William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TeX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

Introduction

These notes (through p. 9.80) are based on my course at Princeton in 1978–79. Large portions were written by Bill Floyd and Steve Kerckhoff. Chapter 7, by John Milnor, is based on a lecture he gave in my course; the ghostwriter was Steve Kerckhoff. The notes are projected to continue at least through the next academic year. The intent is to describe the very strong connection between geometry and low-dimensional topology in a way which will be useful and accessible (with some effort) to graduate students and mathematicians working in related fields, particularly 3-manifolds and Kleinian groups.

Much of the material or technique is new, and more of it was new to me. As a consequence, I did not always know where I was going, and the discussion often tends to wanter. The countryside is scenic, however, and it is fun to tramp around if you keep your eyes alert and don't get lost. The tendency to meander rather than to follow the quickest linear route is especially pronounced in chapters 8 and 9, where I only gradually saw the usefulness of "train tracks" and the value of mapping out some global information about the structure of the set of simple geodesic on surfaces.

I would be grateful to hear any suggestions or corrections from readers, since changes are fairly easy to make at this stage. In particular, bibliographical information is missing in many places, and I would like to solicit references (perhaps in the form of preprints) and historical information.

Contents

CHAPTER 1

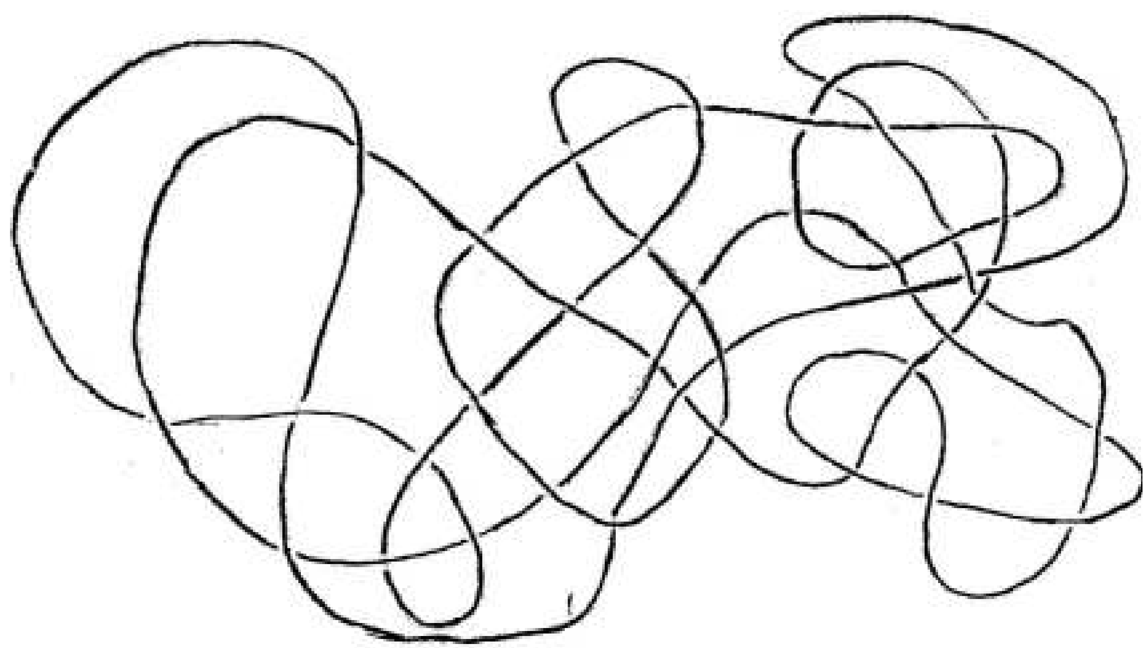
Geometry and three-manifolds

The theme I intend to develop is that topology and geometry, in dimensions up through 3, are very intricately related. Because of this relation, many questions which seem utterly hopeless from a purely topological point of view can be fruitfully studied. It is not totally unreasonable to hope that eventually all three-manifolds will be understood in a systematic way. In any case, the theory of geometry in three-manifolds promises to be very rich, bringing together many threads. 1.1

Before discussing geometry, I will indicate some topological constructions yielding diverse three-manifolds, which appear to be very tangled.

0. Start with the three sphere S^3, which may be easily visualized as $\mathbb{R}^3$, together with one point at infinity.

1. Any knot (closed simple curve) or link (union of disjoint closed simple curves) may be removed. These examples can be made compact by removing the interior of a tubular neighborhood of the knot or link.



1.2

The complement of a knot can be very enigmatic, if you try to think about it from an intrinsic point of view. Papakyriakopoulos proved that a knot complement has fundamental group $\mathbb{Z}$ if and only if the knot is trivial. This may seem intuitively clear, but justification for this intuition is difficult. It is not known whether knots with homeomorphic complements are the same.

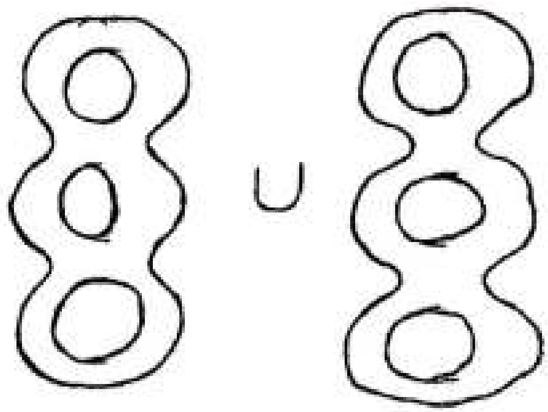
2. Cut out a tubular neighborhood of a knot or link, and glue it back in by a different identification. This is called *Dehn surgery*. There are many ways to do this, because the torus has many diffeomorphisms. The generator of the kernel of the inclusion map $\pi_1(T^2) \to \pi_1$ (solid torus) in the resulting three-manifold determines the three-manifold. The diffeomorphism can be chosen to make this generator an arbitrary primitive (indivisible non-zero) element of $\mathbb{Z} \oplus \mathbb{Z}$. It is well defined up to change in sign.

Every oriented three-manifold can be obtained by this construction (Lickorish) . It is difficult, in general, to tell much about the three-manifold resulting from this construction. When, for instance, is it simply connected? When is it irreducible? (*Irreducible* means every embedded two sphere bounds a ball).

Note that the homology of the three-manifold is a very insensitive invariant. The homology of a knot complement is the same as the homology of a circle, so when Dehn surgery is performed, the resulting manifold always has a cyclic first homology group. If generators for $\mathbb{Z} \oplus \mathbb{Z} = \pi_1(T^2)$ are chosen so that $(1, 0)$ generates the homology of the complement and $(0, 1)$ is trivial then any Dehn surgery with invariant $(1, n)$ yields a homology sphere. 3. *Branched coverings*. If L is a link, then any finite-sheeted covering space of $S^3 - L$ can be compactified in a canonical way by adding circles which cover L to give a closed manifold, M. M is called a 1.3
branched covering of S^3 over L. There is a canonical projection $p : M \to S^3$, which is a local diffeomorphism away from $p^{-1}(L)$. If $K \subset S^3$ is a knot, the simplest branched coverings of S^3 over K are then n-fold cyclic branched covers, which come from the covering spaces of $S^3 - K$ whose fundamental group is the kernel of the composition $\pi_1(S^3 - K) \to H_1(S^3 - K) = \mathbb{Z} \to \mathbb{Z}_n$. In other words, they are unwrapping S^3 from K n times. If K is the trivial knot the cyclic branched covers are S^3. It seems intuitively obvious (but it is not *known*) that this is the only way S^3 can be obtained as a cyclic branched covering of itself over a knot. Montesinos and Hilden (independently) showed that every oriented three-manifold is a branched cover of S^3 with 3 sheets, branched over some knot. These branched coverings are not in general regular: there are no covering transformations.

The formation of irregular branched coverings is somehow a much more flexible construction than the formation of regular branched coverings. For instance, it is not hard to find many different ways in which S^3 is an irregular branched cover of itself.

5. *Heegaard decompositions.* Every three-manifold can be obtained from two handlebodies (of some genus) by gluing their boundaries together.



1.4

The set of possible gluing maps is large and complicated. It is hard to tell, given two gluing maps, whether or not they represent the same three-manifold (except when there are homological invariants to distinguish them).

6. *Identifying faces of polyhedra.* Suppose $P_1, \dots, P_k$ are polyhedra such that the number of faces with K sides is even, for each K.

Choose an arbitrary pattern of orientation-reversing identifications of pairs of two-faces. This yields a three-complex, which is an oriented manifold except near the vertices. (Around an edge, the link is automatically a circle.)

There is a classical criterion which says that such a complex is a manifold if and only if its Euler characteristic is zero. We leave this as an exercise.

In any case, however, we may simply remove a neighborhood of each bad vertex, to obtain a three-manifold with boundary.

The number of (at least not obviously homeomorphic) three-manifolds grows very quickly with the complexity of the description. Consider, for instance, different ways to obtain a three-manifold by gluing the faces of an octahedron. There are

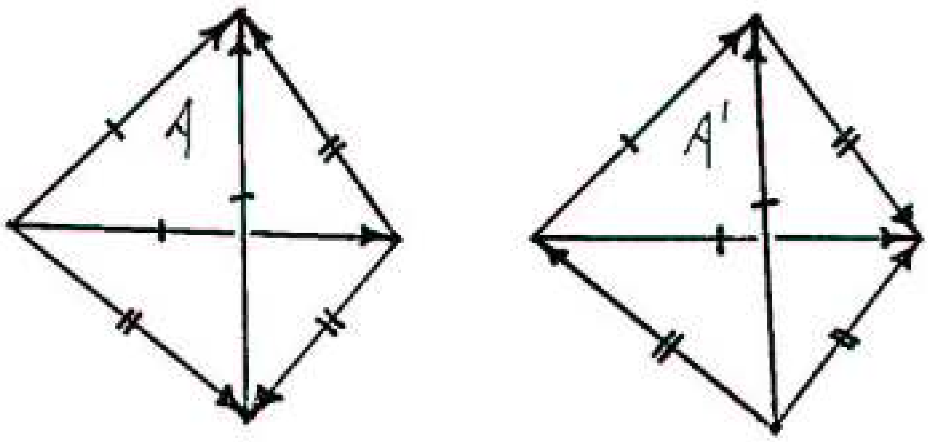
$$\frac{8!}{2^4 \cdot 4!} \cdot 3^4 = 8{,}505$$

possibilities. For an icosahedron, the figure is 38,661 billion. Because these polyhedra are symmetric, many gluing diagrams obviously yield homeomorphic results—but this reduces the figure by a factor of less than 120 for the icosahedron, for instance.

In two dimensions, the number of possible ways to glue sides of $2n$-gon to obtain an oriented surface also grows rapidly with n: it is $(2n)!/(2^n n!)$. In view of the amazing fact that the Euler characteristic is a complete invariant of a closed oriented surface, huge numbers of these gluing patterns give identical surfaces. It seems unlikely that 1.5

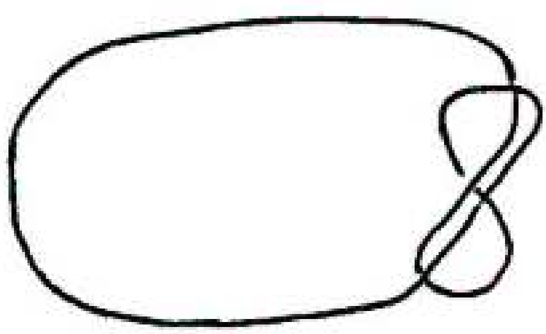
such a phenomenon takes place among three-manifolds; but how can we tell?

Example. Here is one of the simplest possible gluing diagrams for a three-manifold. Begin with two tetrahedra with edges labeled:



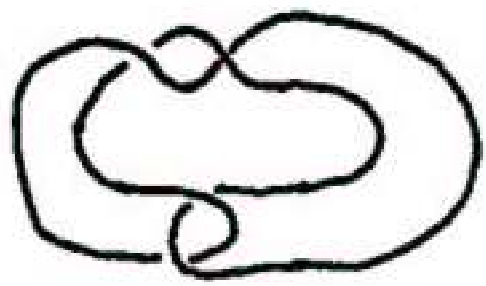
There is a unique way to glue the faces of one tetrahedron to the other so that arrows are matched. For instance, A is matched with A'. All the $\not\longrightarrow$ arrows are identified and all the $/\!\!\!\not\longrightarrow$ arrows are identified, so the resulting complex has 2 tetrahedra, 4 triangles, 2 edges and 1 vertex. Its Euler characteristic is $+1$, and (it follows that) a neighborhood of the vertex is the cone on a torus. Let M be the manifold obtained by removing the vertex.

It turns out that this manifold is homeomorphic with the complement of a figure-eight knot.



"Figure eight knot."

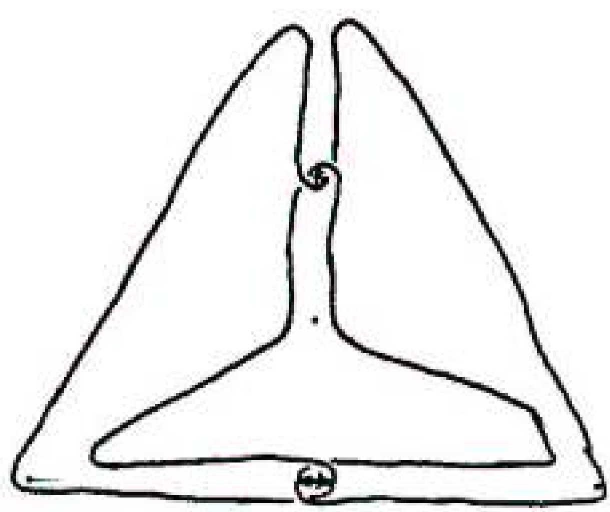
1.6



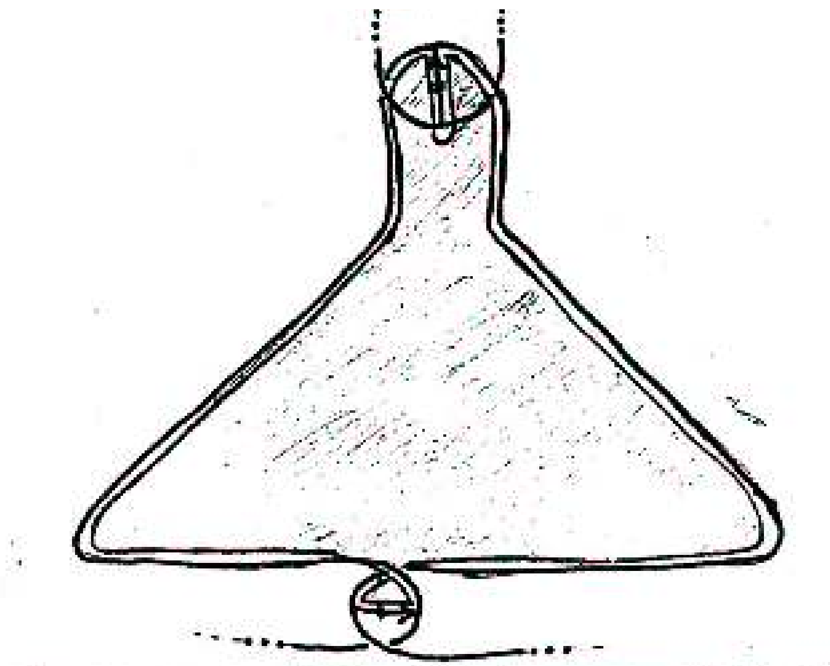
Another view of the figure-eight knot

This knot is familiar from extension cords, as the most commonly occurring knot, after the trefoil knot

In order to see this homeomorphism we can draw a more suggestive picture of the figure-eight knot, arranged along the one-skeleton of a tetrahedron. The knot can be



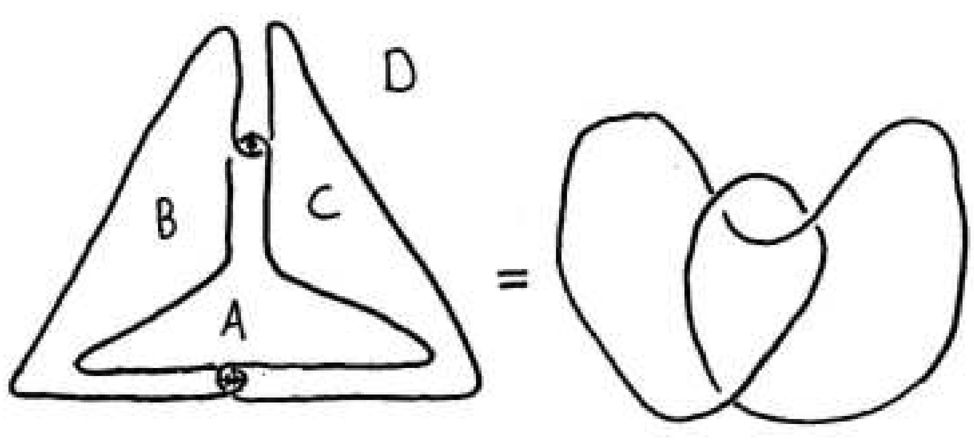
Tetrahedron with figure-eight knot, viewed from above spanned by a two-complex, with two edges, shown as arrows, and four two-cells, one for each face of the tetrahedron, in a more-or-less obvious way: 1.7



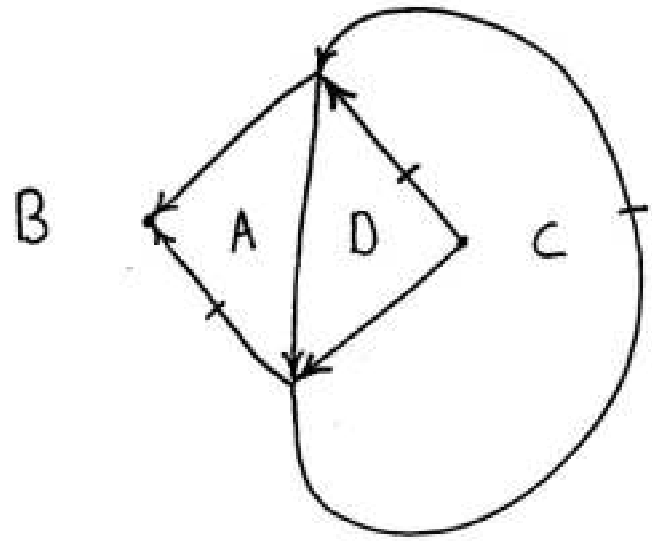
This pictures illustrates the typical way in which a two-cell is attached. Keeping in mind that the knot is not there, the cells are triangles with deleted vertices. The two complementary regions of the two-complex are the tetrahedra, with deleted vertices.

We will return to this example later. For now, it serves to illustrate the need for a systematic way to compare and to recognize manifolds.

NOTE. Suggestive pictures can also be deceptive. A trefoil knot can similarly be arranged along the one-skeleton of a tetrahedron: 1.8



From the picture, a cell-division of the complement is produced. In this case, however, the three-cells are not tetrahedra.



The boundary of a three-cell, flattened out on the plane.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TeX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

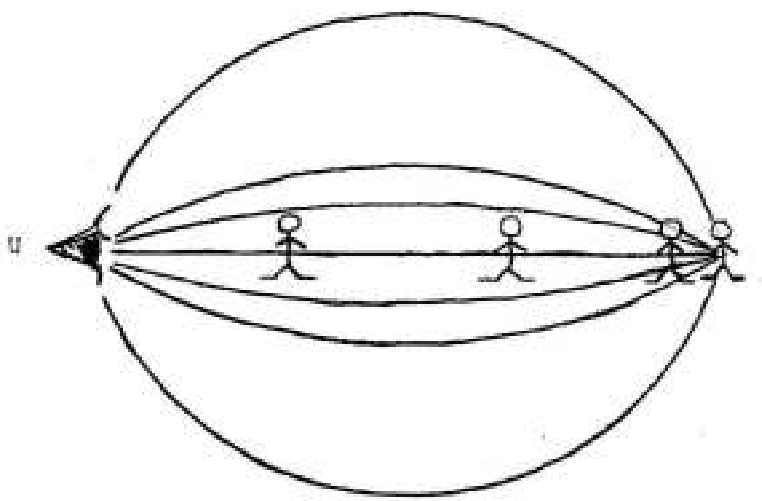
CHAPTER 2

Elliptic and hyperbolic geometry

There are three kinds of geometry which possess a notion of distance, and which look the same from any viewpoint with your head turned in any orientation: these are elliptic geometry (or spherical geometry), Euclidean or parabolic geometry, and hyperbolic or Lobachevskiian geometry. The underlying spaces of these three geometries are naturally Riemannian manifolds of constant sectional curvature $+1$, 0, and -1, respectively.

Elliptic n-space is the n-sphere, with antipodal points identified. Topologically it is projective n-space, with geometry inherited from the sphere. The geometry of elliptic space is nicer than that of the sphere because of the elimination of identical, antipodal figures which always pop up in spherical geometry. Thus, *any* two points in elliptic space determine a unique line, for instance.

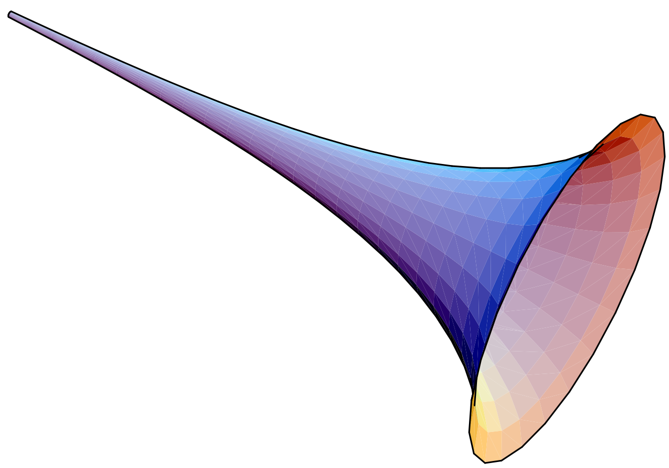
In the sphere, an object moving away from you appears smaller and smaller, until it reaches a distance of $\pi/2$. Then, it starts looking larger and larger and optically, it is in focus behind you. Finally, when it reaches a distance of π, it appears so large that it would seem to surround you entirely.



2.2

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 you would always be confronted with an image of yourself, turned inside out, upside down and filling out the entire background of your field of view. Euclidean space is familiar to all of us, since it very closely approximates the geometry of the space in which we live, up to moderate distances. Hyperbolic space is the least familiar to most people. Certain surfaces of revolution in $\mathbb{R}^3$ have constant curvature -1 and so give an idea of the local picture of the hyperbolic plane. 2.3

The simplest of these is the pseudosphere, the surface of revolution generated by a tractrix. A tractrix is the track of a box of stones which starts at $(0, 1)$ and is dragged by a team of oxen walking along the x-axis and pulling the box by a chain of unit length. Equivalently, this curve is determined up to translation by the property that its tangent lines meet the x-axis a unit distance from the point of tangency. The pseudosphere is not complete, however—it has an edge, beyond which it cannot be extended. Hilbert proved the remarkable theorem that *no* complete C^2 surface with curvature -1 can exist in $\mathbb{R}^3$. In spite of this, convincing physical models can be constructed.



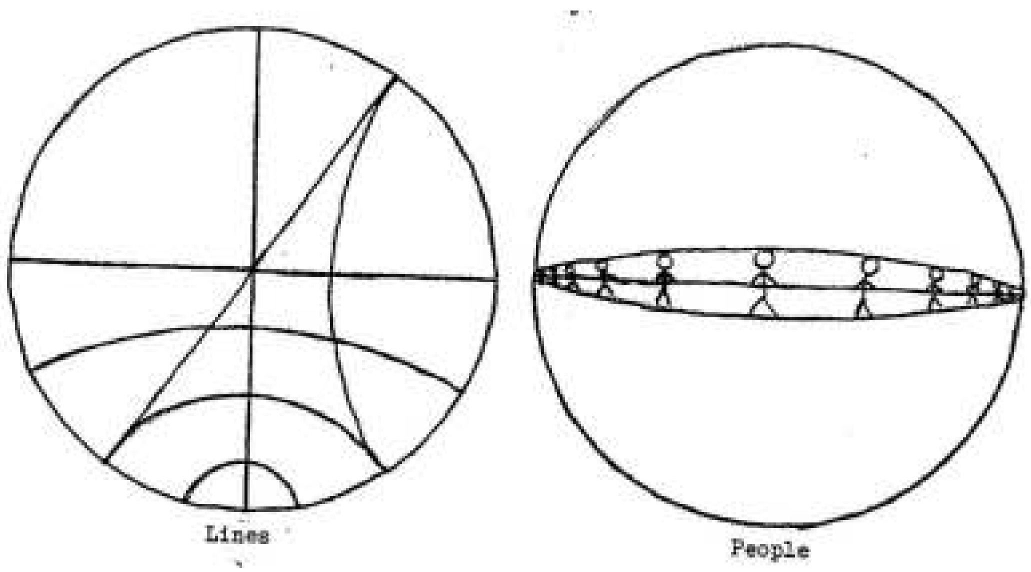
We must therefore resort to distorted pictures of hyperbolic space. Just as it is convenient to have different maps of the earth for understanding various aspects of its geometry: for seeing shapes, for comparing areas, for plotting geodesics in navigation; so it is useful to have several maps of hyperbolic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

2.1. The Poincaré disk model.

Let D^n denote the disk of unit radius in Euclidean n-space. The interior of D^n can be taken as a map of hyperbolic space H^n. A hyperbolic line in the model is any Euclidean circle which is orthogonal to ∂D^n; a hyperbolic two-plane is a Euclidean sphere orthogonal to ∂D^n; etc. The words "circle" and "sphere" are here used in the extended sense, to include the limiting case of a line or plane. This model is conformally correct, that is, hyperbolic angles agree with Euclidean angles, but distances are greatly distorted. Hyperbolic arc length $\sqrt{ds^2}$ is given by the formula 2.4

$$ds^2 = \left(\frac{1}{1-r^2}\right)^2 dx^2,$$

where $\sqrt{dx^2}$ is Euclidean arc length and r is distance from the origin. Thus, the Euclidean image of a hyperbolic object, as it moves away from the origin, shrinks in size roughly in proportion to the Euclidean distance from ∂D^n (when this distance is small). The object never actually arrives at ∂D^n, if it moves with a bounded hyperbolic velocity.



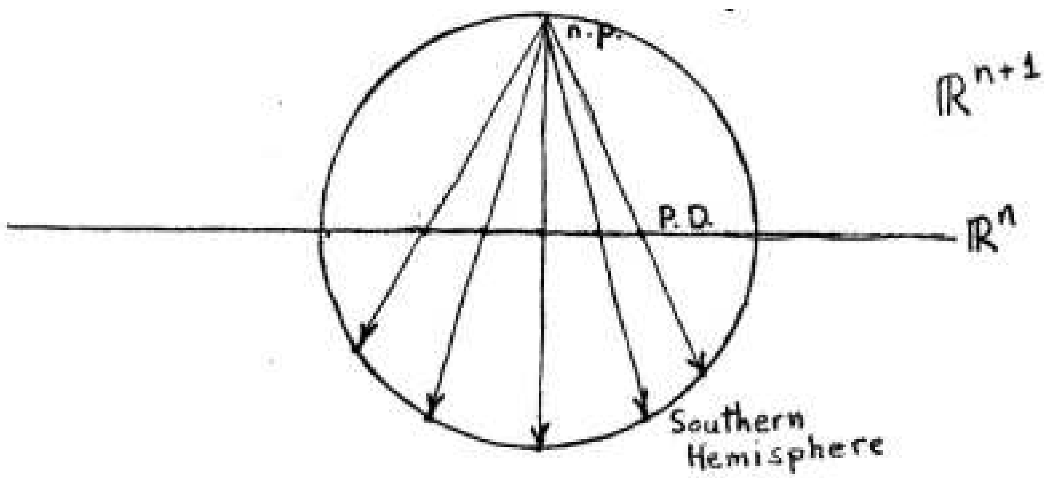
The sphere ∂D^n is called the *sphere at infinity.* It is not actually in hyperbolic space, but it can be given an interpretation purely in terms of hyperbolic geometry, as follows. Choose any base point p_0 in H^n. Consider any geodesic ray R, as seen from p_0. R traces out a segment of a great circle in the visual sphere at p_0 (since p_0 and R determine a two-plane). This visual segment converges to a point in the visual sphere. If we translate H^n so that p_0 is at the origin of the Poincaré disk 2.5

model, we see that the points in the visual sphere correspond precisely to points in the sphere at infinity, and that the end of a ray in this visual sphere corresponds to its Euclidean endpoint in the Poincaré disk model.

2.2. The southern hemisphere.

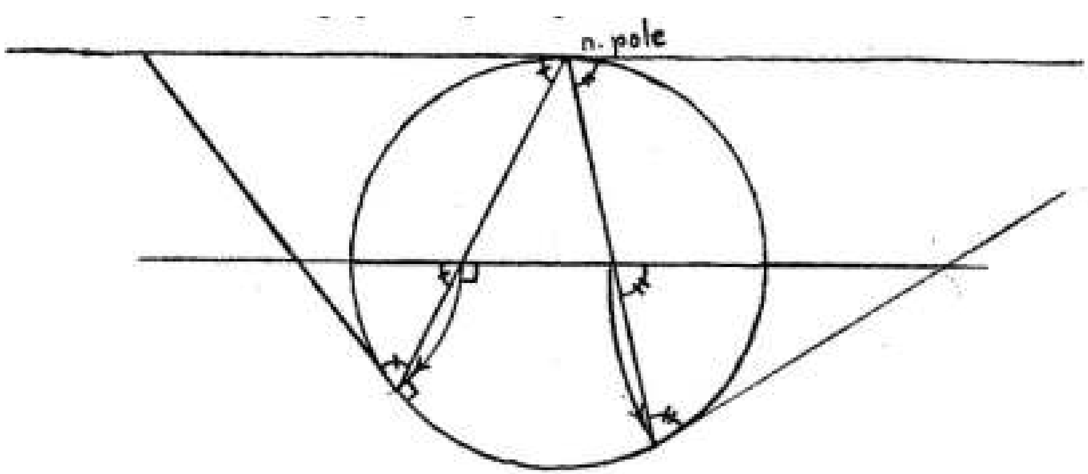
The Poincaré disk $D^n \subset \mathbb{R}^n$ is contained in the Poincaré disk $D^{n+1} \subset \mathbb{R}^{n+1}$, as a hyperbolic n-plane in hyperbolic $(n+1)$-space.

Stereographic projection (Euclidean) from the north pole of ∂D^{n+1} sends the Poincaré disk D^n to the southern hemisphere of D^{n+1}.



