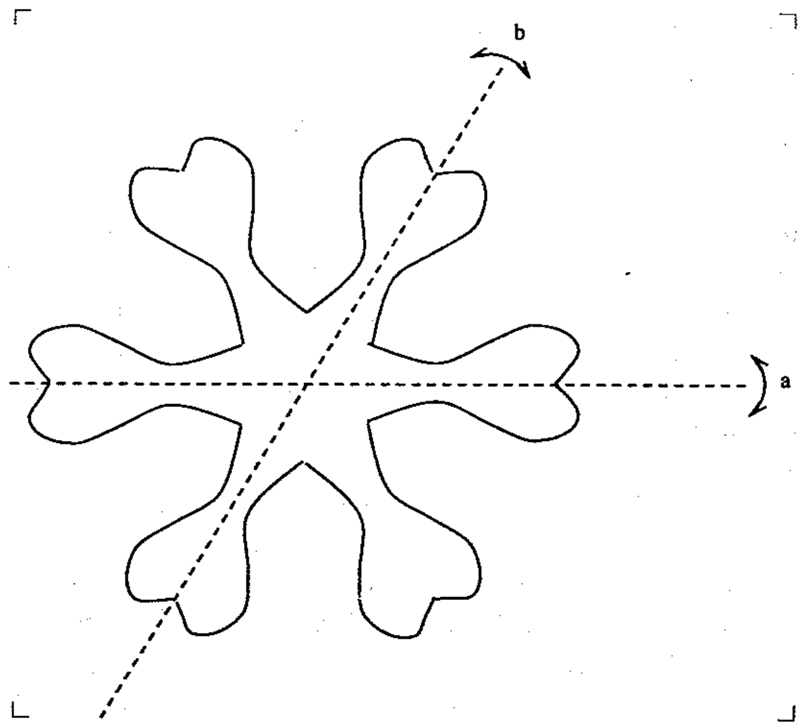
iii) *Corner reflectors of order n: $\mathbf{R}^2/\mathbf{D}_{2n}$ (where $\mathbf{D}_{2n}$ is the dihedral group of order $2n$). The action of $\mathbf{D}_{2n}$ is generated by reflections in two lines which meet at an angle of π/n (figure 5.13).*

Proof of 5.4.2: These are the only three types of finite subgroups of $O(2)$. [5.4.2]

It follows that the underlying space of a two-dimensional orbifold is always a topological surface, possibly with boundary. This makes it easy to enumerate all two-dimensional orbifolds, by enumerating surfaces together with combinatorial information which determines an orbifold structure. It is not very easy with strictly topological methods to determine which of these orbifolds are good. By studying orbifolds with geometric structures, however, we shall find (theorem 5.5.3) that we can sort them out in a straightforward way.

If X is a real analytic manifold and G is a group of analytic diffeomorphisms of X, then an orbifold with a (G, X) structure has a developing map, just like a manifold (section 3.6).

Proposition "orbifold development"
% holonomy exists
% DEF CMPLT
% closed implies complete
Proposition "closed (G, X) orbifolds are complete"



dihedral

Figure 5.13. The dihedral group D_{12}. The group D_{12} is illustrated. It is generated by reflections in lines a and b, which meet in a 60° angle.

Proposition 5.4.3 (orbifold development). *Let G be a group of real analytic diffeomorphisms of a real analytic manifold X. Every (G, X) orbifold Q is good, and there is a developing map*

$$D : \widetilde{Q} \to X$$

from the universal covering of Q to X. The map D is a local (G, X) homeomorphism, and is unique up to compositions of the form $g \circ D$ $(g \in G)$.

The *holonomy* homomorphism

$$H : \pi_1(Q) \to G$$

is defined just as for manifolds (see proposition ??). A (G, X) orbifold is *complete* if the developing map D is a covering map (compare definition ??).

Here is another useful proposition, which parallels ?? for manifolds. This is essentially a case of Poincaré's famous theorem on fundamental polyhedra ([Poi82, Poi83], and [Mas71]):

Proposition 5.4.4 (closed (G, X) orbifolds are complete). *If G is a group of analytic diffeomorphisms acting transitively on X such that each isotropy group G_X is compact, then every compact (G, X) orbifold without boundary is complete.*

% orbifold development
% The developing map
% covering orbifold
% closed (G, X) orbifolds are complete
% closed implies complete

Proof of 5.4.3: The reason that the existence of the developing map does not immediately reduce to what everyone knows about analytic continuation (in contrast to section 3.6) is that one does not know ahead of time that $\widetilde{Q}$ is a manifold, so one does not know ahead of time that analytic continuation of a map from $\widetilde{Q}$ to X works even locally.

To form a picture that will give a quick proof, consider the set $G(Q)$ of all germs of local (G, X) covering maps from X to Q. A *germ* of a map from a space A to another space B is an object determined by a source point $a \in A$ and a map f to B defined on some neighborhood of a. Two maps f_1 and f_2 have the same germ at a if their restrictions to any sufficiently small neighborhood of a are the same. The point $f(a) \in B$ is called the target of the germ.

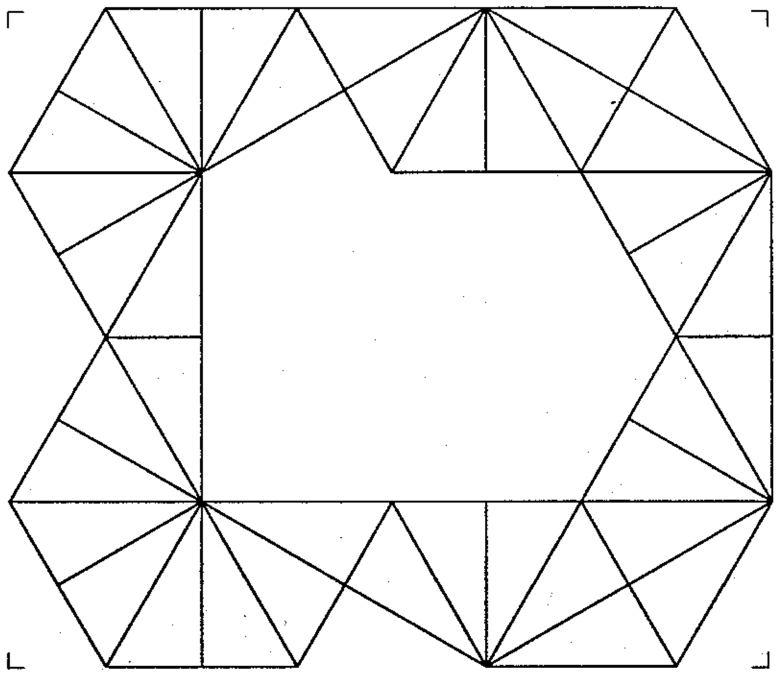
To help develop the picture, consider the case that Q is X itself. The space $G(X)$ is $G \times X$, which we can parametrize so that $x \in X$ determines the target and $g \in G$ determines the map. (Hence, the source point is $g^{-1}x$.) In general, if $U = \widetilde{U}/\Gamma$ is a local coordinate system in Q with $\widetilde{U} \subset X$, then $G(U) = (G \times \widetilde{U})/\Gamma$, with $\gamma \in \Gamma$ acting on (g, u) by sending it to $(\gamma g, \gamma u)$. Observe that the source point is unchanged by this action, so that the local covering map over U is unchanged. In particular, note that if G is a Lie group, $G(Q)$ is a manifold, since the action of Γ on $G \times \widetilde{U}$ is a free action. $G(Q)$ has a kind of foliation (actually a foliation provided G is a Lie group) whose leaves map as local coverings to Q. A leaf of this foliation is defined locally as the set of germs which come from a single local (G, X) covering map. Formally, we define the leaf topology on $G(Q)$, with a neighborhood basis consisting of sets of germs determined by a single local covering map from an open set of X to Q, and a leaf is a component in this topology. By definition, $G(Q)$ with this topology maps to Q at least as a local covering map. Because G is analytic, the leaves are Hausdorff. From the description of $G(U)$, it follows that the map to Q is an even covering (see definition 5.3.1(ii)). Therefore, each leaf of $G(Q)$ is a covering space of Q. On the other hand, each leaf maps as a local covering to X; this map is obtained by projecting the germs to the sources, rather than the targets. Thus we construct the developing map, and at the same time show that Q is good.

In summary: it is awkward to construct the developing map in isolation, so we have instead constructed all possible developing maps at once. The union of the graphs of all possible developing maps (thought of in the form of multi-valued maps from Q to X) can be nicely visualized as a foliation of $G(Q)$, and their properties are easy to verify formally. 5.4.3

Proof of 5.4.4: The proof that closed (G, X) orbifolds are complete is the same as in the case of manifolds (proposition ??). First construct a Riemannian metric for X which is invariant by G. This produces a Riemannian metric for every (G, X) orbifold (i.e. for every neighborhood $U = \widetilde{U}/\Gamma$ an inner product on the tangent space of $\widetilde{U}$ invariant by Γ. In particular the orbifold inherits a distance function from X). If Q is a compact (G, X) orbifold without

% closed (G,X) orbifolds are complete
% 236develop
% closed (G,X) orbifolds are complete

boundary, then there is some ε such that the ε−neighborhood of any point in $\widetilde{Q}$ develops homeomorphically to X. Therefore the map to X is an even covering, hence a covering map. [5.4.4]



236develop

Figure 5.14. The development of (*236) **.** The $90°, 60°, 30°$ Euclidean triangle is an orbifold. Its developing map consists of repeatedly reflecting the triangle through its sides. By proposition 5.4.4, the pattern is discrete.

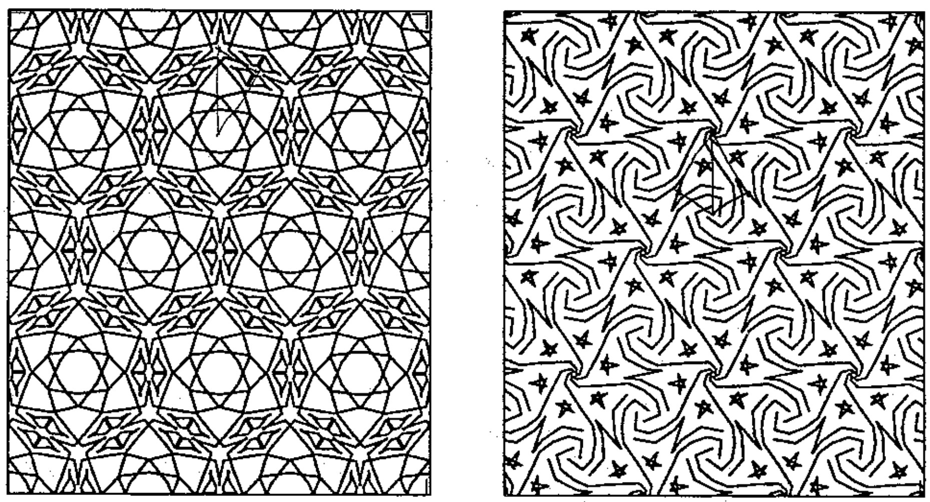
To illustrate proposition 5.4.4, consider a triangle whose sides are mirrors and whose vertices are corner reflectors of orders 2, 3 and 6. This triangle has a Euclidean structure as a $30°, 60°, 90°$ triangle. The developing map is shown in figure 5.14. Proposition 5.4.4 implies that if one begins with a $30°, 60°, 90°$ triangle placed in the Euclidean plane and begins flipping it over, always keeping at least one side in contact with the plane, then whenever it returns to its original location it is also in its original orientation.

Section "The geometric classification of 2-dimensional orbifolds"
% billiard

5.5. The geometric classification of 2-dimensional orbifolds

We will use a handy notation proposed by Conway to describe 2-dimensional orbifolds in terms of their underlying topology plus the combinatorial structure of the singularities. We have already mentioned it in a few places: for instance, the billiard table of figure 5.3 is (*2222). The * here denotes a string of one or more mirrors connected by corner reflectors, and the numbers indicate the orders of the corner reflectors. Numbers which do not follow a * denote cone points. Thus (2222) is an orbifold with four order-2 cone points.

More generally, an orbifold is described by first listing its orientable features, then listing its unorientable features, separated by a vertical bar which is usually optional. For example, (39|*24) = (39*24) has underlying space a disk, with cone points of orders 3 and 9, and a string of two mirrors meeting at corner reflectors of orders 2 and 4. The cone points and the mirror strings can be listed in an arbitrary order (as long as the cone points are listed first), and the order of the corner reflectors in a string of mirrors can be cyclically permuted.



walla

Figure 5.15. Wallpaper groups: a. These patterns show the symmetry of the full and half triangle groups (*632) and (632). Notice the six lines of reflection passing through the six-pointed stars in the figure at left. On the right, there are no orientation-reversing symmetries. Six stars whirl out from each point of order six rotational symmetry. These patterns, and the other wallpaper patterns in this section, were generated using *kali*, a program written by Annamaria Amenta

The default topology for the underlying space of an orbifold is S^2. Each * causes a disk to be removed, to form a boundary component of the underlying space. Thus, for example, (**) is a silvered annulus.

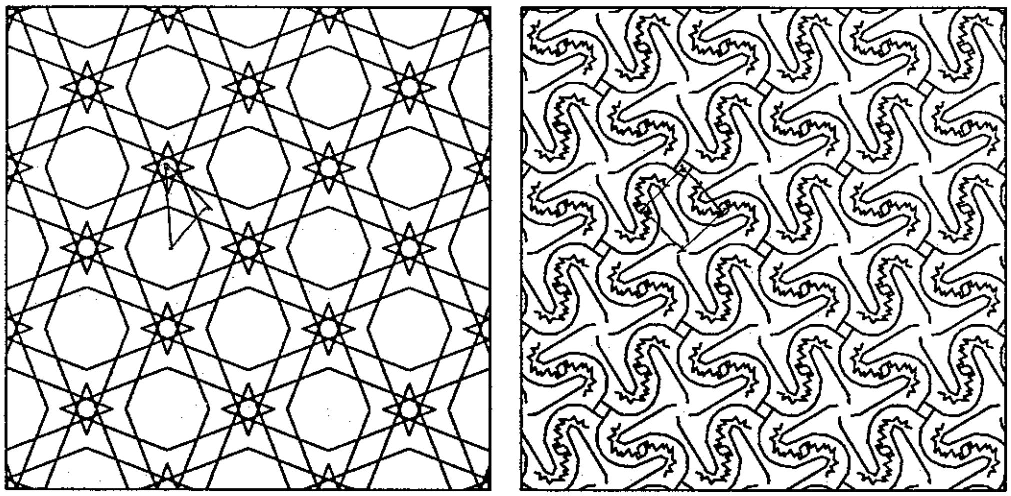
Additional topology is indicated by the symbol ° (pronounced 'circle'). Before the bar, the circle denotes a handle, while after the bar it denotes a

% paperdolls
% waves

cross-cap. This is in line with the general pattern, that each non-orientable feature is related to some orientable feature, its orientable double cover.

The vertical bar can usually be made redundant by putting all handles first, and all cross-caps last. If there are any intervening cone points or mirror strings, this is sufficient. Only in the case that the orbifold is a manifold is the bar then necessary. Thus $(^\circ|)$ is a torus, while $(|^\circ)$ is a projective plane and $(|^{\circ\circ})$ is a Klein bottle. A silvered Möbius band is $(|^{*\circ}) \doteqdot (^{*\circ})$.

The basic identity for surfaces becomes $(^\circ|^\circ) = (|^{\circ\circ\circ})$.



wallb

Figure 5.16. Wallpaper groups: b. These patterns show the symmetry of the full and half triangle groups (*442) and (442). The reflections in the pattern at left force a very different effect than the rotations at right.

Wherever a number can appear, the symbol ∞ can also appear. If it appears before the bar, it means that one point is removed from the orbifold. If it appears in a string of mirrors, it means one point is removed to separate one mirror from the next. Thus, $(^*\infty\infty)$ is topologically a disk, minus two points on the boundary; the remainder of its boundary is silvered. This is equivalent to the mirrored infinite strip (figure 5.8). Similarly, the quotient orbifold of a pattern which repeats under the action of $\mathbf{Z} = \mathbf{C}_\infty$ acting by translations is an infinite cylinder, denoted $(\infty\infty)$ (figure 5.17)

waves

Figure 5.17. Infinite cyclic symmetry. The quotient orbifold for the symmetries of this wave pattern is an infinite cylinder $(\infty\infty)$.

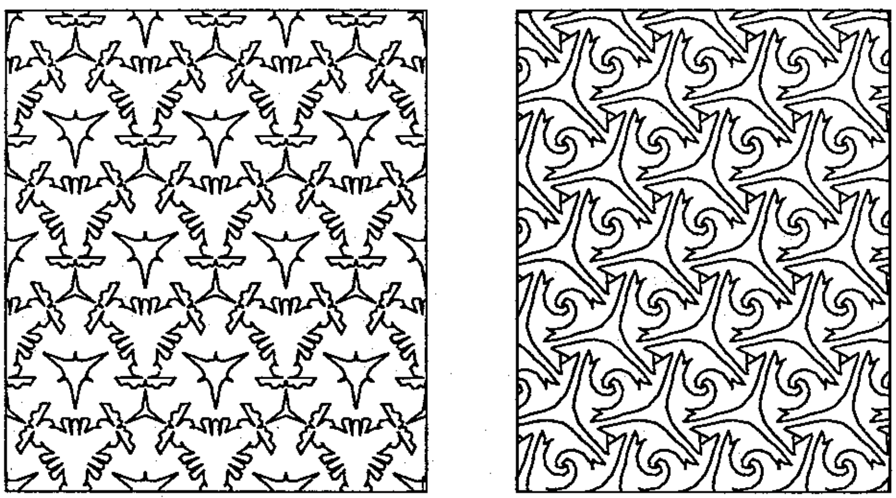
The Euler number for a manifold generalizes to orbifolds. This is especially useful in sorting out 2-orbifolds.

Definition "Euler number of orbifold" ::Euler number ::sheet number Proposition "Euler number multiplies"

Definition 5.5.1 (Euler number of orbifold). Let Q be an orbifold, and suppose that its underlying space has a cell-division such that each open cell is in the same stratum of the singular locus. (That is, the local group associated to the interior points of any cell is constant.) Then the *Euler number* $\chi(Q)$ is defined by the formula

$$\chi(Q) = \sum_{c_i} (-1)^{dim(c_i)} \frac{1}{|\Gamma(c_i)|}$$

where c_i ranges over cells and $|\Gamma(c_i)|$ is the order of the group $\Gamma(c_i)$ associated to the cell.



wallc

Figure 5.18. Wallpaper groups: c. These patterns exhibit symmetry of the full and half triangle groups (*333) and (333). Even with lots of decorations, the pattern at left appears much more regimented.

The Euler number is not always an integer. The definition is concocted for the following reason. Define the *sheet number* of a covering to be the number of preimages of any point in the base which is not a singular point.

Proposition 5.5.2 (Euler number multiplies). *If $\widetilde{Q} \to Q$ is a covering map with sheet number k, then*

$$\chi(\widetilde{Q}) = k\chi(Q)$$

Proof of 5.5.2: The sheet number of a cover can be computed in terms of the behavior of the covering above any point x as the sum

$$\text{sheet number} = \sum_{\{y | p(y)=x\}} \frac{|\Gamma_x|}{|\Gamma_{\widetilde{x}}|}.$$

The formula for the Euler number of a cover follows immediately. $\boxed{5.5.2}$

% 235orb

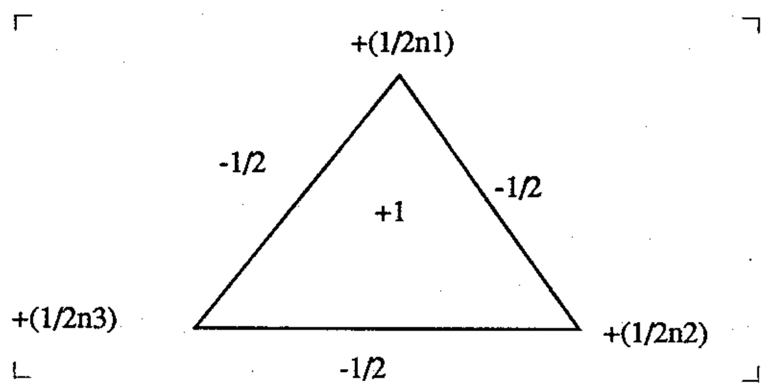
Symbol	full value	half value
°	2	1
*	—	1
m	$(m-1)/m$	$(m-1)/2m$

Table 5.1. Values for Conway's symbols

The Euler number of a 2-orbifold is 2 minus the sum of the values of symbols as indicated above. For instance, (°°23*5*°) is $2-2-2-1/2-2/3-1-2/5-1-1 = -6-29/30$.

To illustrate, a triangle orbifold (*abc) has Euler number

$$\chi((abc)) = \frac{1}{2}(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}-1).$$



eulertriang

Figure 5.19. Computing the Euler number of a triangle orbifold. Computing the Euler number of a triangle orbifold.

Thus, for example (*532) has Euler number $\frac{1}{60}$. Its universal cover is S^2, with deck transformations the group of symmetries of the dodecahedron. (See figure 5.10). This group has order $120 = \frac{2}{1/60}$. On the other hand, the orbifold (*632) has Euler number 0, while (*732) has Euler number $\frac{-1}{84}$. It follows that they cannot be covered by S^2.

The Euler number of a 2-orbifold may be readily computed from Conway's name for it. Each symbol has a certain value, such that the Euler number is 2 minus the sum of these values. The symbol takes full value if it comes before the bar, and half-value afterwards.

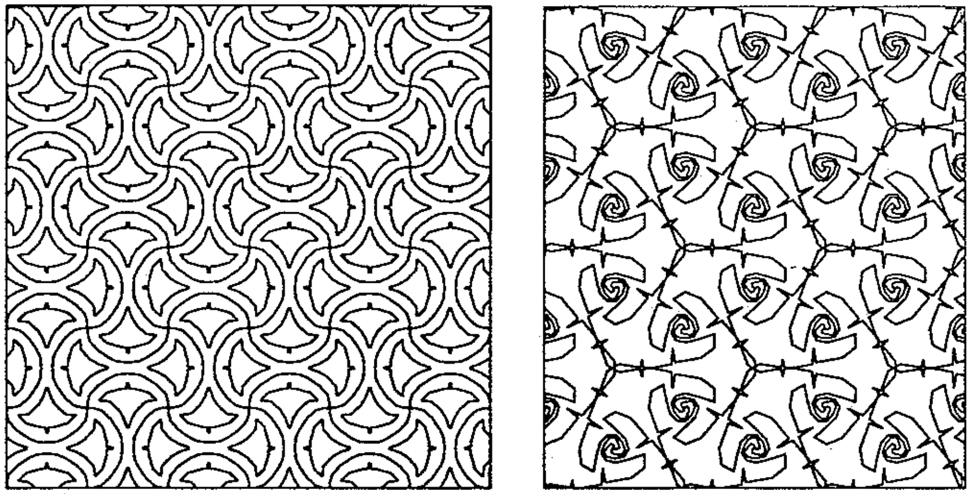
Warning: this orb is not done with the orb macro because of tex in caption difficulties. –wpt

If Q is equipped with a metric coming from invariant Riemannian metrics on the local models $\widetilde{U}$, then one can derive the Gauss-Bonnet formula

$$\int_Q K dA = 2\pi\chi(Q)$$

Theorem "classification of 2-orbifolds"

% orbifold table
% classification of 2-orbifolds



walld

Figure 5.20. Wallpaper groups: d. These patterns exhibit symmetry of the two orbifolds (4*2) and (3*3). These orbifolds are both cones, with one corner reflector on the boundary. Each pattern has one type of point with only rotational symmetry about it, and another type of point through which at least two lines of reflection pass. The pattern at left occurs commonly in tilings of floors, when there are square tiles with grain that alternates between vertical and horizontal.

This may be deduced from the Gauss-Bonnet formula for surfaces by excising from X_Q a small neighborhood of Σ_Q. Observe that along any mirror of Q, the geodesic curvature is 0 (i.e mirrors are geodesics), that corner reflectors of type n have angles of π/n, and that a small circle which encloses an elliptic point of order n has total curvature approximately $2\pi/n$.

It follows that if Q has an elliptic, Euclidean, or hyperbolic structure, it is necessary that $\chi(Q)$ be respectively positive, zero, or negative. Furthermore, if Q is elliptic or hyperbolic, then $area(Q) = 2\pi|\chi(Q)|$.

Theorem 5.5.3 (classification of 2-orbifolds). *A closed, two-dimensional orbifold has an elliptic, Euclidean or hyperbolic structure if and only if it is good. All bad orbifolds have positive Euler number, and the type of geometric structure for a good orbifold is determined by the sign of its Euler number.*

All closed orbifolds which are not hyperbolic are listed in table 5.2.

Proof of 5.5.3: It is routine to list all orbifolds with non-negative Euler number, as shown in the table.

We have already indicated one easy, direct argument to show that the orbifolds are bad when the table claims they are bad, by computing information about the fundamental group. Here is another proof that they are bad, involving the Euler number. First, if the underlying space is D^2, we reduce to the case that the underlying space is S^2 by taking a double cover. If there are two cone points, we may assume their orders are relatively prime, by passing to a cover. The Euler number of (_) is $1 + 1/n$, while the Euler number of

% badorb
% walld

X_Q	Bad	Elliptic	Euclidean
Sphere	$(\_)$ $(\_m)\ n<m$	$(\vert)$ $(\_\_)$ $(22\_)$ (233) (234) (235)	(632) (442) (333) (2222)
Disk	$(^*\_)$ $(^*\_m)\ n<m$	$(^*)$ $(^*\_\_)$ $(^*\_22)$ $(^*332)$ $(^*432)$ $(^*532)$ $(\_^*)$ (2^*m) (3^*2)	$(^*632)$ $(^*442)$ $(^*333)$ $(^*2222)$ (2^*22) (3^*3) (4^*2) (2^*2)
Other		$(\vert^\circ)$ $(\_^\circ)$	(22°) $(^\circ\vert)$ $(\vert^{\circ\circ})$ $(^{**})$ $(^{*\circ})$

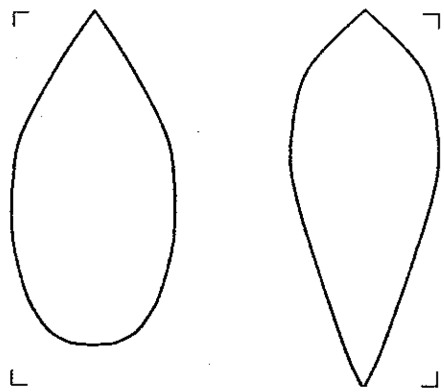
Table 5.2. Classification of non-hyperbolic 2-orbifolds
This table shows all closed orbifolds with positive Euler number, or equivalently, all that do not have a hyperbolic structure. Each is either bad, elliptic, or Euclidean.

$(\_m)$ is $\frac{(n+m)}{nm}$. In either case, the Euler number does not divide 2, so it cannot have S^2 as its universal cover. But these orbifolds have Riemannian metrics of strictly positive curvature (figure 5.21), so by an elementary argument of Riemannian geometry, its universal cover cannot be a noncompact manifold–therefore, it cannot be a manifold.

5.5.4Question: What is the best pinching constant for Riemannian metrics on these orbifolds?

I think this has been answered, either by Calabi or by someone else. One metric is P^3 / C_k.

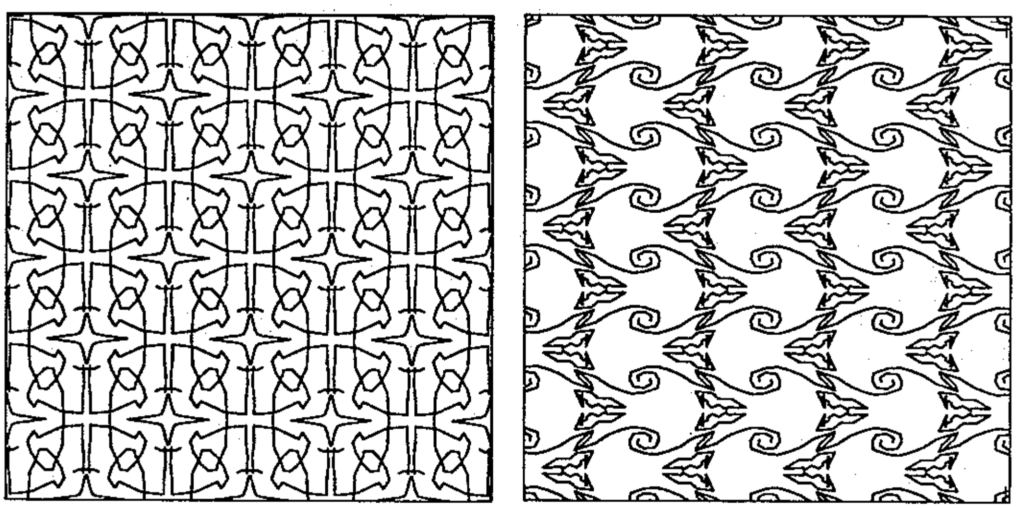
The elliptic or Euclidean structures for the orbifolds may be constructed by either of two methods: they may be identified as the quotient of S^2 or E^2 by a discrete group, or the structure may be directly constructed on the orbifold. For instance, the Euclidean structure on (4^*2) can be constructed by gluing together two adjacent edges of a square. Alternatively, it can be constructed as a discrete group (figure 5.20). The trickiest Euclidean orbifold is (22°), which can be constructed by identifying the boundary of a square by the antipodal map. Among the elliptic orbifolds, many are constructed as polygons on the sphere (for instance $(^*\_22)$ or $(^*\_\_)$), or as the doubles



badorb

Figure 5.21. Bad orbifolds have positively curved metrics. The teardrops $(_{-})$ and spindles $(_{-}m)$ have Riemannian metrics of strictly positive curvature, easily constructed as surfaces of revolution. The half-tears $(^*{}_{-})$ and half-spindles $(^*{}_{-}m)$ also inherit positively curved metrics, since the deck transformations of $(_{-}) \to (^*{}_{-})$ and $(_{-}m) \to (^*{}_{-}m)$ act as isometries.

of these polygons along their boundary. Somewhat trickier are $(2^*{}_{-})$ (a lune $(^*{}_{-}{}_{-})$ modulo a rotation of order 2) and (3^*2) (the triangle $(^*222)$ modulo an order 3 rotation). The fundamental group of (3^*2) acts as the group of orientation preserving symmetries of a regular tetrahedron, cross C_2 acting as the antipodal map (sending the tetrahedron to another tetrahedron).

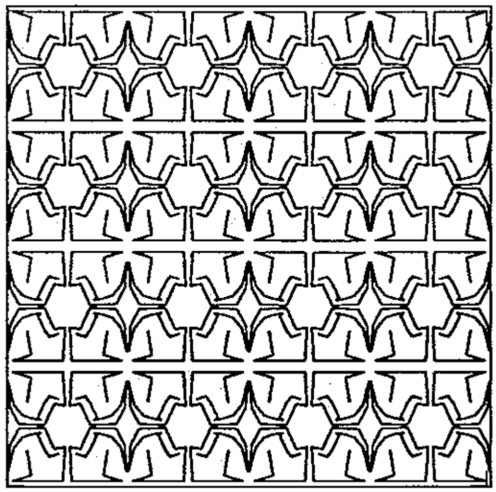


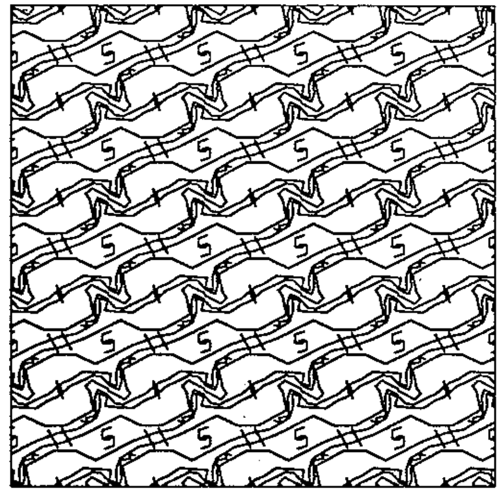
walle

Figure 5.22. Wallpaper groups: e. These patterns exhibit symmetry of the orbifolds (2^*22) (paper hat) and (22^*) (pillow case). These are the first two of five patterns with rotations of order 2, but no other rotations.

The 17 Euclidean orbifolds correspond to the 17 "wallpaper groups" (i.e. the 17 cocompact discrete 2-dimensional Euclidean isometry groups). [reference!!!!!!!] See the figures throughout this section for illustrations. Readers should unfold a sampling of these orbifolds for themselves, to appreciate their

% The Teichmüller space of a surface
% Indecomposable pieces of orbifolds
% orbpieces





wallf

Figure 5.23. Wallpaper groups: f. These patterns exhibit symmetry of the orbifolds (*2222) (rectangle) and (2222) (pillow).

beauty. Another pleasant exercise is to identify the orbifolds associated with some of Escher's prints.

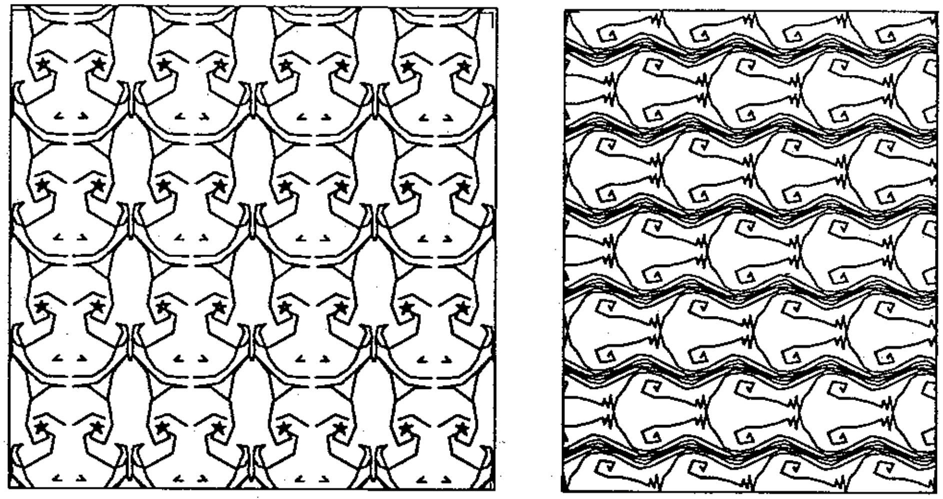
Hyperbolic structures can be found, and classified, for orbifolds with negative Euler characteristics by decomposing them into primitive pieces, in a manner analogous to our analysis of Teichmüller space for a surface (section 3.8) . Given an orbifold Q with $\chi(Q) < 0$, we first cut it into simpler pieces using a maximal family of one-dimensional suborbifolds (namely, circles and mI's) having the property that every component of the complement of the union has negative Euler number. Each resulting piece must have an orientable underlying space, since every non-orientable surface contains a simple curve with a non-orientable neighborhood. Cutting along such a curve does not create any new components, and leaves the Euler number unaltered. Note that an interval which joins two order 2 elliptic points is a suborbifold, and cutting along it creates a circular boundary component. Intervals joining other types of elliptic points are not suborbifolds. The other possible mI suborbifolds to cut along are intervals joining two mirrors, intervals joining a mirror and an order two elliptic point, intervals joining a mirror to itself, and intervals along a mirror between two different order two corner reflectors. 'Cutting' along such a suborbifold amounts to rubbing off the silver, and treating the interval now as boundary. Circular cuts come in three types: the circle may have an annulus neighborhood, a Möbius band neighborhood, or it may be a mirror.

All the possible pieces which can result after a maximal cutting as above are tabulated in ?? and illustrated in figure 5.25. Each double line in the figure indicates part of the boundary.

Verification that these are all the cases is straightforward although slightly tedious.

The indecomposable pieces:

% orbpieces
% The hyperboloid model and the Klein model
% generaltriangle
Teichmüller space of a surface
Lemma "hyperbolic pieces"
% orbpieces



wallg

Figure 5.24. Wallpaper groups: g. These patterns exhibit symmetry of the orbifolds (**) (silvered annulus) and (*°) (silvered Möbius band).

	0	1	2	3
Triangular pillows	$(\_mp)$	$(\_m\infty)$	$(\_\infty\infty)$	$(\infty\infty\infty)$
Triangles	$(^*\_mp)$	$(^*\_m\infty)$	$(^*\_\infty\infty)$	$(\infty\infty\infty)$
Triangular pillowcases	$(\_^*m)$	$(\_^*\infty)$	(∞^*m)	$(\infty^*\infty)$

Table 5.3. Indecomposable pieces of orbifolds

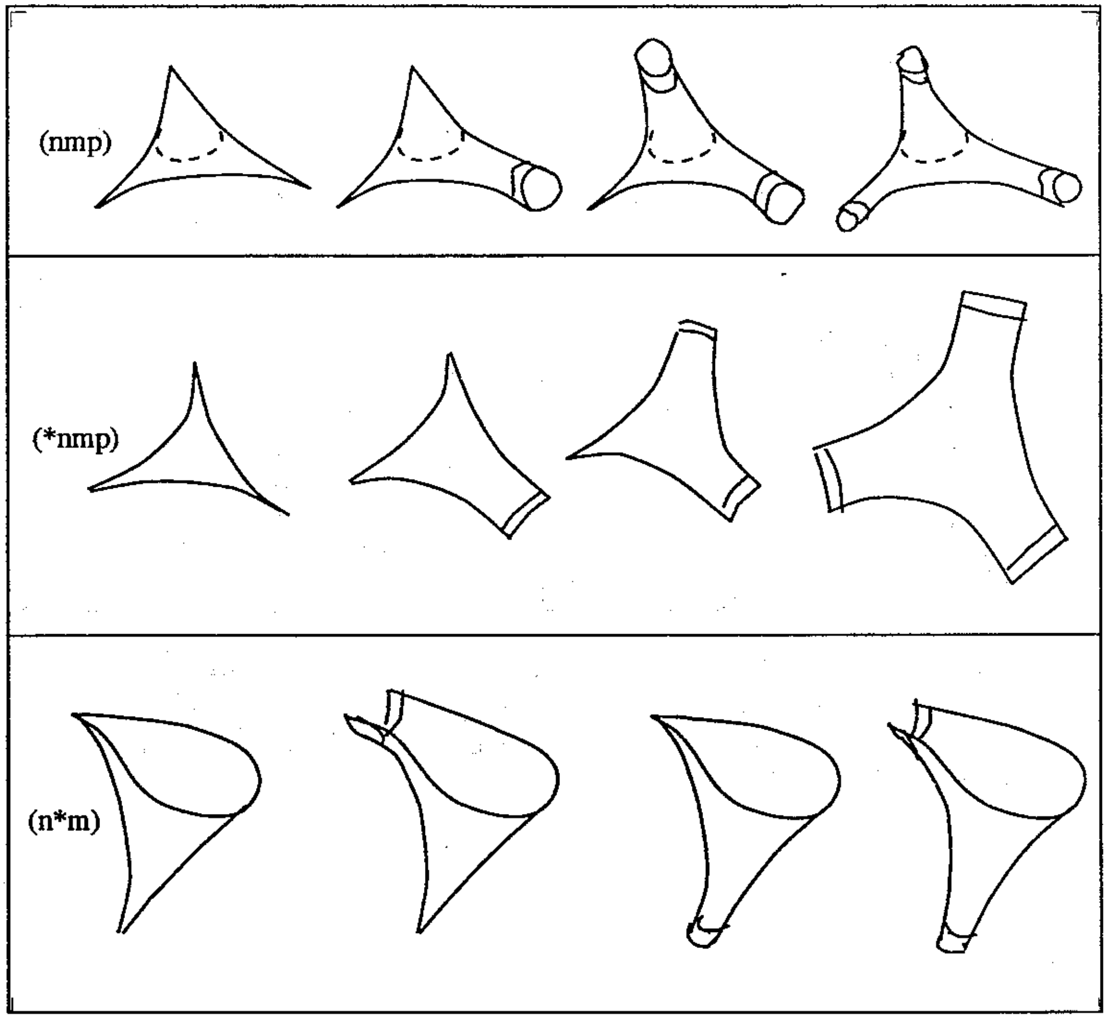
When a 2-dimensional orbifold of negative Euler number is maximally split along closed one-dimensional suborbifolds into pieces of negative Euler number, only the 12 atoms listed remain. The symbols here refer to the orbifold obtained by deleting the boundary, that is, ∞ refers to a stretch of boundary.

To classify hyperbolic structures for the orbifolds of Figure 5.25 (requiring any boundary component to be a geodesic suborbifold), we consider the generalized triangle – a figure obtained from a triangle in the projective model of $\mathbf{H}^2$ (whose vertices may lie outside the disk), by intersecting with its dual triangle. Recall from section 2.3 that the sides of the dual triangle are orthogonal to the sides of the original triangle that intersect them, and observe that each picture is either a generalized triangle or can be formed from two such generalized triangles— see figure 5.27 for an example of this. Using arguments analogous to the proof of theorem 3.8.8 (Teichmüller space of a surface), we obtain

Lemma 5.5.5 (hyperbolic pieces). *Each of the orbifolds of Figure 5.25 has a hyperbolic structure, and the set of all hyperbolic structures is parametrized by the choice of an arbitrary positive real number as the length of each double line.*

Appropriately chosen hyperbolic structures on the pieces can be reassembled to give a hyperbolic structure on the original orbifold. 5.5.3

% classification of 2-orbifolds
Theorem "Teichmüller space of orbifolds"



orbpieces

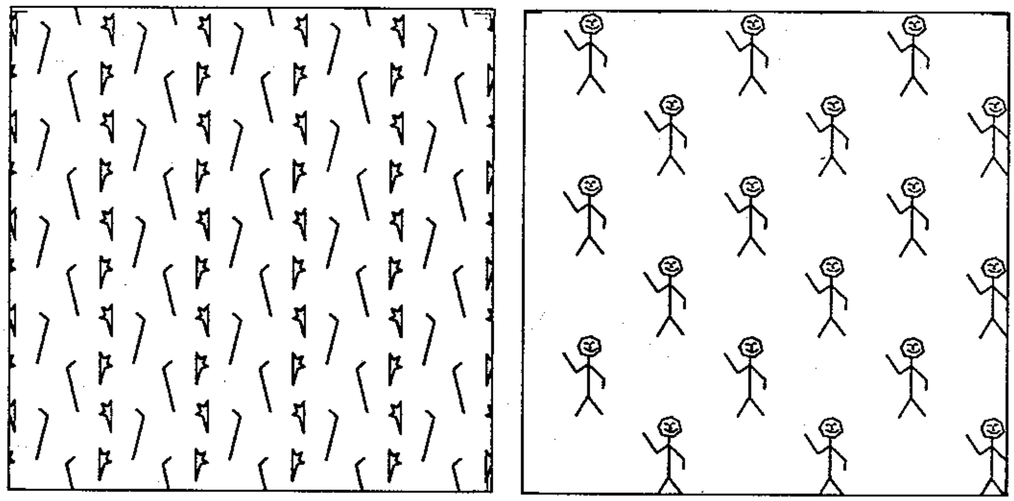
Figure 5.25. Indecomposable pieces of 2-dimensional orbifolds. These are the 12 types of pieces which can be obtained by cutting a 2-dimensional orbifold of negative Euler number along a maximal set of essential, non-parallel 1-sub-orbifolds.

From the proof of 5.5.3, we also obtain a description of the the Teichmüller space:

Theorem 5.5.6 (Teichmüller space of orbifolds). *The Teichmüller space $\mathcal{T}(Q)$ of an orbifold Q with $\chi(Q) < 0$ is homeomorphic to Euclidean space of dimension $-3\chi(X_Q) + 2k + l$, where k is the number of elliptic points and l is the number of corner reflectors.*

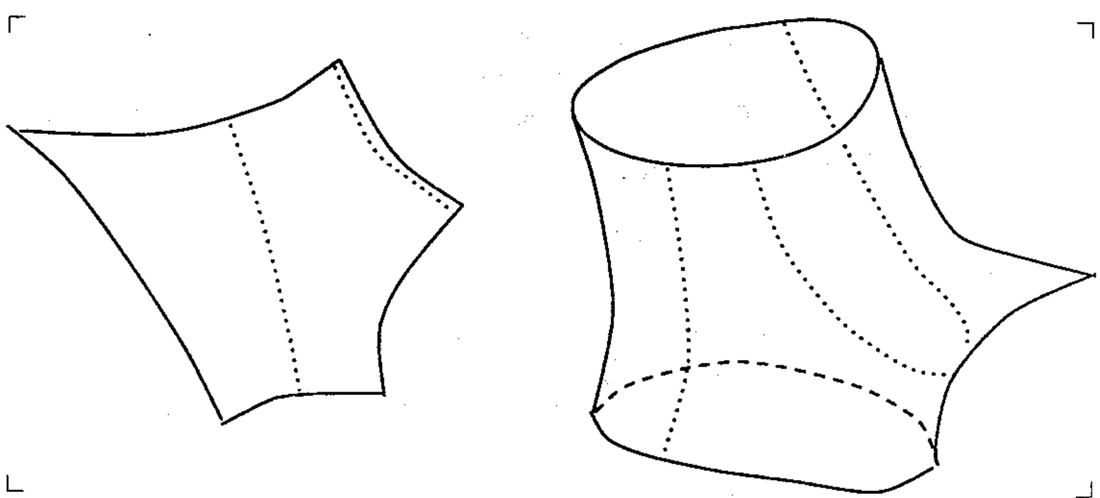
In other words, the dimension of the Teichmüller space depends only on the combinatorial topology, and not the orders, associated with features. If all orders are 3, then the dimension is $3\chi(Q)$.

Proof of 5.5.6: First suppose that all cone points and corner reflectors have order 3. Allocate the parameters associated with the 1-dimensional cuts equally to the abutting pieces. Thus, a circle with a neighborhood which is an annu-



wallh

Figure 5.26. Wallpaper groups: h. These patterns show the symmetry of (|°°) (the Klein bottle) and (°|) (the torus). In the figure at left, notice the two vertical glide axes, showing the structure of the Klein bottle as the union of two Möbius bands along their boundary.



generaltriangle

Figure 5.27. Generalized triangles. Construction of (*_222) and (_**) from generalized triangles.

% Indecomposable pieces of orbifolds

lus has a length parameter and a twist parameter associated, so each side is allocated 1. A circle which has a Möbius band neighborhood or which lies on a mirror has a single length parameter. Thus in all three cases, the abutting pieces each have an allocation of 1 parameter. An mI always has 1 parameter, so its two sides each have an allocation of $1/2$ parameter.

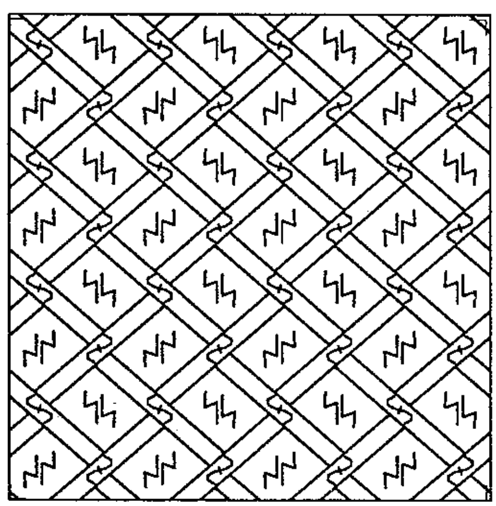
In ??, the possible indecomposable pieces are listed, with boundary removed. An ∞ before the * represents a circle, so it carries 1 parameter, while an ∞ after the star is a silvered interval, with an allocation of $1/2$ parameter. The Euler number of the orbifold, in each instance, is 3 times the number of allocated parameters, establishing the dimension count in the order 3 case.