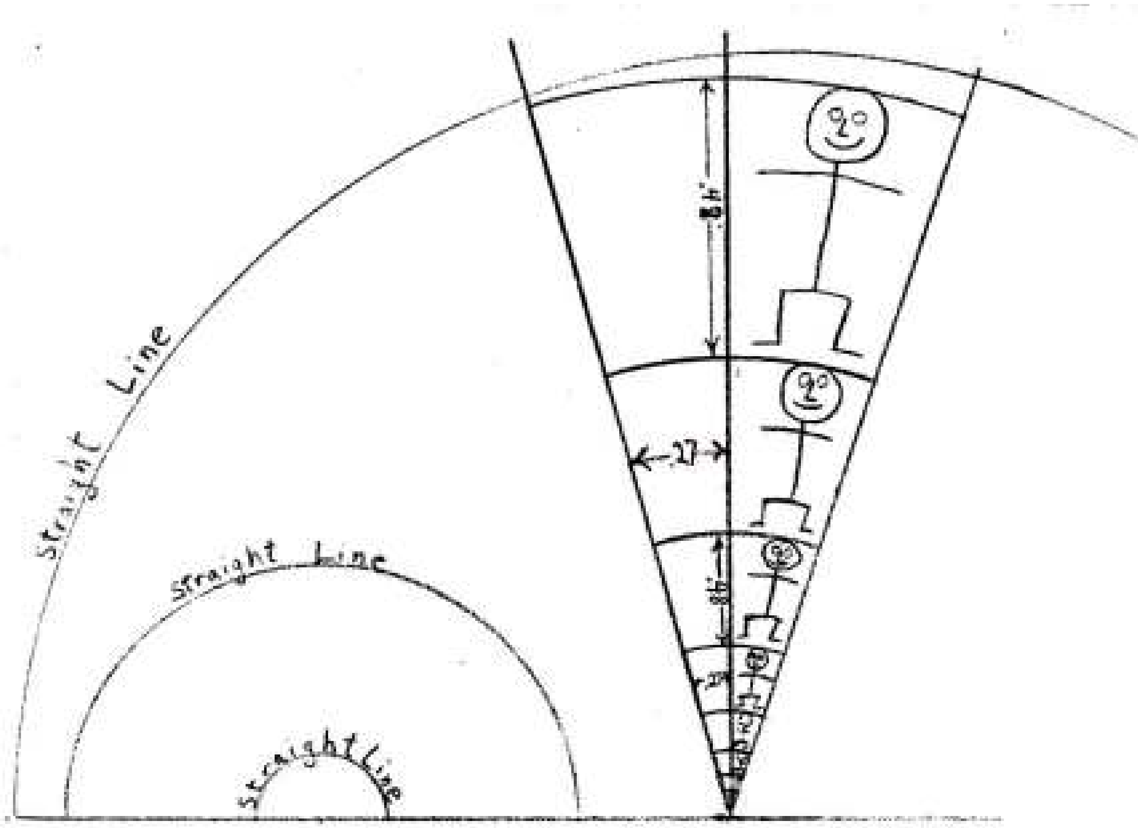
Thus hyperbolic lines in the Poincaré disk go to circles on S^n orthogonal to the equator S^{n-1}.

There is a more natural construction for this map, using only hyperbolic geometry. For each point p in $H^n \subset H^{n+1}$, consider the hyperbolic ray perpendicular to H^n at p, and downward normal. This ray converges to a point on the sphere at infinity, 2.6
which is the same as the Euclidean stereographic image of p.



2.3. The upper half-space model.

This is closely related to the previous two, but it is often more convenient for computation or for constructing pictures. To obtain it, rotate the sphere S^n in $\mathbb{R}^{n+1}$ so that the southern hemisphere lies in the half-space $x_n \geq 0$ is $\mathbb{R}^{n+1}$. Now stereographic projection from the top of S^n (which is now on the equator) sends the southern hemisphere to the upper half-space $x_n > 0$ in $\mathbb{R}^{n+1}$. 2.7



A hyperbolic line, in the upper half-space, is a circle perpendicular to the bounding plane $\mathbb{R}^{n-1} \subset \mathbb{R}^n$. The hyperbolic metric is $ds^2 = (1/x_n)^2\, dx^2$. Thus, the Euclidean image of a hyperbolic object moving toward $\mathbb{R}^{n-1}$ has size precisely proportional to the Euclidean distance from R^{n-1}.

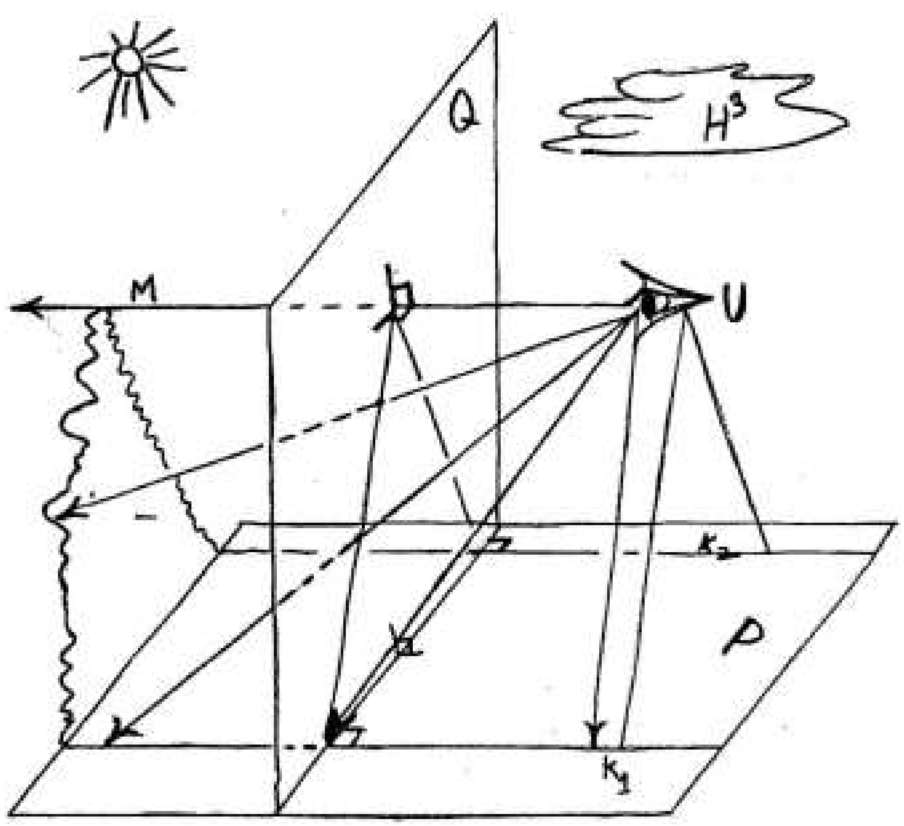
2.4. The projective model.

This is obtained by Euclidean orthogonal projection of the southern hemisphere of S^n back to the disk D^n. Hyperbolic lines become Euclidean line segments. This model is useful for understanding incidence in a configuration of lines and planes. Unlike the previous three models, it fails to be conformal, so that angles and shapes are distorted.

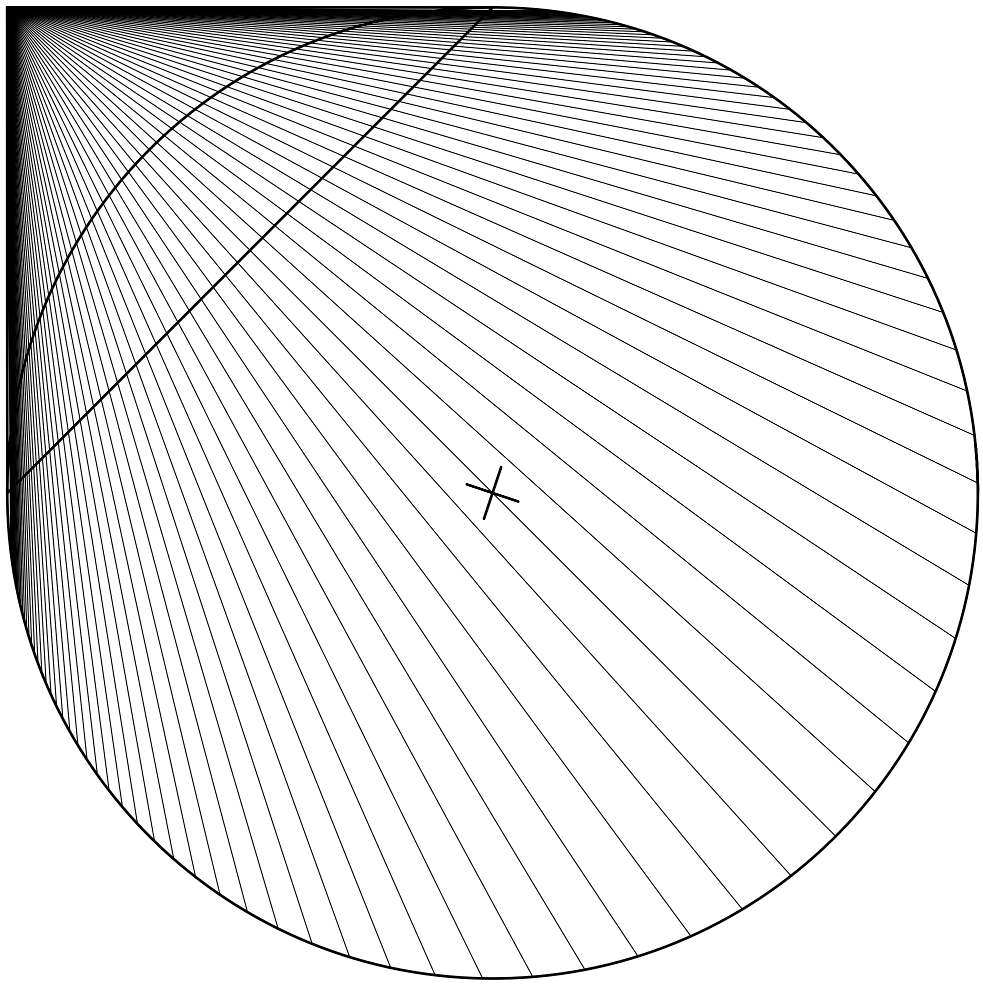
It is better to regard this projective model to be contained not in Euclidean space, but in projective space. The projective model is very natural from a point of 2.8
view inside hyperbolic $(n+1)$-space: it gives a picture of a hyperplane, H^n, in true perspective. Thus, an observer hovering above H^n in H^{n+1}, looking down, sees H^n

as the interior of a disk in his visual sphere. As he moves farther up, this visual disk shrinks; as he moves down, it expands; but (unlike in Euclidean space), the visual radius of this disk is always strictly less than $\pi/2$. A line on H^2 appears visually straight.

It is possible to give an intrinsic meaning within hyperbolic geometry for the points outside the sphere at infinity in the projective model. For instance, in the two-dimensional projective model, any two lines meet somewhere. The conventional sense of meeting means to meet inside the sphere at infinity (at a *finite* point). If the two lines converge in the visual circle, this means that they meet on the circle at infinity, and they are called *parallels*. Otherwise, the two lines are called *ultraparallels*; they have a unique common perpendicular L and they meet in some point x in the Möbius band outside the circle at infinity. *Any other line perpendicular to L passes through x, and any line through x is perpendicular to L.* 2.9



To prove this, consider hyperbolic two-space as a plane $P \subset H^3$. Construct the plane Q through L perpendicular to P. Let U be an observer in H^3. Drop a perpendicular M from U to the plane Q. Now if K is any line in P perpendicular

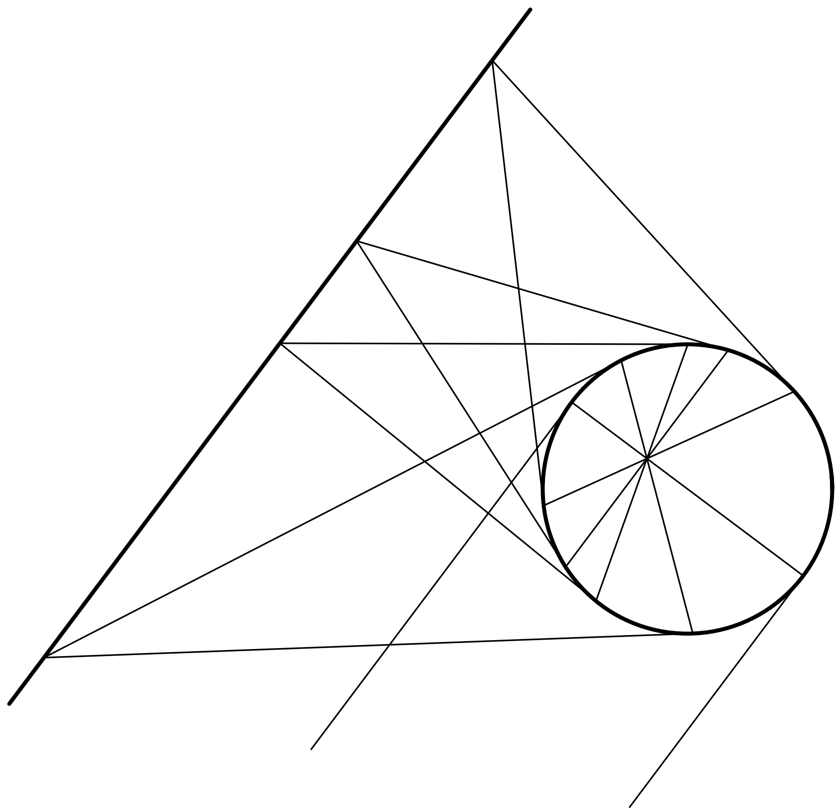


Evenly spaced lines. The region inside the circle is a plane, with a base line and a family of its perpendiculars, spaced at a distance of .051 fundamental units, as measured along the base line shown in perspective in hyperbolic 3-space (or in the projective model). The lines have been extended to their imaginary meeting point beyond the horizon. U, the observer, is directly above the X (which is .881 fundamental units away from the base line). To see the view from different heights, use the following table (which assumes that the Euclidean diameter of the circle in your printout is about 5.25 inches or 13.3cm):

To see the view of U at a height of	hold the picture a distance of	To see the view of U at a height of	hold the picture a distance of
2 units	11″ (28 cm)	5 units	17′ (519 cm)
3 units	27″ (69 cm)	10 units	2523′ (771 m)
4 units	6′ (191 cm)	20 units	10528.75 miles (16981 km)

For instance, you may imagine that the fundamental distance is 10 meters. Then the lines are spaced about like railroad ties. Twenty units is 200 meters: U is in a hot air balloon.

to L, the plane determined by U and K is perpendicular to Q, hence contains M; hence the visual line determined by K in the visual sphere of U passes through the visual point determined by K. The converse is similar. 2.10

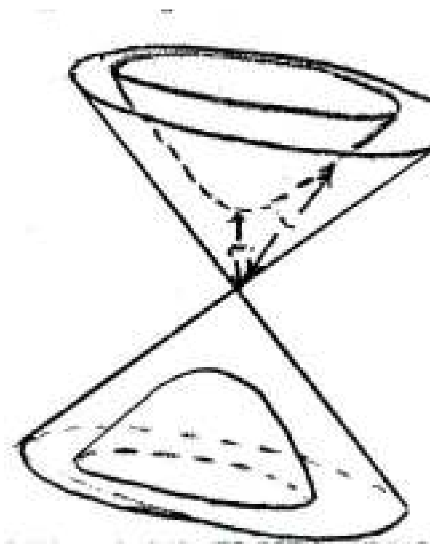


This gives a one-to-one correspondence between the set of points x outside the sphere at infinity, and (in general) the set of hyperplanes L in H^n. L corresponds to the common intersection point of all its perpendiculars. Similarly, there is a correspondence between points in H^n and hyperplanes outside the sphere at infinity: a point p corresponds to the union of all points determined by hyperplanes through p.

2.5. The sphere of imaginary radius.

A sphere in Euclidean space with radius r has constant curvature $1/r^2$. Thus, hyperbolic space should be a sphere of radius i. To give this a reasonable interpretation, we use an indefinite metric $dx^2 = dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2$ in R^{n+1}. The sphere of radius i about the origin in this metric is the hyperboloid

$$x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1.$$



2.11

The metric dx^2 restricted to this hyperboloid is positive definite, and it is not hard to check that it has constant curvature -1. Any plane through the origin is dx^2-orthogonal to the hyperboloid, so it follows from elementary Riemannian geometry that it meets the hyperboloid in a geodesic. The projective model for hyperbolic space is reconstructed by projection of the hyperboloid from the origin to a hyperplane in $\mathbb{R}^n$. Conversely, the quadratic form $x_1^2 + \cdots + x_n^2 - x_{n+1}^2$ can be reconstructed from the projective model. To do this, note that there is a unique quadratic equation of the form

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j = 1$$

defining the sphere at infinity in the projective model. Homogenization of this equation gives a quadratic form of type $(n, 1)$ in $\mathbb{R}^{n+1}$, as desired. Any isometry of the quadratic form $x_1^2 + \cdots + x_n^2 - x_{n+1}^2$ induces an isometry of the hyperboloid, and hence any projective transformation of $\mathbb{P}^n$ that preserves the sphere at infinity induces an isometry of hyperbolic space. This contrasts with the situation in Euclidean geometry, where there are many projective self-homeomorphisms: the affine transformations. In particular, hyperbolic space has no similarity transformations except isometries. This is true also for elliptic space. This means that there is a well-defined unit of measurement of distances in hyperbolic geometry. We shall later see how this is related to three-dimensional topology, giving a measure of the "size" of manifolds. 2.12

2.6. Trigonometry.

Sometimes it is important to have formulas for hyperbolic geometry, and not just pictures. For this purpose, it is convenient to work with the description of hyperbolic space as one sheet of the "sphere" of radius i with respect to the quadratic form

$$Q(X) = X_1^2 + \cdots + X_n^2 - X_{n+1}^2$$

in $\mathbb{R}^{n+1}$. The set $\mathbb{R}^{n+1}$, equipped with this quadratic form and the associated inner product

$$X \cdot Y = \sum_{i=1}^{n} X_1 Y_1 - X_{n+1} Y_{n+1},$$

is called $E^{n,1}$. First we will describe the geodesics on level sets $S_r = \{X : Q(X) = r^2\}$ of Q. Suppose that X_t is such a geodesic, with speed

$$s = \sqrt{Q(\dot{X}_t)}.$$

We may differentiate the equations

$$X_t \cdot X_t = r^2, \qquad \dot{X}_t \cdot \dot{X}_t = s^2,$$

to obtain

$$X_t \cdot \dot{X}_t = 0, \qquad \dot{X}_t \cdot \ddot{X}_t = 0,$$

and

$$X_t \cdot \ddot{X}_t = -\dot{X}_t \cdot \dot{X}_t = -s^2.$$

Since any geodesic must lie in a two-dimensional subspace, $\ddot{X}_t$ must be a linear combination of X_t and $\dot{X}_t$, and we have

2.6.1. $$\ddot{X}_t = -\left(\frac{s}{r}\right)^2 X_t.$$

This differential equation, together with the initial conditions 2.13

$$X_0 \cdot X_0 = r^2, \qquad \dot{X}_0 \cdot \dot{X}_0 = s^2, \qquad X_0 \cdot \dot{X}_0 = 0,$$

determines the geodesics.

Given two vectors X and Y in $E^{n,1}$, if X and Y have nonzero length we define the quantity

$$c(X, Y) = \frac{X \cdot Y}{\|X\| \cdot \|Y\|},$$

where $\|X\| = \sqrt{X \cdot X}$ is positive real or positive imaginary. Note that

$$c(X, Y) = c(\lambda X, \mu Y),$$

where λ and μ are positive constants, that $c(-X, Y) = -c(X, Y)$, and that $c(X, X) = 1$. In Euclidean space E^{n+1}, $c(X, Y)$ is the cosine of the angle between X and Y. In $E^{n,1}$ there are several cases.

We identify vectors on the positive sheet of S_i ($X_{n+1} > 0$) with hyperbolic space. If Y is any vector of real length, then Q restricted to the subspace $Y^\perp$ is indefinite of type $(n-1,\,1)$. This means that $Y^\perp$ intersects H^n and determines a hyperplane.

We will use the notation $Y^\perp$ to denote this hyperplane, *with the normal orientation determined by* Y. (We have seen this correspondence before, in 2.4.)

2.6.2. If X and $Y \in H^n$, then $c(X,Y) = \cosh d\,(X,Y)$,

where $d\,(X,Y)$ denotes the hyperbolic distance between X and Y.

To prove this formula, join X to Y by a geodesic X_t of unit speed. From 2.6.1 we have 2.14

$$\ddot{X}_t = X_t, \qquad X_t \cdot \dot{X}_0 = 0,$$

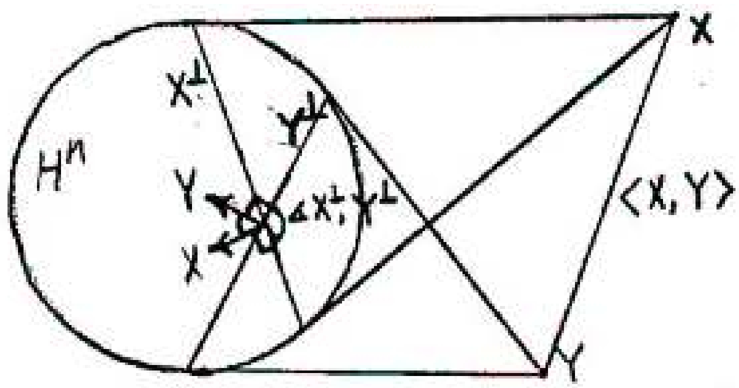
so we get $c(\ddot{X}_t, X_t) = c(X_t, X_t)$, $c(\dot{X}_0, X_0) = 0$, $c(X, X_0) = 1$; thus $c(X, X_t) = \cosh t$. When $t = d\,(X,Y)$, then $X_t = Y$, giving 2.6.2.

If $X^\perp$ and $Y^\perp$ are distinct hyperplanes, then

2.6.3.

$X^\perp$ and $Y^\perp$ intersect

$$\begin{aligned} &\iff Q \text{ is positive definite on the subspace } \langle X,Y\rangle \text{ spanned by } X \text{ and } Y \\ &\iff c(X,Y)^2 < 1 \\ &\implies c(X,Y) = \cos\angle(X,Y) = -\cos\angle(X^\perp, Y^\perp). \end{aligned}$$



To see this, note that X and Y intersect in $H^n \iff Q$ restricted to $X^\perp \cap Y^\perp$ is indefinite of type $(n-2,1) \iff Q$ restricted to $\langle X,Y\rangle$ is positive definite. ($\langle X,Y\rangle$ is the normal subspace to the $(n-2)$ plane $X^\perp \cap Y^\perp$). 2.15

There is a general elementary formula for the area of a parallelogram of sides X and Y with respect to an inner product:

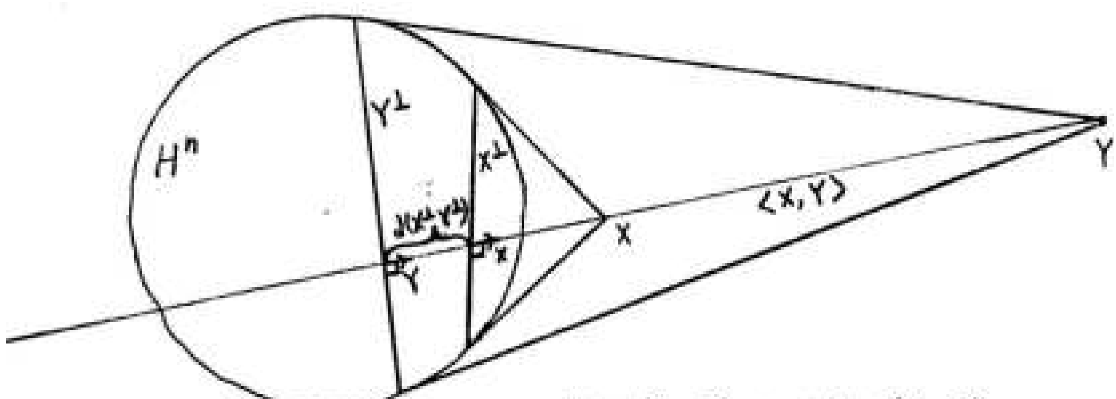
$$\text{area } = \sqrt{X \cdot X\, Y \cdot Y - (X \cdot Y)^2} = \|X\| \cdot \|Y\| \cdot \sqrt{1 - c(X,Y)^2}.$$

This area is positive real if X and Y span a positive definite subspace, and positive imaginary if the subspace has type $(1,1)$. This shows, finally, that $X^\perp$ and $Y^\perp$ intersect $\iff$ $c\,(X,Y)^2 < 1$. The formula for $c(X,Y)$ comes from ordinary trigonometry.

2.6.4.

$$\begin{aligned} X^\perp \text{ and } Y^\perp \text{ have a common perpendicular } &\iff Q \text{ has type } (1,1) \text{ on } \langle X, Y\rangle \\ &\iff c(X,Y)^2 > 1 \\ &\implies c(X,Y) = \pm\cosh\bigl(d\,(X^\perp, Y^\perp)\bigr). \end{aligned}$$

The sign is positive if the normal orientations of the common perpendiculars coincide, and negative otherwise.



$0 < c\,(X\,,\,Y) = \cosh\ d(X\,,\,Y)$ 2.16

The proof is similar to 2.6.2. We may assume X and Y have unit length. Since $\langle X, Y\rangle$ intersects H^n as the common perpendicular to $X^\perp$ and $Y^\perp$, Q restricted to $\langle X, Y\rangle$ has type $(1,1)$. Replace X by $-X$ if necessary so that X and Y lie in the same component of $S_1 \cap \langle X, Y\rangle$. Join X to Y by a geodesic X_t of speed i. From 2.6.1, $\ddot{X}_t = X_t$. There is a dual geodesic Z_t of unit speed, satisfying $Z_t \cdot X_t = 0$, joining $X^\perp$ to $Y^\perp$ along their common perpendicular, so one may deduce that

$$c,(X,Y) = \pm\tfrac{d\,(X,Y)}{i} = \pm d\,(X^\perp, Y^\perp).$$

There is a limiting case, intermediate between 2.6.3 and 2.6.4:

2.6.5. $X^\perp$ and $Y^\perp$ are parallel

$$\begin{aligned} &\iff Q \text{ restricted to } \langle X, Y\rangle \text{ is degenerate} \\ &\iff c(X,Y)^2 = 1. \end{aligned}$$

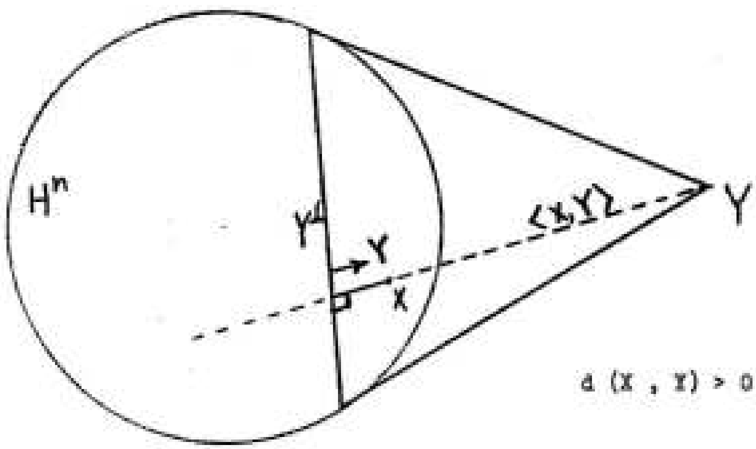
In this case, we say that $X^\perp$ and $Y^\perp$ form an angle of 0 or π. $X^\perp$ and $Y^\perp$ actually have a distance of 0, where the distance of two sets U and V is defined to be the infimum of the distance between points $u \in U$ and $v \in V$.

There is one more case in which to interpret $c(X, Y)$:

2.6.6. If X is a point in H^n and $Y^\perp$ a hyperplane, then

$$c(X,Y) = \frac{\sinh\big(d\,(X, Y^\perp)\big)}{i},$$

where $d\,(X, Y^\perp)$ is the *oriented* distance. 2.17



The proof is left to the untiring reader.

With our dictionary now complete, it is easy to derive hyperbolic trigonometric formulae from linear algebra. To solve triangles, note that the edges of a triangle with vertices u, v and w in H^2 are $U^\perp$, $V^\perp$ and $W^\perp$, where U is a vector orthogonal to v and w, etc. To find the angles of a triangle from the lengths, one can find three vectors u, v, and w with the appropriate inner products, find a dual basis, and calculate the angles from the inner products of the dual basis. Here is the general formula. We consider triangles in the projective model, with vertices inside or outside the sphere at infinity. Choose vectors v_1, v_2 and v_3 of length i or 1 representing these points. Let $\epsilon_i = v_i \cdot v_i$, $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and $c_{ij} = c(v_i, v_j)$. Then the matrix of inner products of the v_i is

$$C = \begin{bmatrix} \epsilon_1 & \epsilon_{12}c_{12} & \epsilon_{13}c_{13} \\ \epsilon_{12}c_{12} & \epsilon_2 & \epsilon_{23}c_{23} \\ \epsilon_{13}c_{13} & \epsilon_{23}c_{23} & \epsilon_3 \end{bmatrix}.$$

The matrix of inner products of the dual basis $\{v^1, v^2, v^3\}$ is C^{-1}. For our purposes, though, it is simpler to compute the matrix of inner products of the basis 2.18

$\{\sqrt{-\det C^i}\}$,

$$- \operatorname{adj} C = (-\det C) \cdot C^{-1} =$$
$$\begin{bmatrix} -\epsilon_2\epsilon_3(1-c_{23}^2) & -\epsilon_{12}\epsilon_3(c_{13}c_{23}-c_{12}) & -\epsilon_{13}\epsilon_2(c_{12}c_{23}-c_{13}) \\ -\epsilon_{12}\epsilon_3(c_{13}c_{23}-c_{12}) & -\epsilon_1\epsilon_3(1-c_{13}^2) & -\epsilon_{23}\epsilon_1(c_{12}c_{13}-c_{23}) \\ -\epsilon_{13}\epsilon_2(c_{12}c_{23}-c_{13}) & -\epsilon_{23}\epsilon_1(c_{12}c_{13}-c_{23}) & -\epsilon_1\epsilon_2(1-c_{12}^2) \end{bmatrix}.$$

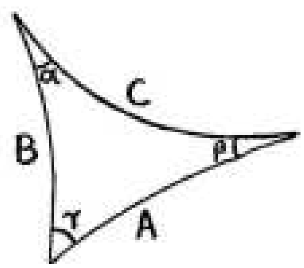
If v^1, v^2, v^3 is the dual basis, and $c^{ij} = c(v^i, v^j)$, we can compute

2.6.7. $$c^{12} = \epsilon \cdot \frac{c_{13}c_{23} - c_{12}}{\sqrt{1-c_{23}^2}\sqrt{1-c_{13}^2}},$$

where it is easy to deduce the sign

$$\epsilon = \frac{-\epsilon_{12}\epsilon_3}{\sqrt{-\epsilon_2\epsilon_3}\sqrt{-\epsilon_1\epsilon_3}}$$

directly. This specializes to give a number of formulas, in geometrically distinct cases. In a real triangle,

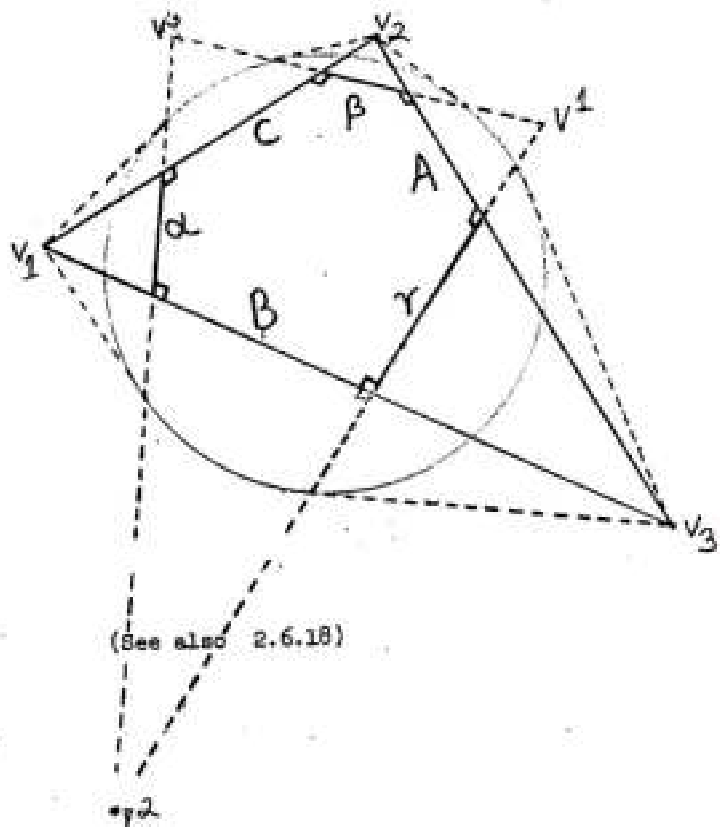


2.6.8. $$\cosh C = \frac{\cos\alpha\cos\beta + \cos\gamma}{\sin\alpha\sin\beta},$$

2.19

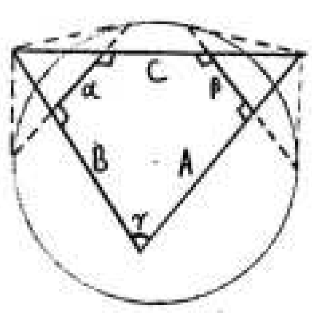
2.6.9. $$\cos\gamma = \frac{\cosh A \cosh B - \cosh C}{\sinh A \sinh B},$$

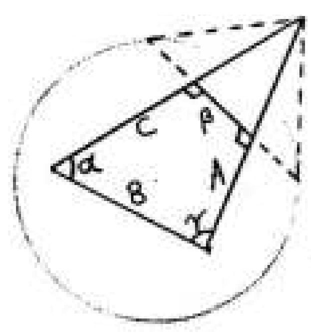
or $\cosh C = \cosh A \cosh B - \sinh A \sinh B \cos c$. (See also 2.6.16.) In an all right hexagon,



2.6.10. $$\cosh C = \frac{\cosh\alpha\cosh\beta + \cosh\gamma}{\sinh\alpha\sinh\beta}.$$

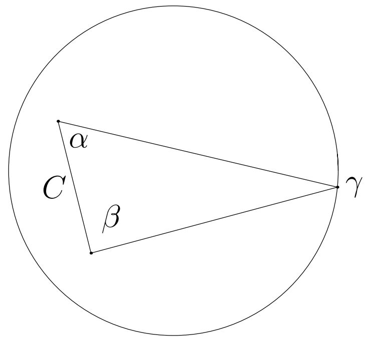
(See also 2.6.18.) Such hexagons are useful in the study of hyperbolic structures on surfaces. Similar formulas can be obtained for pentagons with four right angles, or quadrilaterals with two adjacent right angles:





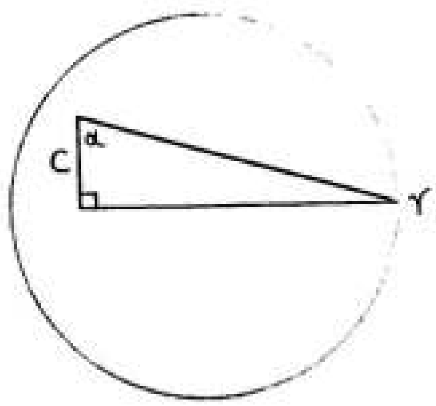
2.20

By taking the limit of 2.6.8 as the vertex with angle γ tends to the circle at infinity, we obtain useful formulas:



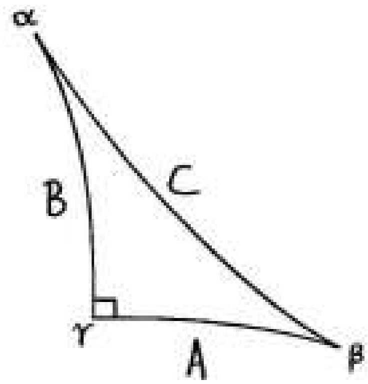
2.6.11.
$$\cosh C = \frac{\cos\alpha\cos\beta + 1}{\sin\alpha\sin\beta},$$

and in particular



2.6.12.
$$\cosh C = \frac{1}{\sin\alpha}.$$

These formulas for a right triangle are worth mentioning separately, since they are particularly simple. 2.21



From the formula for $\cos\gamma$ we obtain the hyperbolic Pythagorean theorem:

2.6.13. $$\cosh C = \cosh A \cosh B.$$

Also,

2.6.14. $$\cosh A = \frac{\cos\alpha}{\sin\beta}.$$

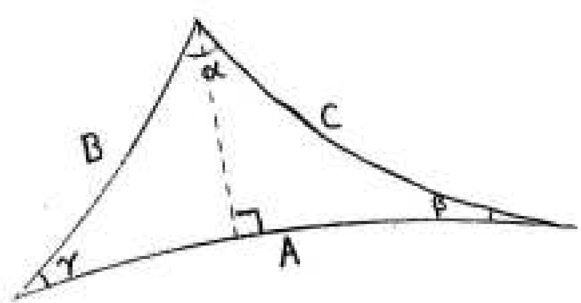
(Note that $(\cos\alpha)/(\sin\beta) = 1$ in a Euclidean right triangle.) By substituting

$$\frac{(\cosh C)}{(\cosh A)}$$

for $\cosh B$ in the formula 2.6.9 for $\cos\alpha$, one finds:

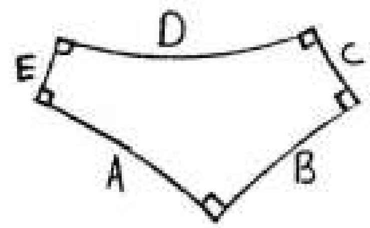
2.6.15. $$\text{In a right triangle, } \sin\alpha = \frac{\sinh A}{\sinh C}.$$

This follows from the general law of sines,



2.6.16. $$\text{In any triangle, } \frac{\sinh A}{\sin\alpha} = \frac{\sinh B}{\sin\beta} = \frac{\sinh C}{\sin\gamma}.$$

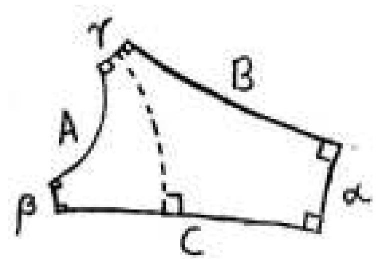
Similarly, in an all right pentagon, 2.22



one has

2.6.17. $$\sinh A \sinh B = \cosh D.$$

It follows that in any all right hexagon,



there is a law of sines:

2.6.18. $$\frac{\sinh A}{\sinh \alpha} = \frac{\sinh B}{\sinh \beta} = \frac{\sinh C}{\sinh \gamma}.$$

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 3

Geometric structures on manifolds

A manifold is a topological space which is locally modelled on $\mathbb{R}^n$. The notion of what it means to be locally modelled on $\mathbb{R}^n$ can be made definite in many different ways, yielding many different sorts of manifolds. In general, to define a kind of manifold, we need to define a set $\mathcal{G}$ of gluing maps which are to be permitted for piecing the manifold together out of chunks of $\mathbb{R}^n$. Such a manifold is called a $\mathcal{G}$-manifold. $\mathcal{G}$ should satisfy some obvious properties which make it a *pseudogroup* of local homeomorphisms between open sets of $\mathbb{R}^n$:

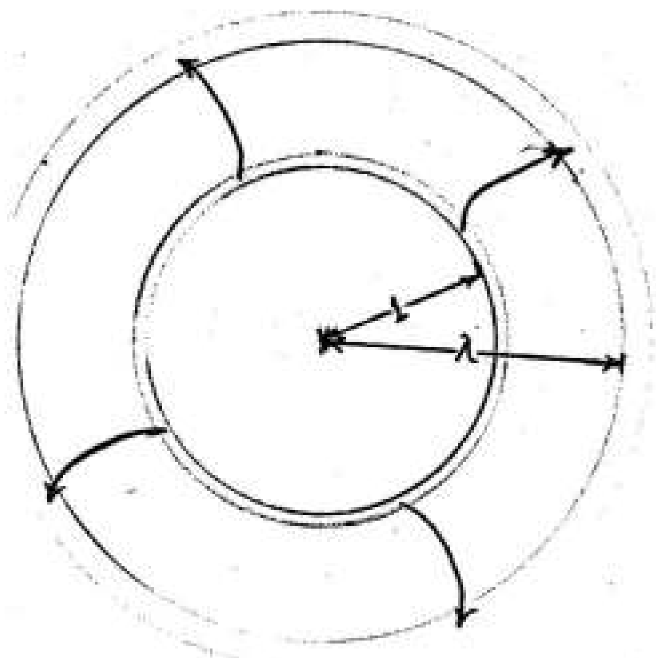
(i) The restriction of an element $g \in \mathcal{G}$ to any open set in its domain is also in $\mathcal{G}$.
(ii) The composition $g_1 \circ g_2$ of two elements of $\mathcal{G}$, when defined, is in $\mathcal{G}$.
(iii) The inverse of an element of $\mathcal{G}$ is in $\mathcal{G}$.
(iv) The property of being in $\mathcal{G}$ is local, so that if $U = \bigcup_\alpha U_\alpha$ and if g is a local homeomorphism $g : U \to V$ whose restriction to each U_α is in $\mathcal{G}$, then $g \in \mathcal{G}$.

It is convenient also to permit $\mathcal{G}$ to be a pseudogroup acting on *any* manifold, although, as long as $\mathcal{G}$ is transitive, this doesn't give any new types of manifolds. See Haefliger, in Springer Lecture Notes #197, for a discussion.

A group G acting on a manifold X determines a pseudogroup which consists of restrictions of elements of G to open sets in X. A (G, X)-manifold means a manifold 3.2
glued together using this pseudogroup of restrictions of elements of G.

Examples. If $\mathcal{G}$ is the pseudogroup of local C^r diffeomorphisms of $\mathbb{R}^n$, then a $\mathcal{G}$-manifold is a C^r-manifold, or more loosely, a differentiable manifold (provided $r \geq 1$).

If $\mathcal{G}$ is the pseudogroup of local piecewise-linear homeomorphisms, then a $\mathcal{G}$-manifold is a PL-manifold. If G is the group of affine transformations of $\mathbb{R}^n$, then a $(G, \mathbb{R}^n)$-manifold is called an *affine* manifold. For instance, given a constant $\lambda > 1$ consider an annulus of radii 1 and $\lambda + \epsilon$. Identify neighborhoods of the two boundary components by the map $x \to \lambda x$. The resulting manifold, topologically, is T^2.



3.3

Here is another method, due to John Smillie, for constructing affine structures on T^2 from any quadrilateral Q in the plane. Identify the opposite edges of Q by the orientation-preserving similarities which carry one to the other. Since similarities preserve angles, the sum of the angles about the vertex in the resulting complex is 2π, so it has an affine structure. We shall see later how such structures on T^2 are intimately connected with questions concerning Dehn surgery in three-manifolds.

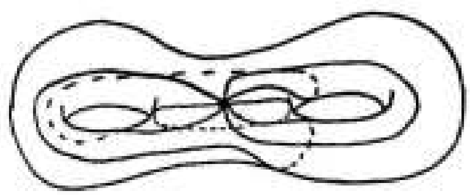
The literature about affine manifolds is interesting. Milnor showed that the only closed two-dimensional affine manifolds are tori and Klein bottles. The main unsolved question about affine manifolds is whether in general an affine manifold has Euler characteristic zero.

If G is the group of isometries of Euclidean space E^n, then a (G, E^n)-manifold is called a *Euclidean* manifold, or often a *flat* manifold. Bieberbach proved that a Euclidean manifold is finitely covered by a torus. Furthermore, a Euclidean structure automatically gives an affine structure, and Bieberbach proved that closed Euclidean manifolds with the same fundamental group are equivalent as affine manifolds. If G is the group $O(n + 1)$ acting on elliptic space $\mathbb{P}^n$ (or on S^n), then we obtain *elliptic* manifolds.

CONJECTURE. Every three-manifold with finite fundamental group has an elliptic structure.

This conjecture is a stronger version of the Poincaré conjecture; we shall see the logic shortly. All known three-manifolds with finite fundamental group certainly have elliptic structures. 3.4

As a final example (for the present), when G is the group of isometries of hyperbolic space H^n, then a (G, H^n)-manifold is a *hyperbolic* manifold. For instance, any surface of negative Euler characteristic has a hyperbolic structure. The surface of genus two is an illustrative example.



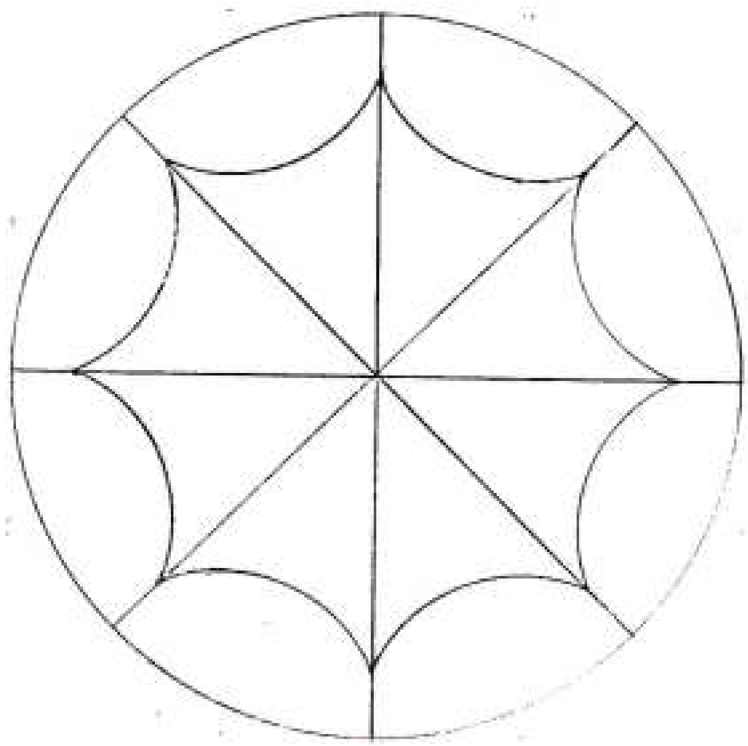
Topologically, this surface is obtained by identifying the sides of an octagon, in the pattern above, for instance. An example of a hyperbolic structure on the surface is obtained form any hyperbolic octagon whose opposite edges have equal lengths and whose angle sum is 2π, by identifying in the same pattern. There is a regular octagon with angles $\pi/4$, for instance. 3.5
3.6

3.1. A hyperbolic structure on the figure-eight knot complement.

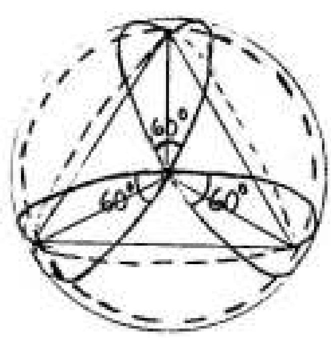
Consider a regular tetrahedron in Euclidean space, inscribed in the unit sphere, so that its vertices are on the sphere. Now interpret this tetrahedron to lie in the projective model for hyperbolic space, so that it determines an ideal hyperbolic simplex: combinatorially, a simplex with its vertices deleted. The dihedral angles of the hyperbolic simplex are 60°. This may be seen by extending its faces to the sphere at infinity, which they meet in four circles which meet each other in 60° angles.

By considering the Poincaré disk model, one sees immediately that the angle made by two planes is the same as the angle of their bounding circles on the sphere at infinity.

Take two copies of this ideal simplex, and glue the faces together, in the pattern described in Chapter 1, using Euclidean isometries, which are also (in this case) hyperbolic isometries, to identify faces. This gives a hyperbolic structure to the resulting manifold, since the angles add up to 360° around each edge. 3.7



A regular octagon with angles $\pi/4$,
whose sides can be identified to give a surface of genus 2.

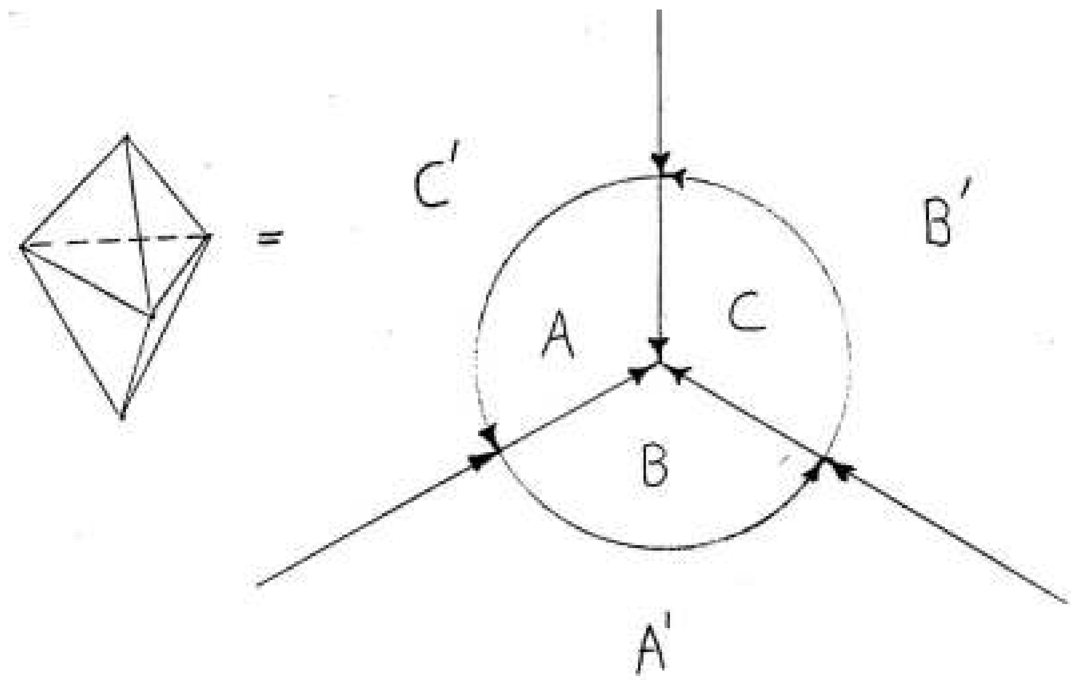


A tetrahedron inscribed in the unit sphere, top view.

According to Magnus, *Hyperbolic Tesselations*, this manifold was constructed by Gieseking in 1912 (but without any relation to knots). R. Riley showed that the figure-eight knot complement has a hyperbolic structure (which agrees with this one). This manifold also coincides with one of the hyperbolic manifolds obtained by an arithmetic construction, because the fundamental group of the complement of the figure-eight knot is isomorphic to a subgroup of index 12 in $\mathrm{PSL}_2(\mathbb{Z}[\omega])$, where ω is a primitive cube root of unity.

3.2. A hyperbolic manifold with geodesic boundary.

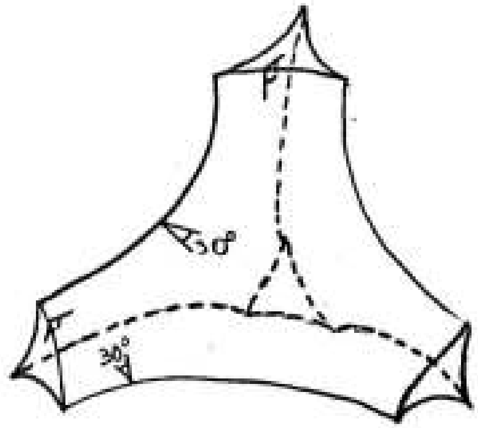
Here is another manifold which is obtained from two tetrahedra. First glue the two tetrahedra along one face; then glue the remaining faces according to this diagram:



3.8

In the diagram, one vertex has been removed so that the polyhedron can be flattened out in the plane. The resulting complex has only one edge and one vertex. The manifold M obtained by removing a neighborhood of the vertex is oriented with boundary a surface of genus 2.

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 go 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 *stunted simplex* with all angles 90° or 30°:

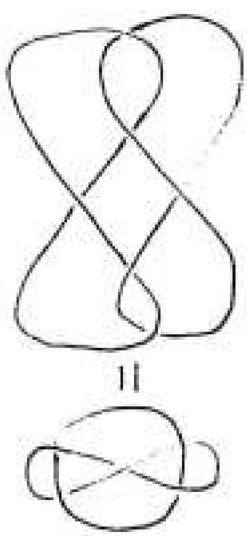


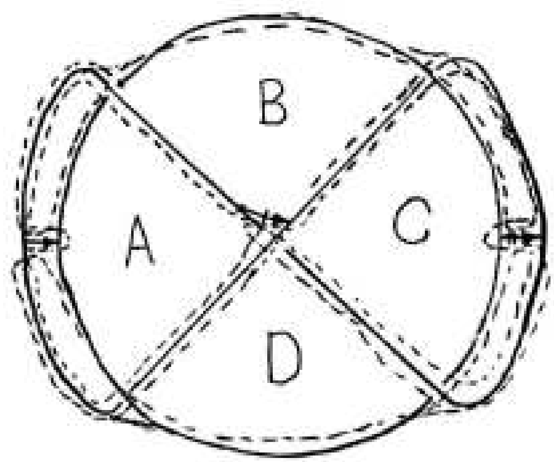
3.9

Two copies glued together give a hyperbolic structure for M, where the boundary of M (which comes from the triangular faces of the stunted simplices) is totally geodesic. 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.

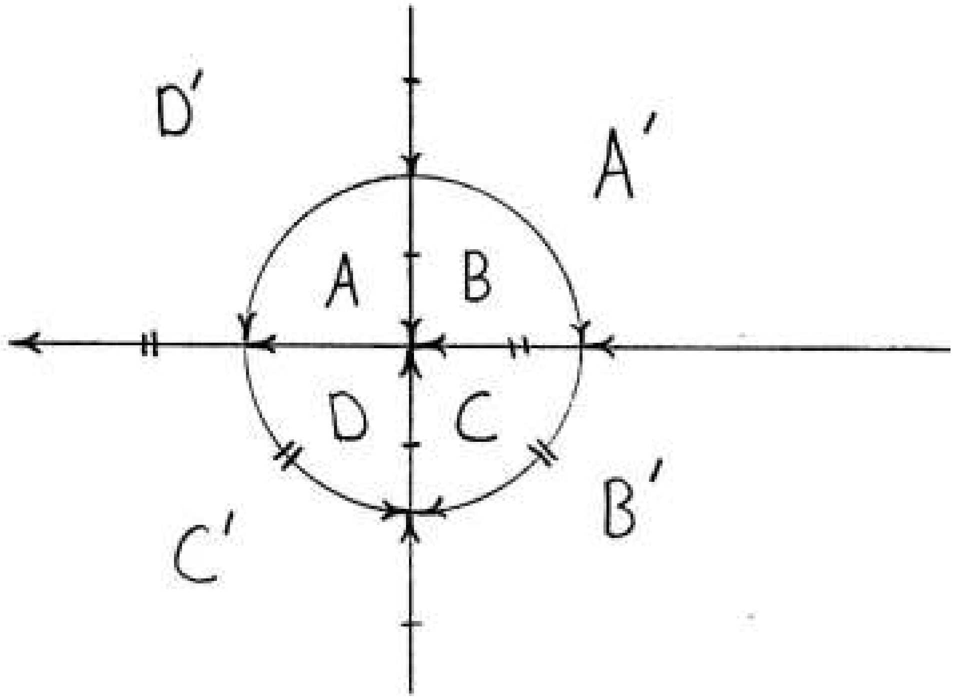
3.3. The Whitehead link complement.

The Whitehead link may be spanned by a two-complex which cuts the complement into an octahedron, with vertices deleted:

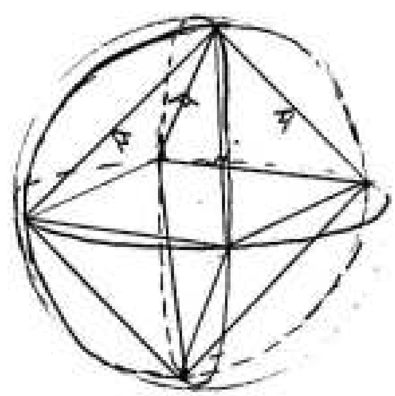




The one-cells are the three arrows, and the attaching maps for the two-cells are indicated by the dotted lines. The three-cell is an octahedron (with vertices deleted), 3.10
and the faces are identified thus:



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°.

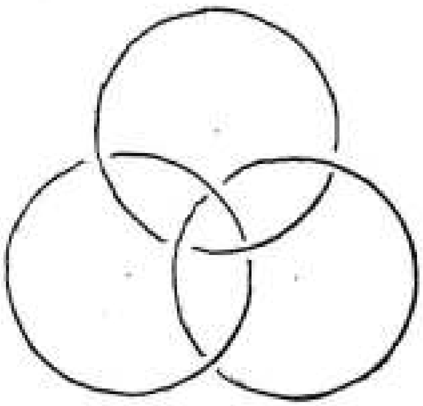


3.11

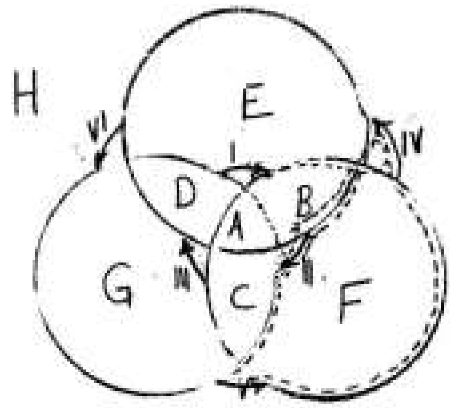
Gluing it in the indicated pattern, again using Euclidean isometries between the faces (which happen to be hyperbolic isometries as well) gives a hyperbolic structure for the complement of the Whitehead link.

3.4. The Borromean rings complement.

This is spanned by a two-complex which cuts the complement into two ideal octahedra:

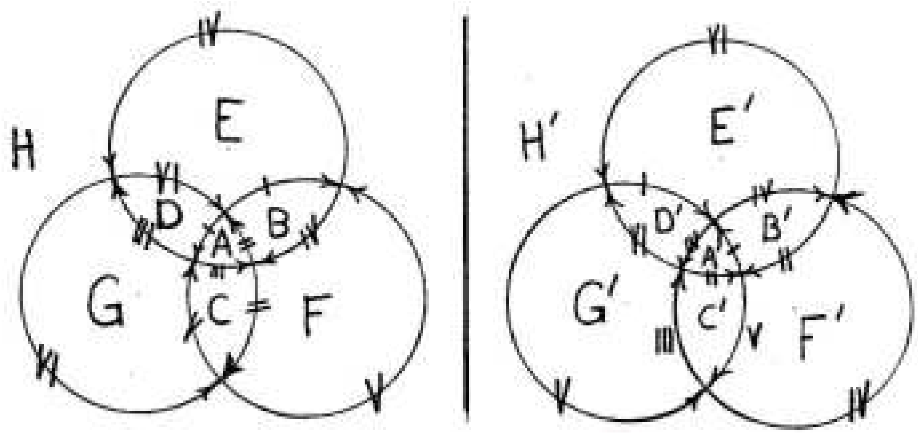


Borromean rings



A spanning 2-complex

Here is the corresponding gluing pattern of two octahedra. Faces are glued to their corresponding faces with 120° rotations, alternating in directions like gears.



3.12

3.5. The developing map.

Let X be any real analytic manifold, and G a group of real analytic diffeomorphisms of X. Then an element of G is completely determined by its restriction to any open set of X.

Suppose that M is any (G, X)-manifold. Let $U_1, U_2, \dots$ be coordinate charts for M, with maps $\phi_i : U_i \to X$ 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.

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.* It follows immediately 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 quotient space of the paths in M.) This map,

$$D : \tilde{M} \to X,$$

is called the *developing map*. D is a local (G, X)-homeomorphism (i.e., it is an immersion inducing the (G, X)-structure on $\tilde{M}$.) D is clearly unique up to composition 3.13
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 this uniqueness property of D, we have in particular that for any covering transformation T_α of $\tilde{M}$ 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

$$H : \alpha \mapsto g_\alpha$$

is a homomorphism, called the *holonomy* of M.

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. M is a *complete* (G, X)-manifold if $D : \tilde{M} \to X$ is a covering map. (In particular, if X is simply-connected, this means D is a homeomorphism.)

If X is similarly connected, then any complete (G, X)-manifold M may easily be reconstructed from the image $\Gamma = H(\pi_1(M))$ of the holonomy, as the quotient space X/Γ.

Here is a useful sufficient condition for completeness.