The only additional complication in general is some degeneracies which occur when some of the orders are 2. For instance, if a cut is made along an mI joining two elliptic points of order 2, only one piece results, which has an additional geodesic circle as boundary. However, a circle can be folded into an mI in a 1-parameter family of different ways, differing by twists. Therefore the circle has two associated parameters, length and twist, but only one piece to receive its allocation.

To take care of this special situation, think of mI as a limiting case of a type of indecomposable piece. In other words, if the orders of the cone points were not both two, the corresponding cut would have been a circle enclosing the two cone points, so we think of cutting on the interval as a limiting case of cutting along an enclosing circle. The degenerate piece has type $(_m\infty)$, and receives an allocation of 1 parameter.

Similarly, a mirrored interval can be thought of as creating a degenerate piece of type $({}^*{}_m\infty)$.

The dimension count then goes through. 5.5.6



walli

Figure 5.28. Wallpaper groups: i. This is perhaps the most interesting of the wallpaper groups, (22°). The quotient space is the projective plane. Notice the two glide axes at right angles, and two types of points of order 2 symmetry.

Exercise "orbifolds on the projective plane"
Exercise "Paper dolls on (*442)"
Exercise "developing Euclidean orbifolds"

Exercise 5.5.7 (orbifolds on the projective plane). The fundamental group of $\mathbf{RP}^2_{(n)}$ contains $\mathbf{C}_n$ with index 2, since $\mathbf{RP}^2_{(n)}$ has $S^2_{(n,n)}$ as a double covering. Is $\pi_1(\mathbf{RP}^2_{(n)}) = D_n$ or $\mathbf{C}_{2n}$?

Exercise 5.5.8 (Paper dolls on (*442)). Fold a square sheet of thin Euclidean paper as a covering of (*442) (using a 32 or greater-fold cover), cut out a pattern – e.g., little half-people – and unfold.

Exercise 5.5.9 (developing Euclidean orbifolds). Invent patterns on (3*3), on (2*22) and on (22°) and sketch their development in $\mathbf{E}^2$.

This section has grown kind of long. I'd like to put in the computation of fundamental groups for 2-orbifolds, but this section already seems overburdened. Also, it would be nice to have some figures with spherical orbifolds, and some sample descriptions of how you actually fold up the patterns to make orbifolds, or unroll the orbifolds to make patterns.

Section "Three-dimensional orbifolds"
% differentiable orbifolds locally orthogonal
% orbifold table
% classification of 2-orbifolds

5.6. Three-dimensional orbifolds

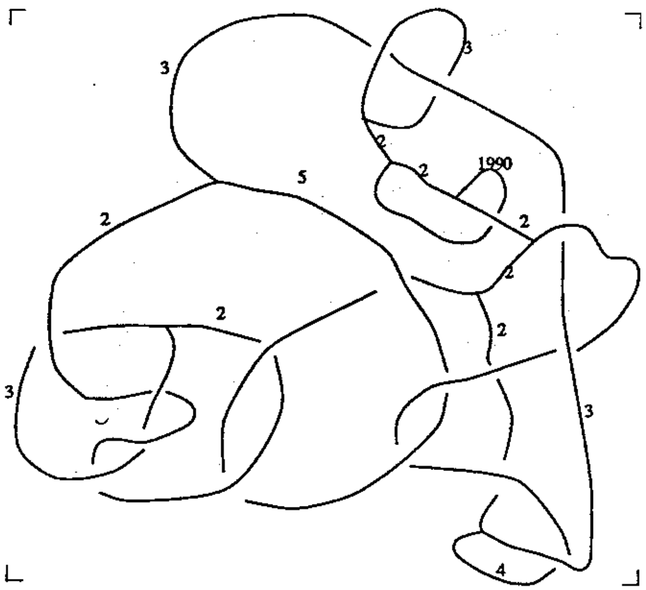
A small neighborhood of any point in a differentiable orbifold is isomorphic to the unit ball in $\mathbf{R}^n$ modulo a discrete subgroup of $O(n)$, by 5.4.1. In other words, the neighborhood is isomorphic to the cone on an (n-1)-dimensional elliptic orbifold (since the unit ball is the cone on S^{n-1}). Conversely, if E^{n-1} is any (n-1)-dimensional elliptic orbifold, then the cone $c(E^{n-1})$ on E^{n-1} is an n-dimensional orbifold, where the cone point is associated with the group $\pi_1(E^{n-1})$.

When $n = 3$, the possible types of local behavior are therefore determined by the elliptic column in table 5.2. The underlying space of a three-dimensional orbifold is topologically a manifold with boundary, except possibly near a finite number of points whose neighborhoods are cones on $(|^\circ) = \mathbf{RP}^2$. (In particular, note that if Q is closed and orientable, then X_Q is a closed manifold.) The orbifold structure is determined by a one-complex in Q, the one-dimensional singular locus together with an integer ≥ 2 labeling each edge. If the edge is an interior edge, then it is elliptic of order n and a neighborhood is the cone on $(\_\_)$, modeled on $\mathbf{R}^3/\mathbf{C}_n$. If the edge is on the boundary of X_Q, then it represents an order n corner reflector and its neighborhood is the cone on $(^*\_\_)$, modeled on $\mathbf{R}^3$ modulo the dihedral group $\mathbf{D}_{2n}$.

The labels must satisfy certain rules at the vertices, determined by theorem 5.5.3. An interior vertex which is not on a $\mathbf{RP}^2$-point must have three ends of edges incident to it, of orders $2,2,n$ or $2,3,3$ or $2,3,4$, or $2,3,5$, corresponding. An interior $\mathbf{RP}^2$ point has at most one end of an edge incident to it: its neighborhood is the cone on either $(|^\circ)$ or $(\_^\circ)$. Finally, a vertex on ∂X_Q must either have three ends of edges on ∂X_Q incident to it, with labels $2,2,n$ or $2,3,3$ or $2,3,4$, or $3,4,5$, or it may be incident to one end of an arbitrarily labeled interior edge (with neighborhood the cone on $(\_^*)$, or it may be incident to the end of one boundary edge and one interior edge, labeled m and 2 or 2 and 3 (with neighborhood the cone on either (2^*m) or (3^*2).)

Since elliptic structures in dimension 2 are determined up to isometry by their topological type and associated integers, this data suffices to determine the local models near the vertices.

There are many, many three-dimensional orbifolds. For example, any knot in S^3 gives rise to an orbifold for each integer n, where the knot is made a cone axis of order n. Any one-complex such that every vertex is incident to 3 ends of edges has at least one labelling describing an orbifold – all edges can be labeled 2. Usually there are many possible labelings obeying the rules Such orbifolds provide a rich source of examples for three-dimensional topology, since it is easy to draw pictures of them.



complicated3orb

Figure 5.29. A three-orbifold. A 3-dimensional orbifold, illustrating some of the possible local structure.

5.7. Fiber bundles

Section "Fiber bundles"
FIBER BUNDLES
Definition "orbifold fiber bundle"
:: fiber bundle
::tangent sphere bundle

There is a natural way to define the tangent space $\mathrm{T}(Q)$ of an orbifold Q. When the universal cover $\tilde{Q}$ is a manifold, then the covering transformations act on $\mathrm{T}(\tilde{Q})$ by their derivatives. $\mathrm{T}(Q)$ is then $\mathrm{T}(\tilde{Q})/\pi_1(Q)$. The tangent space at a singular point is not a vector space, but rather a vector space modulo a finite group. In the general case, Q is made up of pieces covered by manifolds, and the tangent space of Q is pieced together from the tangent space of the pieces. Similarly, any of the fiber bundle which are naturally associated to manifolds gives rise to something over orbifolds.

Silvio, this T op for tangent space doesn't act right.

Definition 5.7.1 (orbifold fiber bundle). A *fiber bundle* E with generic fiber F, over an orbifold Q is an orbifold with a projection

$$p : X_E \longrightarrow X_Q$$

between the underlying spaces, such that each point $x \in Q$ has a neighborhood $U = \tilde{U}/\Gamma$ (with $\tilde{U} \subset \mathbf{R}^n$) such that for some action of Γ on F, $p^{-1}(U) = (\tilde{U} \times F)/\Gamma$ (where Γ acts by the diagonal action). The product structure should of course be consistent with p: the diagram below must commute.

$$\begin{array}{ccc} \tilde{U} \times F & \longrightarrow & p^{-1}(U) \\ \downarrow & & \downarrow \\ \tilde{U} & \longrightarrow & U \end{array}$$

With this definition, natural fiber bundles over manifolds give rise to natural fiber bundles over orbifolds. For example, the *tangent sphere bundle* $\mathrm{TS}(Q)$ is the fiber bundle over Q with generic fiber the sphere of rays through 0 in a regular fiber of $\mathrm{T}(Q)$. When Q is Riemannian, this is identified with the unit tangent bundle $\mathrm{T}_1(M)$. When Q is a two-dimensional orbifold whose local groups preserve orientation, then $\mathrm{TS}(Q)$ is a three-manifold, because the action of the local groups on the unit tangent bundle is free.

Caution is in order in generalizing certain uses of fiber bundles to orbifolds. For example, it makes sense to speak of differentiable functions on an orbifold: they come from differentiable functions on the local model spaces. But what does the directional derivative mean? Most tangent vectors over singular points have several lifts in the local model space. The result is that the direcional derivative of an invariant differentiable function is equal in all these directions. As a consequence, the directional derivatives along certain tangent vectors is always 0 (for example, in any direction at any cone point of a 2-orbifold).

A related phenomenon is that a continuous vector field on an orbifold is tangent to each stratum of the singular locus. On the other hand, a continuos line field need not be tangent: for instance, any parallel line field in the $\mathbf{E}^2$ descends to a continous line field on (*2222).

::Seifert fiber space
Theorem "Euclidean 3-manifolds fiber over orbifolds"
% Three-dimensional Euclidean manifolds
% two-dimensional groups, three-dimensional manifolds
Theorem "elliptic 3-manifolds fiber over orbifolds"
% Elliptic three-manifolds
% classification of elliptic manifolds
% 2DIM DESCRIP
% fibrations over bad orbifolds
Theorem "fibered geometry orbifolds fiber"

A three-manifold is called a *Seifert fiber space* if it fibers over a two-dimensional orbifold. The term is generally used only in the case that the only singular points of the base are elliptic. Seifert fiber spaces (in the narrower sense) were introduced, thoroughly analyzed and used to advantage by Seifert, beginning with [Seifert, 1932].

There are non-orientable manifolds which fiber over two-orbifolds which have mirrors. These have topological properties much like Seifert fiber spaces in the narrower sense, and perhaps the term should be extended to include them. An example is the mapping torus M_ϕ of a diffeomorphism ϕ of the surface of genus 2, where ϕ acts as the deck transformation of the two-fold covering over $m(T^2-D^2)$. The mapping torus M_ϕ fibers over $m(T^2-D^2)$, with generic fiber S^1. We will avoid the question of what is the best terminology by sticking with the language of orbifolds.

In the next section we will analyze in detail the behavior of fiber bundles over orbifolds, including an extension of Conway's notation for 2-orbifolds that will completely describe the structure of 3-orbifolds that fiber over 2-orbifolds. First, though, we will develop some more general facts.

We have already seen a number of examples and classes of three-manifolds which fiber over two-orbifolds or one-orbifolds, but we have expressed them in the language of group actions. For convenience, we will collect the results and rephrase them in terms of orbifolds.

Theorem 5.7.2 (Euclidean 3-manifolds fiber over orbifolds). *(from section 4.4)*

a) Every closed Euclidean three-manifold fibers over a Euclidean two-orbifold and also fibers over a one-orbifold, with fiber a Euclidean two-orbifold.

b) (Problem 4.4.16) Every closed orientable Euclidean three-manifold is TS(Q)*, where Q is a closed Euclidean two-orbifold without mirrors.*

Theorem 5.7.3 (elliptic 3-manifolds fiber over orbifolds). *([Seifert, 1932].) Every elliptic three-manifold fibers (in at least one way) over an elliptic two-orbifold.*

Proof of 5.7.3: This follows from 4.5, particularly 4.5.11, the preceding discussion and exercise ??. [5.7.3]

Note: Many elliptic three-manifolds fiber also over bad orbifolds. (See Exercise 5.7.10.)

Theorem 5.7.4 (fibered geometry orbifolds fiber). *An orbifold Q of finite volume modeled on any of the five fibered geometries itself fibers over a one- or two-dimensional orbifold. More precisely, if Q is modeled*

% The five fibered geometries
% cubic
% cubic
% icosquotient
Definition "orbifold connection"
:: connection

on $S^2 \times \mathbf{E}^1$:	*it fibers over a one-orbifold with fiber an elliptic two-orbifold. (It also fibers topologically over an elliptic two-orbifold.)*
$\mathbf{H}^2 \times \mathbf{E}^1$:	*it fibers over a hyperbolic two-orbifold, and it also fibers over a one-orbifold with fiber a hyperbolic two-orbifold.*
on $SL_2\mathbf{R}$:	*it fibers over a hyperbolic two-orbifold.*
on nilgeometry:	*it fibers over a two-dimensional Euclidean orbifold.*
on solvegeometry:	*it fibers over a one-orbifold with fiber a Euclidean two-orbifold.*

Proof of 5.7.4: This is a restatement of conclusions from ??. [5.7.4]

We have seen enough examples of discrete groups acting on $\mathbf{E}^3$ and S^3 to know that Euclidean and elliptic orbifolds do not, in general, fiber over any one-orbifold or two-orbifold. This can be rigorously proved just by studying neighborhoods of vertices. For example, the group of orientation-preserving symmetries of the cubic tiling of $\mathbf{E}^3$ has as fundamental domain two of the tetrahedra in the barycentric subdivision of this tiling (figure 5.30). The quotient orbifold has underlying space S^3, with singular locus shown in figure 5.30. It does not fiber even locally near a $2,3,4$ vertex (if it fibered over a two-orbifold, the lift of the origin would give a 1-dimensional subspace of $\mathbf{R}^3$ invariant by the action of the $(2,3,4)$-group. But this is just the octahedral group, which has no such invariant line. Similarly, if it fibered over a 1-orbifold the lift of the origin would give a 2-dimensional subspace invariant by the octahedral group).

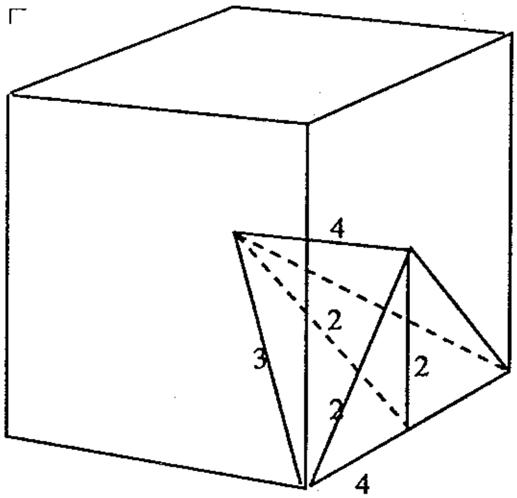
Similarly, $\mathbf{RP}^3/(I \times I)$, where I is the icosahedral group $\pi_1(235)$ is an orbifold with underlying space S^3 and singular locus given in 5.31, so it cannot fiber for a similar reason.

There is also a converse construction: we will show how to construct a geometric structure for any 3-orbifold which fibers over a 2-orbifold with 1-dimensional fibers.

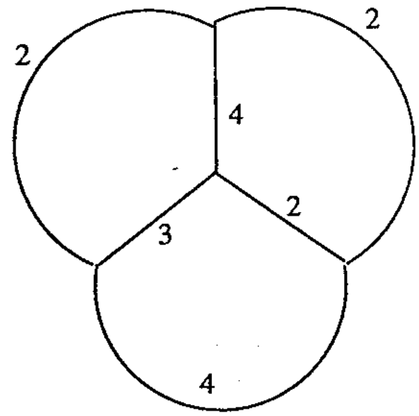
First we need an invariant to help distinguish fibered orbifolds with two-dimensional bases. Suppose that P is a three-orbifold which fibers over a two-orbifold B with projection $p : P \to B$. Choose an arbitrary C^∞ Riemannian metric g_0 on P. The length l of the fiber through any point of P which lies above a regular point of B extends to a C^∞ function on P; the metric $g_1 = (1/l)g_0$ has the property that all generic fibers have length 1 .

Definition 5.7.5 (orbifold connection). An (isometric) *connection* for P is a field τ of two-planes on P such that

a) The two-plane $\tau^2_p \subset \mathrm{T}_p(P)$ at each point p is transverse to the fiber through p, and



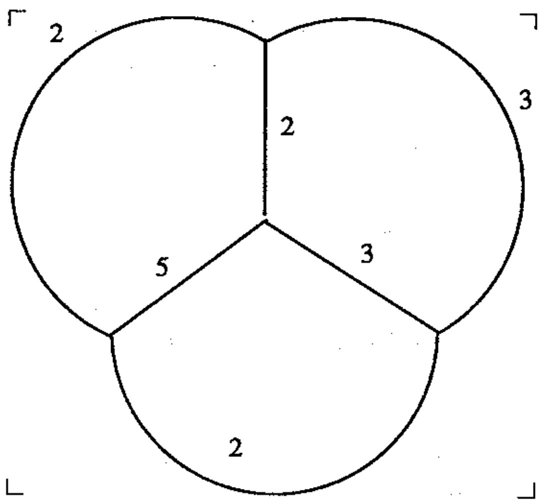
(a)



(b)

cubic

Figure 5.30. The quotient by the symmetries of the cubic tiling . The quotient of $\mathbb{E}^3$ by the group of symmetries of the cubic tiling. It does not fiber as an orbifold.



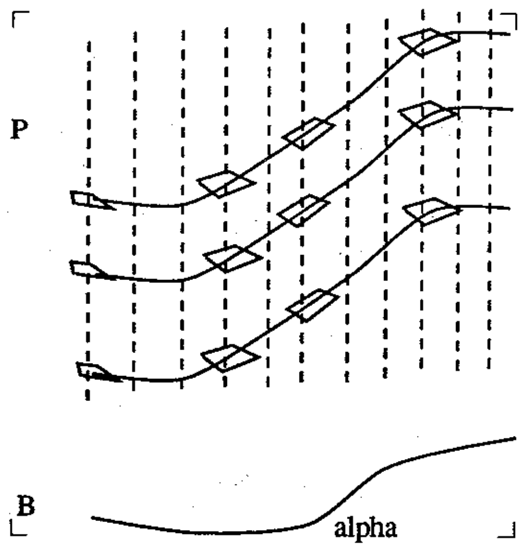
icosquotient

Figure 5.31. Quotient orbifold of the icosahedron. The singular locus of the $\mathbb{R}P^3/(I \times I)$, where I is the icosahedral group $\pi_1(235)$. The quotient of the action of I alone is the Poincarè dodecahedral space (1.4.4), for which a dodecahedron is a fundamental domain. The additional factor of I rolls up this space by the orientation-preserving symmetries of the dodecahedron, giving a double-tetrahedron 3-orbifold as shown

Proposition "existence of isometric connection"

b) the "combing process" defined by τ preserves the metric on the fibers. More precisely, if α is any short arc on B the line field defined by τ on $p^{-1}(\alpha)$ has nearby integral curves a constant distance apart as measured along the fibers.

Recall that a "two-plane" may actually be the quotient of an invariant two-plane by the local group. The interpretation given to the definition should be the same as an invariant connection on the local model for the fiber bundle, modulo the local group.



combing

Figure 5.32. The combing defined by a field of two-planes. A field τ of two-planes on P defines a "combing process" that allows paths on the base B to be lifted.

Proposition 5.7.6 (existence of isometric connection). *The fiber bundle $p : P \to B$ always has a connection respecting the metric g_1 restricted to the fibers.*

Proof of 5.7.6: First we can reduce the construction of a connection to a local problem, as follows. We can think of a two-plane field transverse to the fibers at a point $x \in P$ as a linear function from $\mathrm{T}_{p(x)}(B)$ to $\mathrm{T}_x(P)$, which projects back to the identity under the map $p_* : \mathrm{T}_x(P) \to \mathrm{T}_{p(x)}(B)$. In this way, it makes sense to take convex combinations of such plane fields. If $\tau_1, \dots, \tau_k$ are connections defined over an open set $U \subset B$ and if $\lambda_1, \dots, \lambda_k$ are non-negative C^∞ functions summing to 1, then the convex combination $\Sigma\lambda_i\tau_i$ is again a connection. Therefore, if we can construct a connection above each coordinate neighborhood $U = \tilde{U}/\Gamma$, we can construct a connection globally by using a partition of unity.

Consider the local model $\tilde{P}_U \to \tilde{U}$ for $P_U \to U = \tilde{U}/\Gamma$ over a coordinate neighborhood. A connection can be found in such a neighborhood by first choosing a section $\tilde{U} \to \tilde{P}_U$. This determines a parametrization of $\tilde{P}_U$ as

::connection form)
::curvature
::Euler class
::Euler number
% Euler class is invariant
% Euler number is rational

$\tilde{U} \times F$, by making the $\tilde{U}$-factors equally spaced along the fibers. The tangent space to the $\tilde{U}$-factors is a connection τ for $\tilde{P}_U \to \tilde{U}$. The average $\frac{1}{|\Gamma|} \Sigma \gamma_* \tau$ over the local group Γ gives a connection invariant by Γ, thus a connection for P over U. 5.7.6

Let's specialize to the case that B is oriented, P is oriented, and therefore the fibers also have a consistent orientation, the "quotient" orientation such that the orientation of the base times the orientation of the fibers gives the orientation of P. There is a unique one-form ω (called the *connection form)* such that ω annihilates τ, and ω evaluated on a positive unit length tangent vector to a fiber is 1 . The exterior derivative $d\omega$ may be understood geometrically by Stokes formula, $\int_C d\omega = \int_{\partial C} \omega$. Applying this in the case that C is a small disk, we see that $\int_{\partial C} \omega$ measures the total amount that the combing process going around ∂C translates the fiber down or up, and in particular, it depends only on the image of ∂C in the base B. Therefore $d\omega$ has the form $d\omega = p^*\Omega$, where the two-form Ω defined on B is called the *curvature* of the connection.

There are many possible connections which respect the given metric on the fibers. If ω' is the connection form for such a connection, then $\omega' - \omega$ vanishes on the tangent space to the fibers, and it can be written in the form $\omega' - \omega = p^*\alpha$. If Ω' is the curvature for ω', so $d\omega' = p^*\Omega'$, then $\Omega' - \Omega = d\alpha$, so Ω' is cohomologous to Ω. Conversely, every 2-form Ω' cohomologous to Ω (i.e., having the same integral over B) is the curvature for some connection respecting the given fiber metric (whose connection form is $\omega + p^*\alpha$ where $\Omega' - \Omega = d\alpha$). The cohomology class of Ω is the (real) *Euler class* of P, and $\int_B \Omega$ is the *Euler number.* Note that there is no factor $1/2\pi$ here, since we made fibers have length 1 instead of length 2π.

The Euler class is also independent of the choice of fiber metric (problem 5.7.11). The Euler number is always a rational number expressible with denominator the least common multiple of the orders of the local groups of the base (problem 5.7.12).

In case the base, the fibers, the total space, or all three fail to be oriented, there is still some notion of curvature.

As long as the fibers are consistently oriented, there is still a connection form ω and a curvature form Ω. If the the fibers are consistently oriented but the base is not orientable, all two-forms on the base are cohomologous, so we do not obtain any non-trivial numerical invariant.