Proposition 3.6. *Let G be a group of analytic diffeomorphisms acting 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.*

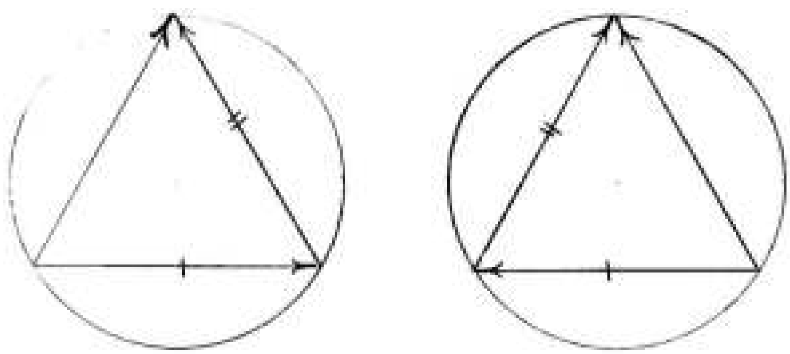
Proof. Let Q be any positive definite on the tangent space $T_x(X)$ of X at some point x. Average the set of transforms $g(Q), g \in G_x$, using Haar measure, to obtain a quadratic form on $T_x(X)$ which is invariant under G_x. Define a Riemannian metric 3.14
$(ds^2)_y = g(Q)$ on X, where $g \in G$ is any element taking x to y. This definition is independent of the choice of g, and the resulting Riemannian metric is invariant under G.

Therefore, this metric pieces together to give a Riemannian metric on any (G, X)-manifold, which is invariant under any (G, X)- map.

If M is any closed (G, X)-manifold, then there is some $\epsilon > 0$ such that the ϵ-ball in the Riemannian metric on M is always convex and contractible. If x is any point in X, then $D^{-1}(B_{\epsilon/2}(x))$ must be a union of homeomorphic copies of $B_{\epsilon/2}(x)$ in $\tilde{M}$. D evenly covers X, so it is a covering projection, and M is complete. $\square$

For example, any closed elliptic three-manifold has universal cover S^3, so any simply-connected elliptic manifold is S^3. Every closed hyperbolic manifold or Euclidean manifold has universal cover hyperbolic three-space or Euclidean space. Such manifolds are consequently determined by their holonomy. 3.15

Even for G and X as in proposition 3.6, the question of whether or not a non- 3.16
compact (G, X)-manifold M is complete can be much more subtle. For example, consider the thrice-punctured sphere, which is obtained by gluing together two triangles minus vertices in this pattern:



A hyperbolic structure can be obtained by gluing two ideal triangles (with all vertices on the circle at infinity) in this pattern. Each side of such a triangle is isometric to the real line, so a gluing map between two sides may be modified by an arbitrary translation; thus, we have a family of hyperbolic structures in the thrice-punctured sphere parametrized by $\mathbb{R}^3$. (These structures need not be, and are not, all distinct.) *Exactly one parameter value yields a complete hyperbolic structure*, as we shall see presently.

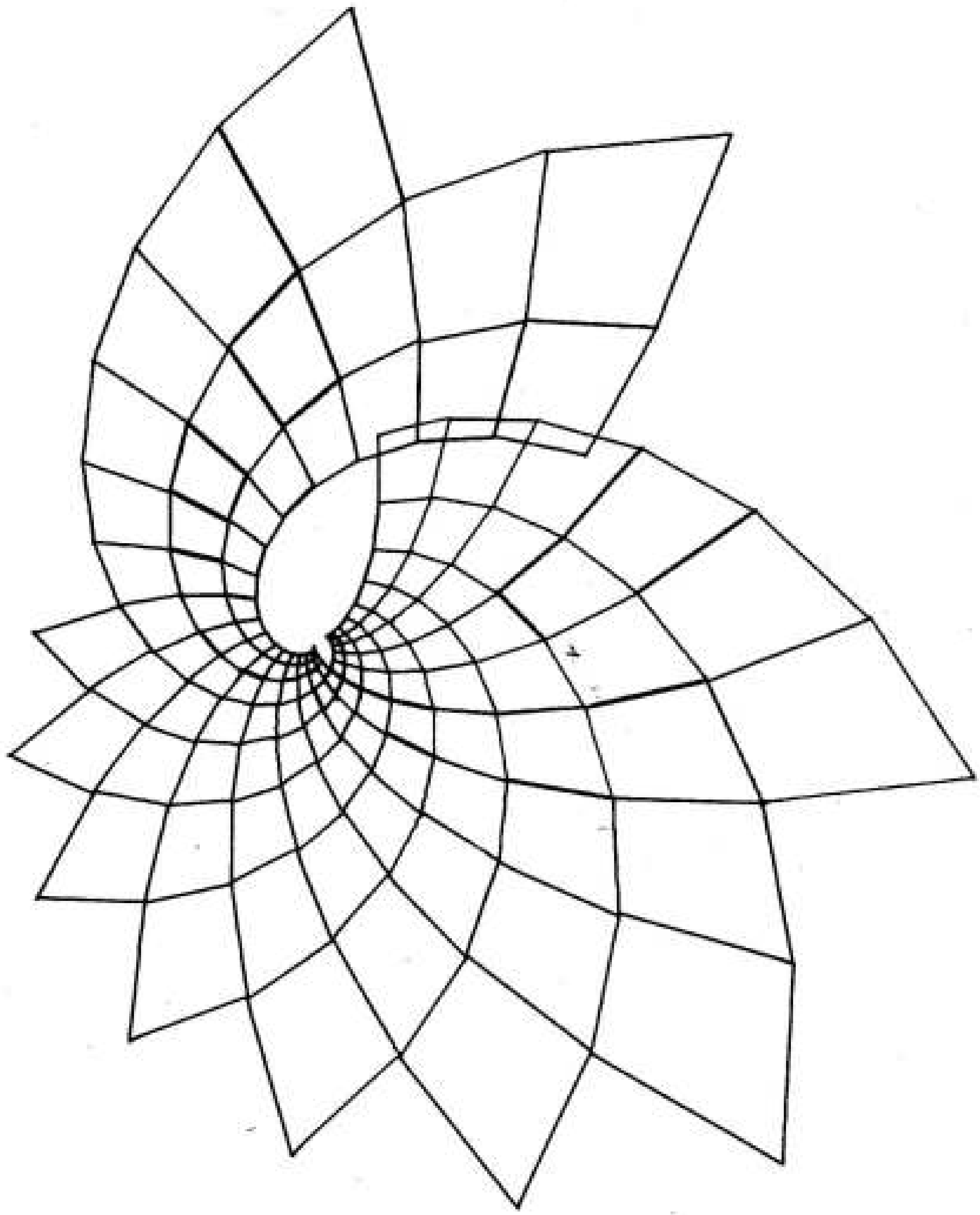
Meanwhile, we collect some useful conditions for completeness of a (G, X)-structure with (G, X) as in 3.6. For convenience, we fix some natural metrics on (G, X)- 3.17
structures.

Labelled 3.7.p to avoid multiple labels

Proposition 3.7. *With (G, X) as above, a (G, X)-manifold M is complete if and only if any of the following equivalent conditions is satisfied.*

(a) *M is complete as a metric space.*
(b) *There is some $\epsilon > 0$ such that each closed ϵ-ball in M is compact.*

3.15



The developing map of an affine torus constructed from a quadrilateral (see p. 3.3). The torus is plainly not complete. *Exercise*: construct other affine toruses with the same holonomy as this one. (Hint: walk once or twice around this page.)

(c) *For every $k > 0$, all closed k-balls are compact.*

(d) *There is a family $\{S_t\}; t \in \mathbb{R}$, of compact sets which exhaust M, such that S_{t+a} contains a neighborhood of radius a about S_t.*

PROOF. Suppose that M is metrically complete. Then $\tilde{M}$ is also metrically complete. We will show that the developing map $D : \tilde{M} \to X$ is a covering map by proving that any path α_t in X can be lifted to $\tilde{M}$. In fact, let $T \subset [0, 1]$ be a maximal connected set for which there is a lifting. Since D is a local homeomorphism, T is open, and because $\tilde{M}$ is metrically complete, T is closed: hence, α can be lifted, so M is complete.

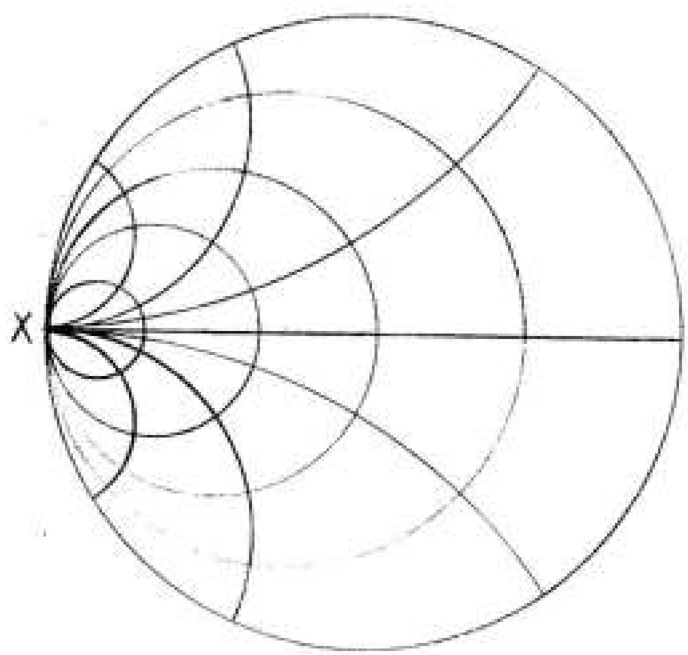
It is an elementary exercise to see that (b) $\iff$ (c) $\iff$ (d) $\implies$ (a). For any point $x_0 \in \tilde{X}$ there is some ϵ such that the ball $B_\epsilon(x)$ is compact; this ϵ works for all $x \in \tilde{X}$ since the group $\tilde{G}$ of (G, X)-diffeomorphisms of $\tilde{X}$ is transitive. Therefore X satisfies (a), (b), (c) and (d). Finally if M is a complete (G, X)-manifold, it is covered by $\tilde{X}$, so it satisfies (b). The proposition follows. □

3.18

3.8. Horospheres.

To analyze what happens near the vertices of an ideal polyhedron when it is glued together, we need the notion of horospheres (or, in the hyperbolic plane, they are called horocycles.) A horosphere has the limiting shape of a sphere in hyperbolic space, as the radius goes to infinity. One property which can be used to determine the spheres centered at a point X is the fact that such a sphere is orthogonal to all lines through X. Similarly, if X is a point on the sphere at infinity, the *horospheres* "centered" at X are the surfaces orthogonal to all lines through X. In the Poincaré disk model, a hyperbolic sphere is a Euclidean sphere in the interior of the disk, and a horosphere is a Euclidean sphere tangent to the unit sphere. The point X of tangency is the center of the horosphere.

Concentric horocycles and orthogonal lines.



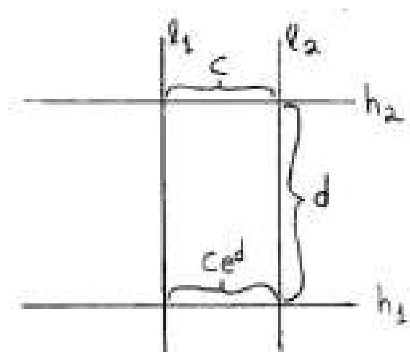
3.19

Translation along a line through X permutes the horospheres centered at X. Thus, *all horospheres are congruent.* The convex region bounded by a horosphere is a *horoball.* For another view of a horosphere, consider the upper half-space model. In this case, hyperbolic lines through the point at infinity are Euclidean lines orthogonal to the plane bounding upper half-space. A horosphere about this point is a horizontal Euclidean plane. From this picture one easily sees that a horosphere in H^n is isometric to Euclidean space E^{n-1}. One also sees that the group of hyperbolic isometries fixing the point at infinity in the upper half-space model acts as the group of similarities of the bounding Euclidean plane. One can see this action internally as follows. Let X be any point at infinity in hyperbolic space, and h any horosphere centered at X. An isometry g of hyperbolic space fixing X takes h to a concentric horosphere h'. Project h' back to h along the family of parallel lines through X. The composition of these two maps is a similarity of h.

Consider two directed lines l_1 and l_2 emanating from the point at infinity in the upper half-space model. Recall that the hyperbolic metric is $ds^2 = (1/x_n^2)\,dx^2$. This means that the hyperbolic distance between l_1 and l_2 along a horosphere is inversely proportional to the Euclidean distance above the bounding plane. The hyperbolic distance between points X_1 and X_2 on l_1 at heights of h_1 and h_2 is $|\log(h_2) - \log(h_1)|$. It follows that for any two concentric horospheres h_1 and h_2 which are a distance d apart, and any pair of lines l_1 and l_2 orthogonal to h_1 and h_2, the ratio of the distance 3.20

FIGURE 1. Horocycles and lines in the upper half-plane

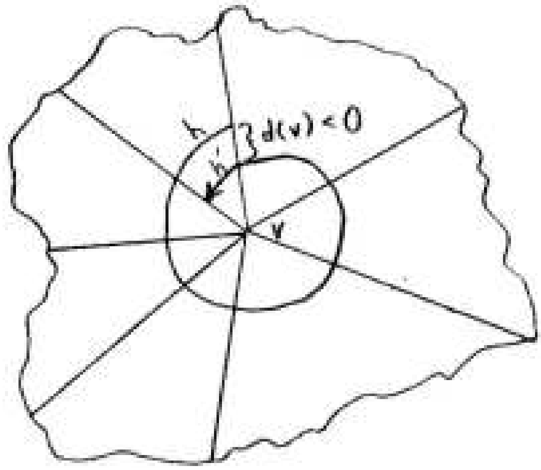
between l_1 and l_2 measured along h_1 to their distance measured along h_2 is $\exp(d)$.



3.9. Hyperbolic surfaces obtained from ideal triangles.

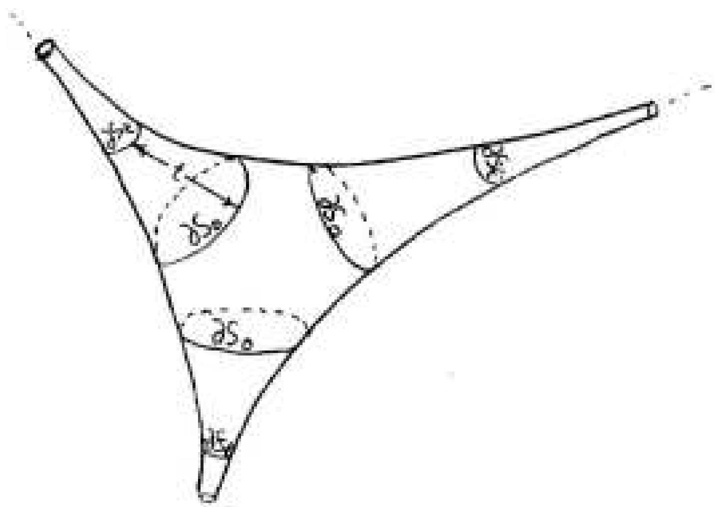
Consider an oriented surface S obtained by gluing ideal triangles with all vertices at infinity, in some pattern. *Exercise: all such triangles are congruent.* (Hint: you can derive this from the fact that a finite triangle is determined by its angles—see 2.6.8. Let the vertices pass to infinity, one at a time.)

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 counter clockwise 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. 3.21



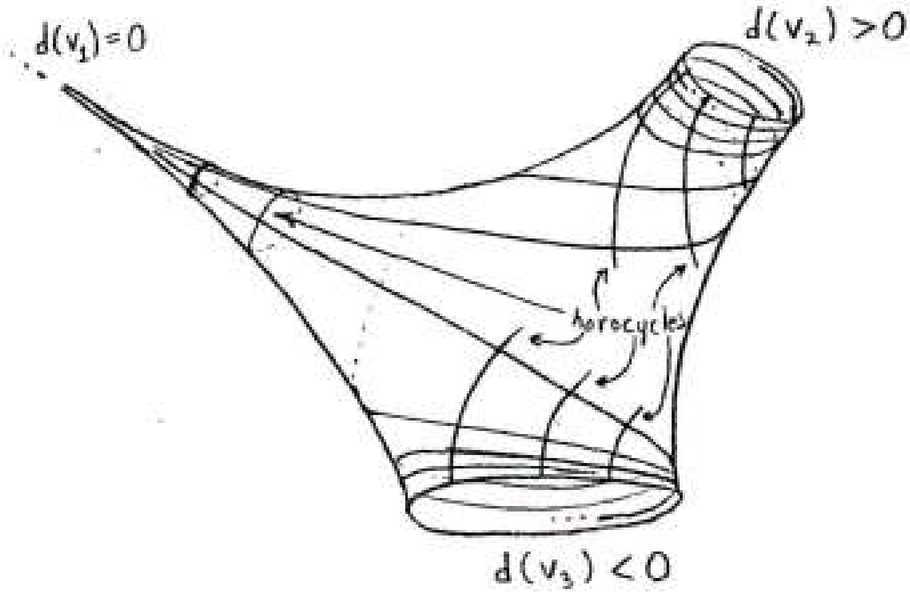
The surface S is complete if and only if all invariants $d(v)$ are 0. Suppose, for instance, that some invariant $d(v) < 0$. Continuing h further round v; the length of each successive circuit around v is reduced by a constant factor < 1, so the total length of h after an infinite number of circuits is bounded. A sequence of points evenly spaced along h is a non-convergent Cauchy sequence.

If all invariants $d(v) = 0$, on the other hand, one can remove horoball neighborhoods of each vertex in K 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 ones. The surfaces S_t satisfy the hypotheses of 3.7(d) 1—hence S is complete.



3.22

For any hyperbolic manifold M, let $\bar{M}$ be the metric completion of M. In general, $\bar{M}$ need not be a manifold. However, if S is a surface obtained by gluing ideal hyperbolic triangles, then $\bar{S}$ is a hyperbolic surface with geodesic boundary. There is one boundary component of length $|d(v)|$ for each vertex v of K such that $d(v) \neq 0$. $\bar{S}$ is obtained by adjoining one limit point for each horocycle which "spirals toward" a vertex v in K. The most convincing way to understand $\bar{S}$ is by studying the picture:



3.10. Hyperbolic manifolds obtained by gluing ideal polyhedra.

Consider now the more general case of a hyperbolic manifold M obtained by gluing together the faces of polyhedra in H^n with some vertices at infinity. Let K 3.23
be the complex obtained by including the ideal vertices. The link of an ideal vertex v is (by definition) the set $L(v)$ of all rays through that vertex. From 3.7 it follows that the link of each vertex has a canonical (similarities of E^{n-1}, E^{n-1}) structure, or similarity structure for short. An extension of the analysis in 3.9 easily shows that *M is complete if and only if the similarity structure on each link of an ideal vertex is actually a Euclidean structure*, or equivalently, if and only if the holonomy of these similarity structures consists of isometries. We shall be concerned mainly with dimension $n = 3$. It is easy to see from the Gauss-Bonnet theorem that any similarity two-manifold has Euler characteristic zero. (Its tangent bundle has a flat orthogonal connection). Hence, if M is oriented, each link $L(v)$ of an ideal vertex is topologically a torus. If $L(v)$ is not Euclidean, then for some $\alpha \in \pi_1 L(v)$, the holonomy $H(\alpha)$ is a contraction, so it has a unique fixed point x_0. Any other element $\beta \in \pi_1(L(v))$ must also fix x_0, since β commutes with α. Translating x_0 to 0, we see that the similarity two-manifold $L(v)$ must be a $(\mathbb{C}^*, \mathbb{C} - 0)$-manifold where $\mathbb{C}^*$ is the multiplicative group of complex numbers. (Compare p. 3.15.) Such a structure

is automatically complete (by 3.6), and it is also modelled on

$$(\tilde{\mathbb{C}}^*, \widetilde{\mathbb{C} - 0}),$$

or, by taking logs, on $(\mathbb{C}, \mathbb{C})$. Here the first $\mathbb{C}$ is an additive group and the second $\mathbb{C}$ is a space. Conversely, by taking exp, any $(\mathbb{C}, \mathbb{C})$ structure gives a similarity structure. $(\mathbb{C}, \mathbb{C})$ structures on closed oriented manifolds are easy to describe, being determined by their holonomy, which is generated by an arbitrary pair (z_1, z_2) of 3.24
complex numbers which are linearly independent over $\mathbb{R}$.

We shall return later to study the spaces $\bar{M}$ in the three-dimensional case. They are sometimes *closed* hyperbolic manifolds obtained topologically by replacing neighborhoods of the vertices by solid toruses.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002
http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 4

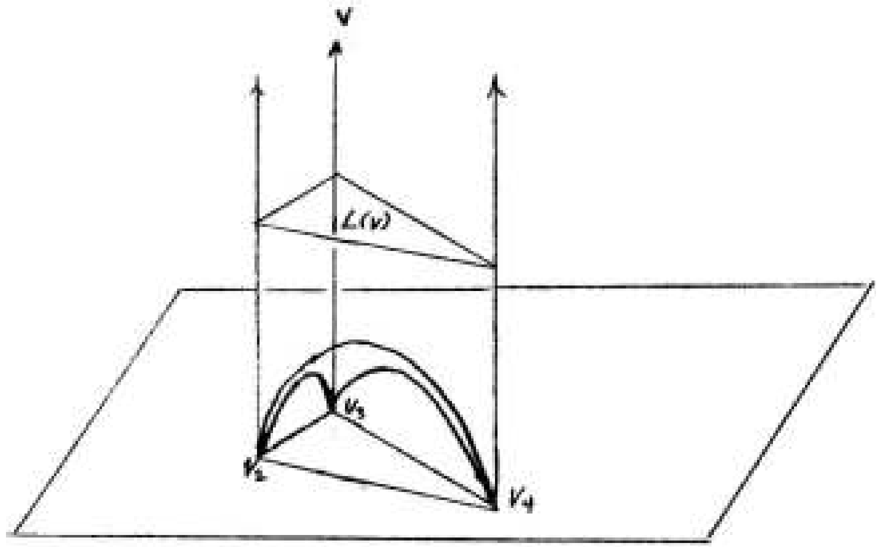
Hyperbolic Dehn surgery

A hyperbolic structure for the complement of the figure-eight knot was constructed in 3.1. This structure was in fact chosen to be complete. The reader may wish to verify this by constructing a horospherical realization of the torus which is the link of the ideal vertex. Similarly, the hyperbolic structure for the Whitehead link complement and the Borromean rings complement constructed in 3.3 and 3.4 are complete.

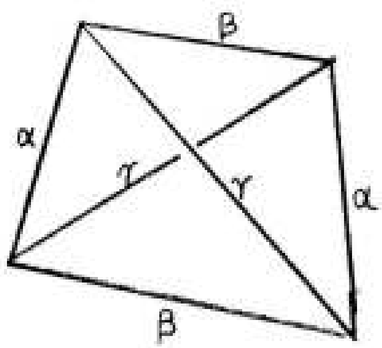
There is actually a good deal of freedom in the construction of hyperbolic structures for such manifolds, although most of the resulting structures are not complete. We shall first analyze the figure-eight knot complement. To do this, we need an understanding of the possible shapes of ideal tetrahedra.

4.1. Ideal tetrahedra in H^3.

The link $L(v)$ of an ideal vertex v of an oriented ideal tetrahedron T (which by definition is the set of rays in the tetrahedron through that vertex) is a Euclidean triangle, well-defined up to orientation-preserving similarity. It is concretely realized as the intersection with T of a horosphere about v. The triangle $L(v)$ actually determines T up to congruence. To see this, picture T in the upper half-space model and arrange it so that v is the point at infinity. The other three vertices of T form a triangle in E^2 which is in the same similarity class as $L(v)$. Consequently, if two 4.2
tetrahedra T and T' have vertices v and v' with $L(v)$ similar to $L(v')$, then T' can be transformed to T by a Euclidean similarity which preserves the plane bounding upper half-space. Such a similarity is a hyperbolic isometry.

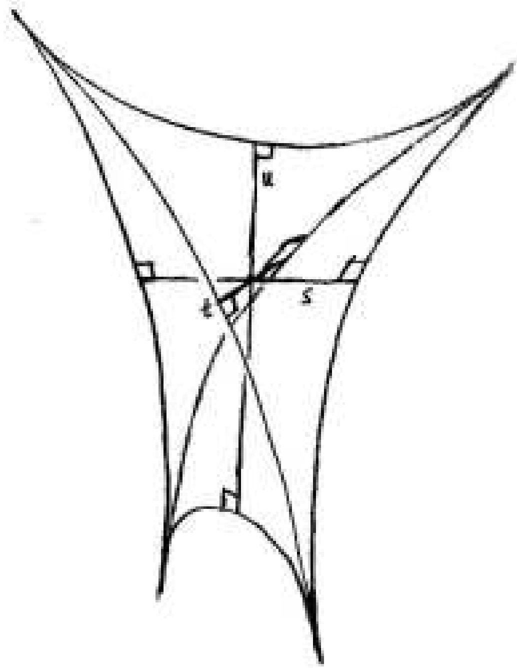


It follows that T is determined by the three dihedral angles α, β and γ of edges incident to the ideal vertex v, and that $\alpha + \beta + \gamma = \pi$. Using similar relations among angles coming from the other three vertices, we can determine the other three dihedral angles:



4.3

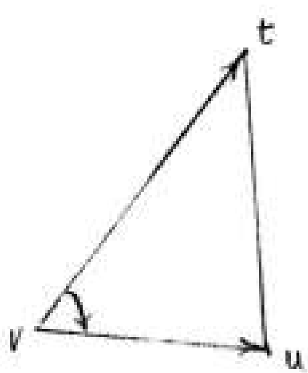
Thus, dihedral angles of opposite edges are equal, and the oriented similarity class of $L(v)$ does not depend on the choice of a vertex v! A geometric explanation of this phenomenon can be given as follows. Any two non-intersecting and non-parallel lines in H^3 admit a unique common perpendicular. Construct the three common perpendiculars s, t and u to pairs of opposite edges of T. Rotation of π about s, for instance, preserves the edges orthogonal to s, hence preserves the four ideal vertices of T, so it preserves the entire figure. It follows that s, t and u meet in a point and that they are pairwise orthogonal. The rotations of π about these three axes are the three non-trivial elements of $z_2 \oplus z_2$ acting as a group of symmetries of T.

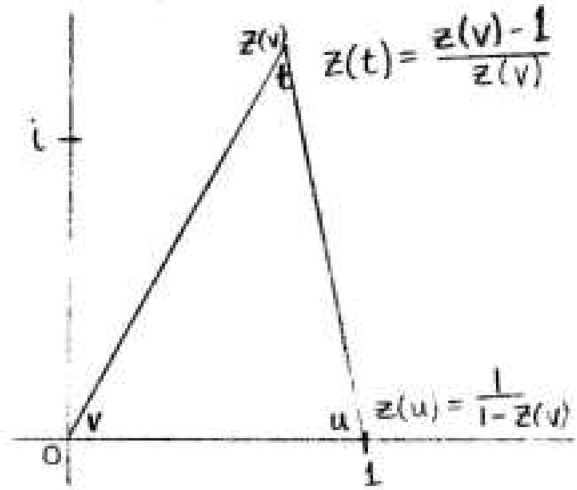


In order to parametrize Euclidean triangles up to similarity, it is convenient to regard E^2 as $\mathbb{C}$. To each vertex v of a triangle $\Delta(t, u, v)$ we associate the ratio 4.4

$$\frac{(t-v)}{(u-v)} = z(v)$$

of the sides adjacent to v. The vertices must be labelled in a clockwise order, so that





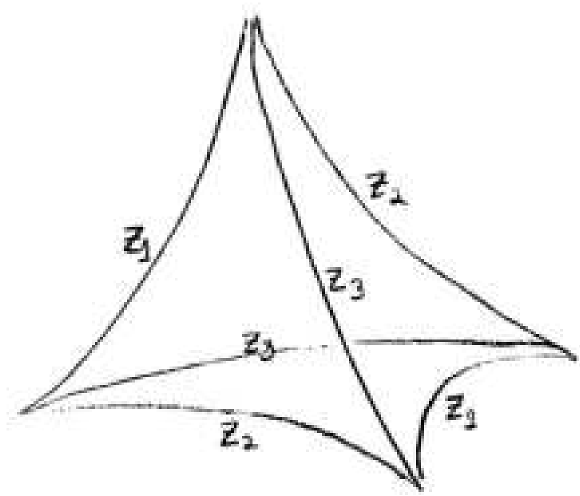
$\operatorname{Im}\big(z(v)\big) > 0$. Alternatively, arrange the triangle by a similarity so that v is at 0 and u at 1; then t is at $z(v)$. The other two vertices have invariants

4.1.1.
$$\begin{aligned} z(t) &= \tfrac{z(v)-1}{z(v)} \\ z(u) &= \tfrac{1}{1-z(v)}. \end{aligned}$$

Denoting the three invariants z_1, z_2, z_3 in clockwise order, with any starting point, we have the identities 4.5

4.1.2.
$$\begin{gathered} z_1\, z_2\, z_3 = -1 \\ 1 - z_1 + z_1 z_2 = 0 \end{gathered}$$

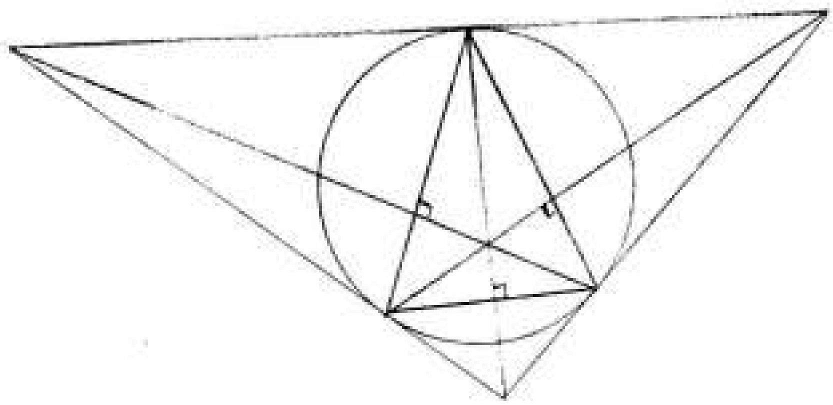
We can now translate this to a parametrization of ideal tetrahedra. Each edge e is labelled with a complex number $z(e)$, opposite edges have the same label, and the three distinct invariants satisfy 4.1.2 (provided the ordering is correct.) Any z_i determines the other two, via 4.1.2.



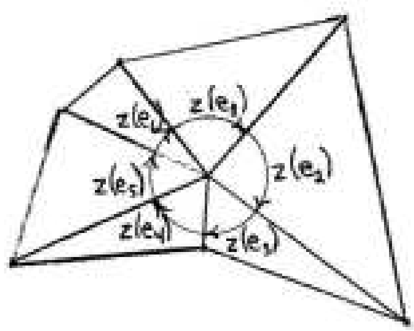
4.2. Gluing consistency conditions.

Suppose that M is a three-manifold obtained by gluing tetrahedra $T_i, \dots, T_j$ and then deleting the vertices, and let K be the complex which includes the vertices. 4.6

Any realization of $T_1, \dots, T_j$ as ideal hyperbolic tetrahedra determines a hyperbolic structure on $(M - (1 - \text{skeleton}))$, since any two ideal triangles are congruent. Such a congruence is uniquely determined by the correspondence between the vertices. (This fact may be visualized concretely from the subdivision of an ideal triangle by its altitudes.)



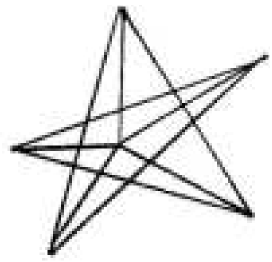
The condition for the hyperbolic structure on $(M - (1 - \text{skeleton}))$ to give a hyperbolic structure on M itself is that its developing map, in a neighborhood of each edge, should come from a local homeomorphism of M itself. In particular, the sum of the dihedral angles of the edges $e_1, \ldots, e_k$ must be 2π. Even when this condition is satisfied, though, the holonomy going around an edge of M might be a non-trivial translation along the edge. To pin down the precise condition, note that for each ideal vertex v, the hyperbolic structure on $M - (1 - \text{skeleton})$ gives a similarity structure to $L(v) - (0 - \text{skeleton})$. The hyperbolic structure extends over an edge e of M if and 4.7
only if the similarity structure extends over the corresponding point in $L(v)$, where v is an endpoint of e. Equivalently, the similarity classes of triangles determined by $z(e_1), \ldots, z(e_k)$ must have representatives which can be arranged neatly around a point in the plane:



The algebraic condition is

4.2.1. $$z(e_1) \cdot z(e_2) \cdot \cdots \cdot z(e_k) = 1.$$

This equation should actually be interpreted to be an equation in the universal cover $\tilde{\mathbb{C}}^*$, so that solutions such as 4.8



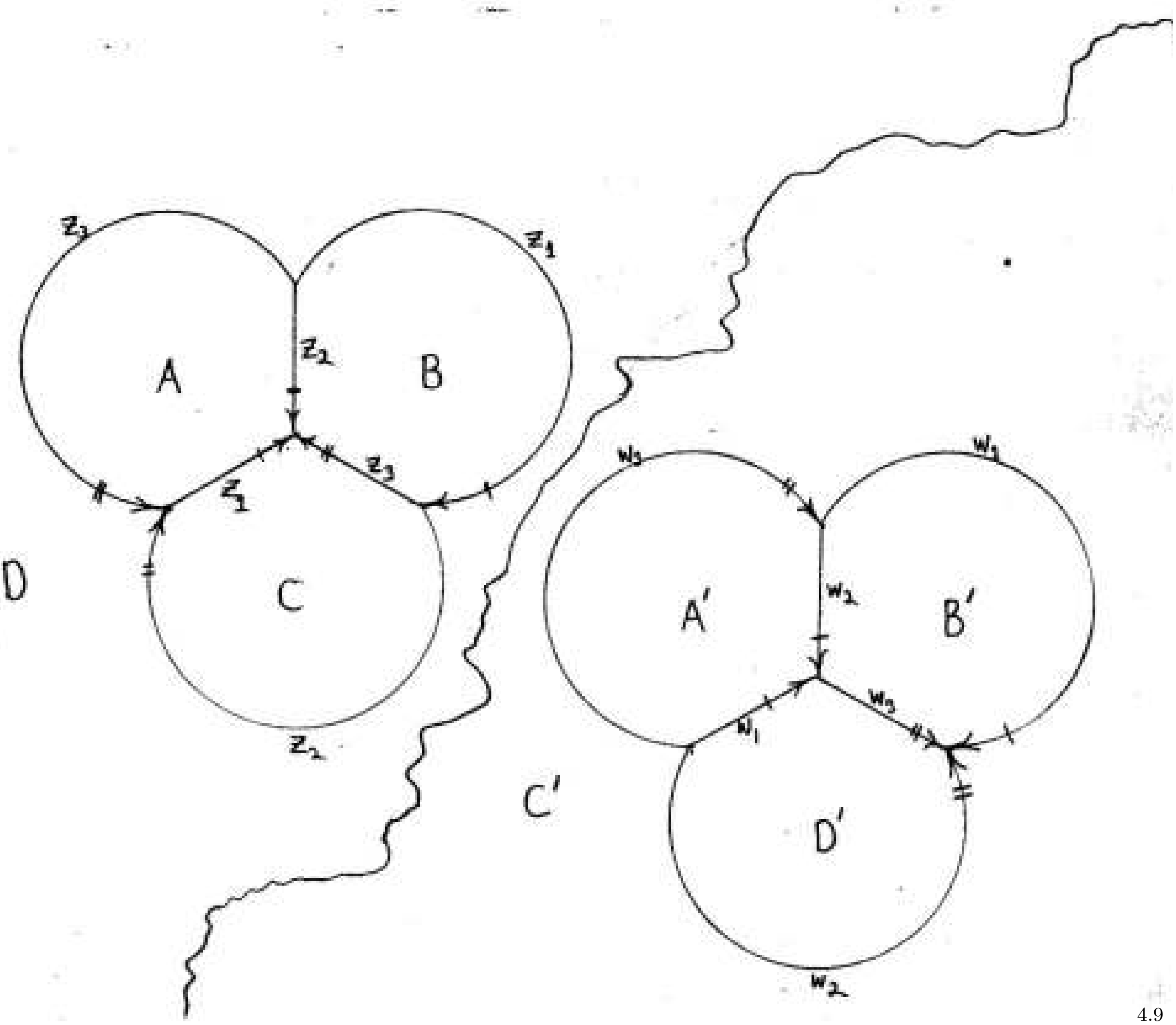
are ruled out. In other words, the auxiliary condition

4.2.2. $$\arg z_1 + \cdots + \arg z_k = 2\pi$$

must also be satisfied, where $0 < \arg z_i \leq \pi$.

4.3. Hyperbolic structure on the figure-eight knot complement.

Consider two hyperbolic tetrahedra to be identified to give the figure eight knot complement:



4.9

We read off the gluing consistency conditions for the two edges:

$$(\not\longrightarrow) z_1^2 z_2 w_1^2 w_2 = 1 \qquad (/\!\!/\!\!\longrightarrow) z_3^2 z_2 w_3^2 w_2 = 1.$$

From 4.1.2, note that the product of these two equations,

$$(z_1 z_2 z_3)^2 (w_1 w_2 w_3)^2 = 1$$

is automatically satisfied. Writing $z = z_1$, and $w = w_1$, and substituting the expressions from 4.1.1 into $(\not\longrightarrow)$, we obtain the equivalent gluing condition,

4.3.1. $$z(z-1)\,w(w-1) = 1.$$

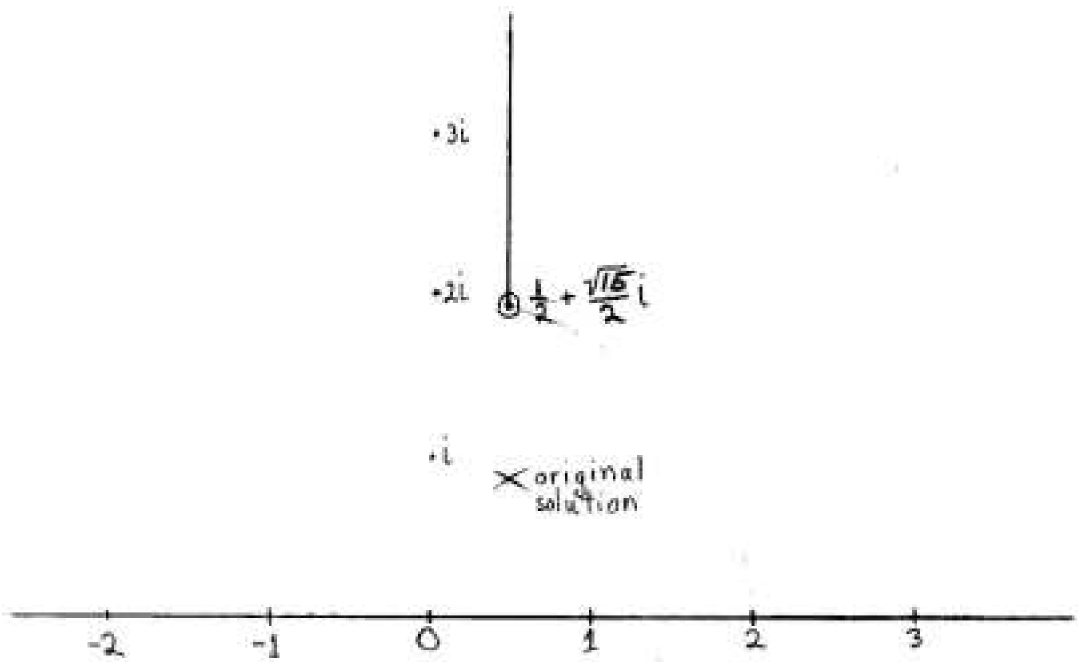
We may solve for z in terms of w by using the quadratic formula.

4.3.2.
$$z = \frac{1 \pm \sqrt{1 + 4/w(w-1)}}{2}$$

We are searching only for geometric solutions

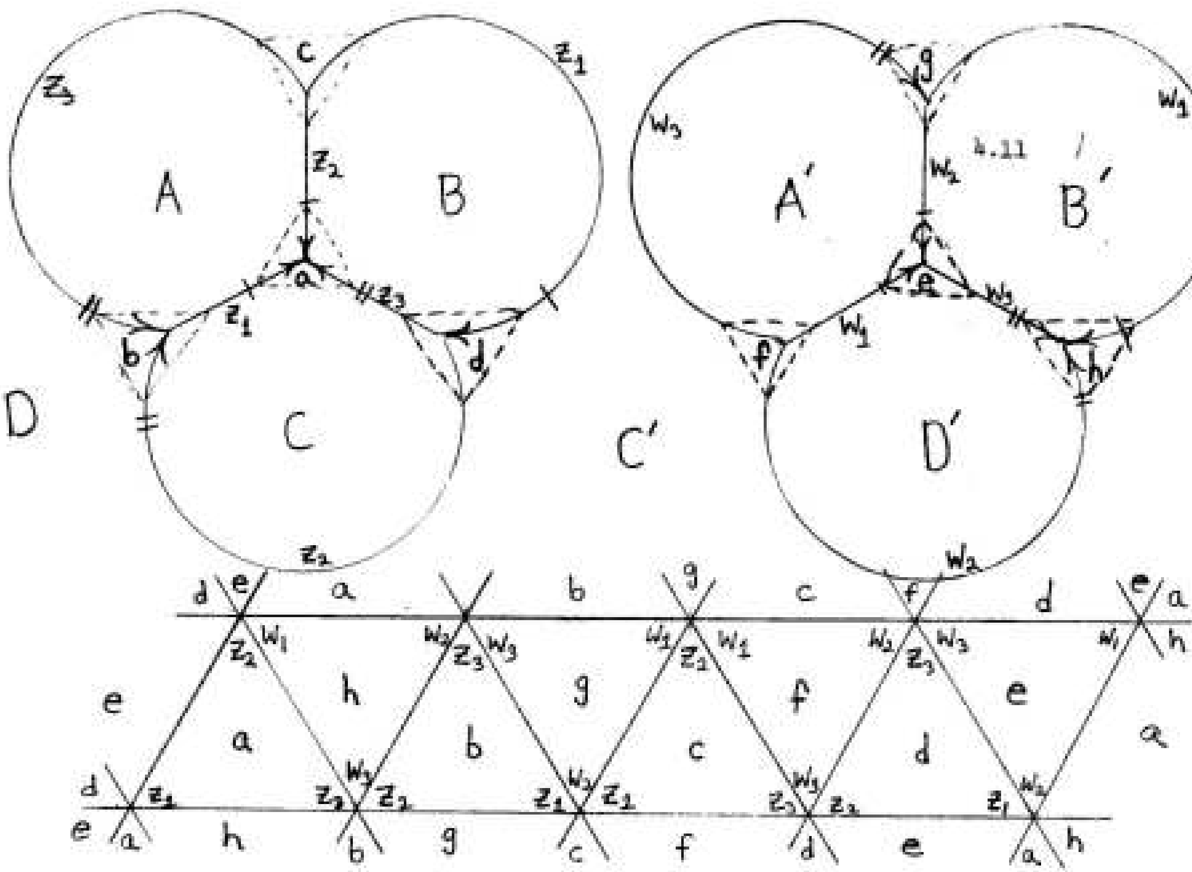
$$\mathrm{Im}(z) > 0 \quad \mathrm{Im}(w) > 0$$

so that the two tetrahedra are non-degenerate and positively oriented. For each value of w, there is at most one solution for z with $\mathrm{Im}(z) > 0$. Such a solution exists provided that the discriminant $1 + 4/w(w-1)$ is not positive real. Solutions are therefore parametrized by w in this region of $\mathbb{C}$: 4.10

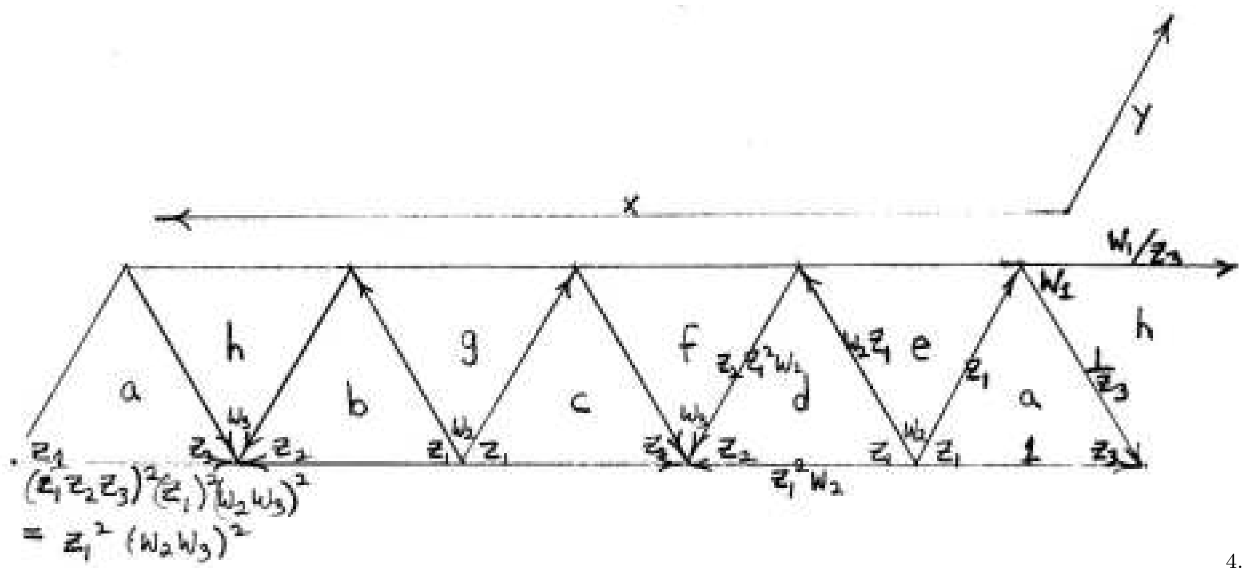


Note that the original solution given in 3.1 corresponds to $w = z = \sqrt[3]{-1} = \frac{1}{2} + \frac{\sqrt{3}}{2}i$.

The link L of the single ideal vertex has a triangulation which can be calculated from the gluing diagram: 4.11



Now let us compute the derivative of the holonomy of the similarity structure on L. To do this, regard directed edges of the triangulation as vectors. The ratio of any two vectors in the same triangle is known in terms of z or w. Multiplying appropriate ratios, we obtain the derivative of the holonomy:



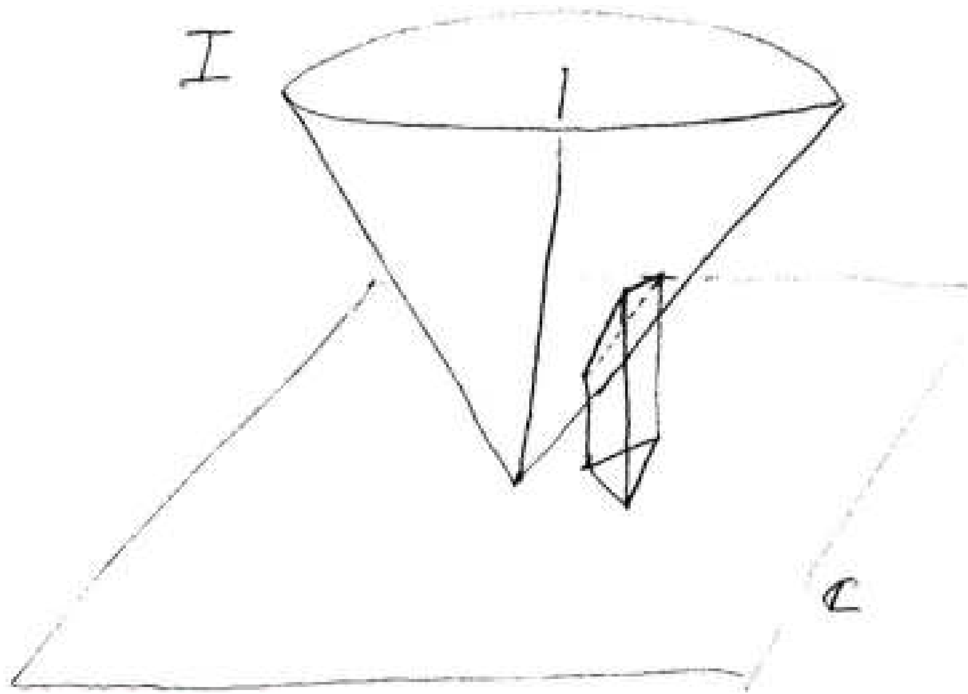
4.12

$$
\begin{aligned}
H'(x) &= z_1^2(w_2w_3)^2 = (\tfrac{z}{w})^2\\
H'(y) &= \tfrac{w_1}{z_3} = w(1-z).
\end{aligned}
$$

Observe that if M is to be complete, then $H'(x) = H'(y) = 1$, so $z = w$. From 4.3.1, $(z(z-1))^2 = 1$. Since $z(z-1) < 0$, this means $z(z-1) = -1$, so that the only possibility is the original solution $w = z = \sqrt[3]{-1}$.

4.4. The completion of hyperbolic three-manifolds obtained from ideal polyhedra.

Let M be any hyperbolic manifold obtained by gluing polyhedra with some vertices at infinity, and let K be the complex obtained by including the ideal vertices. The completion $\bar{M}$ is obtained by completing a deleted neighborhood $\mathcal{N}(v)$ of each ideal vertex v in k, and gluing these completed neighborhoods $\bar{\mathcal{N}}(v)$ to M. The developing map for the hyperbolic structure on $\mathcal{N}(v)$ may be readily understood in terms of the developing map for the similarity structure on $L(v)$. To do this, choose coordinates so that v is the point at infinity in the upper half-space model. The developing images of corners of polyhedra near v are "chimneys" above some polygon in the developing image of $L(v)$ on $\mathbb{C}$ (where $\mathbb{C}$ is regarded as the boundary of upper half-space.) If M is not complete near v, we change coordinates if necessary by a translation of $\mathbb{R}^3$ so that the developing image of $L(v)$ is $\mathbb{C} - 0$. The holonomy for $\mathcal{N}(v)$ now consists of similarities of $\mathbb{R}^3$ which leave invariant the z-axis and the $x - y$ plane ($\mathbb{C}$). Replacing $\mathcal{N}(v)$ by a smaller neighborhood, we may assume that the developing image I of $\mathcal{N}(v)$ is a solid cone, minus the z-axis. 4.13



The completion of I is clearly the solid cone, obtained by adjoining the z-axis to I. It follows easily that the completion of

$$\widetilde{\mathcal{N}(v)} = \tilde{I}$$

is also obtained by adjoining a single copy of the z-axis.

The projection

$$p : \widetilde{\mathcal{N}(v)} \to \mathcal{N}(v)$$

extends to a surjective map $\bar{p}$ between the completions. [$\bar{p}$ exists because p does not increase distances. $\bar{p}$ is surjective because a Cauchy sequence can be replaced by a Cauchy path, which lifts to $\mathcal{N}\tilde{(v)}$.] Every orbit of the holonomy action of $\pi_1\big(\mathcal{N}(v)\big)$ on the z-axis is identified to a single point. This action is given by

$$H(\alpha) : (0, 0, z) \mapsto |H'(\alpha)| \cdot (0, 0, z)$$

where the first $H(\alpha)$ is the hyperbolic holonomy and the second is the holonomy of $L(v)$. There are two cases: 4.14

Case 1. The group of moduli $\{|H'(\alpha)|\}$ is dense in $\mathbb{R}_+$. Then the completion of $\mathcal{N}(v)$ is the one-point compactification.

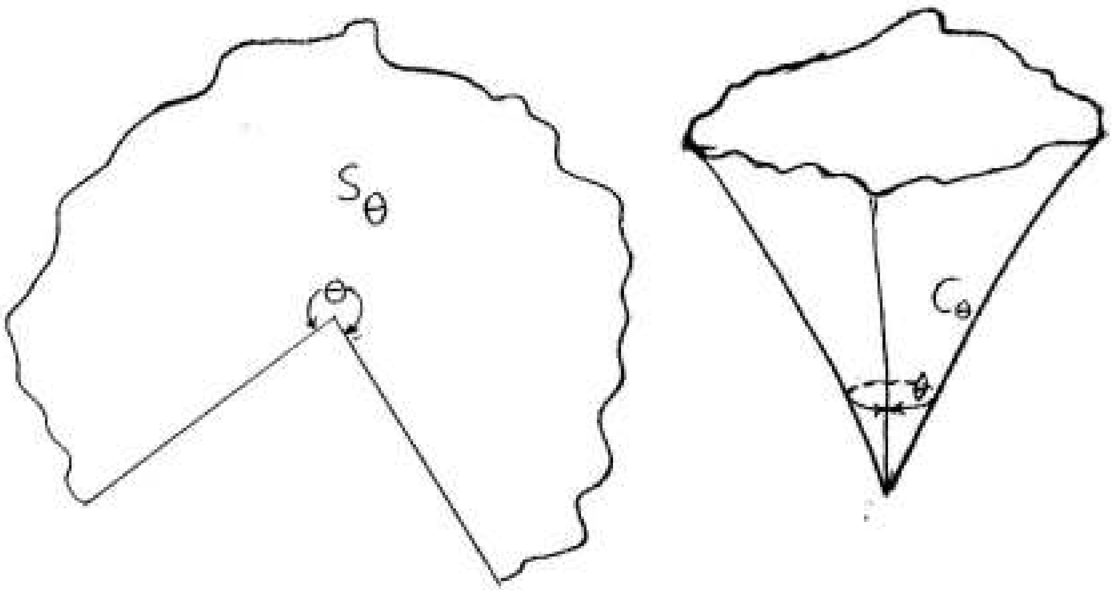
Case 2. The group of moduli $\{|H'(\alpha)|\}$ is a cyclic group. Then the completion

$$\overline{\mathcal{N}(v)}$$

is *topologically* a manifold which is the quotient space $\widetilde{(\overline{\mathcal{N}})}/H$, and it is obtained by adjoining a circle to $\mathcal{N}(v)$. Let $\alpha_1 \in \pi_1(L(v))$ be a generator for the kernel of $\alpha \mapsto |H'(\alpha)|$ and let $1 < |H'(\alpha_2)|$ generate the image, so that α_1 and α_2 generate $\pi_1(L(v)) = \mathbb{Z} \oplus \mathbb{Z}$. Then the added circle in

$$\overline{\mathcal{N}(v)}$$

has length $\log |H'(\alpha_2)|$. A cross-section of $\bar{\mathcal{N}}(v)$ perpendicular to the added circle is a cone C_θ, obtained by taking a two-dimensional hyperbolic sector S_θ of angle θ, $[0 < \theta < \infty]$ and identifying the two bounding rays:



It is easy to make sense of this even when $\theta > 2\pi$. The cone angle θ is the argument 4.15
of the element $\tilde{H}'(\alpha_2) \in \tilde{\mathbb{C}}^*$. In the special case $\theta = 2\pi$, C_θ is non-singular, so

$$\overline{\mathcal{N}(v)}$$

is a hyperbolic manifold. $\overline{\mathcal{N}(v)}$ may be seen directly in this special case, as the solid cone $I \cup (z - \text{axis})$ modulo H.

4.5. The generalized Dehn surgery invariant.

Consider any three-manifold M which is the interior of a compact manifold $\hat{M}$ whose boundary components $P_1, \ldots, P_k$ are tori. For each i, choose generators a_i, b_i for $\pi_1(P_i)$. If M is identified with the complement of an open tubular neighborhood of a link L in S^3, there is a standard way to do this, so that a_i is a *meridian* (it bounds a disk in the solid torus around the corresponding component of L) and b_i is a *longitude* (it is homologous to zero in the complement of this solid torus in S^3). In this case we will call the generators m_i and l_i.

We will use the notation $M_{(\alpha_1,\beta_1),\dots,(\alpha_k,\beta_k)}$ to denote the manifold obtained by gluing solid tori to M so that a meridian in the i-th solid torus goes to $\alpha_i, a_i + \beta_i b_i$. If an ordered pair (α_i, β_i) is replaced by the symbol ∞, this means that nothing is glued in near the i-th torus. Thus, $M = M_{\infty,\dots,\infty}$.

These notions can be refined and extended in the case M has a hyperbolic struc- 4.16
ture whose completion $\bar{M}$ is of the type described in 4.4. (In other words, if M is not complete near P_i, the developing map for some deleted neighborhood $\mathcal{N}_i$ of P_i should be a covering of the deleted neighborhood I of radius r about a line in H^3.) The developing map D of $\mathcal{N}_i$ can be lifted to $\tilde{I}$, with holonomy $\tilde{H}$. The group of isometries of $\tilde{I}$ is $\mathbb{R} \oplus \mathbb{R}$, parametrized by (translation distance, angle of rotation); this parametrization is well-defined up to sign.

Definition 4.5.1. The generalized Dehn surgery invariants (α_i, β_i) for $\bar{M}$ are solutions to the equations

$$\alpha_i \tilde{H}(a_i) + \beta_i \tilde{H}(b_i) = (\text{rotation by } \pm 2\pi),$$

(or, $(\alpha_i, \beta_i) = \infty$ if M is complete near P_i).

Note that (α_i, β_i) is unique, up to multiplication by -1, since when M is not complete near P_i, the holonomy $\tilde{H} : \pi_1(\mathcal{N}_i) \to \mathbb{R} \oplus \mathbb{R}$ is injective. We will say that $\bar{M}$ is a hyperbolic structure for

$$M_{(\alpha_1,\beta_1),\dots,(\alpha_k,\beta_k)}.$$

If all (α_i, β_i) happen to be primitive elements of $\mathbb{Z} \oplus \mathbb{Z}$, then $\bar{M}$ is the topological manifold $M_{(\alpha_1,\beta_1),\dots,(\alpha_k,\beta_k)}$ with a non-singular hyperbolic structure, so that our extended definition is compatible with the original. If each ratio α_i/β_i is the rational number p_i/q_i in lowest terms, then $\bar{M}$ is topologically the manifold $M_{(p_1,q_1),\dots,(p_k,q_k)}$. The hyperbolic structure, however, has singularities at the component circles of $\bar{M} - M$ with 4.17
cone angles of $2\pi(p_i/\alpha_i)$ [since the holonomy $\tilde{H}$ of the primitive element $p_i a_i + q_i b_i$ in $\pi_1(P_i)$ is a pure rotation of angle $2\pi(p_i/\alpha_i)$].

There is also a topological interpretation in case the $(\alpha_i, \beta_i) \in \mathbb{Z} \oplus \mathbb{Z}$, although they may not be primitive. In this case, all the cone angles are of the form $2\pi/n_i$, where each n_i is an integer. Any branched cover of $\bar{M}$ which has branching index n_i around the i-th circle of $\bar{M} - M$ has a non-singular hyperbolic structure induced from $\bar{M}$.

4.6. Dehn surgery on the figure-eight knot.

For each value of w in the region R of $\mathbb{C}$ shown on p. 4.10, the associated hyperbolic structure on $S^3 - K$, where K is the figure-eight knot, has some Dehn surgery invariant $d(w) = \pm(\mu(w), \lambda(w))$. The function d is a continuous map from R to the one-point compactification $\mathbb{R}^2/\pm 1$ of $\mathbb{R}^2$ with vectors v identified to $-v$. Every primitive element (p, q) of $\mathbb{Z} \oplus \mathbb{Z}$ which lies in the image $d(R)$ describes a closed manifold $(S^3 - K)_{(p,q)}$ which possesses a hyperbolic structure.

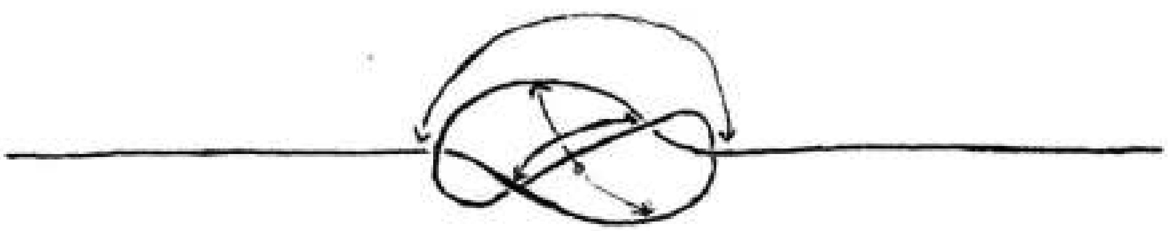
Actually, the map d can be lifted to a map $\tilde{d} : R \to \hat{\mathbb{R}}^2$, by using the fact that the sign of a rotation of

$$\widetilde{(H^3 - z\text{-axis})}$$

is well-defined. (See §4.4. The extra information actually comes from the orientation of the z-axis determined by the direction in which the corners of tetrahedra wrap around it). $\tilde{d}$ is defined by the equation $\tilde{d}(w) = (\mu, \lambda)$ where 4.18

$$\mu \tilde{H}(m) + \lambda \tilde{H}^2(l) = (\text{a rotation by } +2\pi)$$

In order to compute the image $\tilde{d}(R)$, we need first to express the generators l and m for $\pi_1(P)$ in terms of the previous generators x and y on p. 4.11. Referring to page 6, we see that a meridian which only intersects two two-cells can be constructed in a small neighborhood of K. The only generator of $\pi_1(L(v))$ (see p. 4.11) which intersects only two one-cells is $\pm y$, so we may choose $m = y$. Here is a cheap way to see what l is. The figure-eight knot can be arranged (passing through the point at infinity) so that it is invariant by the map $v \mapsto -v$ of $\hat{\mathbb{R}}^3 = S^3$.