If the fibers are not consistently oriented, then a connection form cannot be defined as simply as before. Instead, we think of a connection form ω as a one-form with values in the tangent space to the fibers. Previously we (perhaps unconsciously) identified this tangent space with $\mathbf{R}$, using the orientation, but now such an identification is not possible. The exterior derivative $d\omega$ is still well-defined, but it has values in the tangent space to the fibers. If the fibers are all circles, so they are orientable over a small neighborhood in the base, then the tangent space to the fibers is the pullback of a 1-dimensional vector

Definition "Euler number of fibration" :: Euler number
Proposition "fibered orbifolds have geometric structures"

bundle over the base, and $d\omega$ is the pullback of a 2-form Ω on the base, with values in this bundle.

If there are any fibers which are mI instead of S^1, then the tangent space to the fibers is not the pullback of a bundle on the base. If the total space of the bundle is oriented, then there is a simple expedient: note that when everything is oriented, the Euler number $\int_B \Omega$ can be calculated alternatively as $\int_P \omega \wedge d\omega$. As long as the total space to the bundle is oriented, $\omega \wedge d\omega$ is a three-form with $\mathbf{R}$ coefficients, and if the base and fiber orientations are simultaneously reversed, $\omega \wedge d\omega$ is unaltered.

Definition 5.7.7 (Euler number of fibration). The *Euler number* $\chi(P)$ is defined for any closed oriented three-orbifold P which fibers over a two-dimensional base as $\int_P \omega \wedge d\omega$, where ω is any connection form. If P is non-orientable, $\chi(P)$ is defined to be 0 .

This definition agrees with the previous definition in the oriented case. The fact that it is independent of choices may be derived from the independence in a four-fold cover in which everything is oriented.

Proposition 5.7.8 (fibered orbifolds have geometric structures). *Every three-orbifold P which fibers over a good closed two-orbifold B admits a geometric structure. The type of structure is given by this table:*

	$\chi(B) > 0$	$\chi(B) = 0$	$\chi(B) < 0$
$\lvert\chi\rvert(P) \neq 0$	S^3	*Nilgeometry*	$\widetilde{\mathrm{PSL}(2,\mathbf{R})}$
$\lvert\chi\rvert(P) = 0$	$S^2 \times \mathbf{E}^1$	$\mathbf{E}^3$	$\mathbf{H}^2 \times \mathbf{E}^1$

Remark: The Euler characteristic of B is independent of orientation, while the sign of $\chi(P)$ depends on the orientation of P. Thus for an orientable but not oriented fiber bundle P, it does not make sense to talk about the case $\chi(P) > 0$ versus the case $\chi(P) < 0$. If we were to stick purely with oriented objects we could make this distinction, and we would also need to distinguish the two orientations for nilgeometry and for $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$, making ten types of geometry instead of eight.

Proof of 5.7.8: Consider first the case that everything is oriented. By theorem 5.5.6, B has a geometric structure; choose such a structure. Choose a fiber-metric for P and a connection form ω. The curvature Ω is cohomologous to a two-form KdA for some constant K, where A is the area form (since the 2-dimensional de Rham cohomology is 1-dimensional). By the preceding discussion we may suppose that ω was chosen so that $\Omega = KdA$. A geometry may now be constructed which serves as a model for P. We can take the total space of such a geometry to be the universal cover $\tilde{P}$, which fibers over $\tilde{B}$. The structure group can be taken as the group of diffeomorphisms of $\tilde{P}$ which preserve ω and take fibers to fibers inducing an isometry of $\tilde{B}$ to itself. That this group is transitive and has compact stabilizer G_x is elementary

% connection-preserving diffeomorphism are transitive
% enlarge geometry
% fibered orbifolds have geometric structures
Proposition "geometric structures for fibrations over 1-orbifolds"
Problem "fibrations over bad orbifolds"
Problem "Euler class is invariant"
Problem "Euler number is rational"

(problem 5.7.13). Further analysis shows that the geometry so constructed is subsumed under the geometry given in the table (problem 5.7.14).

In the non-oriented case, choose a geometric structure for the base and a connection. Now pass to the two- or four-fold connected cover $\widehat{P}$ which orients everything, where the connection is represented by an (ordinary) connection form $\widehat{\omega}$ with curvature an (ordinary) 2-form $\widehat{\Omega}$ on $\widehat{B}$. The deck transformations which reverse fiber orientation take $\widehat{\omega}$ to $-\widehat{\omega}$, while the others preserve $\widehat{\omega}$. We know that there is some one-form α on $\widehat{B}$ such that $\widehat{\omega}+p^*\alpha$ is a connection form having constant curvature. The form α may not transform properly under the group Δ of deck transformations of $\widehat{P}$ over P. However, if we form the average $\beta = \frac{1}{|\Delta|}\sum_{\delta\in\Delta}\varepsilon(\delta)\delta^*\alpha$, where $\varepsilon(\delta)$ is the sign of the action on orientations of the fibers is a connection form $\widehat{\omega}' = \widehat{\omega} + p^*\beta$ which has constant curvature and does transform correctly under Δ.

Note that if any element δ of Δ reverses orientation of $\widehat{P}$, then on the one hand $\widehat{\omega}' \wedge d\widehat{\omega}' = -\delta^*(\widehat{\omega}' \wedge d\widehat{\omega}')$ (because $\widehat{\omega}'$ on the fibers is equal to the length element and $d\widehat{\omega}' = p^*K\,dA$, so that $\widehat{\omega}' \wedge d\widehat{\omega}'$ is a constant multiple of the volume form of $\widehat{P}$), while on the other hand, $\widehat{\omega}' \wedge d\widehat{\omega}' = (\varepsilon(\delta))^2 \cdot \delta^*(\widehat{\omega}' \wedge d\widehat{\omega}') = \delta^*(\widehat{\omega}' \wedge d\widehat{\omega}')$. This implies that the constant curvature K is zero, justifying our definition in the non-orientable case. 5.7.8

There is a result similar to 5.7.8 for three-orbifolds which fiber over a 1-orbifold, provided the fiber is elliptic or Euclidean:

Proposition 5.7.9 (geometric structures for fibrations over 1-orbifolds). *If Q is a 3-orbifold which fibers over a 1-orbifold with fiber an elliptic 2-orbifold, then Q admits an $S^2 \times \mathbf{E}^1$ structure.*

If Q fibers over a 1-orbifold with fiber a Euclidean 2-orbifold, then Q admits either a Euclidean structure, a nilgeometry structure, or a solvegeometry structure.

Proof of 5.7.9: 5.7.9

Insert proof. Maybe use harmonic forms, or averaging maps, or some other clever device. Otherwise, use that homotopy implies diffeomorphism in dimension two.

Problem 5.7.10 (fibrations over bad orbifolds). a) Show that the three-sphere fibers over every possible bad two-orbifold. (Hint: consider the actions of S^1 on S^3.)

b) What bad orbifolds do lens spaces fiber over?

Problem 5.7.11 (Euler class is invariant). a) Show that a one-form on P is a connection form respecting some fiber metric iff ω is positive on positive tangent vectors to the fibers and $d\omega$ has the form $p^*\Omega$, for some two-form Ω on B.

b) Show that if ω_1 and ω_2 are connection forms respecting (possibly different) fiber metrics but with the same fiber length of 1, then $\omega_1 - \omega_2$ has the form $df + p^*\alpha$, where f is a function on P and α a one-form on B.

c) The Euler class is an invariant up to fiber-preserving and orientation-preserving diffeomorphism.

Problem 5.7.12 (Euler number is rational). Describe a more topological construction for $\chi(P)$ which yields a rational number expressible with denominator the least common multiple of the orders of the local groups.

Problem "connection-preserving diffeomorphism are transitive"
Problem "enlarge geometry"
% connection-preserving diffeomorphisms are transitive
% The five fibered geometries
Problem "Euler number of orbifold is that of sphere bundle"

(Hint: consider multiple-valued sections of $P \to B$ over the complement of a regular point of B.)

Problem 5.7.13 (connection-preserving diffeomorphism are transitive). Show that for a circle bundle or $\mathbf{E}^1$-bundle X over a two-dimensional geometry B, if ω is a connection form having constant curvature, then the group of diffeomorphisms of X which preserve ω and project to isometries of B acts transitively and with compact stabilizers.

(Hint: Given an isometry $f : B \to B$ and an isometry $g : p^{-1}(x_0) \to p^{-1}(f(x_0))$ for some $x_0 \in B$, consider how to extend g along a path beginning on $p^{-1}(x_0)$.)

Problem 5.7.14 (enlarge geometry). Show that the group of diffeomorphisms of X, which was constructed in **??** can be enlarged to give the geometry listed in the table of **??**.

Problem 5.7.15 (Euler number of orbifold is that of sphere bundle). Prove that for any closed two-orbifold B, $\chi(\mathrm{TS}(B)) = \chi(B)$.

Problem 5.7.16. Show that an orientable three-orbifold which fibers over a non-compact hyperbolic orbifold of finite area has both a $\widetilde{\mathrm{PSL}(2,\mathbf{R})}$ structure and an $\mathbf{H}^2 \times \mathbf{E}^1$ structure.

Section "Some examples of three-orbifolds which fiber over two-orbifolds."
% tangent sphere bundles of polygonal orbifolds
Example "tangent sphere bundle of mD^2"
% z2hopf

5.8. Some examples of three-orbifolds which fiber over two-orbifolds.

We will build up a description and naming system for three-orbifolds which fiber over two-orbifolds, starting with local information and working up in scale.

The first distinction to note is the two possibilities for a generic fiber. If the generic fiber is a circle, the name will be enclosed in round parentheses, for instance, $(2_1 5_{-}{}^{-}2.2{:}43_1 6)$. A name enclosed in square brackets, such as $[2_0{}^{+}2.2{:}]$, means the generic fiber is an interval.

How about some complete names for a few things, followed by adding more complexities. Start with everythingoriented: base and total space. Just cone points. Then add case of oriented total space. Then oriented fibers. Finally the general.

We'll first describe what can happen when the generic fiber is a circle.

Above the neighborhood of a cone point p of order m on a two-orbifold, the bundle p is the quotient of $D^2 \times S^1$ by the cyclic group $\mathbf{C}_m$. If the $\mathbf{C}_m$ preserves orientation, then it acts on S^1 by a k/m rotation of the circle. Given a choice of an orientation for the fiber over p, k is well-defined modulo m. This behavior is denoted m_k. The group $\mathbf{C}_m$ can reverse orientation only when m is even; this behavior is denoted mN. All orientation reversing isometries of the circle are conjugate, so there are no further local distinctions.

Above a mirror, the bundle has the form $D^2 \times S^1/\mathbf{C}_2$, where $\mathbf{C}_2$ acts on D^2 by a reflection. There are three possibilites for the action on S^1: the non-trivial element of $\mathbf{C}_2$ may act as the identity, as a rotation of order 2, or as a reflection. These three behaviors are indicated, respectively, by ., by :, or by a blank, in the position corresponding to that mirror.

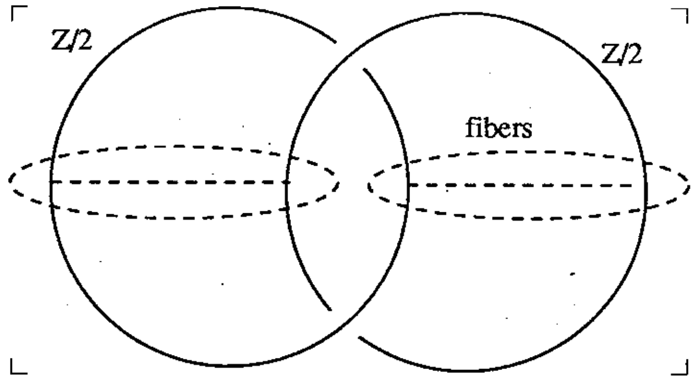
At a corner reflector, the local fundamental group of the base is $\mathbf{D}_{2m}$, generated by the reflections a and b in the two adjacent mirrors. If m is odd, then a is conjugate to b, and therefore the behavior of the fiber bundles above the two adjacent mirrors must be the same. Since neither the identity and an order two rotation have any other conjugates, the action of the $\mathbf{D}_{2m}$ on the circle is determined up to conjugacy by the conjugacy classes of the action of a and of b if either preserves orientation. Additional information only needs to be supplied for a corner reflector when both a and b reverse the orientation of the fiber. (This is the same as the case that the total space of the bundle is locally orientable.) In this case, ab preserves orientation, and it might have any rotation number k/m, recorded as m_k. Geometrically, the universal cover of the fiber bundle over a neighborhood of the corner reflector is an infinite solid cylinder, $D^2 \times \mathbf{R}$.

This is a working front, for finishing putting in 3-d notation. –wpt

If Q is a polygonal orbifold, that is, a two-orbifold $({}^*abc\dots z)$ whose underlying space D^2 and singular locus ∂D^2, then it turns out that every orientable three-orbifold P which fibers over Q has underlying space S^3. (See 5.8.5)

Here are some examples.

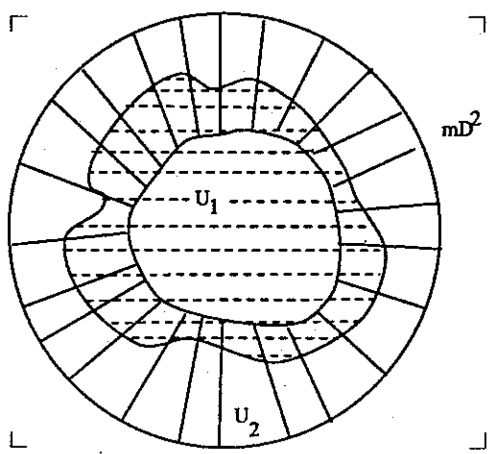
Example 5.8.1 (tangent sphere bundle of mD^2). The tangent sphere bundle of mD^2 is S^3, with singular locus two $\mathbf{C}_2$–Hopf circles (figure 5.33).



z2hopf

Figure 5.33. The tangent sphere bundle of mD^2 is S^3

One way to derive this is to begin with the universal cover. We know that $\mathrm{TS}(S^2)$ is $\mathbf{RP}^3$, so its universal covering is S^3, and $\mathrm{TS}(S^2)$ is a double covering of $\mathrm{TS}(mD^2)$. Therefore $\mathrm{TS}(mD^2)$ is S^3 modulo a group of isometries of order 4. The quotient orbifold has two circles of $\mathbf{C}_2$−singular loci, consisting of the clockwise tangents to ∂D^2 and the counterclockwise tangents. The group must be $(\mathbf{C}_2)^2$, generated by two 180° rotations about two Hopf circles which are dual to each other. (Alternate generators are the antipodal map, together with a 180° rotation about one of the Hopf circles.) The quotient space is obtained by constructing a fundamental domain at the intersection of two hemispheres containing the two Hopf circles on their boundaries, and folding it up.



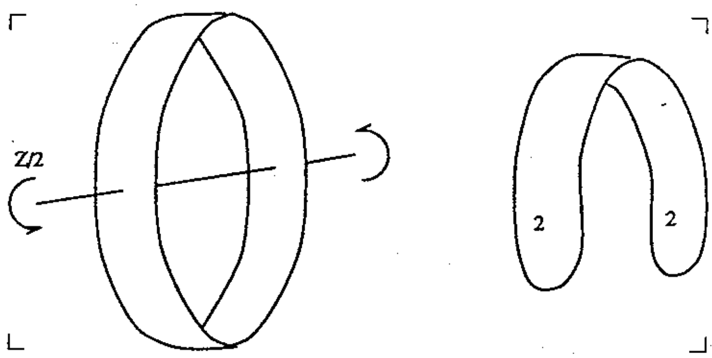
fiberpatch

Figure 5.34.

Alternatively, one may construct the picture by piecing it together from $\mathrm{TS}(U_i)$, where $\{U_i\}_{i=1,2}$ is an open cover of the base. Let U_1 be an open disk in the interior, and U_2 be a neighborhood of ∂D^2. $\mathrm{TS}(U_1)$ is an open solid torus $U_1 \times S^1$ since the tangent bundle of U is trivial. We can think of U_2 as $(m[0,1/2)) \times S^1 = ((-1/2,1/2) \times S^1/\mathbf{C}_2)$. Its tangent circle bundle is $((-1/2,1/2) \times S^1 \times S^1)/\mathbf{C}_2$, where $\mathbf{C}_2$ acts as a reflection in the first and second factors. In other words, it acts on these two factors as a 180° rotation

Exercise "fibration of twisted band"
Example "tangent to billiards"
% billiardbundle

of a cylinder about an axis of symmetry intersecting the cylinder in two points. The quotient is an open disk, with two C_2—elliptic points. Therefore $TS(U_2)$ is an open solid torus, with two parallel C_2—elliptic axes running lengthwise.



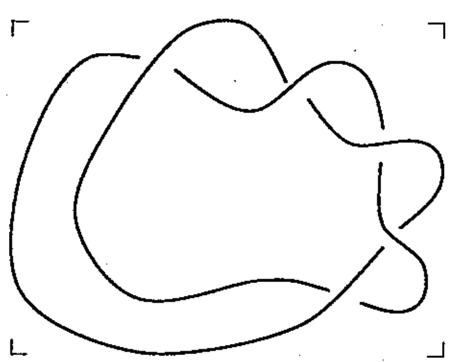
flipstrip

Figure 5.35.

The two solid tori are glued together by gluing a longitude of the first to a meridian of the second, and longitude of the second to a meridian plus longitude of the first. This imparts the 360° twist which links the two C_2—axes.

Exercise 5.8.2 (fibration of twisted band). For each integer n, let P_n be the orbifold with underlying space S^3 and singular locus the boundary of an unknotted band with n half-twists, labeled 2. Show that P_n fibers over mD^2 with Euler number $n/2$.

(Hint: To get the fibration, think first about fibering the twisted band by itself. Given the fibration, construct a connection form ω above a neighborhood of the mirror locus using a trivialization of the bundle there. Consider any extension of ω over the rest of P_n. The integral of Ω over mD^2 equals the integral of $d\omega$ over a section above the interior of mD^2 , which can be computed using Stokes formula.)

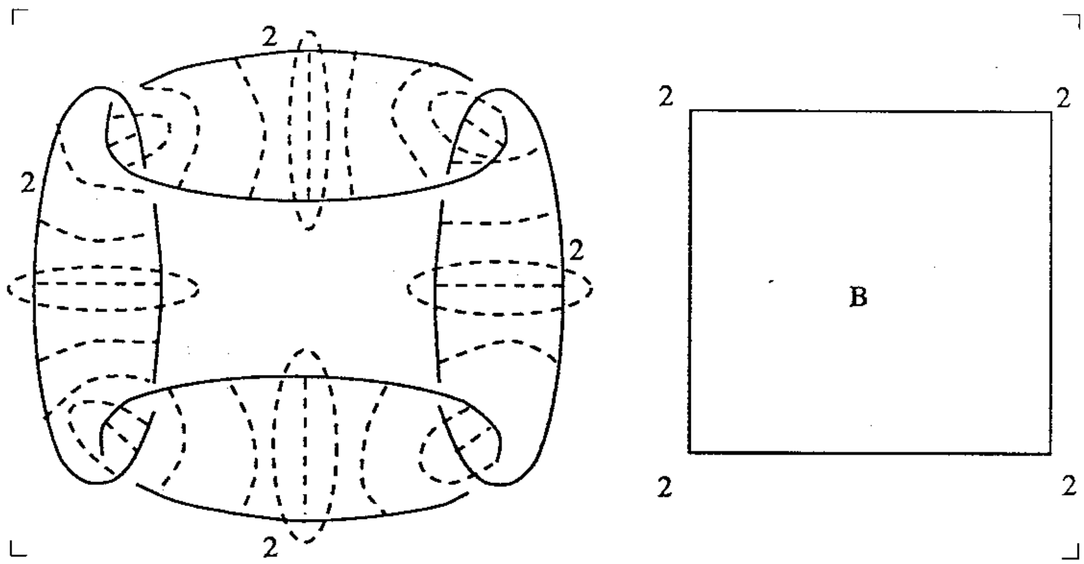


twistorb

Figure 5.36. An orbifold with a twisted singular locus. The orbifold $P_n, n = +5$.

Example 5.8.3 (tangent to billiards). A rectangular billiard table B has the tangent circle bundle pictured in figure 5.37.

% tangent sphere bundle of mD^2
% billiardbundle
% billiardbundle
% Euler number of orbifold is that of sphere bundle
% borromean orbifold
Problem "billiard foliation"
% billiardbundle
Example "tangent sphere bundles of polygonal orbifolds"



billiardbundle

Figure 5.37. Tangent circle bundle of a billiard table. The tangent circle bundle of a rectangular billiard table.

The picture of $\mathrm{TS}(B)$ may be constructed using the double covering $mC \to B$. Arguing as for example 5.8.1, $\mathrm{TS}(mC)$ is $S^1 \times (S^2_{(2,2,2,2;\ )})$. Dividing by the $\mathbf{C}_2$ action gives 5.37. The $\mathbf{C}_2$-cover is reconstructed from 5.37 by cutting open along two $D^2_{(\ ;2,2)}$'s spanning the upper and lower loops, and then doubling. From problem 5.7.15 we know that the tangent circle bundle should have a connection with 0 curvature, and one can see this directly by choosing a connection form which vanishes on the tangent space of a suborbifold $mC \subset \mathrm{TS}(B)$, where C is a cylinder connecting the top and bottom circle.

Note the two-fold symmetry in $\mathrm{TS}(B)$, which comes from rotating B 180° and then reversing the image of each vector. The quotient space is the same as Example 5.1.8.

Problem 5.8.4 (billiard foliation). The motion of a billiard ball on B (ignoring affects of friction or spin) determines a flow on $\mathrm{TS}(B)$. Describe the trajectories in terms of figure 5.37.

(Hint: there is a codimension-one foliation of $\mathrm{TS}(B)$ with two leaves which are $mC's$, while the other leaves are tori.)

Example 5.8.5 (tangent sphere bundles of polygonal orbifolds). $\mathrm{TS}(D^2_{(\ ;n_1,n_2,\ldots,n_k)})$.

[/u/gt3m/3mbook/pictures/chap5/8/figur1.ps not found]

figur1

Figure 5.38. 2

$\mathrm{TS}(D^2_{(\ ;n_1,n_2,\ldots,n_k)})$ is constructed by piecing together models of the tangent circle bundle above neighborhoods of each corner, edge, and above the interior. Over the non-singular part of $D^2_{(\ ;n_1,n_2,\ldots,n_k)}$, we have a solid torus. Over an

% WEAVE

edge e, we have $mI \times e$, with generic fibers folded once to become mI; nearby fibers go once around these $mI's$. The picture above a corner reflector of order n is described by Figure 5.39. The tangent circle bundle over the disk cover of a neighborhood of the corner reflector is drawn as a solid torus $D^2 \times S^1$, on which the action of the dihedral group is generated by a rotation of π around the axis R, and a combined rotation of $2\pi/n$ in D^2 and around S^1. The fundamental domain of this action is a solid cylinder with both bottom and top faces identified to themselves by folding around C_2 axes, situated at an angle of π/n from each other. Thus, the generic fiber is folded with multiplicity $2n$ in mI. The fibers above the nearby edges weave up and down n times, and nearby circles wind around $2n$ times.

/u/gt3m/3mbook/pictures/chap5/8/WEAVE.ps not found

WEAVE

Figure 5.39. 2

When the pieces are assembled, we obtain this knot or link:

/u/gt3m/3mbook/pictures/chap5/8/figur2.ps not found

figur2

Figure 5.40. 2

When $D^2_{(\ ;n_1,n_2,\ldots,n_k)}$, is elliptic, then all geodesics are closed, having length dividing 2π, and the geodesic flow comes from a circle action. It follows that $T_1(D^2_{(\ ;n_1,n_2,\ldots,n_k)})$, is a fibration in a second way, by projecting to the quotient space by the geodesic flow. For instance, the singular locus of $T_1(D^2_{(2,3,5)})$ can be rearranged to fit on the surface of a torus, as a (3,5) curve. Therefore, it also fibers over $S^2_{(2,3,5)}$.

figur3

figur3

Figure 5.41. figur3

A more complete discussion describing and classifying fibered orbifolds in terms of the Euler number and other invariants may be found in [Dun] or [BS85]. Dunbar also has illustrations of all the Euclidean and elliptic orbifolds which have underlying space S^3. Bonahon and Siebenmann analyze Conway's notion of "algebraic" knots and links by considering the C_2-orbifold which the link determines and studying decompositions of these orbifolds as unions of fiber bundles, when such a decomposition exists.

Section "Tetrahedral orbifolds"
TETRAHEDRAL ORBIFOLDS
% Three-dimensional orbifolds
Section "GEOM ON SIMP"

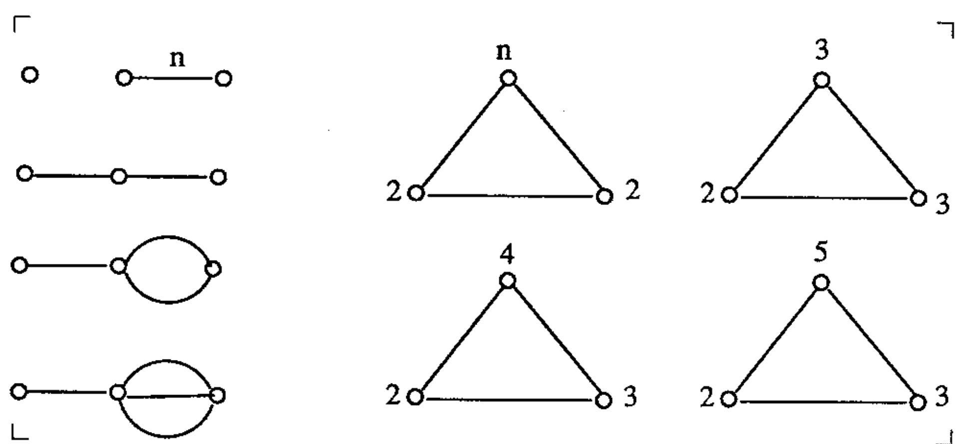
5.9. Tetrahedral orbifolds

It is beyond our scope to give a geometric classification of all orbifolds, but we will analyze some special classes of orbifolds which illustrate the beauty and wide adaptability of geometric structures.

We will begin by analyzing a special, but significant, classical set of orbifolds which have the combinatorial type of tetrahedron. The next project will be to classify orbifolds whose underlying space is a three-manifold with boundary, and whose singular locus is the boundary. In particular, the case when X_Q is the three-disk is interesting – the fundamental group of such an orbifold (if it is good) is called a *reflection group.*

There is a system of notation, called the *Coxeter diagram,* which is efficient for describing n-orbifolds of the type of a simplex. The Coxeter diagram is a graph, whose vertices are in correspondence with the (n-1)-faces of the simplex. Each pair of (n-1)-faces meet on an (n-2)-face which is a corner reflector of some order k. The corresponding vertices of the Coxeter graph are joined by (k-2)-edges, or alternatively, a single edge labeled with the integer k-2. The notation is efficient because the most commonly occurring corner reflector has order 2, and it is not mentioned. Sometimes this notation is extended to describe more complicated orbifolds with $X_Q = D^n$ and $\Sigma_Q \subset \partial D^n$, by using dotted lines to denote faces which are not incident. However, for a complicated polyhedron – even the dodecahedron – this becomes quite unwieldy.

A necessary and sufficient condition for a graph with (n+1)-vertices to determine an orbifold (of the type of a n-simplex) is that each complete subgraph on n vertices is the Coxeter diagram for an elliptic (n-1)-orbifold, since the group generated by any n mirrors is the local fundamental group at one of the vertices of the simplex (see the discussion in section 5.6).

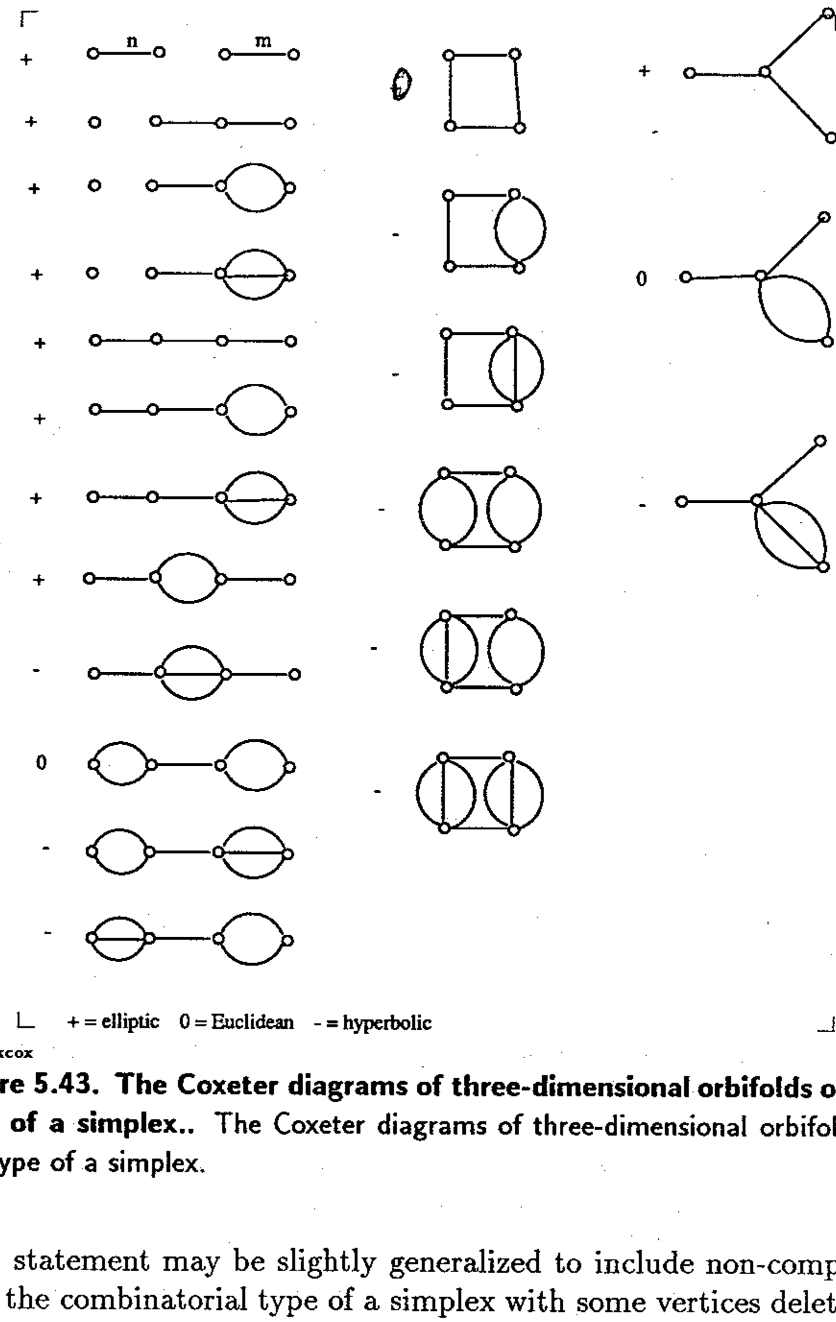


ellipticcox

Figure 5.42. The Coxeter diagrams for the elliptic triangle orbifolds..
The Coxeter diagrams for the elliptic triangle orbifolds.

% simplexcox
Section "GEOM ON DELSIMP"
% delsimpcox
% GEOM ON SIMP
% GEOM ON DELSIMP

Theore 5.9.1. *Every n-orbifold of the type of a simplex has either an elliptic, Euclidean or hyperbolic structure. The types in the three-dimensional case are listed in figure 5.43.*



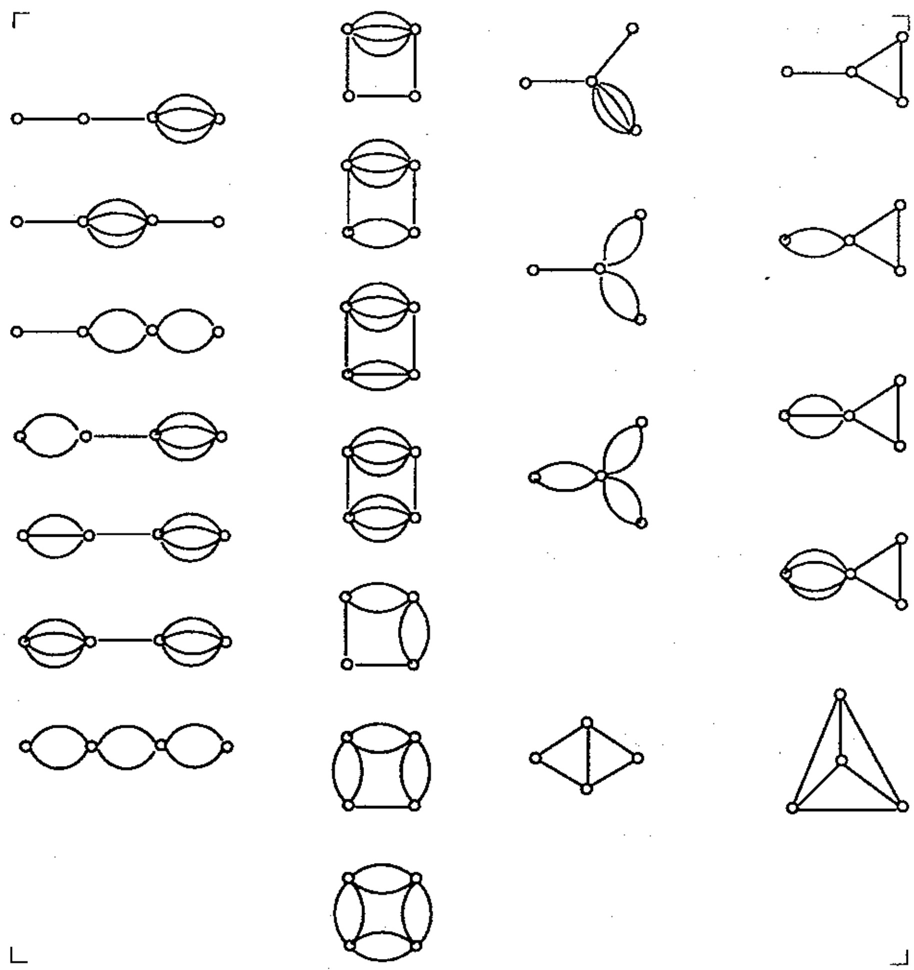
simplexcox

Figure 5.43. The Coxeter diagrams of three-dimensional orbifolds of the type of a simplex.. The Coxeter diagrams of three-dimensional orbifolds of the type of a simplex.

This statement may be slightly generalized to include non-compact orbifolds of the combinatorial type of a simplex with some vertices deleted.

Theore 5.9.2. *Every n-orbifold which has the combinatorial type of a simplex with some deleted vertices, such that the "link" of each deleted vertex is a Euclidean orbifold, and whose Coxeter diagram is connected, admits a complete hyperbolic structure of finite volume. The three-dimensional examples are listed in figure 5.44.*

Proof of 5.9.1 and 5.9.2:



delsimpcox

Figure 5.44. Orbifolds of simplex type with deleted vertices. Coxeter diagrams for the three-orbifolds of the combinatorial type of a simplex with deleted vertices

The method is to describe a simplex in terms of the quadratic form models. Thus, an n-simplex σ^n on S^n has (n+1)-hyperfaces. Each face is contained in the intersection of a codimension-one subspace of E^{n+1} with S^n. Let $V_0, ..., V_n$ be unit vectors orthogonal to these subspaces in the direction away from σ^n. Clearly, the V_i are linearly independent (otherwise the faces would all intersect in a single point). Note that $V_i \cdot V_i = 1$, and when $i \neq j$, $V_i \cdot V_j = -\cos\alpha_{ij}$, where α_{ij} is the angle between face i and face j. Similarly, each face of an n-simplex in H^n is contained in the intersection of a subspace of $E^{n,1}$ with the sphere of imaginary radius $X_1^2 + ... + X_n^2 - X_{n+1}^2 = -1$ (with respect to the standard inner product $X \cdot Y = \sum_{i=1}^n X_i \cdot Y_i - X_{n+1} \cdot Y_{n+1}$ on $E^{n,1}$). Outward vectors $V_0, ..., V_n$ orthogonal to these subspaces have real length, so they can be normalized to have length 1 . Again, the V_i are linearly independent and $V_i \cdot V_j = -\cos\alpha_{ij}$ when $i \neq j$. For an n-simplex σ^n in Euclidean n-space, let $V_0, ..., V_n$ be outward unit vectors in directions orthogonal to the faces on σ^n. Once again, $V_i \cdot V_j = -\cos\alpha_{ij}$.

% GEOM ON SIMP
% GEOM ON DELSIMP

Given a collection $\{\alpha_{ij}\}$ of angles, we now try to construct a simplex. Form a matrix M of presumed inner products, with 1's down the diagonal and $m_{ij} = -\cos\alpha_{ij}$ off the diagonal. If the quadratic form represented by M is positive definite or of type $(n, 1)$, then we can find an equivalence to E^{n+1} or $E^{n,1}$, which sends the basis vectors to vectors $V_0, ..., V_n$ having the specified inner product matrix. (Recall that $S^tMS = D$, where D is the diagonal quadratic form of $\mathbf{E}^{n+1}$ or $\mathbf{E}^{n,1}$ and the columns of S are scaled versions of M's eigenvectors. Our vectors are the rows of S.) The intersection of the half-spaces $X \cdot V_i \leq 0$ is a cone, which must be non-empty since the $\{V_i\}$ are linearly independent. In the positive definite case, the cone intersects S^n in a simplex, whose dihedral angles β_{ij} satisfy $\cos\beta_{ij} = \cos\alpha_{ij}$, hence $\beta_{ij} = \alpha_{ij}$. In the type $(n, 1)$ case, the cone determines a simplex in $\mathbf{RP}^n$, but the simplex may not be contained in $H^n \subset \mathbf{RP}^n$. To determine the positions of the vertices, observe that each vertex v_i, the intersection of the planes corresponding to $V_0, \ldots, \hat{V_i}, \ldots, V_n$, determines a one-dimensional subspace whose orthogonal subspace is spanned by $V_0, \ldots, \hat{V_i}, \ldots, V_n$. The vertex v_i is on H^n, on the sphere at infinity, or outside infinity according to whether the quadratic form restricted to this orthogonal subspace is positive definite, degenerate, or of type $(n\text{-}1, 1)$. Thus, the angles $\{\alpha_{ij}\}$ are the angles of an ordinary hyperbolic simplex iff M has type $(n, 1)$, and for each i the submatrix obtained by deleting the i^{th} row and the corresponding column is positive definite. They are the angles of an ideal hyperbolic simplex (with vertices in H^n or S^{n-1}_∞) iff all such submatrices are either positive definite, or have rank n-1.

By similar considerations, the angles $\{\alpha_{ij}\}$ are the angles of a Euclidean n-simplex iff M is positive semidefinite of rank n.

When the angles $\{\alpha_{ij}\}$ are derived from the Coxeter diagram of an orbifold, then each submatrix of M obtained by deleting the i^{th} row and the i^{th} column corresponds to an elliptic orbifold of dimension $n - 1$, hence it is positive definite. The full matrix must be either positive definite, or type $(n, 1)$ or positive semidefinite with rank n. It is an easy exercise to list the examples in any dimension, and compute the determinant of their corresponding matrix to ascertain the geometry in which they are realized. We have thus proven theorem 5.9.1.

Notice now that in the Euclidean case, the subspace of vectors of zero length with respect to M is spanned by $a = (a_0, ..., a_n)$, where a_i is the $(n\text{-}1)$-dimensional area of the i^{th} face of σ (the j^{th} entry of Ma is the sum of the oriented areas of the projections of faces onto the plane determined by the j^{th} face, which is clearly 0).

To establish 5.9.2, first consider any submatrix M_i of rank n-1 which is obtained by deleting the i^{th} row and i^{th} column so the link of the i^{th} vertex is

Euclidean. Change basis so that M_i becomes

$$\begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 0 \end{pmatrix}$$

using $(a_0, \dots, \hat{a}_i, \dots, a_n)$ (which is just $\sum_{i \neq j} a_i V_i$) as the last basis vector. When the basis vector V_i is put back, the quadratic form determined by M becomes

$$\begin{pmatrix} 1 & & & 0 & * \\ & \ddots & & & * \\ & & 1 & & * \\ 0 & & & 0 & -C \\ * & * & * & -C & 1 \end{pmatrix}$$

where $-C = -\sum_{j \neq i} a_i \cos \alpha_{ij}$ is negative since the Coxeter diagram was supposed to be connected. It follows that M has type $(n, 1)$, which implies that the orbifold is hyperbolic. 5.9.1 and 5.9.2

Remark: The material in this section is essentially due to [Cox34], [Lan50] and [CW50]. An interesting survey of the subject is given by [Vin85].

I'm not sure the second reference to Coxeter is necessary but he seems to think that some of his results in the second reference preceeded Lannner's-Dick

Section "Andre'ev's theorem and generalizations" ::n-disk quotient Definition "incompressible" ::compressible ::incompressible ::compression disk quotient % incompressible ::parallel Theorem "Andre'ev 1" Theorem "Andre'ev 2"

5.10. Andre'ev's theorem and generalizations

The somewhat subtle classification of orbifolds combinatorially equivalent to a simplex would seem to indicate that the analysis for a general polyhedron would be hopelessly complicated. Amazingly, the situation in dimension 3 becomes much simpler and cleaner once the polyhedron is more complicated than a simplex. Before stating the theorem, due to [?, ?], we need some definitions.

An orbifold Q with boundary is said to be an *n-disk quotient* if its universal cover is the n-disk D^n.

Definition 5.10.1 (incompressible). A two-dimensional compact suborbifold P of a three-orbifold Q is *compressible* if one of the following two conditions is satisfied.

a) P is the boundary of a suborbifold of Q, which is a three-disk quotient R.

b) There is a one-dimensional suborbifold $S \subset P$ which is *not* the boundary of a two-disk quotient in the boundary of Q and which *is* the boundary of a two-disk quotient R which is a proper suborbifold of Q.

An *incompressible* two-dimensional compact suborbifold of Q is a compact suborbifold which is not compressible. It would be possible also to talk of non-compact compressible and incompressible surfaces, but we will not do so here. In each case, the suborbifold R is called a *compression disk quotient.*

Note in particular that according to the definition any bad suborbifold is automatically incompressible. The two different types of compressible surfaces in definition 5.10.1 are of Euler characteristic greater than zero and less than or equal to zero respectively. The definitions can also be adapted to one-dimensional suborbifolds of two-dimensional orbifolds. For example, consider the mirrored annulus. An arc going from a mirror to itself is compressible, because it bounds a suborbifold whose double cover is a two-disk. An arc going from one of the mirrors of the annulus to the other is incompressible. If we multiply by S^1 we get an example of a compressible suborbifold and an incompressible suborbifold.

Bill, put in the new symbolic name

Two proper suborbifolds S_0 and S_1 of Q are said to be *parallel* if there is a suborbifold $P \times I$ of Q, such that $P \times 0 = S_0$ and $P \times 1 = S_1$.

Theorem 5.10.2 (Andre'ev 1). *Let Q be a three-orbifold with X_Q homeomorphic to B^3 and $\Sigma_Q = \partial B^3$. Suppose that Q has at least five faces. Then Q has a hyperbolic structure if and only if every incompressible compact suborbifold P of Q has $\chi(P) < 0$.*

Andre'ev also proved a generalization to the finite volume case:

Theorem 5.10.3 (Andre'ev 2). *Let Q be a three-orbifold with X_Q homeomorphic to B^3(−finitely many points) which is combinatorially a polyhedron with some vertices deleted, and suppose that Σ_Q is the boundary of B^3(−finitely*

% Andre'ev 1
% Andre'ev 2
% Andre'ev 1
% Andre'ev 2
% Classification of non-hyperbolic 2-orbifolds
% suborbifolds and the singular locus
% 3VERTEXPOSS

many points). Suppose Q has at least five faces. Then Q has a hyperbolic structure with finite volume if and only if

a) *The link of each missing vertex is a Euclidean two-orbifold, and*

b) *each incompressible compact suborbifold $P \subset Q$ with $\chi(P) \geq 0$ is parallel to the link of a missing vertex.*

We will prove these theorems later, as a consequence of more general results. In the meantime we can at least prove the theorem's easy direction. The proof refers ahead to some basic three-manifold topology discussed in chapter ???. These forward references could have been avoided by rephrasing the statements of theorem 5.10.2 and theorem 5.10.3 in a purely combinatorial way, thus avoiding the mention of incompressible suborbifolds, whose definitions are useful only because of Alexander's theorem and the loop theorem. We have chosen this formulation because it is conceptually clearer than the combinatorial formulation.

Proof of necessity in 5.10.2 and 5.10.3: Suppose Q has a hyperbolic structure. Then Q looks like a convex polyhedron in $\mathbf{H}^3$. We first take a look at ?? and see that each vertex of Q lies on exactly three edges, because its link is a spherical orbifold, except that a missing vertex may have four incident edges. Let P be a compact incompressible suborbifold P with $\chi(P) \geq 0$. If P is a sphere or a torus, then it must be disjoint from the singular set, by problem 5.2.12. If it is a sphere, it bounds a three-ball by Alexander's theorem (see [?]), and so it is compressible. If it is a torus, Dehn's lemma (see [Hem76]) proves that it is compressible. To show that P cannot be an annulus, note that its boundary is disjoint from Σ^1_Q. But each face of Q is simply connected, and so each boundary curve bounds a compression disk in $\Sigma_Q - \Sigma^1_Q$. So we know that P is combinatorially a disk, with 0,1,2,3 or 4 vertices.

If P has no vertices, then its boundary lies on a plane, and bounds a disk on that plane (by the Jordan curve theorem and by Alexander's theorem P bounds a B^3/C_2). P cannot have one vertex, because a hyperbolic plane never meets itself at an angle. If P has two vertices, they must both lie on the line of intersection of two planes, and P bounds a B^3/D_n.

If P has three vertices, then the sides of P lie on three planes which intersect in a point p in $\mathbf{H}^3$ if $\chi(P) > 0$ or on S^2_∞ if $\chi(P) = 0$. We claim that p is in Q, or in its closure if it is on S^2_∞. To see this, consider any vertex v of Q between P and p and as far as possible from p. One first shows that v must lie on an edge e containing one of the vertices p_1 of P, and p_1, v and p are collinear. But a spherical orbifold which is a mirrored triangle has each side of length at most $\pi/2$. This means that the edges at v other than e have the other endpoint further from p than v, contradicting the definition of v. (see figure 5.45). Therefore, P bounds a B^3/Γ or is parallel to the link of a missing cusp vertex.

I don't know what the figure 3VERTEXPOSS looks like, but it is unlikely to be fully appropriate for my proof. dbae

If P has four vertices, then they must all be D_2 vertices. Consider the four planes determined by the four sides of P. Either they all meet at ∞, or

% 4VERTEXPOSS
% Andre'ev 1
% Andre'ev 2
Corollary "circle pack"
% Andre'ev 1
% circle pack

/u/gt3m/3mbook/pictures/chap5/10/3VERTEXPOSS.ps not found

3VERTEXPOSS

Figure 5.45. . Missing picture

two of them meet in an edge and the other two are perpendicular to the edge (see figure 5.46). In either case, we apply an argument as in the three-vertex case to deduce that P is compressible or parallel to a missing cusp vertex.

necessity in 5.10.2 and 5.10.3

/u/gt3m/3mbook/pictures/chap5/10/4VERTEXPOSS.ps not found

4VERTEXPOSS

Figure 5.46. 2. The two possible ways that four planes can meet at 90°

Before proceeding to generalize and prove 5.10.2 and 5.10.3, we will give two beautiful applications to Euclidean geometry.