This map can be made an isometry of the complete hyperbolic structure constructed for $S^3 - K$. (This can be seen directly; it also follows immediately from Mostow's Theorem, ...). This hyperbolic isometry induces an isometry of the Euclidean structure on $L(v)$ which takes m to m and l to $-l$. Hence, a geodesic representing l must be orthogonal to a geodesic representing m, so from the diagram on the bottom of 4.19
p. 4.11 we deduce that the curve $l = +x + 2y$ is a longitude. (Alternatively, it is not hard to compute m and l directly).

From p. 4.12, we have

4.6.1.
$$\begin{aligned} H(m) &= w(1-z) \\ H(l) &= z^2(1-z)^2 \end{aligned}$$

The behavior of the map $\tilde{d}$ near the boundary of R is not hard to determine. For instance, when w is near the ray $\mathrm{Im}(w) = 0$, $\mathrm{Re}(w) > 1$, then z is near the ray $\mathrm{Im}(z) = 0$, $\mathrm{Re}(z) < 0$. The arguments of $\tilde{H}(m)$ and $\tilde{H}(l)$ are easily computed by analytic continuation from the complete case $w = z = \sqrt[3]{-1}$ (when the arguments are 0) to be

$$\arg \tilde{H}(m) = 0 \quad \arg \tilde{H}(l) \approx \pm 2\pi.$$

Consequently, (μ, λ) is near the line $\lambda = +1$. As $w \to 1$ we see from the equation

$$z(1-z)\, w(1-w) = 1$$

that

$$|z|^2 \cdot |w| \to 1$$

so (μ, λ) must approach the line $\mu + 4\lambda = 0$. Similarly, as $w \to +\infty$, then $|z|\,|w|^2 \to 1$, so (μ, λ) must approach the line $\mu - 4\lambda = 0$. Then the map $\tilde{d}$ extends continuously to send the line segment

$$\overline{1, +\infty}$$

to the line segment 4.20

$$\overline{(-4, +1), (+4, +1)}.$$

There is an involution τ of the region R obtained by interchanging the solutions z and w of the equation $z(l-z)\, w(l-w) = 1$. Note that this involution takes $H(m)$ to $1/H(m) = H(-m)$ and $H(l)$ to $H(-l)$. Therefore $\tilde{d}(\tau w) = -\tilde{d}(w)$. It follows that $\tilde{d}$ extends continuously to send the line segment

$$\overline{-\infty, 0}$$

to the line segment

$$\overline{(+4, -1), (-4, -1)}.$$

When $|w|$ is large and $0 < \arg(w) < \pi/2$, then $|z|$ is small and

$$\arg(z) \approx \pi - 2\arg(w).$$

Thus $\arg \tilde{H}(m) \approx \arg w$, $\arg \tilde{H}(l) \approx 2\pi - 4\arg w$ so $\mu \arg w + \lambda(2\pi - 4\arg w) = 2\pi$. By considering $|H(m)|$ and $|H(l)|$, we have also $\mu - 4\lambda \approx 0$, so $(\mu, \lambda) \approx (4, 1)$.

There is another involution σ of R which takes w to

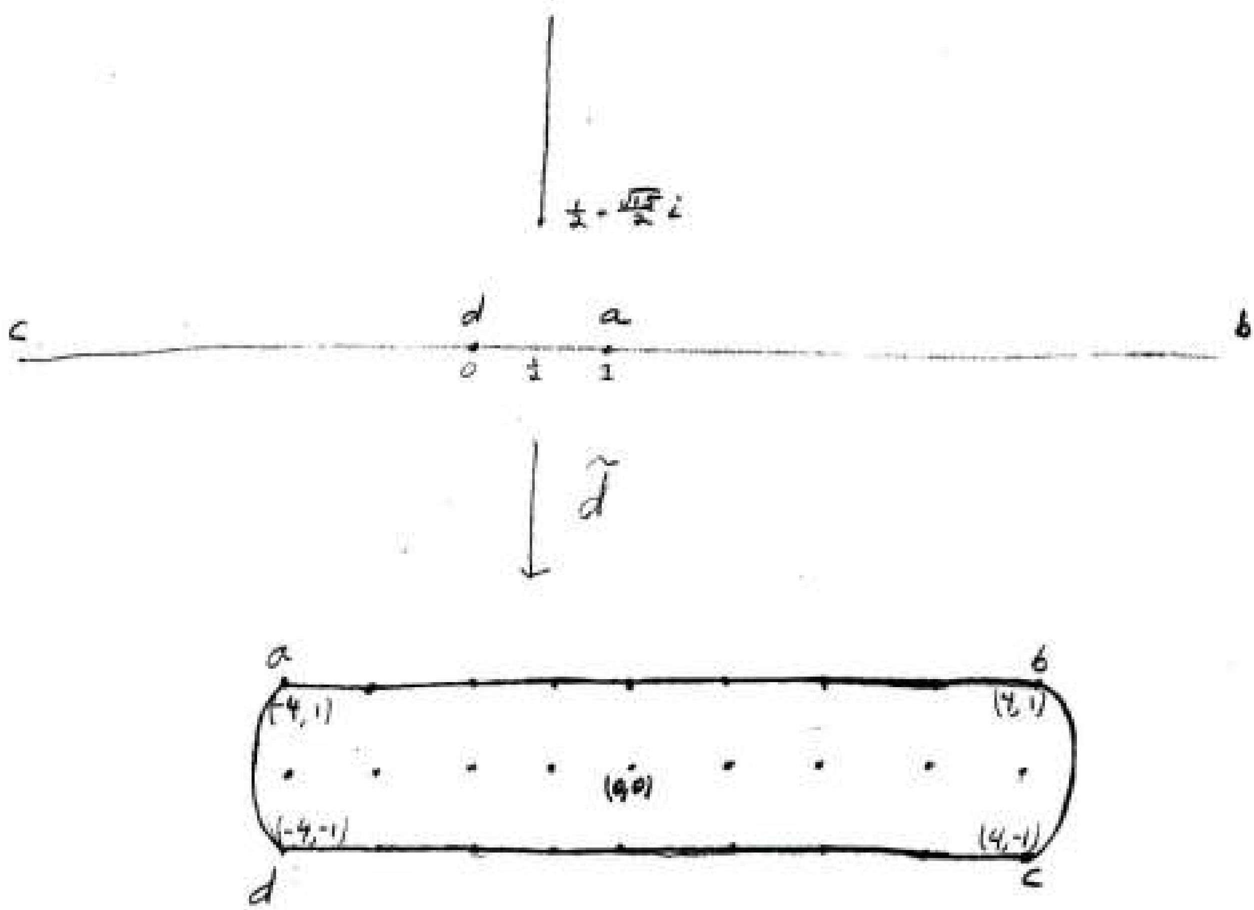
$$\overline{1 - w}$$

(and z to $\overline{1-z}$). From 4.6.1 we conclude that if $\tilde{d}(w) = (\mu, \lambda)$, then $\tilde{d}(\sigma w) = (\mu, -\lambda)$. With this information, we know the boundary behavior of $\tilde{d}$ except when w or τw is near the ray r described by

$$\mathrm{Re}(w) = \tfrac{1}{2}, \quad \mathrm{Im}(w) \geq \tfrac{\sqrt{15}}{2}\, i.$$

The image of the two sides of this ray is not so neatly described, but it does not represent a true boundary for the family of hyperbolic structures on $S^3 - K$, as w crosses r from right to left, for instance, z crosses the real line in the interval $(0, \frac{1}{2})$. For a while, a hyperbolic structure can be constructed from the positively oriented simplex determined by w and the negatively oriented simplex determined by z, by cutting the z-simplex into pieces which are subtracted from the w-simplex to leave a polyhedron P. P is then identified to give a hyperbolic structure for $S^3 - K$. 4.21

For this reason, we give only a rough sketch of the boundary behavior of $\tilde{d}$ near r or $\tau(r)$:



Since the image of $\tilde{d}$ in $\hat{\mathbb{R}}^2$ does not contain the origin, and since $\tilde{d}$ sends a curve winding once around the boundary of R to the curve $abcd$ in $\hat{\mathbb{R}}^2$, it follows that the image of $\tilde{d}(R)$ contains the exterior of this curve. 4.22

In particular

THEOREM 4.7. *Every manifold obtained by Dehn surgery along the figure-eight knot K has a hyperbolic structure, except the six manifolds:*

$$(S^3 - K)_{(\mu,\lambda)} = (S^3 - K)_{(\pm\mu,\pm\lambda)}$$

where (μ, λ) *is* $(1, 0)$, $(0, 1)$, $(1, 1)$, $(2, 1)$, $(3, 1)$ *or* $(4, 1)$.

The equation

$$(S^3 - K)_{(\alpha,\beta)} = (S^3 - K_{(-\alpha,\beta)}$$

follows from the existence of an orientation reversing homeomorphism of $S^3 - K$.

I first became interested in studying these examples by the beautiful ideas of Jørgensen (compare Jørgensen, "Compact three-manifolds of constant negative curvature fibering over the circle," *Annals* 106 (1977) 61-72). He found the hyperbolic structures corresponding to the ray $\mu = 0$, $\lambda > 1$, and in particular, the integer and half-integer (!) points along this ray, which determine discrete groups.

The statement of the theorem is meant to suggest, but not imply, the true fact that the six exceptions do *not* have hyperbolic structures. Note that at least

$$S^3 = (S^3 - K)_{(1,0)}$$

does not admit a hyperbolic structure (since $\pi_1(S^3)$ is finite). We shall arrive at an understanding of the other five exceptions by studying the way the hyperbolic structures are degenerating as (μ, λ) tends to the line segment

$$\overline{(-4, 1), (4, 1)}.$$

4.8. Degeneration of hyperbolic structures.

DEFINITION 4.8.1. A codimension-k foliation of an n-manifold M is a $\mathcal{G}$-structure, on M, where $\mathcal{G}$ is the pseudogroup of local homeomorphisms of $\mathbb{R}^{n-k} \times \mathbb{R}^k$ which 4.23
have the local form

$$\phi(x, y) = (f(x, y),\, g(y)).$$

In other words, $\mathcal{G}$ takes horizontal $(n - k)$-planes to horizontal $(n - k)$-planes. These horizontal planes piece together in M as $(n - k)$-submanifolds, called the *leaves* of the foliation. M, like a book without its cover, is a disjoint union of its leaves.

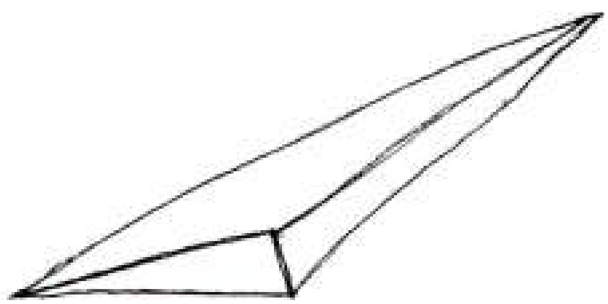
For any pseudogroup $\mathcal{H}$ of local homeomorphisms of some k-manifold N^k, the notion of a codimension-k foliation can be refined:

Definition 4.8.2. An $\mathcal{H}$-*foliation* of a manifold M^n is a $\mathcal{G}$-structure for M^n, where $\mathcal{G}$ is the pseudogroup of local homeomorphisms of $\mathbb{R}^{n-k} \times N^k$ which have the local form

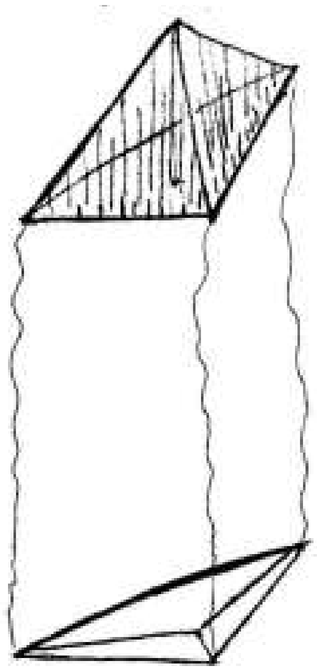
$$\phi(x, y) = (f(x, y),\, g(y))$$

with $g \in \mathcal{H}$. If $\mathcal{H}$ is the pseudogroup of local isometries of hyperbolic k-space, then an $\mathcal{H}$-foliation shall, naturally, be called a codimension-k hyperbolic foliation. A hyperbolic foliation determines a hyperbolic structure for each k-manifold transverse to its leaves.

When w tends in the region $R \subset \mathbb{C}$ to a point $\mathbb{R} - [0, 1]$, the w-simplex and the z-simplex are both flattening out, and in the limit they are flat: 4.24



If we regard these flat simplices as projections of nondegenerate simplices A and B (with vertices deleted), this determines codimension-2 foliations on A and B, whose leaves are preimages of points in the flat simplices:



4.25

A and B glue together (in a unique way, given the combinatorial pattern) to yield a hyperbolic foliation on $S^3 - K$. You should satisfy yourself that the gluing consistency conditions for the hyperbolic foliation near an edge result as the limiting case of the gluing conditions for the family of squashing hyperbolic structures.

The notation of the developing map extends in a straightforward way to the case of an $\mathcal{H}$-foliation on a manifold M, when $\mathcal{H}$ is the set of restrictions of a group J of real analytic diffeomorphisms of N^k; it is a map

$$D : \tilde{M}^n \to N^k.$$

Note that D is not a local homeomorphism, but rather a local projection map, or a submersion. The holonomy

$$H : \pi_1(M) \to J$$

is defined, as before, by the equation

$$D \circ T_\alpha = H(\alpha) \circ D.$$

Here is the generalization of proposition 3.6 to $\mathcal{H}$-foliations. For simplicity, assume that the foliation is differentiable:

PROPOSITION 4.8.1. *If J is transitive and if the isotropy subgroups J_x are compact, then the developing map for any $\mathcal{H}$-foliation $\mathcal{F}$ of a closed manifold M is a fibration* labeled 4.8.1def 4.26

$$D : \tilde{M}^n \to N^k.$$

PROOF. Choose a plane field τ^k transverse to $\mathcal{F}$ (so that τ is a complementary subspace to the tangent space to the leaves of $\mathcal{F}$, called $T\mathcal{F}$, at each point). Let h be an invariant Riemannian metric on N^k and let g be any Riemannian metric on M. Note that there is an upper bound K for the difference between the g-length of a nonzero vector in τ and the k-length of its local projection to N^k.

Define a horizontal path in $\tilde{M}$ to be any path whose tangent vector always lies in τ. Let $\alpha : [0, 1] \to N$ be any differentiable path, and let $\tilde{\alpha}_0$ be any point in the preimage $D^{-1}(\alpha_0)$. Consider the problem of lifting α to a horizontal path in $\tilde{M}$ beginning at $\tilde{\alpha}_0$. Whenever this has been done for a closed interval (such as $[0, 0]$), it can be obviously extended to an open neighborhood. When it has been done for an open interval, the horizontal lift $\tilde{\alpha}$ is a Cauchy path in $\tilde{M}$, so it converges. Hence, by "topological induction", α has a (unique) global horizontal lift beginning at $\tilde{\alpha}_0$. Using horizontal lifts of the radii of disks in N, local trivializations for $D : \tilde{M} \to N$ are obtained, showing that D is a fibration. □

DEFINITION. An $\mathcal{H}$-foliation is *complete* if the developing map is a fibration.

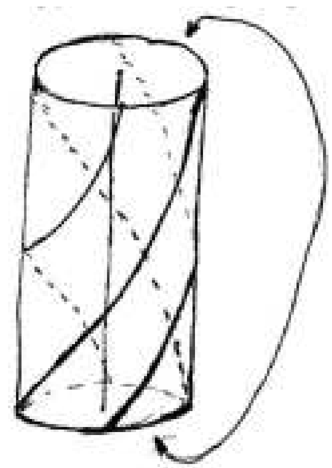
4.27

Any three-manifold with a complete codimension-2 hyperbolic foliation has universal cover $H^2 \times \mathbb{R}$, and covering transformations act as global isometries in the first coordinate. Because of this strong structure, we can give a complete classification of such manifolds. A *Seifert fibration* of a three-manifold M is a projection $p : M \to B$ to some surface B, so that p is a submersion and the preimages of points are circles in M. A Seifert fibration is a fibration except at a certain number of singular points $x_1, \ldots, x_k$. The model for the behavior in $p^{-1}(N_\epsilon(x_i))$ is a solid torus with a foliation having the core circle as one leaf, and with all other leaves winding p times around the short way and q times around the long way, where $1 < p < q$ and $(p, q) = 1$.

The projection of a meridian disk of the solid torus to its image in B is q-to-one, 4.28
except at the center where it is one-to-one.

A group of isometries of a Riemannian manifold is *discrete* if for any x, the orbit of x intersects a small neighborhood of x only finitely often. A discrete group Γ of orientation-preserving isometries of a surface N has quotient N/Γ a surface. The projection map $N \to N/\Gamma$ is a local homeomorphism except at points x where the isotropy subgroup Γ_x is nontrivial. In that case, Γ_x is a cyclic group $\mathbb{Z}/q\mathbb{Z}$ for some $q > 1$, and the projection is similar to the projection of a meridian disk cutting across a singular fiber of a Seifert fibration.

THEOREM 4.9. *Let $\mathcal{F}$ be a hyperbolic foliation of a closed three-manifold M. Then either*



A meridian disk of the solid torus wraps q times around its image disk. Here $p = 1$ and $q = 2$.

(a) *The holonomy group $H(\pi_1 M)$ is a discrete group of isometries of H^2, and the developing map goes down to a Seifert fibration*

$$D_{/\pi_1 M} : M \to H^2/H(\pi_1 M),$$

or

(b) *The holonomy group is not discrete, and M fibers over the circle with fiber a torus.*

The structure of $\mathcal{F}$ and M in case (b) will develop in the course of the proof. 4.29

Proof. (a) If $H(\pi_1 M)$ is discrete, then $H^2/H(\pi_1 M)$ is a surface. Since M is compact the fibers of the fibration $D : \tilde{M}^3 \to H^2$ are mapped to circles under the projection $\pi : \tilde{M}^3 \to M^3$. It follows that $D/H(\pi_1 M) : M^3 \to H^2/H(\pi_1 M)$ is a Seifert fibration.

(b) When $H(\pi_1 M)$ is not discrete, the proof is more involved. First, let us assume that the foliation is oriented (this means that the leaves of the foliation are oriented, or in other words, it is determined by a vector field). We choose a π_1 M-invariant Riemannian metric g in $\tilde{M}^3$ and let τ be the plane field which is perpendicular to the fibers of $D : \tilde{M}^3 \to H^2$. We also insist that along τ, g be equal to the pullback of the hyperbolic metric on H^2.

By construction, g defines a metric on M^3, and, since M^3 is compact, there is an infimum I to the length of a nontrivial simple closed curve in M^3 when measured with respect to g. Given $g_1, g_2 \in \pi_1 M$, we say that they are *comparable* if there is a $y \in \tilde{M}^3$ such that

$$d\big(D(g_1(y)), D(g_2(y))\big) < I,$$

where $d(\ ,\ )$ denotes the hyperbolic distance in H^2. In this case, take the geodesic in H^2 from $D(g_1(y))$ to $D(g_2(y))$ and look at its horizontal lift at $g_2(y)$. Suppose its other endpoint e where $g_1(y)$. Then the length of the lifted path would be equal to the length of the geodesic in H^2, which is less than I. Since $g_1 g_2^{-1}$ takes $g_2(y)$ to $g_1(y)$, the path represents a nontrivial element of $\pi_1 M$ and we have a contradiction. Now if we choose a trivialization of $H^2 \times \mathbb{R}$, we can decide whether or not $g_1(x)$ is greater than e. If it is greater than e we say that g_1 is greater than g_2, and write 4.30
$g_1 > g_2$, otherwise we write $g_1 < g_2$. To see that this local ordering does not depend on our choice of y, we need to note that

$$U(g_1, g_2) = \{x \mid d\big(H(g_1(x)),\ H(g_2(x))\big) < I\}$$

is a connected (in fact convex) set. This follows from the following lemma, the proof of which we defer.

Lemma 4.9.1. *$f_{g_1,g_2}(x) = d(g_1 x, g_2 x)$ is a a convex function on H^2.*

One useful property of the ordering is that it is invariant under left and right multiplication. In other words $g_1 < g_2$ if and only if, for all g_3, we have $g_3 g_1 < g_3 g_2$ and $g_1 g_3 < g_2 g_3$. To see that the property of comparability is equivalent for these three pairs, note that since $H(\pi_1 H^3)$ acts as isometries on H^2,

$$d(Dg_1 y, Dg_2 y) < I \quad \text{implies that} \quad d(Dg_3 g_1 y, Dg_3 g_2 y) < I.$$

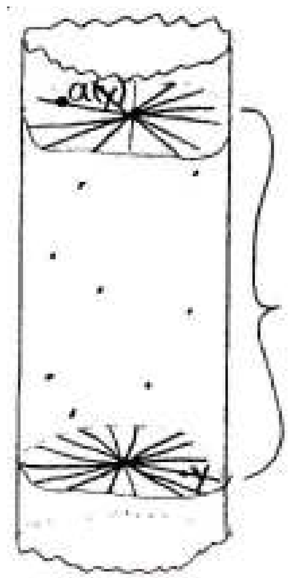
Also, if $d(Dg_1 y, Dg_2 y)$ then $d(Dg_3 g_1(g_3^{-1} y), Dg_3 g_2(g_3^{-1} y)) < I$, so that $g_1 g_3$ and $g_2 g_3$ are comparable. The invariance of the ordering easily follows (using the fact that $\pi_1 M$ preserves orientation of the $\mathbb{R}$ factors).

For a fixed $x \in H^2$ we let $G_\epsilon(X) \subset \pi_1 M$ be those elements whose holonomy acts on x in a way $C^1 - \epsilon$-close to the identity. In other words, for $g \in G_\epsilon(x)$, $d(x, H_g(x)) < \epsilon$ and the derivative of $H_g(x)$ parallel translated back to x, is ϵ-close to the identity.

Proposition 4.9.2. *There is an ϵ_0 so that for all $\epsilon < \epsilon_0$ $[G_\epsilon, G_\epsilon] \subset G_\epsilon$.*

Proof. For any Lie group the map $[*,*] : G \times G \to G$ has derivative zero at $(\mathrm{id}, \mathrm{id})$. Since for any $g \in G$, $(g, \mathrm{id}) \mapsto \mathrm{id}$ and $(\mathrm{id}, g) \mapsto 1$. The tangent spaces of $G \times \mathrm{id}$ and $\mathrm{id} \times G$ span the tangent space to $G \times G$ at $(\mathrm{id}, \mathrm{id})$. Apply this to the group of isometries of H^2. □

From now on we choose $\epsilon < I/8$ so that any two words of length four or less in G_ϵ are comparable. We claim that there is some $\beta \in G_\epsilon$ which is the "smallest" element in G_ϵ which is $>$ id. In other words, if $\mathrm{id} < \alpha \in G_\epsilon$, $a \neq \beta$, then $\alpha > \beta$. This can be seen as follows. Take an ϵ-ball B of $x \in H^2$ and look at its inverse image $\tilde{B}$ under D. Choose a point y in $\tilde{B}$ and consider y and $\alpha(y)$, where $\alpha \in G_\epsilon$. We can truncate $\tilde{B}$



There are only finitely many translates of y in this region.

by the lifts of B (using the horizontal lifts of the geodesics through x) through y and $\alpha(y)$. Since this is a compact set there are only a finite number of images of y under $\pi_1 M$ contained in it. Hence there is one $\beta(y)$ whose $\mathbb{R}$ coordinate is the closest to that of y itself. β is clearly our minimal element. 4.32

Now consider $\alpha > \beta > 1$, $\alpha \in G_\epsilon$. By invariance under left and right multiplication, $\alpha^2 > \beta_\alpha > \alpha$ and $\alpha > \alpha^{-1}\beta\alpha > 1$. Suppose $\alpha^{-1}\beta\alpha < \beta$. Then $\beta > \alpha^{-1}\beta\alpha > 1$ so that $1 > \alpha^{-1}\beta\alpha\beta^{-1} > \beta^{-1}$. Similarly if $\alpha^{-1}\beta\alpha > \beta > 1$ then $\beta > \alpha\beta\alpha^{-1} > 1$ so that $1 > \alpha\beta\alpha^{-1}\beta^{-1} > \beta^{-1}$. Note that by multiplicative invariance, if $g_1 > g_2$ then $g_2^{-1} = g_1^{-1}g_1g_2^{-1} > g_1^{-1}g_2g_2^{-1} = g_1^{-1}$. We have either $1 < \beta\alpha^{-1}\beta^{-1}\alpha < \beta$ or $1 < \beta\alpha\beta^{-1}\alpha^{-1} < \beta$ which contradicts the minimality of β . Thus $\alpha^{-1}\beta\alpha = \beta$ for all $\alpha \in G_\epsilon$.

We need to digress here for a moment to classify the isometries of H^2. We will prove the following:

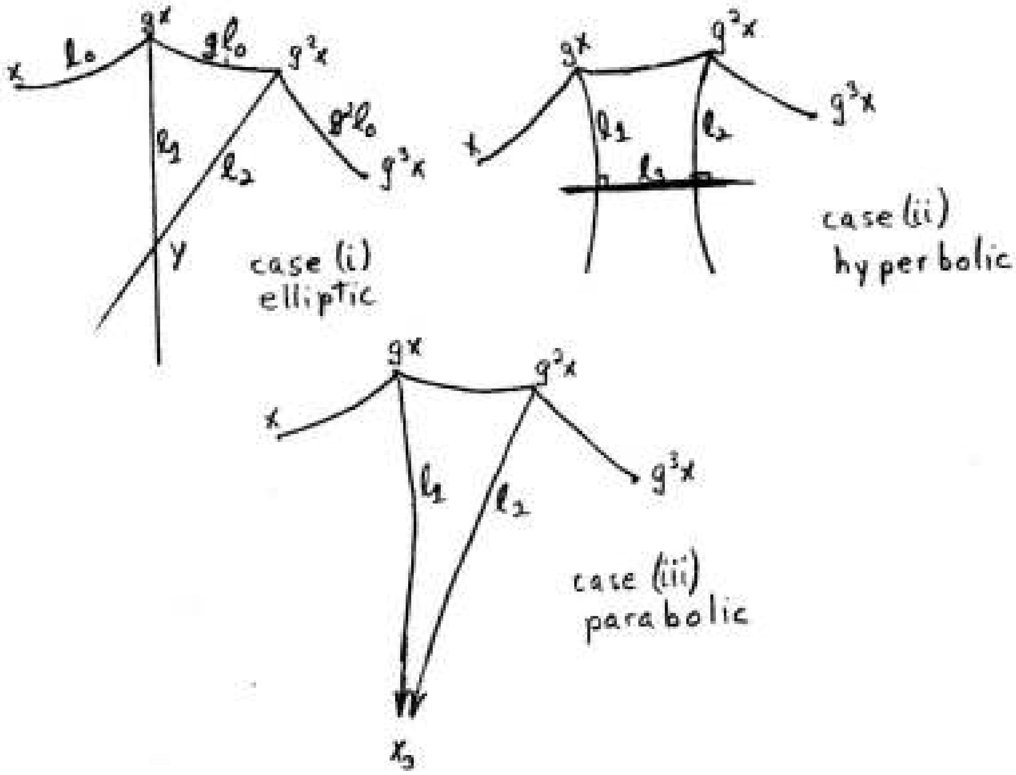
Proposition 4.9.3. *If $g : H^2 \to H^2$ is a non-trivial isometry of H^2 which preserves orientation, then exactly one of the following cases occurs:*

(i) *g has a unique interior fixed point or*
(ii) *g leaves a unique invariant geodesic or*
(iii) *g has a unique fixed point on the boundary of H^2.*

Case (i) is called *elliptic*, case (ii) *hyperbolic*, case (iii) *parabolic.*

PROOF. *This follows easily from linear algebra, but we give a geometric proof.* Pick an interior point $x \in H^2$ and connect x to gx by a geodesic l_0. Draw the geodesics l_1, l_2 at gx and g^2x which bisect the angle made by l_0 and gl_0, gl_0 and g^2l_0 respectively. There are three cases: 4.33

(i) l_1 and l_2 intersect in an interior point y
(ii) There is a geodesic l_3 perpendicular to l_1, l_2
(iii) l_1, l_2 are parallel, i.e., they intersect at a point at infinity x_3.



In case (i) the length of the arc gx, y equals that of g^2x, y since $\Delta(gx, g^2x, y)$ is an isoceles triangle. It follows that y is fixed by g.

In case (ii) the distance from gx to l_3 equals that from g^2x to l_3. Since l_3 meets l_1 and l_2 in right angles it follows that l_3 is invariant by g. 4.34

Finally, in case (iii) g takes l_1 and l_2, both of which hit the boundary of H^2 in the same point x_3. It follows that g fixes x_3 since an isometry takes the boundary to itself.

Uniqueness is not hard to prove. □

Using the classification of isometries of H^2, it is easy to see that the centralizer of any non-trivial element g in $\mathrm{isom}(H^2)$ is abelian. (For instance, if g is elliptic with fixed point x_0, then the centralizer of g consists of elliptic elements with fixed point x_0). It follows that the centralizer of β in $\pi_1(M)$ is abelian; let us call this group N.

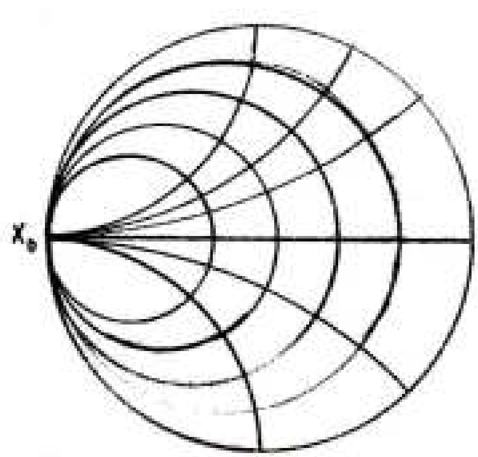
Although $G_\epsilon(x)$ depends on the point x, for any point $x' \in H^2$, if we choose ϵ' small enough, then $G_{\epsilon'}(x') \subset G_\epsilon(x)$. In particular if $x = H(g)x$, $g \in \pi_1 M$, then all elements of $G_{\epsilon'}(x')$ commute with β. It follows that N is a normal subgroup of $\pi_1(M)$.

Consider now the possibility that β is elliptic with fixed point x_0 and $n \in N$ fixes x_0 we see that all of $\pi_1 M$ must fix x_0. But the function $f_{x_0} : H^2 \to \mathbb{R}^+$ which measures the distance of a point in H^2 from x_0 is $H(\pi_1 M)$ invariant so that it lifts to a function f on M^3. However, M^3 is compact and the image of $\tilde{f}$ is non-compact, which is impossible. Hence β cannot be elliptic.