A *packing* of disks in $\mathbf{E}^2$ (or $\mathbf{H}^2$ or S^2) means a collection of closed disks with disjoint interiors. The *nerve* of a packing is then a one-complex, whose vertices correspond to the disks, and whose edges correspond to pairs of disks which intersect. This graph has a canonical embedding in the plane, by mapping the vertices to the centers of the disks and the edges to straight line segments which will pass through points of tangency of the disks.

Corollary 5.10.4 (circle pack). *Let γ be any graph in $\mathbf{R}^2$, such that each has distinct ends and no two vertices are joined by more than one edge. Then there is a packing of circles in $\mathbf{E}^2$ whose nerve is isotopic to γ. If γ is the one-skeleton of a triangulation of S^2, then this disk packing is unique up to Moebius transformations.*

Proof of 5.10.4: We transfer the problem to S^2 by stereographic projection. Add an extra vertex in each non-triangular region of $S^2 - \gamma$, and edges connecting it to neighboring vertices, so that γ becomes the one-skeleton of a triangulation T of S^2. Let P be the polyhedron (meaning only a topological structure of faces, edges, and vertices) obtained by cutting off neighborhoods of the vertices of T, down to the middle of each edge of T, and replacing them by new polygonal faces.

Let Q be the orbifold with underlying space $X_Q = D^3 -$ vertices of P, and $\Sigma^1_Q =$ edges of P, each modeled on $\mathbf{R}^3/D_2$. For any compact incompressible suborbifold Q' with $\chi(Q') \geq 0$, $\partial X_{Q'}$ must be a curve crossing at most four edges of P; but the only such curves are the curves which circumnavigate a vertex. These cross exactly four edges and bound Euclidean suborbifolds. Thus, Q satisfies the hypotheses of 5.10.2, and Q has a hyperbolic structure. This means that P is realized as an ideal polyhedron in $\mathbf{H}^3$, with all dihedral angles equal to 90°. The planes of the new faces of P (those resulting from truncating vertices of T) intersect S^2_∞ in circles. Two of the circles are tangent whenever the two faces meet at an ideal vertex of P. This is the packing required by 5.10.4.

fig1

fig1

Figure 5.47. fig1. A circle packing of S^2

fig2

fig2

Figure 5.48. fig2. γ extended to a triangulation of S^2

fig3

fig3

Figure 5.49. fig3. P is obtained from T by "truncating" at the vertices

% MOSTOWGEO Corollary "circumpoly" % circle pack % The hyperboloid model and the Klein model

Mostow's theorem ?? states that the hyperbolic structure on Q is unique up to Moebius transformations. Since the hyperbolic structure on Q may be reconstructed from the packing of disks on S^2_∞, our circle packing is also unique up to Moebius transformations. To make the reconstruction, observe that any three pairwise tangent disks have a unique common orthogonal circle. The set of planes determined by the packing of disks on S^2_∞, together with extra circles orthogonal to the triples of tangent circles coming from vertices of the triangular regions of $S^2 - \gamma$ cut out a polyhedron of finite volume combinatorially equivalent to P, which gives a hyperbolic structure for Q. 5.10.4

Remark: Andre'ev also gave a proof of uniqueness of a hyperbolic polyhedron with assigned concave angles, so the reference to Mostow's theorem is not essential.

Corollary 5.10.5 (circumpoly). *Let T be any triangulation of S^2. Then there is a convex polyhedron in $\mathbf{R}^3$, combinatorially equivalent to T whose one-skeleton is circumscribed about the unit sphere (i.e., each edge of T is tangent to the unit sphere). Furthermore, this polyhedron is unique up to a projective transformation of $\mathbf{R}^3 \subset \mathbf{RP}^3$ which preserves the unit sphere.*

Proof of 5.10.5: Construct the ideal polyhedron P, as in the proof of 5.10.4. Embed $\mathbf{H}^3$ in $\mathbf{RP}^3$, as the projective model. The intersection of the half-spaces(in $\mathbf{RP}^3$) determined by old faces of P (i.e those coming from faces of T) forms a polyhedron in $\mathbf{RP}^3$, combinatorially equivalent to T. Notice that the vertices of this polyhedron are dual (in the sense of section 2.3) to the new faces of P. Adjust by a projective transformation if necessary so that this polyhedron is in $\mathbf{R}^3$. (To do this, transform P so that the origin is in its interior.) 5.10.5

Remarks: Note that the dual cell-division T^* to T is also a convex polyhedron in $\mathbf{R}^3$, with one-skeleton of T^* circumscribed about the unit sphere. The intersection $T \cap T^* = P$.

These three polyhedra may be projected from the north pole of $S^2 \subset \mathbf{RP}^3$ along straight lines to $\mathbf{R}^2 \subset \mathbf{RP}^2$. Stereographic projection is conformal on S^2, so the edges of T project to lines orthogonal to the packing of circles, while the edges of T^* project to tangents to these circles. It follows that the vertices of T project to the centers of the circles. Thus the image of the one-skeleton of T is the geometric embedding in $\mathbf{R}^2$ of the nerve γ of the circle packing.

/u/gt3m/3mbook/pictures/chap5/10/fig4.ps not found

fig4

Figure 5.50. 1.5. A circle packing and its nerve

The existence of other geometric patterns of circles in $\mathbf{R}^2$ may also be deduced from Andre'ev's theorem. For instance, it gives necessary and sufficient condition for the existence of a family of circles meeting only orthogonally in a

Theorem "Andre'ev on torus"
Theorem "Andre'ev on surface"
% Andre'ev on torus
% Andre'ev on surface
% Andre'ev 1
% Andre'ev 2

certain pattern, or meeting at 60° angles. This connection will be made clearer in the next two sections.

One might also ask about the existence of packing of circles on surfaces of constant curvature other than S^2. The answers are corollaries of the following theorems:

Theorem 5.10.6 (Andre'ev on torus). *Let Q be an orbifold such that $X_Q \cong T^2 \times [0,\infty)$, (possibly with some deleted vertices on $T^2 \times 0$ having Euclidean links) and $\Sigma_Q = \partial X_Q$. Then Q admits a complete hyperbolic structure of finite volume if and only if each compact incompressible suborbifold $P \subset Q$ with $\chi(P) \geq 0$ is a Euclidean suborbifold parallel to the link of a missing vertex or to the link of the point at infinity corresponding to $T^2 \times \infty$.*

Theorem 5.10.7 (Andre'ev on surface). *Let M^2 be a closed surface, with $\chi(M^2) < 0$. An orbifold Q such that $X_Q = M^2 \times [0,1]$ (possibly with some deleted vertices on $M^2 \times 1$ having Euclidean links), $\Sigma_Q = \partial X_Q$ and Σ_Q^1, the one-skeleton of Σ_Q, is contained in $\subset M^2 \times 1$. Then Q has a hyperbolic structure if and only if each compact incompressible suborbifold $P \subset Q$ with $\chi(P) \geq 0$ is a Euclidean suborbifold parallel to the link of a missing vertex.*

fig5

fig5

Figure 5.51. fig5

Proof of necessity in 5.10.6 and 5.10.7: We will not carry out the proof in detail, because it is almost identical with the proof of necessity in 5.10.2 and 5.10.3. We confine ourselves to discussing the one situation which is different here from the preceding proof—that is, we show that if Q is a hyperbolic orbifold, then any incompressible mirrored annulus P is parallel to a missing vertex. The universal cover of Q is $\mathbf{H}^3$. The inverse image of the singular set is a set of hyperbolic planes, which cuts the inverse image of the annulus into components. Let C be one of these components, with disjoint boundary components $\partial_0 C$ and $\partial_1 C$, lying on hyperbolic planes H_0 and H_1. Since the boundary curves on the annulus P are mirrors, they cannot intersect the one-skeleton of Σ_Q.

Bill, do you want to have the formal nomenclature for these 2-orbifolds as they arise? dbae

% circle pack
% circumpoly
% Andre'ev on torus
% Andre'ev on surface
% circumpoly
Example "Borromean rings"
% borromean orbifold
% BOROCUBE
% BOROCUBE

Therefore, if $\partial_0 C$ and $\partial_1 C$ are simple closed curves, then they bound disks in H_0 and H_1 respectively, and these disks are disjoint from the one-skeleton. This means that the boundary curves of P bound compression disks in $\Sigma_Q - \Sigma_Q^1$.

If $\partial_0 C$ is an infinite curve then the corresponding generator γ of $\pi_1 P$ is parabolic or hyperbolic, and its action preserves H_0 and H_1. If $H_0 = H_1$, then there is an annulus A in $\Sigma_Q - \Sigma_Q^1$ with the same boundary curves as P. If we move A a little, we obtain a suborbifold of Q which is isotopic to P, keeping the boundary of the annulus fixed. Without loss of generality, $P = A$, and then it is easy to see there is a compression disk quotient.

So we may assume that $H_0 \neq H_1$. If γ is parabolic, then we put the parabolic point at infinity in the upper half space model, and see that H_0 and H_1 must be parallel vertical half planes. The surface called C above is an infinite strip, invariant under γ. Since each complementary component of the singular set is a convex subset of $\mathbf{H}^3$, everything in $\mathbf{H}^3$ lying above C is in this component. This shows that P is parallel to the vertex of a missing cusp point.

The final situation to consider is when γ is hyperbolic and $H_0 \neq H_1$. Then its axis must equal to $H_0 \cap H_1$, which lies in the one-skeleton of the singular set, and we have already seen that $\partial_0 C$ and $\partial_1 C$ are disjoint from this. We obtain a wedge of hyperbolic space bounded by C, a strip of H_0 and a strip of H_1, which covers the product of a triangle with two mirrored sides with S_1. The interior of the wedge is disjoint from the singular set. Consider a triangle with one edge on C, one edge on H_0 and one edge on H_1. This is a compression disk quotient for P. some pictures would be helpful

| necessity in 5.10.6 and 5.10.7 |

Exercise 5.10.8. Formulate and deduce the analogues of corollaries 5.10.4 and 5.10.5 for theorems 5.10.6 and 5.10.7. (Hint: Consider how $\pi_1(Q)$ acts on S^2_∞, to get the right pictures for the circle-packing corollary. To get a good picture for 5.10.5, make use of horospheres and surfaces which are a constant distance from the hyperbolic plane.)

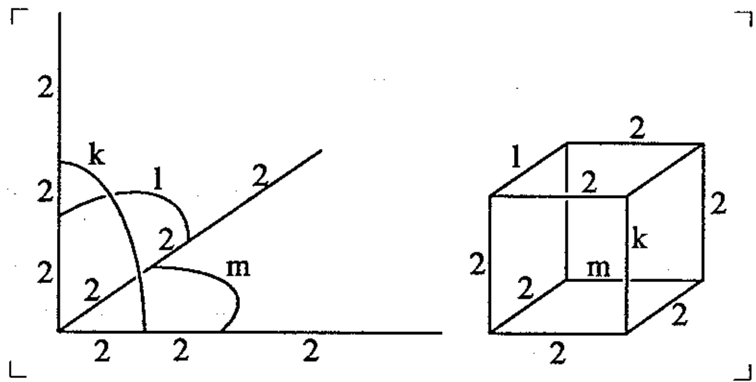
Example 5.10.9 (Borromean rings). We have seen (5.1.8) that the Borromean rings are the singular locus for a Euclidean orbifold, in which they are elliptic axes of order 2. With the aid of Andre'ev's theorem, we may find all hyperbolic orbifolds which have underlying space S^3 and the Borromean rings as singular locus. The rings can be arranged so they are invariant by reflection in three orthogonal great spheres in S^3 (see figure 5.52). We can form a new orbifold Q by making the rings elliptic axes of orders k, l and m, and by acting on it with the reflections in the great spheres we see that it is an eight-fold covering of another orbifold, which has the combinatorial type of a cube (see 5.52). By Andre'ev's theorem, such an orbifold has a hyperbolic structure if and only if k,l and m are all greater than 2. If k is 2, for example, then there is a sphere in S^3 separating the elliptic axes of order l and m and intersecting the elliptic axis of order 2 in four points. This forms an incompressible Euclidean suborbifold of Q, which breaks Q into two halves, each fibering over two-orbifolds with boundary, but in incompatible ways (unless l or m is 2).

When $k = l = m = 4$, the fundamental domain, for $\pi_1(Q)$ acting on $\mathbf{H}^3$ is a regular right angled dodecahedron!

Any of the numbers k,l or m can be permitted to take the value ∞ in this discussion, denoting a parabolic cusp. When $l = m = \infty$, for instance, then Q has a k-fold cover which is the complement of the untwisted $2k$-link chain D_{2k}.

Problem 5.10.10. Which hyperbolic orbifolds have D_{2k} as singular locus?

difficulty level random
dbae



BOROCUBE

Figure 5.52. 1.5

/u/gt3m/3mbook/pictures/chap5/10/fig6.ps not found

fig6

Figure 5.53. 1.5. Base spaces of the fibrations

Section "Constructing patterns of circles"
PATTERNS OF CIRCLES
% ANDREEV1 ON SURF
% ANDREEV2 ON SURF
% ANDREEV1
% ANDREEV2

5.11. Constructing patterns of circles

We will formulate a precise statement about patterns of circles on surfaces of non-positive Euler characteristic which gives theorems ?? and ?? as immediate consequences. We will later use these when we return to the proof of Andre'ev's theorems (?? and ??).

If S is a complete surface of constant curvature and if $\mathcal{D}$ is a family of immersed geometric disks on S, we can describe the pattern of intersection in terms of a graph $G(\mathcal{D})$, whose vertices are in correspondence with the elements of $\mathcal{D}$ and whose edges are in correspondence to the "intersections" of elements of $\mathcal{D}$. Here we mean *an intersection* not in the mathematical sense but in a more visual sense which we can define mathematically to be a homotopy class of paths $p : [0,1] \to S$ from the center of a disk $D \in \mathcal{D}$ to the center of another disk $D' \in \mathcal{D}$ (not necessarily different from D) such that when p is lifted to the universal cover $\widetilde{S}$, the lifts of the closed disks D and D' centered at its endpoints have a non-empty intersection.

Let E denote the set of edges of $G(\mathcal{D})$. There is a function $\Theta : E \to [0, \pi]$ which measures angles of intersection of the boundaries of the disks. We measure the (unsigned) angle between the clockwise tangent of one disk and the counterclockwise tangent of the others, so that an angle of 0 means an external tangency and π means an internal tangency.

There is a canonical geodesic map $g : G(\mathcal{D}) \to S$, provided the curvature is non-positive, which takes each vertex to the center of its disk and each edge to the geodesic in its homotopy class.

Associated with $\mathcal{D}$ there is a hyperbolic manifold $P(\mathcal{D})$ with polyhedral boundary. When S has positive curvature, we scale the metric to make it coincide with S^2_∞, and obtain $P(\mathcal{D})$ by chopping out the hyperbolic half-spaces which intersect S^2_∞, in the elements of $\mathcal{D}$. When S has curvature 0, we identify its universal cover $\widetilde{S}$ with the boundary of upper half-space. We remove hyperbolic half-spaces corresponding to the disks on $\widetilde{S}$, and divide by $\pi_1(S)$ to obtain $P(\mathcal{D})$. If the curvature is negative, we identify $\widetilde{S}$ with the northern hemisphere of S^2_∞, chop out hyperbolic half-spaces corresponding to the disks on $\widetilde{S}$. We also remove the lower half-ball, and divide by $\pi_1(S)$ to obtain $P(\mathcal{D})$.

We specialize to the case that the angles do not exceed $\pi/2$, and in the spherical case that each disk is smaller than a hemisphere. Then, an edge e of $g(G(\mathcal{D}))$ with $\Theta(E) > 0$ can only intersect the two relevant disks, so the edges of the polyhedral manifold $P(\mathcal{D})$ correspond to edges e of $G(\mathcal{D})$ with $\Theta(e) > 0$, and the dihedral angles of $P(\mathcal{D})$ are given by Θ. This means also, that $g : G(\mathcal{D}) \to S$ is an embedding; except that the following pattern may occur:

Let G' be the graph embedded in S obtained by deleting all edges of $g(G(\mathcal{D}))$ which intersect other edges. We will assume that $S - G'$ is contractible. Denote its edge set E'.

% ANDREEV1
% ANDREEV2
Section "CIRCLE COND"
Section "CIRCLE PAT"
% CIRCLE COND
% ANDREEV1 ON SURF
% ANDREEV2 ON SURF
% CIRCLE PAT
% CIRCLE COND
Section "3 CIRC CONFIG"

Arguing as in the necessity proof for ?? and ??, we see that Θ (transferred to G') must satisfy certain conditions:

5.11.1Conditions:

a) No edge forms a null-homotopic closed loop.

b) The only null-homotopic closed loops made of two edges have the form $e * e^{-1}$.

c) If $e_1 * e_2 * e_3$ $(e_i \in E')$ forms a null-homotopic closed loop, and if $\sum_{i=1}^{3} \Theta(e_i) \geq \pi$, then these three edges go around the boundary of a triangular component of $S - G'$.

d) If $e_1 * e_2 * e_3 * e_4$ $(e_i \in E)$ is a null-homotopic closed loop, and if $\sum_{i=1}^{4} \Theta(e_i) = 2\pi$, then this loop either goes around the boundary of a quadrilateral of $S - G'$ or goes around two adjacent triangles.

In fact, these conditions are sufficient.

Theore 5.11.2. *Let S be a closed surface with $\chi(S) \leq 0$. Let $G' \subset S$ be any embedded graph such that $S - G'$ is contractible, and $\Theta : E' \to [0, \pi/2]$ any function (where E' is the edge set of G') so that G' and Θ satisfy conditions 5.11.1.*

Then there is a metric of constant curvature on S and a family F of disks on S so that G' and Θ arise from F. The metric together with the disks are unique, up to a change of metric by a constant followed by an isometry between metrics which is isotopic to the density.

This theorem establishes ?? and ??, since if we take $G(\mathcal{D})$ to be dual to Σ_1^Q and choose Θ to reflect the required cone angles, then $P(\mathcal{D})$ is the desired orbifold.

Proof of 5.11.2: First, we can augment the graph G' by adding edges which subdivide it into triangles with the value of $\Theta = 0$ on these edges. Conditions 5.11.1 remain valid for the new graph, which we will call G. Let τ be the generalized triangulation having G as its one-skeleton.

The idea is to solve for the *radii* of the circles C_{v_i}. Given an arbitrary set of radii, we shall construct a Riemannian metric on S with cone type singularities at the vertices of τ, which has a family of circles of the given radii meeting at the given angles. We adjust the radii until S lies flat at each vertex. Thus, the proof is closely related to the idea that one can make a conformal change of any given Riemannian metric on a surface until it has constant curvature. Recall that a conformal map is one which takes infinitesimal circles to infinitesimal circles; the conformal factor is the ratio of the radii of the target and source circles.

Lemm 5.11.3. *For any three non-obtuse angles Θ_1, Θ_2 and $\Theta_3 \in [0, \pi/2]$ and any three positive numbers R_1, R_2 and R_3, there is a configuration of three circles in both hyperbolic and Euclidean geometry, unique up to isometry having radii R_i and meeting in angles Θ_i.*

% CIRCLE PAT
% 3 CIRC CONFIG

Proof of 5.11.3: The length l_k of a side of the hypothetical triangle of centers of the circles is determined as the side opposite the obtuse angle $\pi - \Theta_k$ in a triangle whose other sides are R_i and R_j. Thus, $\sup\,(R_i, R_j) < l_k \leq R_i + R_j$. The three numbers l_1, l_2 and l_3 obtained in this way clearly satisfy the triangle inequalities $l_k < l_i + l_j$. Hence, one can construct the appropriate triangle, which gives the desired circles. $\boxed{5.11.3}$

Continuation of proof of 5.11.2: Let V denote the set of vertices of τ. For every element $R \in \mathbf{R}_+^V$ (i.e., if we choose a radius for the circle about each vertex), there is a singular Riemannian metric, which is pieced together from the triangles of centers of circles with given radii and angles of intersection as in 5.11.3. The triangles are taken in $\mathbf{H}^2$ or $\mathbf{E}^2$ depending on whether $\chi(S) < 0$ or $\chi(S) = 0$. The edge lengths of cells of τ match whenever they are glued together, so we obtain a metric, with singularities only at the vertices, and constant curvature 0 or -1 everywhere else.

The notion of curvature can easily be extended to Riemannian surfaces with certain sorts of singularities. The curvature form KdA becomes a measure κ on such a surface. Tailors are of necessity familiar with curvature as a measure. Thus, a seam has geodesic curvature $(k_1 - k_2) \cdot v$, where v is one-dimensional Lebesgue measure and k_1 and k_2 are the geodesic curvatures of the two halves. (The effect of gathering is more subtle – it is obtained by putting two lines infinitely close together, one with positive curvature and one with balancing negative curvature. Another instance of this is the boundary of a lens.)

More to the point for us is the curvature concentrated at the apex of a cone: it is $2\pi - \alpha$, where α is the cone angle (computed by slitting the cone to the apex and laying it flat). It is easy to see that this is the unique value consistent with the Gauss-Bonnet theorem, which then generalizes to give:

$$\sum_{v \in C} \kappa(v) + \int_{\partial S} \kappa_g ds + \int_S K dA = 2\pi\chi(S) \qquad 5.11.4$$

where S is a surface with boundary ∂S and C is the set of cone points.

Formally, we now have a map

$$F : \mathbf{R}_+^V \to \mathbf{R}^V.$$

Given an element $R \in \mathbf{R}_+^V$, we construct the singular Riemannian metric on S, as above; $F(R)$ describes the discrete part of the curvature measure κ_R on S, in other words, $F(R)(v) = \kappa_R(V)$. Our problem is to show that 0 is in the image of F, for then we will have a non-singular metric with the desired pattern of circles built in.

When $\chi(S) = 0$, then the shapes of the Euclidean triangles do not change when we multiply R by a constant, so $F(R)$ also does not change. Thus we may as well normalize so that $\sum_{v \in V} R(v) = 1$. Let $\Delta \subset \mathbf{R}_+^V$ be this locus – Δ is the interior of the standard $|V| - 1$ simplex. Observe, by the Gauss-Bonnet

% CIRCLE PAT Section "VARY ONE RADIUS" % CIRCLE PAT % VARY ONE RADIUS

theorem, that $\sum_{v \in V} \kappa_R(v) = 0$. Let $Z \subset \mathbf{R}^V$ be the locus defined by this equation.

If $\chi(S) < 0$, then changing R by a constant does make a difference in κ. In this case, let $\Delta \subset \mathbf{R}_+^V$ denote the set of R such that the associated metric on S has total area $2\pi|\chi(S)|$. By the Gauss-Bonnet theorem, $\Delta = F^{-1}(Z)$ (with Z as above). We will prove that Δ intersects each ray through 0 in a unique point, so Δ is a simplex in this case also. This fact is easily deduced from the following lemma, which will also prove the uniqueness of the solution obtained in 5.11.2;

Lemm 5.11.5. *Let C_1, C_2 and C_3 be circles of radii R_1, R_2 and R_3 in hyperbolic or Euclidean geometry, meeting pairwise in non-obtuse angles. If C_2 and C_3 are held constant but C_1 is varied in such a way that the angles of intersection are constant but R_1 decreases, then the center of C_1 moves toward the interior of the triangle of centers.*

Thus we have

$$\frac{\partial \alpha_1}{\partial R_1} < 0, \quad \frac{\partial \alpha_2}{\partial R_1} > 0, \quad \frac{\partial \alpha_3}{\partial R_1} > 0,$$

where the α_i are the angles of the triangle of centers.