If β were hyperbolic, the same reasoning would imply that $H(\pi_1 M)$ leaves invari- 4.35
ant the invariant geodesic of β. In this case we could define $f_l : H^2 \to \mathbb{R}$ to be the distance of a point from l. Again, the function lifts to a function on M^3 and we have a contradiction.

The case when β is parabolic actually does occur. Let x_0 be the fixed point of β on the circle at infinity. N must also fix x_0. Using the upper half-plane model for H^2 with x_0 at ∞, β acts as a translation of $\mathbb{R}^2$ and N must act as a group of similarities; but since they commute with β, they are also translations. Since N is normal, $\pi_1 M$ must act as a group of similarities of $\mathbb{R}^2$ (preserving the upper half-plane).

Clearly there is no function on H^2 measuring distance from the point x_0 at infinity. If we consider a family of finite points $x_\tau \to X$, and consider the functions f_{x_τ}, even though f_{x_τ} blows up, its derivative, the closed 1-form df_{x_τ}, converges to a closed 1 form ω. Geometrically, ω vanishes on tangent vectors to horocycles about x_0 and takes the value 1 on unit tangents to geodesics emanating from x_0.



4.36

The non-singular closed 1-form ω on H^2 is invariant by $H(\pi_1 M)$, hence it defines a non-singular closed one-form $\bar{\omega}$ on M. The kernel of $\bar{\omega}$ is the tangent space to the leaves of a codimension one foliation $\mathfrak{F}$ of M. The leaves of the corresponding foliation $\tilde{\mathfrak{F}}$ on $\tilde{M}$ are the preimages under D of the horocycles centered at x_0. The group of periods for ω must be discrete, for otherwise there would be a translate of the horocycle about x_0 through x close to x, hence an element of G_ϵ which does not commute with β. Let p_0 be the smallest period. Then integration of ω defines a map from M to $S^1 = \mathbb{R}/\langle p_0 \rangle$, which is a fibration, with fibers the leaves of $\mathfrak{F}$. The fundamental group of each fiber is contained in N, which is abelian, so the fibers are toruses.

It remains to analyze the case that the hyperbolic foliation is not oriented. In this case, let M' be the double cover of M which orients the foliation. M' fibers over S^1 with fibration defined by a closed one-form ω. Since ω is determined by the unique fixed point at infinity of $H(\pi_1 M')$, ω projects to a non-singular closed one-form on M. This determines a fibration of M with torus fibers. (Klein bottles cannot occur even if M is not required to be orientable.) $\square$

We can construct a three-manifold of type (b) by considering a matrix

$$A \in SL(2, \mathbb{Z})$$

which is hyperbolic, i.e., it has two eigenvalues λ_1, λ_2 and two eigenvectors V_1, V_2. Then $AV_1 = \lambda_1 V_1, AV_2 = \lambda_2 V_2$ and $\lambda_2 = 1/\lambda_1$.

Since $A \in SL(2, \mathbb{Z})$ preserves $\mathbb{Z} \oplus \mathbb{Z}$ its action on the plane descends to an action on the torus T^2. Our three-manifold M_A is the mapping torus of the action of A 4.37
on T^2. Notice that the lines parallel to V_1 are preserved by A so they give a one-dimensional foliation on M_A. Of course, the lines parallel to V_2 also define a foliation. The reader may verify that both these foliations are hyperbolic. When

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

then M_A is the manifold $(S^3 - K)_{(D, \pm 1)}$ obtained by Dehn surgery on the figure-eight knot. The hyperbolic foliations corresponding to $(0, 1)$ and $(0, -1)$ are distinct, and they correspond to the two eigenvectors of

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

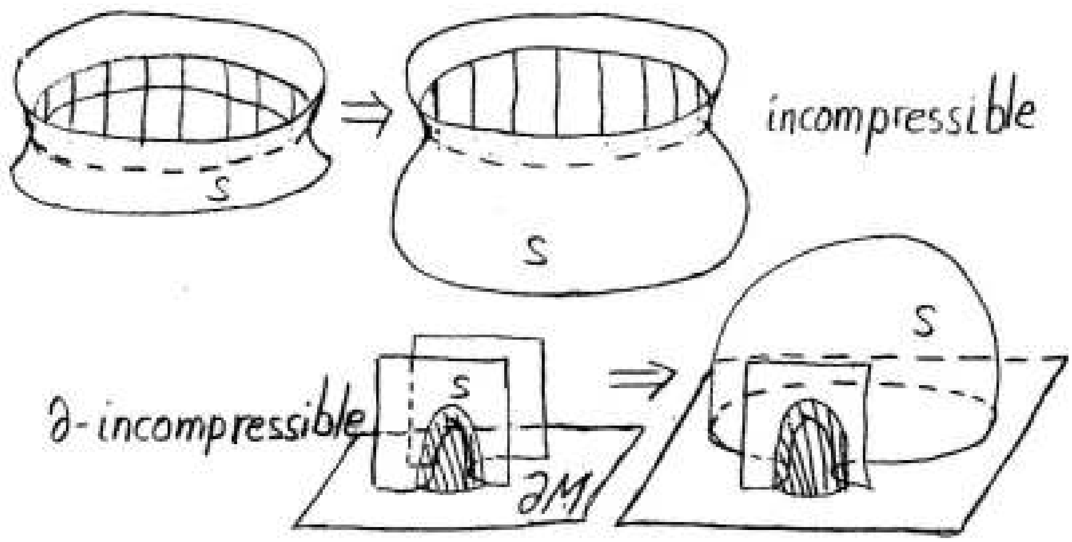
All codimension-2 hyperbolic foliations with leaves which are not closed are obtained by this construction. This follows easily from the observation that the hyperbolic foliation restricted to any fiber is given by a closed non-singular one-form, together with the fact that a closed non-singular one-form on T^2 is determined (up to isotopy) by its cohomology class.

The three manifolds $(S^3 - K)_{(1,1)}$, $(S^3 - K)_{(2,1)}$ and $(S^3 - K)_{(3,1)}$ also have codimension-2 hyperbolic foliations which arise as "limits" of hyperbolic structures.

Since they are rational homology spheres, they must be Seifert fiber spaces. A Seifert fiber space cannot be hyperbolic, since (after passing to a cover which orients the fibers) a general fiber is in the center of its fundamental group. On the other hand, the centralizer of an element in the fundamental group of a hyperbolic manifold is abelian. 4.38

4.10. Incompressible surfaces in the figure-eight knot complement.

Let M^3 be a manifold and $S \subset M^3$ a surface with $\partial S \subset \partial M$. Assume that $S \neq S^2$, IP^2, or a disk D^2 which can be pushed into ∂M. Then S is *incompressible* if every loop (simple closed curve) on S which bounds an (open) disk in $M - S$ also bounds a disk in S. Some people prefer the alternate, stronger definition that S is (*strongly*) *incompressible* if $\pi_1(S)$ injects into $\pi_1(M)$. By the loop theorem of Papakyriakopoulos, these two definitions are equivalent if S is two-sided. If S has boundary, then S is also ∂-*incompressible* if every arc α in S (with $\partial(\alpha) \subset \partial S$) which is homotopic to ∂M is homotopic in S to ∂S.

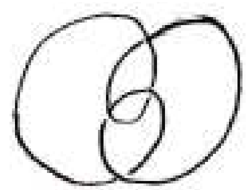


If M is oriented and irreducible (every two-sphere bounds a ball), then M is *sufficiently large* if it contains an incompressible and ∂-incompressible surface. A 4.39
compact, oriented, irreducible, sufficiently large three-manifold is also called a *Haken-manifold*. It has been hard to find examples of three-manifolds which are irreducible but can be shown not to be sufficiently large. The only previously known examples are certain Seifert fibered spaces over S^2 with three exceptional fibers. In what follows we give the first known examples of compact, irreducible three-manifolds which are not Haken-manifolds and are not Seifert fiber spaces.

NOTE. If M is a compact oriented irreducible manifold $\neq D^3$, and either $\partial M \neq \emptyset$ or $H^1(M) \neq 0$, then M is sufficiently large. In fact, $\partial M \neq 0 \Rightarrow H^1(M) \neq 0$. Think of a non-trivial cohomology class α as dual to an embedded surface; an easy argument using the loop theorem shows that the simplest such surface dual to α is incompressible and ∂-incompressible.

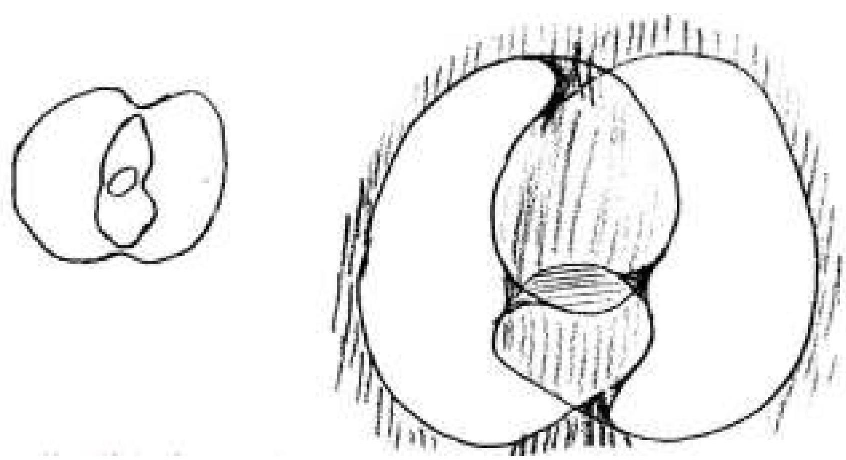
The concept of an incompressible surface was introduced by W. Haken (International Congress of Mathematicians, 1954), (*Acta. Math.* 105 (1961), *Math A.* 76 (1961), *Math Z* 80 (1962)). If one splits a Haken-manifold along an incompressible and ∂-incompressible surface, the resulting manifold is again a Haken-manifold. One can continue this process of splitting along incompressible surfaces, eventually arriving (after a bounded number of steps) at a union of disks. Haken used this to give algorithms to determine when a knot in a Haken-manifold was trivial, and when two knots were linked.

Let K be a figure-eight knot, $M = S^3 - \mathcal{N}(K)$. M is a Haken manifold by 4.40
the above note [M is irreducible, by Alexander's theorem that every differentiable two-sphere in S^3 bounds a disk (on each side)].



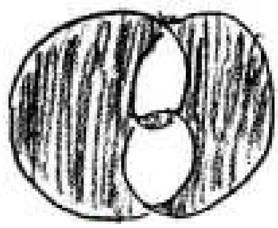
Here is an enumeration of the incompressible and ∂-incompressible surfaces in M. There are six reasonably obvious choices to start with;

- S_1 is a torus parallel to ∂M,
- $S_2 = T^2$-disk = Seifert surface for K. To construct S_2, take 3 circles lying above the knot, and span each one by a disk. Join

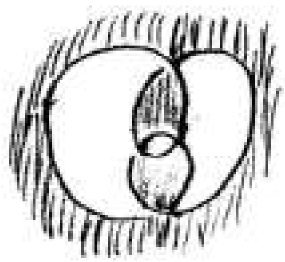


the disks by a twist for each crossing at K to get a surface S_2 with boundary the longitude $(0, \pm 1)$. S_2 is oriented and has Euler characteristic $= -1$, so it is T^2-disk.

- $S_3 =$ (Klein bottle-disk) is the unoriented surface pictured.



- $S_4 = \partial$ (tubular neighborhood of S_3) $= T^2 - 2$ disks. $\partial S_4 = (\pm 4, 1)$, (depending on the choice of orientation for the meridian).
- $S_5 =$ (Klein bottle-disk) is symmetric with S_3.



- $S_6 = \partial$ (tubular neighborhood of S_5) $= T^2 - 2$ disks. $\partial S_6 = (\pm 4, 1)$.

It remains to show that

Theorem 4.11. *Every incompressible and ∂-incompressible connected surface in M is isotopic to one of S_1 through S_6.*

Corollary. *The Dehn surgery manifold $M_{(m,l)}$ is* irreducible*, and it is a Haken-manifold if and only if $(m, l) = (0, \pm 1)$ or $(\pm 4, \pm 1)$.*

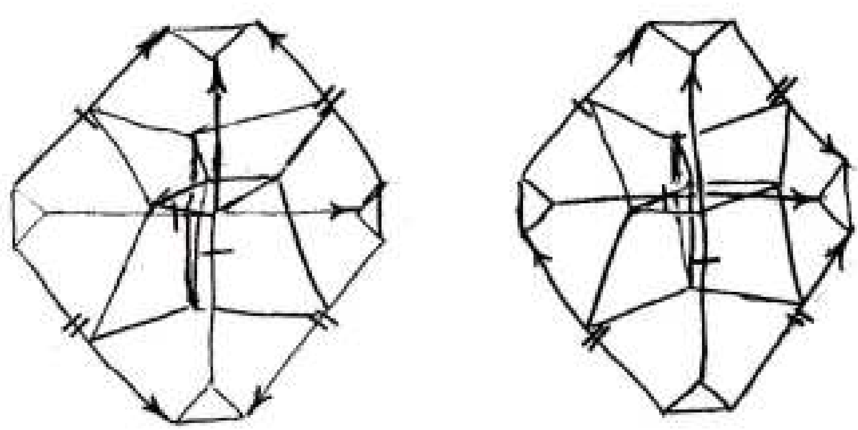
In particular, none of the hyperbolic manifolds obtained from M by Dehn surgery is sufficiently large. (Compare 4.7.) Thus we have an infinite family of examples of oriented, irreducible, non-Haken-manifolds which are not Seifert fiber spaces. It seems likely that Dehn surgery along other knots and links would yield many more examples. 4.42

Proof of corollary from theorem. Think of $M_{(m,l)}$ as M union a solid torus, $D^2 \times S^1$, the solid torus being a thickened core curve. To see that $M_{(m,l)}$ is irreducible, let S be an embedded S^2 in $M_{(m,l)}$, transverse to the core curve α (S intersects the solid torus in meridian disks). Now isotope S to minimize its intersections with α. If S doesn't intersect α then it bounds a ball by the irreducibility

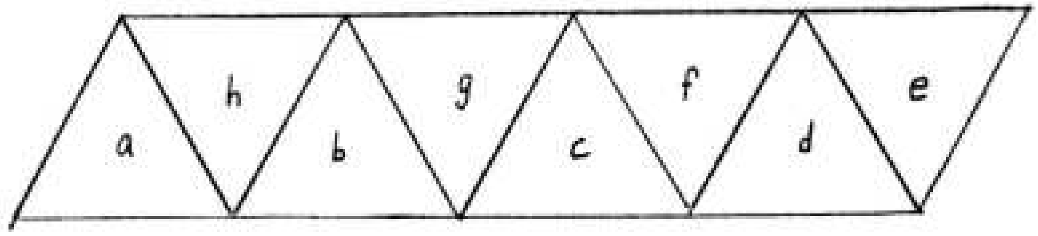
of M. If it does intersect α we may assume each component of intersection with the solid torus $D^2 \times S^1$ is of the form $D^2 \times x$. If $S \cap M$ is not incompressible, we may divide S into two pieces, using a disk in $S \cap M$, each of which has fewer intersections with α. If S does not bound a ball, one of the pieces does not bound. If $S \cap M$ is ∂-incompressible, we can make an isotopy of S to reduce the number of intersections with α by 2. Eventually we simplify S so that if it does not bound a ball, $S \cap M$ 4.43
is incompressible and ∂-incompressible. Since none of the surfaces $S_1, \dots, S_6$ is a submanifold of S^2, it follows from the theorem that S in fact bounds a ball.

The proof that $M_{(m,l)}$ is not a Haken-manifold if $(m, l) \neq (0, \pm 1)$ or $(\pm 4, \pm 1)$ is similar. Suppose S is an incompressible surface in $M_{(m,l)}$. Arrange the intersections with $D^2 \times S^1$ as before. If $S \cap M$ is not incompressible, let D be a disk in M with $\partial D \subset S \cap M$ not the boundary of a disk in $S \cap M$. Since S in incompressible, $\partial D = \partial D'$ for some disk $D' \subset S$ which must intersect α. The surface S' obtained from S by replacing D' with D is incompressible. (It is in fact isotopic to S, since M is irreducible; but it is easy to see that S' is incompressible without this.) S' has fewer intersections with α than does S. If S is not ∂-incompressible, an isotopy can be made as before to reduce the number of intersections with α. Eventually we obtain an incompressible surface (which is isotopic to S) whose intersection with M is incompressible and ∂-incompressible. S cannot be S_1 (which is not incompressible in $M_(m, l)$), so the corollary follows from the theorem. $\square$

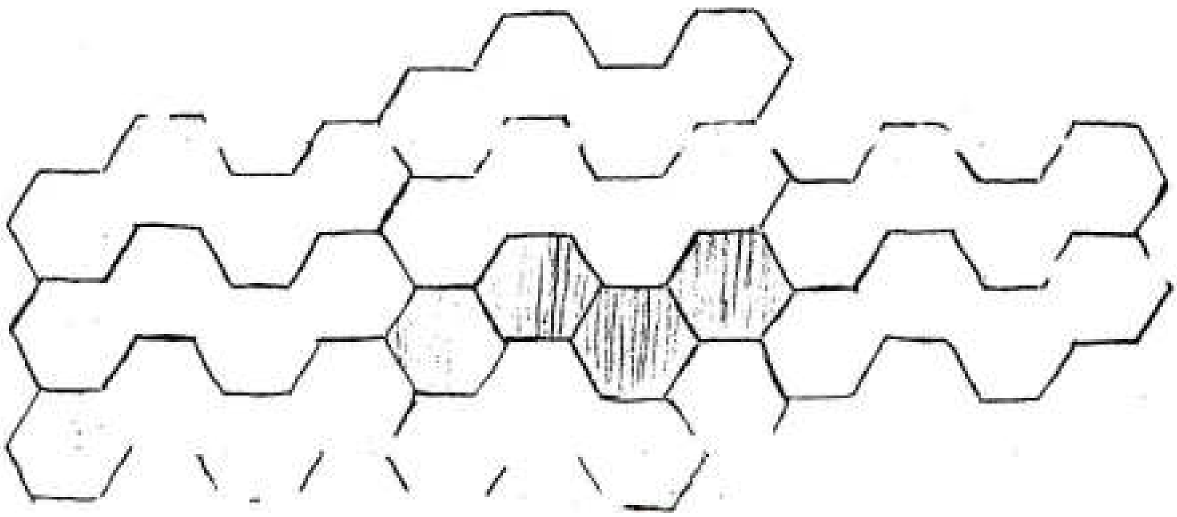
Proof of theorem 4.11. Recall that $M = S^3 - \mathcal{N}(K)$ is a union of two tetrahedra-without-vertices. To prove the theorem, it is convenient to use an alternate description of M at $T^2 \times I$ with certain identifications on $T^2 \times \{1\}$ (compare Jørgensen, "Compact three-manifolds of constant negative curvature fibering over the 4.44
circle", *Annals of Mathematics* **106** (1977), 61–72, and R. Riley). One can obtain this from the description of M as the union of two tetrahedra with corners as follows. Each tetrahedron = (corners) $\times\ I$ with certain identifications on (corners) $\times \{1\}$.



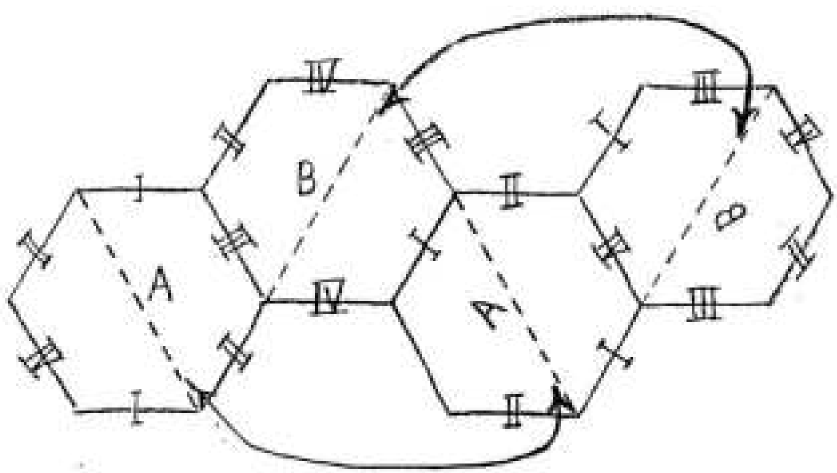
This "product" structure carries over to the union of the two tetrahedra. The boundary torus has the triangulation (p. 4.11)



4.45



$T^2 \times \{1\}$ has the dual subdivision, which gives T^2 as a union of four hexagons. The diligent reader can use the gluing patters of the tetrahedra to check that the identifications on $T^2 \times \{1\}$ are



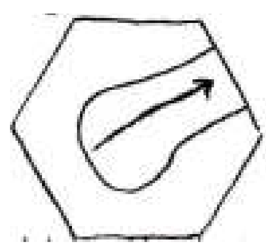
where we identify the hexagons by flipping through the dotted lines.

The complex $N = T^2 \times \{1\}$/identifications is a spine for N. It has a cell subdivision with two vertices, four edges, and two hexagons. N is embedded in M, and its complement is $T^2 \times [0, 1)$.

If S is a connected, incompressible surface in M, the idea is to simplify it with respect to the spine N (this approach is similar in spirit to Haken's). First isotope S 4.46
so it is transverse to each cell of N. Next isotope S so that it doesn't intersect any hexagon in a simple closed curve. Do this as follows.

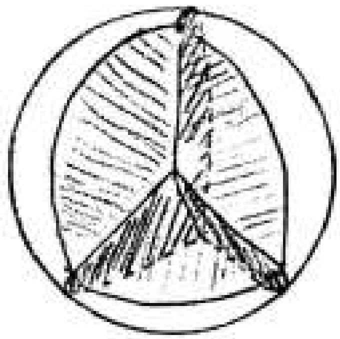
If $S \cap$ hexagon contains some loops, pick an innermost loop α. Then α bounds an open disk in $M^2 - S$ (it bounds one in the hexagon), so by incompressibility it bounds a disk in S. By the irreducibility of M we can push this disk across this hexagon to eliminate the intersection α. One continues the process to eliminate all such loop intersections. This does not change the intersection with the one-skeleton $N_{(1)}$.

S now intersects each hexagon in a collection of arcs. The next step is to isotope S to minimize the number of intersections with $N_{(1)}$. Look at the preimage of $S \cap N$. We can eliminate any arc which enters and leaves a hexagon in the same edge by pushing the arc across the edge.



If at any time a loop intersection is created with a hexagon, eliminate it before proceeding. 4.47

Next we concentrate on corner connections in hexagons, that is, arcs which connect two adjacent edges of a hexagon. Construct a small ball $\mathcal{B}$ about each vertex,

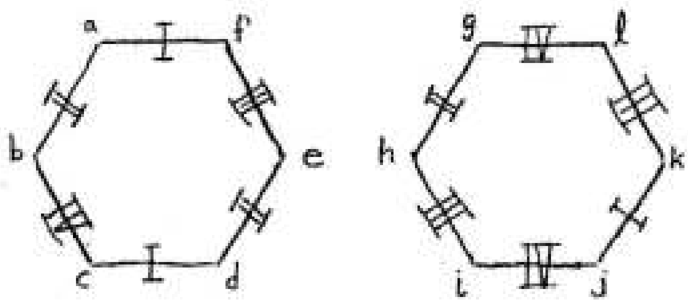


and push S so that the corner connections are all contained in $\mathcal{B}$, and so that S is transverse to $\partial\mathcal{B}$. S intersects $\partial\mathcal{B}$ in a system of loops, and each component of intersection of S with $\mathcal{B}$ contains at least one corner connection, so it intersects $N_{(1)}$ at least twice. If any component of $S \cap \mathcal{B}$ is not a disk, there is some "innermost" such component S_i; then all of its boundary components bound disks in $\mathcal{B}$, hence in S. Since S is not a sphere, one of these disks in S contains S_i. Replace it by a disk in $\mathcal{B}$. This can be done without increasing the number of intersections with $N_{(1)}$, since every loop in $\partial\mathcal{B}$ bounds a disk in $\mathcal{B}$ meeting $N_{(1)}$ at most twice.

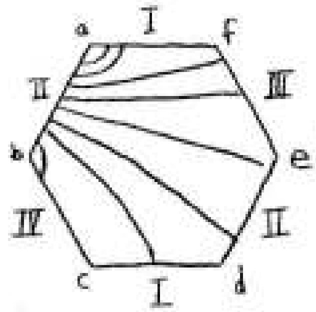
Now if there are any two corner connections in $\mathcal{B}$ which touch, then some component of $S \cap \mathcal{B}$ meets $N_{(1)}$ at least three times. Since this component is a disk, it can 4.48
be replaced by a disk which meets $N_{(1)}$ at most twice, thus simplifying S. (Therefore at most two corners can be connected at any vertex.)

Assume that S now has the minimum number of intersections with $N_{(1)}$ in its isotopy class. Let I, II, III, and IV denote the number of intersections of S with edges I, II, III, and IV, respectively (no confusion should result from this). It remains to analyze the possibilities case by case.

Suppose that none of I, II, III, and IV are zero. Then each hexagon has connections at two corners. Here are the possibilities for corner connections in hexagon A.



If the corner connections are at a and b then the picture in hexagon A is of the form



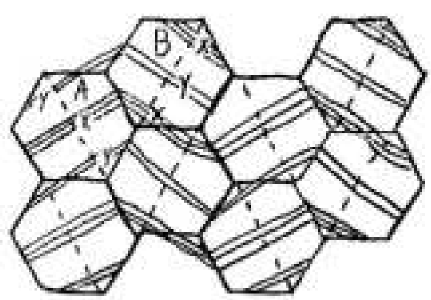
4.49

This implies that $\mathrm{II} = \mathrm{I} + \mathrm{III} + \mathrm{II} + \mathrm{I} + \mathrm{IV}$, which is impossible since all four numbers are positive in this case. A similar argument also rules out the possibilities c-d, d-e, a-f, b-f, and c-e in hexagon, and h-i, i-j, k-l, g-l, g-k and h-j in hexagon B.

The possibility a-c cannot occur since they are adjacent corners. For the same reason we can rule out a-e, b-d, d-f, g-i, i-k, h-l, and j-l.

Since each hexagon has at least two corner connections, at each vertex we must have connections at two opposite corners. This means that knowing any one corner connection also tells you another corner connection. Using this one can rule out all possible corner connections for hexagon A except for a-d.

If a-d occurs, then $\mathrm{I} + \mathrm{IV} + \mathrm{II} = \mathrm{I} + \mathrm{III} + \mathrm{II}$, or $\mathrm{III} = \mathrm{IV}$. By the requirement of opposite corners at the vertices, in hexagon B there are corner connections at i and l, which implies that $\mathrm{I} = \mathrm{II}$. Let $x = \mathrm{III}$ and $y = \mathrm{I}$. The picture is then

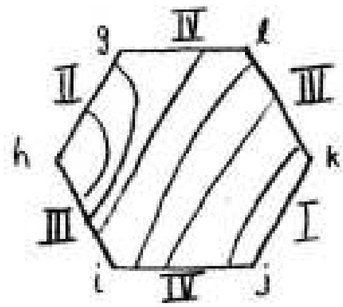


4.50

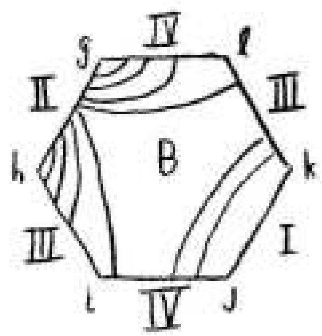
We may reconstruct the intersection of S with a neighborhood of N, say $\mathcal{N}(N)$, from this picture, by gluing together $x + y$ annuli in the indicated pattern. This yields $x+y$ punctured tori. If an x-surface is pushed down across a vertex, it yields a y-surface, and similarly, a y-surface can be pushed down to give an x-surface. Thus, $S \cap \mathcal{N}(N)$ is $x + y$ parallel copies of a punctured torus, which we see is the fiber of a fibration of $\mathcal{N}(N) \approx M$ over S^1. We will discuss later what happens outside $\mathcal{N}(N)$. (Nothing.)

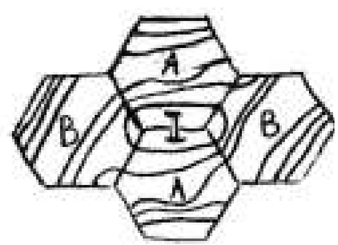
Now we pass on to the case that at least one of I, II, III, and IV are zero. The case $\mathrm{I} = 0$ is representative because of the great deal of symmetry in the picture.

First consider the subcase $\mathrm{I} = 0$ and none of II, III, and IV are zero. If hexagon B had only one corner connection, at h, then we would have $\mathrm{III} + \mathrm{IV} = \mathrm{II} + \mathrm{IV} + \mathrm{III}$,



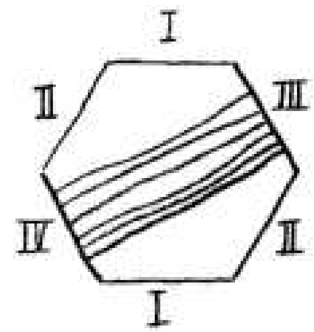
contradicting $\mathrm{II} > 0$. By the same reasoning for all the other corners, we find that hexagon B needs at least two corner connections. At most one corner connection can occur in a neighborhood of each vertex in N, since no corner connection can involve the edge I. Thus, hexagon B must have exactly two corner connections, and hexagon A has no corner connections. By checking inequalities, we find the only possibility is corner connections at g-h. If we look at the picture in the pre-image $T^2 \times \{1\}$ near I we see that there is a loop around I. This loop bounds a disk in S by incompressibility, 4.51





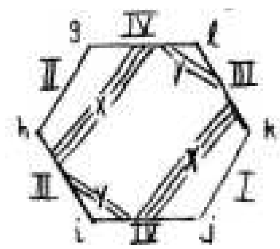
and pushing the disk across the hexagons reduces the number of intersections with $N_{(1)}$ by at least two (you lose the four intersections drawn in the picture, and gain possibly two intersections, above the plane of the paper). Since S already has minimal intersection number with $N_{(1)}$ already, this subcase cannot happen.