Proof of 5.11.5: Consider first the Euclidean case. Let l_1, l_2 and l_3 denote the lengths of the sides of the triangle of centers. The partial derivatives $\frac{\partial l_2}{\partial R_1}$ and $\frac{\partial l_3}{\partial R_1}$ can be computed geometrically. If v_1 denotes the center of C_1, then $\frac{\partial v_1}{\partial R_1}$ is determined as the vector whose orthogonal projections to sides 2 and 3 are $\frac{\partial l_2}{\partial R_1}$ and $\frac{\partial l_3}{\partial R_1}$. Thus, $R_1 \frac{\partial v_1}{\partial R_1}$ is the vector from v_1 to the intersection of the lines joining the pairs of intersection points of two circles.

When all angles of intersection of circles are acute, no circle meets the opposite side of the triangle of centers: It follows that $\frac{\partial v_1}{\partial R_1}$ points to the interior of $\Delta v_1 v_2 v_3$.

The hyperbolic proof is similar, except that some of it takes place in the tangent space to $\mathbf{H}^2$ at v_1. [5.11.5]

Continuation of proof of 5.11.2. From lemma 5.11.5 it follows that when all three radii are increased, the new triangle of centers can be arranged to contain the old one. Thus, the area of S is monotone, for each ray in $\mathbf{R}_+^V$. The area near 0 is near 0, and near ∞ it approaches $\pi \times (\#\textit{triangles}) = -2\pi\chi(S - V) > -2\pi\chi(S)$; thus the ray intersects $\Delta = F^{-1}(Z)$ in a unique point.

It is now easy to prove that $F : \Delta \to Z$ is one-to-one. In fact, consider any two distinct points R and $R' \in \Delta$. Let $V^- \subset V$ be the set of v where $R'(v) < R(v)$. Clearly V^- is a proper subset. Let τ_{V^-} be the subcomplex of τ spanned by V^-. (τ_{V^-} consists of all cells whose vertices are contained in V^-.) Let S_{V^-} be a small neighborhood of τ_{V^-}. We compare the geodesic curvature ∂S_{V^-} in the two metrics. To do this, we may arrange ∂S_{V^-} to be orthogonal to each edge it meets. Each arc of intersection of ∂S_{V^-} with a triangle having one vertex in V^- contributes approximately α_i to the total curvature, while

% VARY ONE RADIUS
% VARY ONE RADIUS
% GAUSS-BONN
% CIRCLE COND
% GBINEQ
% LIM GB
% ANDREEV1 ON SURF
% ANDREEV2 ON SURF
% CIRCLE PAT

each arc of intersection having two vertices in V^- contributes approximately $\beta_i + \gamma_i - \pi$.

In view of lemma 5.11.5, an angle such as α_1 increases in the R' metric. The change in β_1 and γ_1 is unpredictable. However, their sum must increase: first, let R_1 and R_2 decrease; $\pi - \delta_1 - (\beta_1 + \gamma_2)$, which is the area of the triangle in the hyperbolic case, decreases or remains constant but δ_1 also decreases, so $\beta_1 + \gamma_1$ must increase. Then let R_3 increase; by 5.11.5, β_1 and γ_1 both increase. Hence, the geodesic curvature of ∂S_{V^-} increases.

From the Gauss-Bonnet formula (5.11.4) applied to S_{V^-}, it follows that the total curvature at vertices in V^- must decrease in the R' metric (Note that the area of S_{V^-} decreases, so if $K = -1$, the third term on the left increases). In particular, $F(R) \neq F(R')$, which shows that F is one-to-one on Δ.

The proof that 0 is in the image of F is based on the same principle as the proof of uniqueness. We can extract information about the limiting behavior of F as R approaches $\partial\Delta$ by studying the total curvature of the subsurface S_{V^0}, where V^0 consists of the vertices v such that $R(v)$ is tending toward 0 . When a triangle of τ has two vertices in V^0 and the third not in V^0, then the sum of the two angles at vertices in V^0 tends toward π. When a triangle of τ has only one vertex in V^0, then the angle at that vertex tends toward the value $\pi - \Theta(e)$, where e is the opposite edge. Thus, the total curvature of ∂S_{V^0} tends toward the value $\sum_{e \in L(\tau_{V0})} (\pi - \Theta(e))$, where $L(\tau_{V^0})$ is the "link of τ_{V^0}." The Gauss-Bonnet formula gives

$$\mathrm{Lim} \sum_{v \in V^0} \kappa(v) = - \sum_{e \in L(\tau_{V0})} (\pi - \Theta(e)) + 2\pi\chi(S_{V^0}) < 0 \qquad 5.11.6$$

(Note that area $(S_{V^0}) \to 0$.) To see that the right-hand side is always negative, it suffices to consider the case that τ_{V^0} is connected. Unless τ_{V^0} has Euler characteristic one, both terms are non-positive, and the sum is negative. If $L(\tau_{V^0})$ has length five or more, then $\sum_{e \in L(\tau_{V0})} \pi - \Theta(E) > 2\pi$, so the sum is negative. The cases when $L(\tau_{V^0})$ has the length three or four are dealt with by conditions 5.11.1(c) and (d).

When V' is any proper subset of V^0 and $R \in \Delta$ is an arbitrary point, we also have an inequality

$$\sum_{v \in V'} \kappa_R(v) > - \sum_{e \in L(\tau_{V'})} (\pi - \Theta(e)) + 2\pi\chi(S_{V'}). \qquad 5.11.7$$

This may be deduced quickly by comparing the R metric with a metric R' in which $R'(V')$ is near 0 . In other words, the image $F(\Delta)$ is contained in the open polyhedron $P \subset Z$ defined by the inequalities 5.11.7. Since $F(\Delta)$ is an open set whose boundary is ∂P (by 5.11.6), $F(\Delta) = interior(P)$. Since $0 \in int(P)$, this completes the proof of ??, ??, and 5.11.2. 5.11.2

Remarks: This proof was based on a practical algorithm for actually constructing patterns of circles. The idea of the algorithm is to adjust, iteratively, the radii of the circles. A change of any single radius affects most strongly the curvature at that vertex, so this process converges reasonably well.

The patterns of circles on surfaces of constant curvature, with singularities at the centers of the circles, have a three-dimensional interpretation. Because of the inclusions $Isom(\mathbf{H}^2) \subset Isom(\mathbf{H}^3)$ and $Isom(\mathbf{E}^2) \subset Isom(\mathbf{H}^3)$, there is associated with such a surface S a hyperbolic three-manifold M_S, homeomorphic to $S \times \mathbf{R}$, with cone type singularities along (the singularities of S) $\times\mathbf{R}$. Each circle on S determines a totally geodesic submanifold (a "plane") in M_S, as in the construction of $P(F)$. These, together with the totally geodesic surface isotopic to S when S is hyperbolic, cut out a submanifold of M_S with finite volume. If the cone angles have the form $2\pi/n$, this is a hyperbolic orbifold.

/u/gt3m/3mbook/pictures/chap5/11/figure1.ps not found

figure1

Figure 5.54. 1. Notice that this is the same configuration we came across in the proof of the necessity of ?? and ??

/u/gt3m/3mbook/pictures/chap5/11/figure2.ps not found

figure2

Figure 5.55. 2. If $\sum_{i=1}^{3} \Theta(e_i) \geq \pi$, then three faces of the $P(F)$ intersect either in $\mathbf{H}^3$ or S^2_∞, thus as in section ?? this intersection point is a vertex of $P(F)$, so the three edges determine a triangular component of $S - G'$.

/u/gt3m/3mbook/pictures/chap5/11/figure3.ps not found

figure3

Figure 5.56. 2. As in section ??, the above are the only two possibilities.

/u/gt3m/3mbook/pictures/chap5/11/figure4.ps not found

figure4

Figure 5.57. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure5.ps not found

figure5

Figure 5.58. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure6.ps *not found*

figure6

Figure 5.59. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure7.ps not found

figure7

Figure 5.60. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure8.ps not found

figure8

Figure 5.61. 2

/u/gt3m/3mbook/pictures/chap5/11/figure9.ps not found

figure9

Figure 5.62. 2. C_3 meets $\overline{v_1 v_2} \Rightarrow C_1$ and C_2 don't meet

/u/gt3m/3mbook/pictures/chap5/11/figure10.ps not found

figure10

Figure 5.63. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure11.ps not found

figure11

Figure 5.64. 1.5

/u/gt3m/3mbook/pictures/chap5/11/figure12.ps not found

figure12

Figure 5.65. 2

Section "Proof of Andre'ev's theorem"
PROOF ANDREEV
% CIRCLE PAT
% ANDREEV1
% ANDREEV2
Section "DISK FAMILY"
% CIRCLE COND
% CIRCLE PAT
% DISK FAMILY
Section "SPEC CASE"
% DISK FAMILY
% SPEC CASE
% OTHER CASE
% PATTERNS OF CIRCLES

5.12. Proof of Andre'ev's theorem

In this section we will prove a result about convex polyhedra which extends theorem 5.11.2 to the case $\chi(S) > 0$, and which immediately implies Andre'ev's theorem (?? and ??).

Theore 5.12.1. *(Andre'ev). Let τ be a triangulation of S^2 with more than four triangles, let G' be its one-skeleton with edge set E' and let $\Theta : E' \to [0, \pi/2]$ be any function such that G' and Θ satisfy conditions 5.11.1. Then there is a family $\mathcal{D}$ of disks on S^2 (in the usual metric) giving rise to G' (up to isotopy) and Θ.*

The method of the preceding section for the proof of 5.11.2 does not directly work for 5.12.1. The proof breaks down, for instance, in the application of the Gauss-Bonnet formula to prove the uniqueness (5.11.12), which was tied up with existence. In fact, the map F is not injective and is not even surjective to a neighborhood of zero, when S is the two-sphere. It is not injective because moving a polyhedron isometrically in $\mathbf{H}^3$ changes the radii of the disks which describe its faces, but does not change the value of $F \equiv 0$. It is not surjective because otherwise one could obtain a Riemannian metric of constant curvature on S^2 except at one isolated cone-type singularity, for instance, an elliptic structure on a teardrop orbifold. This is impossible for any cone angle $\alpha \neq 2\pi$.

Nonetheless, there are some special cases of Andre'ev's theorem which can be derived from the preceding method after first converting patterns of circles on S^2 to patterns in $\mathbf{E}^2$ using stereographic projection.

Lemm 5.12.2. *Theorem 5.12.1 holds provided there is at least one triangle of τ whose three edges e_1, e_2 and e_3 satisfy $\Theta(e_1) + \Theta(e_2) + \Theta(e_3) \leq \pi$.*

Proof of 5.12.2: 5.12.2 (See also exercise 5.12.3) The hypothesis that the angle sum is $\leq \pi$ means that if there is a solution $\mathcal{D}$, there is some point p on S^2 not in the interior of any disk in $\mathcal{D}$. If the inequality is strict, p can be chosen not to be in any closed disk of $\mathcal{D}$. Stereographic projection from p transforms the problem to $\mathbf{E}^2$: we think of G' as a triangulation of a triangle in $\mathbf{E}^2$.

The method of 5.11 is modified as follows. First choose three radii for the "big" disks, whose centers are vertices of the assumed triangle. If $\sum_{i=1}^3 \Theta(e_i) < \pi$, we can take the three radii to be 1 . If $\sum_{i=1}^3 \Theta(e_i) = \pi$, the three circles on S^2 become straight lines in $\mathbf{E}^2$, and we give the three radii the limiting value ∞. These radii are to be held constant for the proof. The map F is defined from $\mathbf{R}_+^{V-V_0}$ to $\mathbf{R}^{V-V_0}$, where V_0 is the set consisting of the three outer vertices.

If the radii of V_0 are finite, we study F on all of $\mathbf{R}_+^{V-V_0}$ (normalization has been replaced by fixing these values). Notice that since each triangle is Euclidean $F(R)|_{V-V_0} \equiv 0$ implies that $F(R) \equiv 0$. We consider $\mathbf{R}_+^{V-V_0}$ to be the interior of a simplex with boundary faces consisting of the coordinate planes and one boundary face (with all its subfaces) "at infinity." The finite

% VARY ONE RADIUS
% VARY ONE RADIUS
% GBINEQ
Section "OTHER CASE"
% SPEC CASE
% CIRCLE PAT
% SPEC CASE
% DISK FAMILY

boundary behaviour was already taken care of by the previous cases. We may now obtain information about the limiting behavior of F as R approaches an infinite boundary face. Let τ_{V_∞} be the subcomplex spanned by all vertices V_∞ whose associated radii approach infinity. As before we thicken τ_{V_∞} to S_{V_∞} and approximate the geodesic curvature of ∂S_{V_∞}. We obtain

$$\operatorname{Lim} \sum_{v \in V_\infty} \kappa(v) = 2\pi\chi(S_{V_\infty}) + \sum_{e \in \partial\tau_{V_\infty}} (\pi - \Theta(e)) > 0$$

where the sum is over all the exposed edges of τ_{V_∞}. As before this leads to the inequalities

$$\sum_{v \in V_\infty} \kappa(v) \leq 2\pi\chi(S_{V_\infty}) + \sum_{e \in \partial\tau_{V_\infty}} (\pi - \Theta(e))$$

We again prove, using lemma 5.11.5, that F is injective. Thus, as before, we have proved that $0 \in im(F)$ since our inequalities tell us the boundary behavior.

If the radii are infinite, then when we multiply $R \in \mathbf{R}^{V-V_0}$ by a constant, F does not change. We can normalize so $\sum_{v \in V-V_0} R(v) = 1$. We can think of this as a limiting case, as above, so we see that $\int_{\partial S} \kappa_g = 2\pi$ where S is a thickening of the subcomplex spanned by all the finite vertices. Thus $\sum_{v \in V-V_0} F(R)(v) = 0$. Lemma 5.11.5 carries over even when $R\,|_{V_0}$ is infinite, and the uniqueness proof carries over. (Note that since $R\,|_{V_0}$ is constant, there are no extraneous boundary effects. The angles and angle sums which "increase" and "decrease" may do so only in the weak sense so that they may remain constant, if some of the vertices involved have infinite radii.) Existence follows from the univalence of F, together with boundary behavior given by 5.11.7. Note that the solution develops appropriately into $\mathbf{E}^2$, even though we have not controlled the lengths of arcs of the big disks. **5.12.2**

Exercis 5.12.3. Derive Lemma 5.12.2 in the case $\Theta(e_1) + \Theta(e_2) + \Theta(e_3) < \pi$ as a formal consequence of theorem 5.11.2.

(Hint: Any convex acute-angled polyhedron which has three faces meeting at angles α, β and γ with $\alpha + \beta + \gamma < \pi$ can be decomposed into two pieces along a common orthogonal plane (to each of the three faces). Glue the problem of lemma 5.12.2 to an appropriately chosen genus-two problem.)

Proof of 5.12.3: 5.12.1 For the general case, we will follow Andre'ev's method, but make use of the special case to bypass some steps.

Dual to the triangulation τ of S^2 is a polyhedron P. The edge set $\mathcal{E}'$ of τ we identify with the set $\mathcal{E}$ of edges of P. The plan is to study the space of hyperbolic polyhedra of the combinatorial type of P. Since τ is a triangulation, only three faces meet at each vertex of P. In case P has symmetries, we understand a combinatorial equivalence to P to be fixed. Let F, E and V denote the numbers of faces, edges and vertices of P. The condition that a set of half-spaces determine a polyhedron combinatorially equivalent to P is an open condition. A polyhedron combinatorially equivalent to P is determined

% CIRCLE COND
% SPEC CASE
% DISK FAMILY
Section "PLUSMIN SEP"

as an intersection of F half-spaces, and the set of half-spaces is the three-manifold $S^2 \times (0, \pi)$. (A half-space is specified by the center and radius of its disk on S^2_∞.) Therefore, the set of polyhedra combinatorially equivalent to P is a manifold of dimension $3F$. The group of isometries of $\mathbf{H}^3$ is six-dimensional, and it acts freely on the space of polyhedra combinatorially equivalent to P, so the space of congruence classes of polyhedra combinatorially equivalent to P is a manifold of dimension $3F-6$. Let $C(P)$ denote the open subset consisting of polyhedra with angles $< \pi/2$, and $\bar{C}(P)$ the polyhedra with angles $\leq \pi/2$.

Lemm 5.12.4. *If there is a $\Theta : \mathcal{E} \to [0, \pi/2]$ satisfying conditions 5.11.1 such that the three edges e_1, e_2, and e_3 of every triangle of τ satisfy $\Theta(e_1) + \Theta(e_2) + \Theta(e_3) > \pi$, then $C(P)$ is not empty.*

Proof of 5.12.4: Let $K > \pi$ be the minimum value of the Θ-sum for the three edges of any triangle of τ. Then $(\pi/K) \cdot \Theta$ satisfies the hypotheses of lemma 5.12.2, so there is a non-compact polyhedron of finite volume combinatorially equivalent to P, with every angle strictly less than $\pi/2$. Push each face plane slightly inward, to get a compact acute-angled polyhedron. $\boxed{5.12.4}$

Continuation of proof of 5.12.1: We will now study the map $A : \bar{C}(P) \to [0, \pi/2]^E$ which describes the dihedral angle at the edges of a polyhedron. Since the Euler characteristic of S^2 is 2 , we have

$$V - E + F = 2$$

$$2E = 3V$$

so

$$E = 3F - 6.$$

Thus, the dimension of the domain of A equals the dimension of the range of A.

Next we will prove that A is one-to-one. This, combined with an analysis of the "boundary behavior" of A, will tell us the range of A.

Lemm 5.12.5. *Let B be a polygon in $\mathbf{H}^2$ with all angles $\leq \pi/2$, and let C be another polygon with the same angles. Label each side of B with a $+, 0$ or $-$ according to whether the corresponding side of C has a larger, equal, or smaller length. If C is not congruent to B, then there is at least one pair of $+$ sides of B which separate a pair of $-$ sides.*

Proof of 5.12.5: Otherwise, there is some common perpendicular m to some pair of sides of B such that all sides of B to the right of m are labeled 0 or $+$, and all sides to the left are labeled 0 or $-$. Consider the length of the altitude of C corresponding to m. Looking at the right half of B with the right half of C, we see that this altitude of C lengthens, but looking at the left halves we see that it shortens, a contradiction. $\boxed{5.12.5}$

This lemma combines nicely with a well-known lemma of Cauchy, which was a key step to the proof of rigidity of convex polyhedra in $\mathbf{E}^3$:

Section "CAUCHY"
% COMB TYPE
Section "A INJECT"
% PLUSMIN SEP
% CAUCHY

Lemm 5.12.6. *(Cauchy). Let P be a polyhedron, meaning a cell-division of S^2 such that two cells touch either along at most one one-cell or along at most one zero-cell. Suppose the edges of P are labeled with $+$, 0 and $-$ in such a way that for each face of P, either all edges are labeled 0, or there is at least one pair of $+$'s which separate at least one pair of $-$'s. Then the labeling is identically 0.*

Proof of 5.12.6: Let V be the union of closed two-cells whose edges are not labeled identically 0. Let V_0 be any component of V. Define a line field with singularities on V_0 which is tangent to any edge labeled 0 or $+$, transverse to any edge labeled $-$, and has at most one singularity which then has negative index in the interior of any face of V_0, as in figure 5.66.

Now collapse each component of ∂V_0 to a point, to obtain a singular line field on S^2. The index at any vertex is negative or 0 , so the total index is negative, which contradicts the fact that the total index must be $+2 = \chi(S^2)$. (Recall that the Poincaré-Hopf index theorem says that the index of any line-field is the same as the Euler characteristic of the space.) 5.12.6

COMB TYPE

COMB TYPE

Figure 5.66. COMB TYPE

Corollar 5.12.7. *The map $A : \bar{C}(P) \to [0, \pi/2]^E$ is injective.*

Proof of 5.12.7: Observe that the corner angles of faces of P are determined by the dihedral angles of P, and they are all $\leq \pi/2$. Apply 5.12.5 and 5.12.6. 5.12.7

% DISK FAMILY
% A INJECT
% CIRCLE COND
% CIRCLE COND

Continuation of proof of 5.12.1: By invariance of domain (or by an elementary application of the inverse function theorem using a first-order version of 5.12.7) the image of A restricted to $C(P)$ is an open set, $U \subset [0,\pi/2)^E$, contained in the open subset V, which is defined by conditions 5.11.1 together with the open condition that for the three edges e_1, e_2 and e_3 of any triangle of τ, $\Theta(e_1) + \Theta(e_2) + \Theta(e_3) > \pi$. To prove that $U = V$, it suffices to prove that $\partial U \subseteq \partial V$. This is sufficient, since V is connected and we already know that $U \subseteq V$. Let $\Theta_b \in \partial U \subseteq \bar{V}$ be any limiting assignment of angles, and let $\{P_i\}$ be a sequence of hyperbolic polyhedra $\in C(P)$ such that the dihedral angles $A(P_i)$ converge to Θ_b. Consider any face F of $\{P_i\}$. If its area tends toward 0, then the sum of its angles tends toward the Euclidean value. But the dihedral angle of the edge which meets F at any corner is not less than the corner angle of F, so we get a circuit of faces circumnavigating F so that Θ_b assigns values which sum to $(f-2)\pi$ (if F has f sides). Clearly, $f \leq 4$, since each angle is no greater than $\pi/2$. Since we assumed that P is not a tetrahedron, this circuit does not circumnavigate a vertex, so Θ_b fails 5.11.1, and must be outside V, hence in ∂V.

If no face has area tending toward 0, and if all edges of $\{P_i\}$ remain bounded in length, then $\{P_i\}$ has a subsequence converging up to congruence. (If any edge tends to 0 in length, it is easy to see that the corner angles of its two incident faces must tend to $\pi/2$, and adjacent edges tend to ∞ in length.) Since Θ_b is assumed to be on ∂U, we conclude that at least one edge length does not remain bounded. Pass to a subsequence where the length of some edge e goes to ∞.

The area of each face is bounded. Therefore in the faces of P_i containing e, e cannot have a standard tubular neighborhood of width ε; in fact, e must nearly parallel other edges (within ε) for all but a bounded portion of its length, so we conclude that the lengths of more edges of $\{P_i\}$ tend to ∞. How long can this continue? Since each edge has an angle of at most $\pi/2$ and the angle sum for such a circuit converges to the Euclidean value (as one sees by considering the intersection with an approximately perpendicular plane to the edges in question) the chain of faces with long, nearly parallel, faces can only have length three or four.

This shows that $\Theta_b \in \partial V$. (Note that if the circuit goes around an edge, it forces the two ends of the edge to have angles $\to (0,\pi/2,\pi/2)$, so Θ_b fails the condition on two triangles of τ. If it goes around a vertex, it similarly fails.) 5.12.3

Exercis 5.12.8. Let P be a cube truncated at four alternate corners. Show that $C(P)$ is empty.

Exercis 5.12.9. * a) Consider the operation of taking two polyhedra which have three edges at each vertex, and performing a kind of connected sum by deleting a neighborhood of a vertex of each, and matching the edges which were incident to one deleted vertex with the edges which were incident to the other, yielding a new polyhedron of the same type. Show that any polyhedron with three edges incident

% CIRCLE COND
% PLUSMIN SEP
Section "CAUCHY-LIKE"
% CAUCHYLIKE
% CAUCHY

to each vertex is obtained in a unique way by combining polyhedra which cannot be decomposed further. ("Unique" must be appropriately interpreted, of course.)

b) Let P be a polyhedron with three edges at each vertex, and let $P_1, \ldots, P_k$ be its indecomposable pieces. Label each vertex of $P_1, \ldots, P_k$ with -1 if it is used for a connected sum operation, and label it $+1$ otherwise. Show that $C(P)$ is empty iff for at least one P_i, the vertices are labeled alternately, that is each edge has a $+1$ vertex and a -1 vertex.

(Hint: If there is a solution for Θ, there is a solution near $\Theta \equiv \pi/3$ which satisfies the closed inequalities. Solutions for Θ near $\pi/3$ form an open convex cone. Such a cone is empty iff the spanning vectors for the dual cone (see 5.11.1) contain 0 in their convex hull. Consider what this means on P in relation to its decomposition; use induction on the tree-like pattern in which the P_i are assembled.)

Exercis 5.12.10. Show that lemma 5.12.5 is sharp, in the sense that for any hyperbolic polygon B with angles $\leq \pi/2$, and any labeling of the sides of B with +'s, 0's, and −'s such that at least one pair of +'s separates a pair of −'s, there is a polygon C with the same angles as B and with its + sides larger, its 0 sides the same length, and its − sides smaller.

Exercis 5.12.11. Show that if B and C are two convex polygons on S^2 with corresponding sides equal in length, and if the vertices of B are labeled with +'s, 0's and −'s according to whether the angle at the corresponding vertex of C is larger, the same, or smaller than the angle of B, then at least one pair of +'s separates at least one pair of −'s.

Exercis 5.12.12. Prove Cauchy's rigidity theorem for convex polyhedra in $\mathbf{E}^3$: if P and Q are two convex polyhedra in $\mathbf{E}^3$, and $f : P \to Q$ is a piecewise linear isometry between them, then f extends to an isometry of $\mathbf{E}^3$.

(Hint: first triangulate P so that f is linear on each triangle. Apply 5.12.11 and 5.12.6.)

::action
::orbits
::quotient
::ambient space
::ambient homeomorphism

†differentiable
::covering map
::covering space
::covering
::covering transformation
::covering group
universal cover
::diffeomorphism
::diffeomorphic
::differentiable of class C^r
::C^r-map
::smooth
::of class C^∞
::of class C^r

Glossary

action [Ber87, vol. I, p. 5]. If G is a group and X is a topological space, an *action* of G on X (or G-action) is a homomorphism from G into the group of homeomorphisms of X; we denote the image of $x \in X$ under the homeomorphism associated to $g \in G$ by $g(x)$. A G-action defines an equivalence relation on X, whereby $x, y \in X$ are equivalent if and only if $y = g(x)$ for some $g \in G$. The equivalence classes of this equivalence are the *orbits* of G. The space of orbits, with the quotient topology, is the *quotient* of X by the G-action.

ambient. When we talk about two spaces (or manifolds, etc.), one contained in the other, the containing space is sometimes called the *ambient space*. A homeomorphism (or diffeomorphism, etc.), of the containing space is then called an *ambient homeomorphism.*

barycentric subdivision. The barycentric subdivision of a triangle is obtained by joining the triangle's barycenter—the average of its three vertices—to the vertices and to the midpoints of the sides. The subdivision of a triangulation is what you get by taking the barycentric subdivision of each triangle in the triangulation.

C^r, C^∞. See †differentiable.

cone on a torus.

covering, covering map, covering space, covering transformation, covering group [Mun75, p. 331–341, 398]. A continuous map $p : \tilde{X} \to X$ between path-connected topological spaces is a *covering map* if every point of X has a neighborhood V such that every connected component of $p^{-1}(V)$ is mapped homeomorphically onto V by p. In this case $\tilde{X}$ is called a *covering space* of X, and $(\tilde{X}, p, X)$ is called a *covering.* A map $\phi : \tilde{X} \to \tilde{X}$ such that $p \circ \phi = p$ is a *covering transformation* for this covering; covering transformations form a group, called the *covering group.* If $\tilde{X}$ is simply connected, it is called the universal cover of X—"the" because it is unique up to homeomorphism.

diffeomorphism, diffeomorphic. If f is bijective and both f and f^{-1} are differentiable, f is called a *diffeomorphism*, and X and Y are said to be *diffeomorphic.*

differentiable map [Hir76, p. 9, 15]. A map $f : U \to \mathbf{R}^m$, where U is open in $\mathbf{R}^m$, is *differentiable of class C^r* (or a *C^r-map*) if f has continuous partial derivatives of order up to r. It is *smooth*, or *of class C^∞*, if it has continuous partial derivatives of all orders. A map $f : X \to \mathbf{R}^m$, where $X \subset \mathbf{R}^n$ is arbitrary, is of class C^r if it can be extended to a C^r map on a neighborhood of X.

A map $f : X \to Y$ between differentiable manifolds or manifolds-with-boundary is *of class C^r* if, for every point of $x \in X$, the expression of f in local coordinates in a neighborhood of x is of class C^r.

::dihedral angle
†immersion
::Euclidean n-space
$\mathbf{E}^n$
::plane
::Euclidean metric
::base point
::fundamental group
::simply connected
::geodesic
projective geometry
Euclidean geometry
hyperbolic geometry
::homothety
::homotopic
::homotopy
::homotoped
::isotopy
::isotopic
::immersion
::embedding
::embedded

differentiable manifold.

dihedral angle. The *dihedral angle* between two intersecting planes in space is the angle between the lines determined on the two planes by a third plane orthogonal to both.

embed, embedding. See †immersion.

Euclidean [Ber87, p. 153, 202]. We get *Euclidean n-space* $\mathbf{E}^n$ (*plane* if $n = 2$) by giving $\mathbf{R}^n$ the *Euclidean metric* $d(x, y) = \left(\sum_i x_i y_i\right)^{1/2}$, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are points in $\mathbf{R}^n$. For $n = 3$ this is the space of our everyday experience.

Euclid's parallel axiom. "Given a line and a point outside the line, there is exactly one parallel to the line through the point." This axiom, which holds in a Euclidean space, has numerous equivalent formulations, such as "The sum of the angles of a triangle equals 180°."

frame, frame bundle.

fundamental group [Mun75, p. 326]. If X is a connected topological space, the set of homotopy classes of loops beginning and ending at a fixed point of X (the *base point*) is a group under concatenation of paths. It is called the *fundamental group* of X. If the fundamental group of X is trivial, we say that X is *simply connected.*

geodesic [dC76, p. ??]. Geodesics are curves that are as straight as possible. More precisely, given a Riemannian manifold X and an interval $I \subset \mathbf{R}$, a curve $\gamma : I \to X$ is a *geodesic* if, for $x, y \in I$ close enough, the shortest path joining $\gamma(x)$ and $\gamma(y)$ in X coincides with $\gamma([x, y])$. Normally we also require that γ be parametrized at constant speed (which can be zero). Often the image $\gamma(I)$, too, is called a geodesic.

geometry. Sometimes we use "geometry" interchangeably with "metric." Other times, "geometry" refers to the set of properties of a space that depend on the metric. Other times yet, it refers to the properties of a space that are invariant under a group of transformations of the space, which depends on the context. Thus projective geometry, Euclidean geometry, hyperbolic geometry.

homothety [Ber87, vol. I, p. 39]. A *homothety* centered at a point $p \in \mathbf{E}^n$ is the map that fixes p and "blows up" or "shrinks down" all of space around p. More precisely, it takes q to $k(q - p) + p$, for $k \in \mathbf{R} \setminus \{0\}$ fixed.

homotopy, homotopic, homotope [Mun75, p. 318–319]. Two continuous maps $f, g : X \to Y$ between topological spaces are *homotopic* if they can be continuously deformed into one another, that is, if there is a continuous map $F : X \times [0, 1] \to Y$ such that $F|_{X\times\{0\}} = f$ and $F|_{X\times\{1\}} = g$. Such a map is called a *homotopy* between f and g, and we say that f can be *homotoped* into g through F. If every stage of the homotopy, that is, every map $F|_{X\times\{t\}}$ for $t \in [0, 1]$, is a homeomorphism onto its image, we say that F is an *isotopy* and that f and g are *isotopic.*

immersion [Hir76, p.21]. A differentiable map $f : M \to N$ between differentiable manifolds is an *immersion* if the derivative df has maximal rank at every point $p \in M$. If, in addition, f is a homeomorphism onto its image, it is called an *embedding*, and we say that M is *embedded* in N. In particular, an embedding is a one-to-one immersion.

induced orientation on the boundary.

intermediate value theorem. A continuous function $f : [a, b] \to \mathbf{R}$ that is positive at a and negative at b must be zero somewhere in (a, b). A trivial consequence of the connectedness of an interval.

†homotopy
$O(g)$
::grows polynomially
†action
::order
::orientation
::orientation-preserving
::orientation-reversing
::orientable
::oriented
::orientation-preserving
::orientation-reversing
†action
::quotient topology
::Riemannian metric
::Riemannian manifold
::similarity
†fundamental group
†differentiable

invariance of domain, theorem on the [Bro12a,]. If a subset $A \subset \mathbf{R}^n$ is homeomorphic to an open subset of $\mathbf{R}^n$, it is itself open. From this it easily follows that an m-dimensional manifold cannot have a subset homeomorphic to $\mathbf{R}^n$, for $n > m$.

isotopy, isotopic, isotope. See †homotopy.

$O(\cdot)$. If f and g are real-valued functions of a real or integer variable x, we say that f is $O(g)$ if f is bounded by a multiple of g for x large enough. For instance, f *grows polynomially* if f is $O(x^n)$ for some n, that is, if f is bounded by a polynomial of order n.

orbit. See †action.

order. The *order* of an element g of a group is the smallest positive integer n (if one exists) such that g^n is the identity.

orientation, oriented, orientable, orientation-preserving, orientation-reversing [Mil65, p. 27]. An *orientation* for a finite-dimensional vector space is an equivalence class of (ordered) bases that can be taken to one another by linear transformations with positive determinant. A linear transformation between oriented vector spaces is *orientation-preserving* or *orientation-reversing* depending on whether its determinant is positive or negative.

A manifold M is *orientable* if the tangent spaces to M at all points can be oriented consistently. This means that M can be covered by coordinate patches such that the derivative map of any coordinate map at any $x \in M$ is an orientation-preserving map between T_xM and $\mathbf{R}^n$ with its standard orientation. We say that M is *oriented* if such a choice of orientations has been made. A local diffeomorphism between oriented manifolds is *orientation-preserving* or *orientation-reversing* according to what its derivative is at each point.

orthogonal trajectories.

polygonal region. A closed set on the plane bounded by a polygon.

quotient by a group action. See †action.

quotient topology [Mun75, p. 134–136]. If X is a topological space and $\sim$ is an equivalence relation on X, we define the *quotient topology* on the set $X/\sim$ of equivalence classes by decreeing that a set of classes is open if and only if the union of their inverse images in X is open.

Riemannian manifold, Riemannian metric [BG88, p. 126]. A *Riemannian metric* on a differentiable manifold X is a rule that gives for each point $p \in X$ an inner product on the tangent space T_pX, in such a way that the inner product varies smoothly with p. Giving an inner product also gives a notion of length for tangent vectors, and consequently for differentiable paths. A *Riemannian manifold* is a manifold with a Riemannian metric; such a manifold is a metric space in a natural way, the distance between two points being the infimum of the lengths of paths joining the two points.

semigroup.

similarity [Ber87, vol. I, p. 183]. A *similarity* of Euclidean space is a transformation that multiplies all distances by the same factor k.

simply connected. See †fundamental group.

smooth. See †differentiable.

star. star of a vertex, star of a simplex in a simplicial complex

symmetry.

::tiling
::tiles
::tiling
†covering

tiling [Ber87, vol. I, p. 11–31]. Roughly speaking, a *tiling* of a space X is a way to fill up X with copies of one or more standard *tiles*, or shapes. There are various ways to restrict and formalize this definition to make it more manageable. We can stick to the following definition: A *tiling* of X consists of a connected compact subset $P \subset X$ and a group G of isometries of X such that the interior of P is non-empty, the union of images of P under G is X, and two images of P coincide if their interiors intersect.

universal cover. See †covering.

Bibliography

[AB27] John W. Alexander and G.B. Briggs. On types of knotted curves. *Annals of Mathematics*, 28:562–586, 1927.

[Ale20] John W. Alexander. Note on Riemann spaces. *Bulletin of the American Mathematical Society*, 26:370–372, 1920.

[Bae28] Reinhold Baer. Isotopien von kurven auf orientierbaren, geschlossenen flächen. *Journal für die Reine und Angewandte Mathematik*, 159:101–116, 1928.

[Bar94] W. Barlow. Über die geometrischen eigenschaften homogener starrer structuren und ihrer structuren und ihre anwendung auf krystalle. *Zeitschrift fur Krystallographie und Mineralogie*, 23:1–63, 1894.

[Ber87] Marcel Berger. *Geometry (2 vols.)*. Springer, Heidelberg, 1987.

[BG88] Marcel Berger and Bernard Gostiaux. *Differential Geometry: Manifolds, Curves and Surfaces.* Springer, New York, 1988.

[Bia92] Luigi Bianchi. Sui gruppi di sostituzioni lineari con coefficienti appartenenti a corpi quadratici immaginari. *Mathematische Annalen*, 40:332–412, 1892.

[Bie11] Bieberbach. Über die bewegungsgruppen der euklidischen räume I. *Mathematische Annalen*, 70:297–336, 1911.

[Bie12] Bieberbach. Über die bewegungsgruppen der euklidischen räume II. *Mathematische Annalen*, 72:400–412, 1912.

[Bin59] R.H. Bing. An alternative proof that 3-manifolds can be triangulated. *Annals of Mathematics*, 69:37–65, 1959.

[Bra49] Auguste Bravais. Memoire sur les polyedres de forme symetrique. *Journal de Mathematiques Pures et Appliquées*, 14:141–180, 1849.

[Bro12a] Luitzen Egbertus Jan Brouwer. Beweis der invarianz des n-dimensionalen gebiets. *Mathematische Annalen*, 71:305–313, 1912.

[Bro12b] Luitzen Egbertus Jan Brouwer. Beweiss des ebenen translationsatzes. *Mathematische Annalen*, 72:37–54, 1912.

[Bro60] E. J. Brody. The topological classification of lens spaces. *Annals of Mathematics*, 71:163–184, 1960.

[BS85] Francis Bonahon and L. Siebenmann. The classification of seifert-fibered 3-orbifolds. In Roger Fenn, editor, *Low-Dimensional Toplogy*, pages 19–85, Cambridge University Press, Cambridge, 1985.

[Bus55] Herbert Busemann. *The Geometry of Geodesics.* Academic Press, New York, 1955.

[BZ85] Gerhard Burde and Heiner Zieschang. *Knots.* W. DeGruyter, Berlin-New York, 1985.

[Cay78] Arthur Cayley. On the theory of groups. *American Journal of Mathematics*, 1:174–176, 1878.

[Cha86] Leonard Charlap. *Bieberbach groups and flat manifolds.* Springer-Verlag, New York, 1986.

[Con70] John Conway. An enumeration of knots and links, and some of their algebraic properties. In John Leech, editor, *Computational Problems in Abstract Algebra*, pages 329–358, Pergammon Press, New York, 1970.

[Cox34] H. S. M. Coxeter. Discrete groups generated by reflections. *Annals of Mathematics*, 35:588–621, 1934.

[CW50] Harold S. M. Coxeter and ?? Whitrow. World structure and non-euclidean honeycombs. *Proceedings of the Royal Society of London Series A*, 201:417–437, 1950.

[dC76] Manfredo P. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, Englewood Cliffs, NJ, 1976.

[Deh38] Max Dehn. Die gruppe der abbildungsklassen. *Acta Mathematica*, 69:135–206, 1938.

[Dun] William Dunbar. Fibered orbifolds and crystallographic groups. Ph. D. thesis, Princeton University, 1981.

[Eps66] David Epstein. Curves on 2-manifolds and isotopies. *Acta Mathematica*, 115:83–107, 1966.

[FR79] Roger Fenn and Colin Rourke. On Kirby's calculus of links. *Topology*, 18:1–15, 1979.

[Hae58] André Haefliger. Structures feuilletées et cohomologie 'a valeur dans un faisceau de grupoïdes. *Commentarii Mathematici Helvetici*, 32:248–329, 1958.

[Hem76] J. Hempel. *3-manifolds.* Princeton University Press, Princeton, N.J., 1976.

[Hes30] J.F.C. Hessel. Krystallometrie oder krystallometrie und krystallographie. In *Gehler's Phys. Wörterbuch, Band 5, Abtheilung 2*, Schwickert, Liepzig, 1830.

[Hic65] Noel Hicks. *Notes on Differential Geometry.* Van Nostrand, Princeton, NJ, 1965.

[Hil01] David Hilbert. über Flächen von constanter gausscher Krümmung. *Transactions of the American Mathematical Society*, 87–99, 1901.

[Hil76] Hugh M. Hilden. Three-fold branched coverings of S^3. *American Journal of Mathematics*, 98:989–997, 1976.

[Hir76] Morris Hirsch. *Differential Topology.* Springer, New York, 1976.

[HLM85] Hugh M. Hilden, Maria T. Lozano, and José M. Montestinos. On knots that are universal. *Topology*, 24:499–504, 1985.

[Hop26] Heinz Hopf. Zum Clifford-Kleinschen raumproblem. *Mathematische Annalen*, 95:313–339, 1926.

[Hum82] Stephen Humphries. Generators for the mapping class group. In Roger Fenn, editor, *Topology of Low-Dimensional Manifolds*, pages 44–47, Springer Verlag, New York, 1982.

[HW35] W. Hantschze and H. Wendt. Dreidimensionale euklidische raumforme. *Mathematische Annalen*, 110:593–611, 1935.

[Kir78] Robion Kirby. A calculus of framed links. *Inventiones Mathematicae*, 35–56, 1978.

[Kui55] Nicolaas Kuiper. On c^1-isometric embeddings ii. *Nederl. Akad. Wetensch. Proc. Ser. A*, 683–689, 1955.

[Lan50] F. Lanner. On complexes with transitive groups of automorphisms. *Math. Univ. Lund.*, 11:1–71, 1950.

[Lic62] W. B. R. Lickorish. A representation of orientable combinatorial 3-manifolds. *Annals of Mathematics*, 76:531–538, 1962.

[Lic65] W. B. R. Lickorish. A foliation for 3-manifolds. *Annals of Mathematics*, 82:414–420, 1965.

[Mag74] Wilhelm Magnus. *Non-Euclidean Tesselations and Their Groups.* Academic Press, New York, 1974.

[Mas71] B. Maskit. On Poincaré's theorem for fundamental polygons. *Advances in Mathematics*, 7:219–230, 1971.

[Mil56] John Milnor. On manifolds homeomorphic to the 7-sphere. *Annals of Mathematics*, 64:399–405, 1956.

[Mil58] John Milnor. On the existence of a connection with curvature zero. *Commentarii Mathematici Helvetici*, 32:215–223, 1958.

[Mil65] John Milnor. *Topology from a Differentiable Viewpoint.* University of Virginia, Charlottesville, 1965.

[MKS66] Wilhelm Magnus, Abraham Karras, and Donald Solitar. *Combinatorial Group Theory: presentations of groups in terms of generators and relations.* Interscience, New York, 1966.

[Moi77] Edwin E. Moise. *Geometric Topology in Dimensions 2 and 3.* Springer Verlag, New York, 1977.

[Mon76] José M. Montesinos. Three-manifolds as 3-fold branched coverings of S^3. *Quarterly Journal of Mathematics Oxford*, 27:85–94, 1976.

[Mun60] James Munkres. Obstructions to the smoothing of piecewise differentiable homeomorphisms. *Annals of Mathematics*, 72:521–554, 1960.

[Mun75] James Munkres. *Topology: A First Course.* Prentice-Hall, Englewood Cliffs, NJ, 1975.

[MZ55] Deane Montgomery and Leo Zippin. *Topological Transformation Groups.* Interscience, New York, 1955.

[New31] M.H.A. Newman. A theorem on periodic transformations of space. *Quarterly Journal of Mathematics Oxford*, 2:1–8, 1931.

[Nov65] S.P. Novikov. Topology of foliations. *Transactions of the Moscow Mathematical Society*, 14:268–305, 1965.

[Now34] Werner Nowacki. Die euklidischen, dreidimensionalen, geschlossenen und offenen raumformen. *Commentarii Mathematici Helvetici*, 7:81–93, 1934.

[ONe66] Barrett O'Neill. *Elementary Differential Geometry.* Academic Press, New York, 1966.

[Poi82] Henri Poincaré. Théorie des groupes fuchsienes. *Acta Mathematica*, 1:1–62, 1882.

[Poi83] Henri Poincaré. Mémoire sur les groupes kleinéens. *Acta Mathematica*, 3:49–92, 1883.

[Poi95] Henri Poincaré. Analysis situs. *J. Ecole Polytechnique*, 1:1–121, 1895.

[Ree52] Georges Reeb. *Sur certains proprietes topologiques des varietes feuilletes.* Hermann, Paris, 1952.

[Ril82] Robert Riley. Seven excellent knots. In R. Brown and T. L. Thickstun, editors, *Low Dimensional Topology*, pages 81–151, Cambridge University Press, Cambridge, 1982.

[Rol76] Dale Rolfsen. *Knots and Links.* Publish or Perish, Wilmington, DE, 1976.

[Rol84] Dale Rolfsen. Rational surgery calculus: extension of Kirby's theorem. *Pacific Journal of Mathematics*, 110:377–386, 1984.

[Sat56] Ichiro Satake. On a generalization of the notion of a manifold. *Proceedings of the National Academy of Sciences U.S.A.*, 42:359–363, 1956.

[Sat57] Ichiro Satake. The gauss-bonnet theorem for v-manifolds. *Journal of the Mathematical Society of Japan*, 9:464–492, 1957.

[Sch91] A. Schoenfliess. *Krystallsysteme und Krystallstruktur.* Teubner Verlag, Liepzig, 1891.

[Sco83] Peter Scott. The geometries of 3-manifolds. *Bulletin of the London Mathematical Society*, 15:401–487, 1983.

[Sma59] Stephen Smale. Diffeomorphism of the 2-sphere. *Proceedings of the American Mathematical Society*, 10:621–626, 1959.

[Soh79] L. Sohncke. *Entwickelung einer Theorie der Krystallstruktur.* Teubner Verlag, Liepzig, 1879.

[Spa66] Edwin H. Spanier. *Algebraic Topology.* McGraw-Hill, New York, 1966.

[ST30] Herbert Seifert and William Threlfall. Topologische untersuchung der discontinuitätsbereiche des dreidimensionalen sphärischen raumes. *Mathematische Annalen*, 104:1–70, 1930.

[ST32] Herbert Seifert and William Threlfall. Topologische untersuchung der discontinuitätsbereiche des dreidimensionalen sphärischen raumes (schluss). *Mathematische Annalen*, 107:543–586, 1932.

[Tak85] Moto-o Takahashi. On the concrete construction of hyperbolic structures of 3-manifolds. *Tsukuba Journal of Mathematics*, 9:41–83, 1985.

[Thu] William P. Thurston. Universal links. preprint.

[Thu76] William P. Thurston. Existence of codimension-one foliations. *Annals of Mathematics*, 104:249–268, 1976.

[vF91] Evgraph Stepanowitsch von Fedorow. Die symmetrie der figuren der regelmäßiger systeme. *Verhandlungen der russisch-kaislerlichen mineralogischen Geschellschaft zu St. Petersburg*, 28:1–146, 1891.

[Vin85] E. B. Vinberg. Hyperbolic reflection groups. *Russian Math. Surveys*, 40:31–75, 1985.

[Wal60] Andrew H. Wallace. Modifications and cobounding manifolds. *Canadian Journal of Mathematics*, 12:503–528, 1960.

[Whi36] Hassler Whitney. Differentiable manifolds. *Annals of Mathematics*, 37:645–680, 1936.

[Wol67] Joseph A. Wolf. *Spaces of Constant Curvature.* McGraw-Hill, New York, 1967.

[Woo69] John W. Wood. Foliations on 3-manifolds. *Annals of Mathematics*, 89:336–358, 1969.

[Zas48] H. Zassenhaus. Uber einen algorithmus, zur bestimmung der räumgruppen. *Commentarii Mathematici Helvetici*, 21:117–141, 1948.