Now consider the subcase $\mathrm{I} = 0$ and $\mathrm{II} = 0$. In hexagon A the picture is



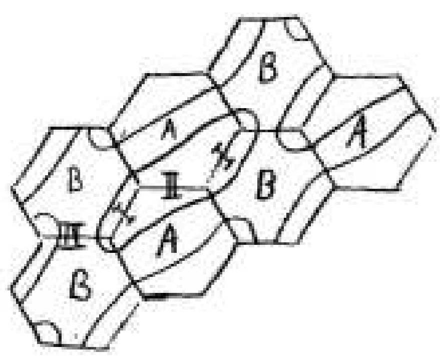
4.52

implying III = IV. The picture in hexagon B is



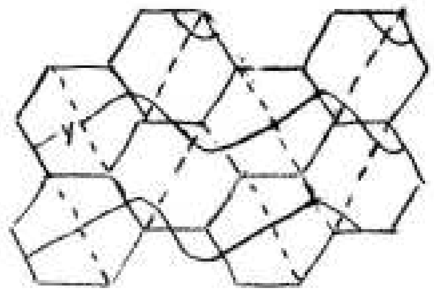
with y the number of corner connections at corner l and $x = \mathrm{IV} - y$. The three subcases to check are x and y both nonzero, $x = 0$, and $y = 0$.

If both x and y are nonzero, there is a loop in S around

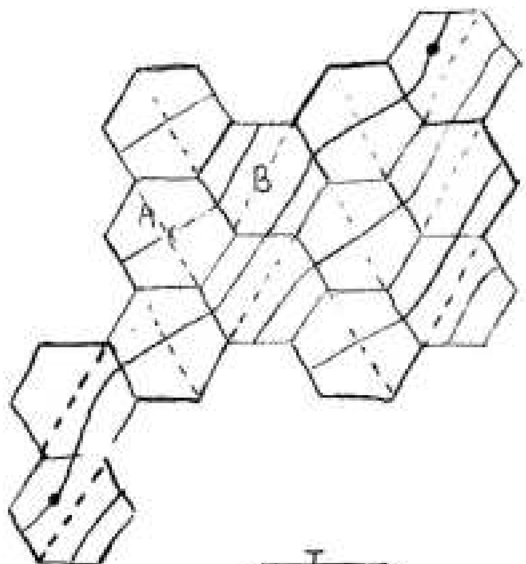


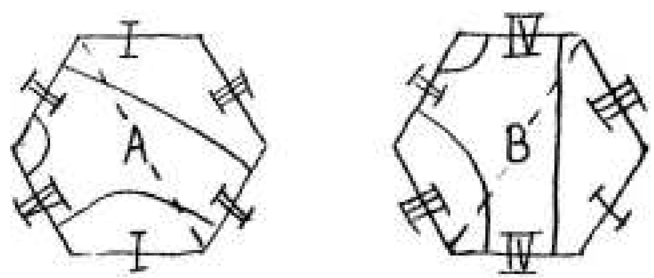
edges I and II. The loop bounds a disk in S, and pushing the disk across the hexagons reduces the number of intersections by at least two, contradicting minimality. So x and y cannot both be nonzero.

If I = II = 0 and $x = 0$, then $S \cap \mathcal{N}(N)$ is y parallel copies of a punctured torus. 4.53

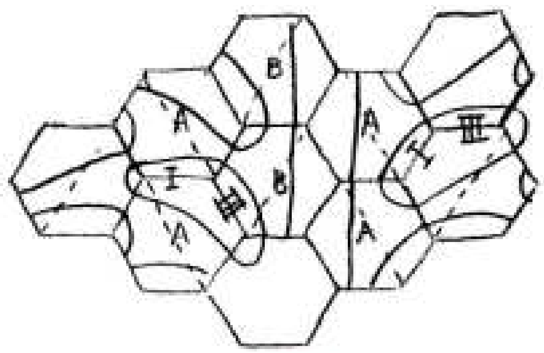


If I = II = 0 and $y = 0$, then $S \cap \mathcal{N}(N)$ consists of $\lfloor x/2 \rfloor$ copies of a twice punctured torus, together with one copy of a Klein bottle if x is odd.

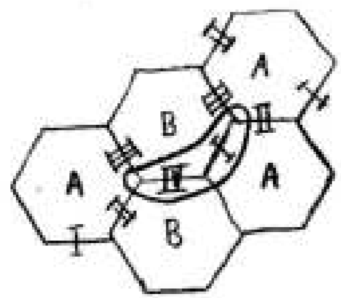




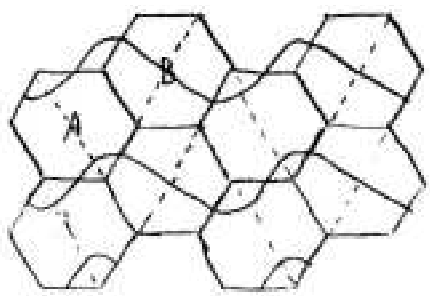
Now consider the subcase I = III = 0. If S intersects the spine N, then II $\neq 0$ 4.54
because of hexagon A and IV $\neq 0$ because of hexagon B. But this means that there is a loop around edges I and III, and S can be simplified further, contradicting minimality.

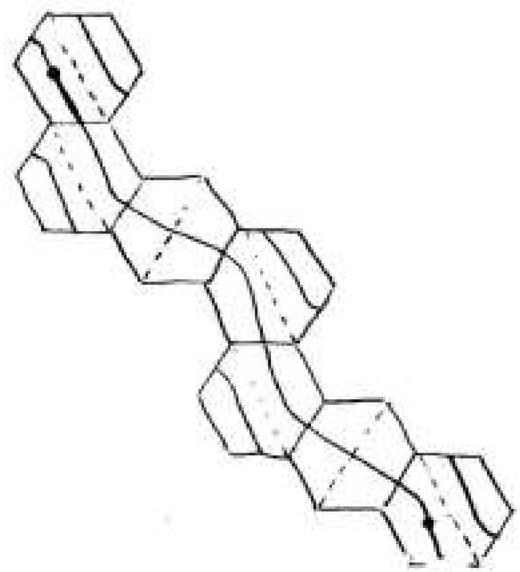


The subcase I = IV = 0 also cannot occur because of the minimality of the number of intersections of S and $N_{(1)}$. Here is the picture.



By symmetric reasoning, we find that only one more case can occur, that III = IV = 0, with I = II. The pictures are symmetric with preceding ones: 4.55





To finish the proof of the theorem, it remains to understand the behavior of S in $M - \mathcal{N}(N) = T^2 \times [0, .99]$. Clearly, $S \cap (T^2 \times [0, .99])$ must be incompressible. (Otherwise, for instance, the number of intersections of S with $N_{(1)}$ could be reduced.) It is not hard to deduce that either S is parallel to the boundary, or else a union of annuli. If one does not wish to assume S is two-sided, this may be accomplished by studying the intersection of $S \cap (T^2 \times [0, .99])$ with a non-separating annulus. 4.56
If any annulus of $S \cap (T^2 \times [0, .99])$ has both boundary components in $T^2 \times .99$, then by studying the cases, we find that S would not be incompressible. It follows that $S \cap (T^2 \times [0, .99])$ can be isotoped to the form (circles $\times$ $[0, .99]$). There are five possibilities (with S connected). Careful comparisons lead to the descriptions of $S_2, \ldots, S_6$ given on pages 4.40 and 4.41. $\square$

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TeX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

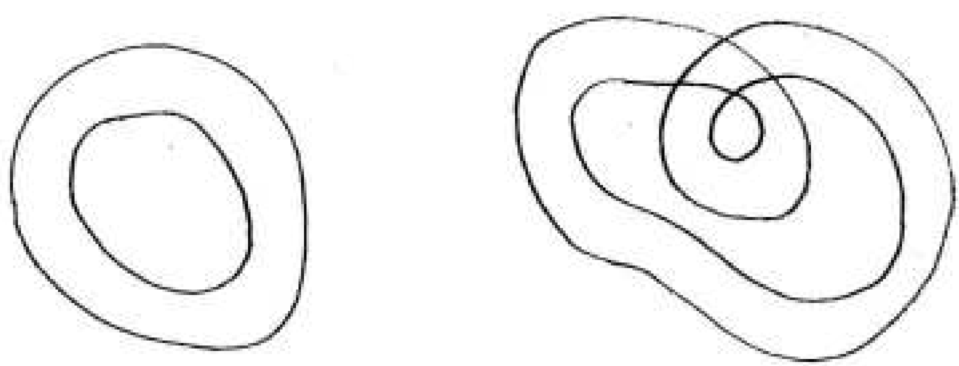
Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 5

Flexibility and rigidity of geometric structures

In this chapter we will consider deformations of hyperbolic structures and of geometric structures in general. By a geometric structure on M, we mean, as usual, a local modelling of M on a space X acted on by a Lie group G. Suppose M is compact, possibly with boundary. In the case where the boundary is non-empty we do not make special restrictions on the boundary behavior. If M is modelled on (X, G) then the developing map $\tilde{M} \xrightarrow{D} X$ defines the holonomy representation $H : \pi_1 M \longrightarrow G$. In general, H does not determine the structure on M. For example, the two immersions of an annulus shown below define Euclidean structures on the annulus which both have trivial holonomy but are not equivalent in any reasonable sense.



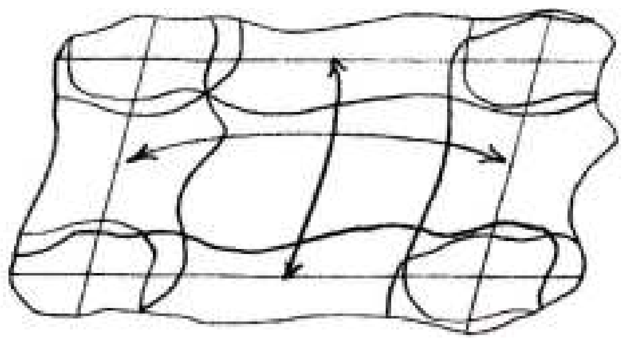
However, the holonomy is a complete invariant for (G, X)-structures on M *near* a given structure M_0, up to an appropriate equivalence relation: two structures M_1 and M_2 near M_0 are equivalent deformations of M_0 if there are submanifolds M_1' and M_2', containing all but small neighborhoods of the boundary of M_1 and M_2, with a (G, X) homeomorphism between them which is near the identity. 5.2

Let M_0 denote a fixed structure on M, with holonomy H_0.

Proposition 5.1. *Geometric structures on M near M_0 are determined up to equivalency by holonomy representations of $\pi_1 M$ in G which are near H_0, up to conjugacy by small elements of G.*

Proof. Any manifold M can be represented as the image of a disk D with reasonably nice overlapping near ∂D. Any structure on M is obtained from the structure induced on D, by gluing via the holonomy of certain elements of $\pi_1(M)$.

Any representation of $\pi_1 M$ near H_0 gives a new structure, by perturbing the identifications on D. The new identifications are still finite-to-one giving a new manifold homeomorphic to M_0.



5.3

If two structures near M_0 have holonomy conjugate by a small element of G, one can make a small change of coordinates so that the holonomy is identical. The two structures then yield nearby immersions of D into X, with the same identifications; restricting to slightly smaller disks gives the desired (G, X)-homeomorphism. $\square$

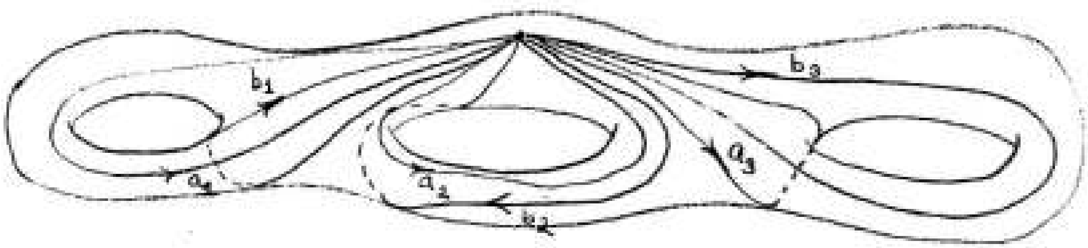
5.2

As a first approximation to the understanding of small deformations we can describe $\pi_1 M$ in terms of a set of generators $\mathcal{G} = \{g_1, \ldots, g_n\}$ and a set of relators $\mathcal{R} = \{r_1, \ldots, r_l\}$. [Each r_i is a word in the g_i's which equals 1 in $\pi_1 M$.] Any representation $\rho : \pi_1 M \to G$ assigns each generator g_i an element in G, $\rho(g_i)$. This embeds the space of representations R in $G^{\mathcal{G}}$. Since any representation of $\pi_1 M$ must respect the relations in $\pi_1 M$, the image under ρ of a relator r_j must be the identity in G. If $p : G^{\mathcal{G}} \to G^{\mathcal{R}}$ sends a set of elements in G to the $|\mathcal{R}|$ relators written with these elements, then D is just $p^{-1}(1, \ldots, 1)$. If p is generic near H_0, (i.e., if the derivative dp is surjective), the implicit function theorem implies that $\mathcal{R}$ is just a manifold of dimension $(|\mathcal{G}| - |\mathcal{R}|) \cdot (\dim G)$. One might reasonably expect this to be the case, provided the generators and relations are chosen in an efficient way. If the action of G on itself by conjugation is effective (as for the group of isometries of hyperbolic space) then generally one would also expect that the action of G on $G^{\mathcal{G}}$ by conjugation, near H_0, has orbits of the same dimension as G. If so, then the space of deformations of 5.4
M_0 would be a manifold of dimension

$$\dim G \cdot (|\mathcal{G}| - |\mathcal{R}| - 1).$$

5.2

Example. Let's apply the above analysis to the case of hyperbolic structures on closed, oriented two-manifolds of genus at least two. G in this case can be taken to be $\mathrm{PSL}(2,\mathbb{R})$ acting on the upper half-plane by linear fractional transformations. $\pi_1(M_g)$ can be presented with $2g$ generators $a_1, b_1, \dots a_g, b_g$ (see below) together with the single relator $\prod_{i=1}[a_i, b_i]$.



Since $\mathrm{PSL}(2,\mathbb{R})$ is a real three-dimensional Lie group the expected dimension of the deformation space is $3(2g-1-1) = 6g-6$. This can be made rigorous by showing directly that the derivative of the map $p : G^{\mathcal{G}} \to G^{\mathcal{R}}$ is surjective, but since we will have need for more global information about the deformation space, we won't make the computation here. 5.5

Example. The initial calculation for hyperbolic structures on an oriented three-manifold is less satisfactory. The group of isometries on H^3 preserves planes which, in the upper half-space model, are hemispheres perpendicular to $\mathbb{C} \cup \infty$ (denoted $\hat{\mathbb{C}}$). Thus the group G can be identified with the group of circle preserving maps of $\hat{\mathbb{C}}$. This is the group of all linear fractional transformations with complex coefficients $\mathrm{PSL}(2,\mathbb{C})$. (All transformations are assumed to be orientation preserving). $\mathrm{PSL}(2,\mathbb{C})$, is a *complex* Lie group with real dimensions 6. M^3 can be built from one zero-cell, a number of one- and two-cells, and (if M is closed), one 3-cell.

If M is closed, then $\chi(M) = 0$, so the number k of one-cells equals the number of two-cells. This gives us a presentation of $\pi_1 M$ with k generators and k relators. The expected (real) dimension of the deformation space is $6(k-k-1) = -6$.

If $\partial M \neq \emptyset$, with all boundary components of positive genus, this estimate of the dimension gives

5.2.1. $$6 \cdot (-\chi(M)) = 3(-\chi(\partial M)).$$

This calculation would tend to indicate that the existence of *any* hyperbolic structure on a closed three-manifold would be unusual. However, subgroups of $\mathrm{PSL}(2,\mathbb{C})$ have many special algebraic properties, so that certain relations can follow from other relations in ways which do not follow in a general group. 5.6

The crude estimate 5.2.1 actually gives some substantive information when $\chi(M) < 0$.

Proposition 5.2.2. *If M^3 possesses a hyperbolic structure M_0, then the space of small deformations of M_0 has dimension at least $6 \cdot (-\chi(M))$.*

Proof. $\mathrm{PSL}(2,\mathbb{C})^{\mathcal{G}}$ is a complex algebraic variety, and the map

$$p : \mathrm{PSL}(2,\mathbb{C})^{\mathcal{G}} \to \mathrm{PSL}(2,\mathbb{C})^{\mathcal{R}}$$

is a polynomial map (defined by matrix multiplication). Hence the dimension of the subvariety $p = (1, \dots, 1)$ is at least as great as the number of variables minus the number of defining equations. □

We will later give an improved version of 5.2.2 whenever M has boundary components which are tori.

5.3

In this section we will derive some information about the global structure of the space of hyperbolic structures on a closed, oriented surface M. This space is called the *Teichmüller space* of M and is defined to be the set of hyperbolic structures on M where two are equivalent if there is an isometry homotopic to the identity between them. In order to understand hyperbolic structures on a surface we will cut the surface up into simple pieces, analyze structures on these pieces, and study the ways they can be put together. Before doing this we need some information about closed geodesics in M.

Proposition 5.3.1. *On any closed hyperbolic n-manifold M there is a unique, closed geodesic in any non-trivial free homotopy class.*

Proof. For any $\alpha \in \pi_1 M$ consider the covering transformation T_α on the universal cover H^n of M. It is an isometry of H^n. Therefore it either fixes some interior point of H^n (elliptic), fixes a point at infinity (parabolic) or acts as a translation on some unique geodesic (hyperbolic). That all isometries of H^2 are of one of these types was proved in Proposition 4.9.3; the proof for H^n is similar.

Note. A distinction is often made between "loxodromic" and "hyperbolic" transformations in dimension 3. In this usage a loxodromic transformation means an isometry which is a pure translation along a geodesic followed by a non-trivial twist, while a hyperbolic transformation means a pure translation. This is usually not a useful distinction from the point of view of geometry and topology, so *we will use the term "hyperbolic" to cover either case.*

Since T_α is a covering translation it can't have an interior fixed point so it can't be elliptic. For any parabolic transformation there are points moved arbitrarily small distances. This would imply that there are non-trivial simple closed curves of arbitrarily small length in M. Since M is closed this is impossible. Therefore T_α translates a unique geodesic, which projects to a closed geodesic in M. Two geodesics corresponding to the translations T_α and T'_α project to the same geodesic in M if 5.8
and only if there is a covering translation taking one to the other. In other words, $\alpha' = g\alpha g^{-1}$ for some $g \in \pi_1 M$, or equivalently, α is free homotopic to α. □

Proposition 5.3.2. *Two distinct geodesics in the universal cover H^n of M which are invariant by two covering translations have distinct endpoints at ∞.*

Proof. If two such geodesics had the same endpoint, they would be arbitrarily close near the common endpoint. This would imply the distance between the two closed geodesics is uniformly $\leq \epsilon$ for all ϵ, a contradiction. □

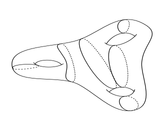
Proposition 5.3.3. *In a hyperbolic two-manifold M^2 a collection of homotopically distinct and disjoint nontrivial simple closed curves is represented by disjoint, simple closed geodesics.*

Proof. Suppose the geodesics corresponding to two disjoint curves intersect. Then there are lifts of the geodesics in the universal cover H^2 which intersect. Since the endpoints are distinct, the pairs of endpoints for the two geodesics must link each other on the circle at infinity. Consider any homotopy of the closed geodesics in M^2. It lifts to a homotopy of the geodesics in H^2. However, no homotopy of the geodesics moving points only a finite hyperbolic distance can move their endpoints; thus the images of the geodesics under such a homotopy will still intersect, and this intersection projects to one in M^2. 5.9

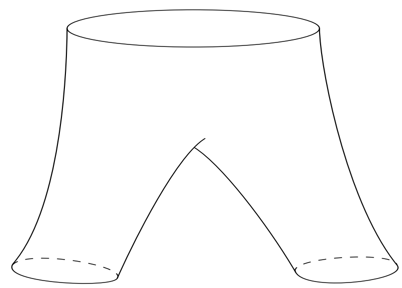
The proof that the closed geodesic corresponding to a simple closed curve is simple is similar. The argument above is applied to two different lifts of the same geodesic. □

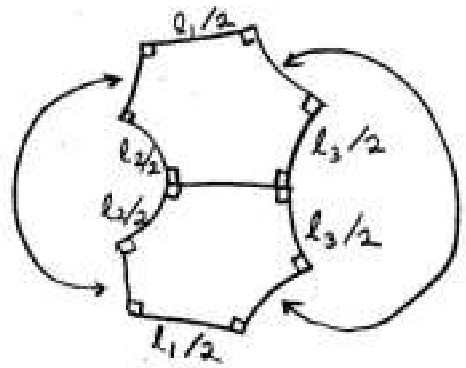
Now we are in a position to describe the Teichmüller space for a closed surface. The coordinates given below are due to Nielsen and Fenchel.

Pick $3g-3$ disjoint, non-parallel simple closed curves on M^2. (This is the maximum number of such curves on a surface of genus g.) Take the corresponding geodesics and cut along them. This divides M^2 into $2g-2$ surfaces homeomorphic to S^2—three disks (called "pairs of pants" from now on) with geodesic boundary.



5.9a





5.10

On each pair of pants P there is a unique arc connecting each pair of boundary components, perpendicular to both. To see this, note that there is a unique homotopy class for each connecting arc. Now double P along the boundary geodesics to form a surface of genus two. The union of the two copies of the arcs connecting a pair of boundary components in P defines a simple closed curve in the closed surface. There is a unique geodesic in its free homotopy class and it is invariant under the reflection which interchanges the two copies of P. Hence it must be perpendicular to the geodesics which were in the boundary of P.

This information leads to an easy computation of the Teichmüller space of P.

Proposition 5.3.4. *$\mathcal{T}(P)$ is homeomorphic to $\mathbb{R}^3$ with coordinates*

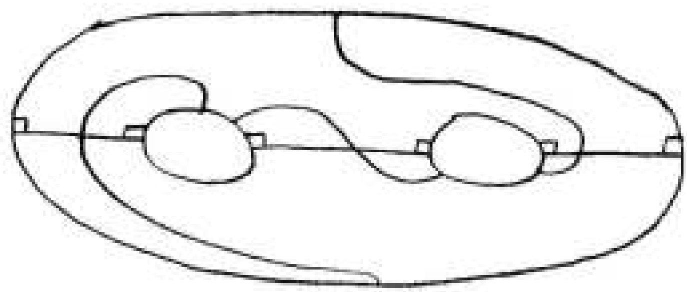
$$(\log l_1, \log l_2, \log l_3),$$

where l_i = length of the i-th boundary component.

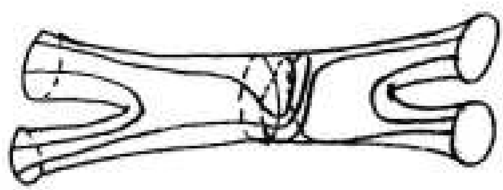
Proof. The perpendicular arcs between boundary components divide P into two right-angled hexagons. The hyperbolic structure of an all-right hexagon is determined by the lengths of three alternating sides. (See page 2.19.) The lengths of the connecting arcs therefore determine both hexagons so the two hexagons are isometric. Reflection in these arcs is an isometry of the hexagons and shows that the boundary curves are divided in half. The lengths $l_i/2$ determine the hexagons; hence they also determine P. Any positive real values for the l_i are possible so we are done. $\square$ 5.11

In order to determine the hyperbolic structure of the closed two-manifold from that of the pairs of pants, some measurement of the twist with which the boundary geodesics are attached is necessary. Find $3g-3$ more curves in the closed manifold which, together with the first set of curves, divides the surface into hexagons.

In the pairs of pants the geodesics corresponding to these curves are arcs connecting the boundary components. However, they may wrap around the components. In P it is possible to isotope these arcs to the perpendicular connecting arcs discussed above. Let $2d_i$ denote the total number of degrees which this isotopy moves the feet of arcs which lie on the i-th boundary component of p.



5.12



Of course there is another copy of this curve in another pair of pants which has a twisting coefficient d_i'. When the two copies of the geodesic are glued together they cannot be twisted independently by an isotopy of the *closed* surface. Therefore $(d_i - d_i') = \tau_i$ is an isotopy invariant.

Remark. If a hyperbolic surface is cut along a closed geodesic and glued back together with a twist of $2\pi n$ degrees (n an integer), then the resulting surface is isometric to the original one. However, the isometry is not isotopic to the identity so the two surfaces represent distinct points in Teichmüller space. Another way to say this is that they are isometric as surfaces but not as *marked* surfaces. It follows that τ_i is a well-defined real number, not just defined up to integral multiples of 2π.

Theorem 5.3.5. *The Teichmüller space $\mathcal{T}(M)$ of a closed surface of genus g is homeomorphic to $\mathbb{R}^{6g-6}$. There are explicit coordinates for $\mathcal{T}(M)$, namely*

$$(\log l_1, \tau_1, \log l_2, \tau_2, \ldots, \log l_{3g-3}, \tau_{3g-3}),$$

where l_i is the length and τ_i the twist coefficient for a system of $3g-3$ simple closed 5.13
geodesics.

In order to see that it takes precisely $3g-3$ simple closed curves to cut a surface of genus g into pairs of pants P_i notice that $\chi(P_i) = -1$. Therefore the number of P_i's is equal to $-\chi(M_g) = 2g-2$. Each P_i has three curves, but each curve appears in two P_i's. Therefore the number of curves is $\frac{3}{2}(2g-2) = 3g-3$. We can rephrase Theorem 5.3.5 as

$$\mathcal{T}(M) \approx \mathbb{R}^{-3\chi(M)}.$$

It is in this form that the theorem extends to a surface with boundary.

The *Fricke space* $\mathcal{F}(M)$ of a surface M with boundary is defined to be the space of hyperbolic structures on M such that the boundary curves are geodesics, modulo isometries isotopic to the identity. A surface with boundary can also be cut into pairs of pants with geodesic boundary. In this case the curves that were boundary curves in M have no twist parameter. On the other hand these curves appear in only one pair of pants. The following theorem is then immediate from the gluing procedures above.

THEOREM 5.3.6. *$\mathcal{F}(M)$ is homeomorphic to $\mathbb{R}^{-3\chi(M)}$.*

5.14

The definition of Teichmüller space can be extended to general surfaces as the space of all metrics of constant curvature up to isotopy and change of scale. In the case of the torus T^2, this space is the set of all Euclidean structures (i.e., metrics with constant curvature zero) on T^2 with area one. There is still a closed geodesic in each free homotopy class although it is not unique. Take some simple, closed geodesic on T^2 and cut along it. The Euclidean structure on the resulting annulus is completely determined by the length of its boundary geodesic. Again there is a real twist parameter that determines how the annulus is glued to get T^2. Therefore there are two real parameters which determine the flat structures on T^2, the length l of a simple, closed geodesic in a fixed free homotopy class and a twist parameter τ along that geodesic.

THEOREM 5.3.7. *The Teichmüller space of the torus is homeomorphic to $\mathbb{R}^2$ with coordinates $(\log l, \tau)$, where l, τ are as above.*

5.4. Special algebraic properties of groups of isometries of H^3.

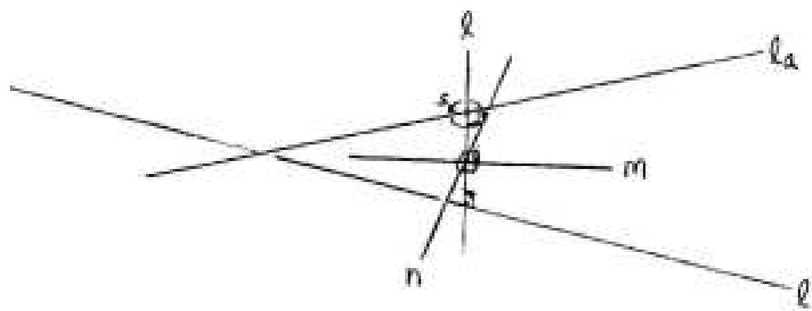
On large open subsets of $\mathrm{PSL}(2,\mathbb{C})^{\mathcal{G}}$, the space of representations of a generating set $\mathcal{G}$ into $\mathrm{PSL}(2,\mathbb{C})$, certain relations imply other relations. This fact was anticipated in the previous section from the computation of the expected dimension of small deformations of hyperbolic structures on closed three manifolds. The phenomenon that dp is not surjective (see 5.3) suggests that, to determine the structure of $\pi_1 M^3$ 5.15
as a discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$, not all the relations in $\pi_1 M^3$ as an abstract group are needed. Below are some examples.

gJørgensen

Proposition 5.4.1 (Jørgensen). *Let a, b be two isometries of H^3 with no common fixed point at infinity. If $w(a, b)$ is any word such that $w(a, b) = 1$ then $w(a^{-1}, b^{-1}) = 1$. If a and b are conjugate (i.e., if* $\operatorname{Trace}(a) = \pm \operatorname{Trace}(b)$ *in* $\mathrm{PSL}(2, \mathbb{C})$ *) then also $w(b, a) = 1$.*

Proof. If a and b are hyperbolic or elliptic, let l be the unique common perpendicular for the invariant geodesics l_a, l_b of a and b. (If the geodesics intersect in a point x, l is taken to be the geodesic through x perpendicular to the plane spanned by l_a and l_b). If one of a and b is parabolic, (say b is) l should be perpendicular to l_a and pass through b's fixed point at ∞. If both are parabolic, l should connect the two fixed points at infinity. In all cases rotation by 180° in l takes a to a^{-1} and b and b^{-1}, hence the first assertion.

If a and b are conjugate hyperbolic elements of $\mathrm{PSL}(2, \mathbb{C})$ with invariant geodesics l_a and l_b, take the two lines m and n which are perpendicular to l and to each other and which intersect l at the midpoint between g_b and l_a. Also, if g_b is at an angle of θ to l_b along l then m should be at an angle of $\theta/2$ and n at an angle of $\theta + \pi/2$. 5.16



Rotations of 180° through m or n take l_a to l_b and vice versa. Since a and b are conjugate they act the same with respect to their respective fixed geodesics. It follows that the rotations about m and n conjugate a to b (and b to a) or a to b^{-1} (and b to a^{-1}).

If one of a and b is parabolic then they both are, since they are conjugate. In this case take m and n to be perpendicular to l and to each other and to pass through the unique point x on l such that $d(x, ax) = d(x, bx)$. Again rotation by 180° in m and n takes a to b or a to b^{-1}. □

Remarks. 1. This theorem fails when a and b are allowed to have a common fixed point. For example, consider

$$a = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix},$$

where $\lambda \in \mathbb{C}^*$. Then

$$(b^{-k}ab^k)^l = b^{-k}a^l b^k = \begin{bmatrix} 1 & l\lambda^{2k} \\ 0 & 1 \end{bmatrix}.$$

5.17

If λ is chosen so that λ^2 is a root of a polynomial over $\mathbb{Z}$, say $1 + 2\lambda^2 = 0$, then a relation is obtained: in this case

$$w(a, b) = (a)(bab^{-1})^2 = I.$$

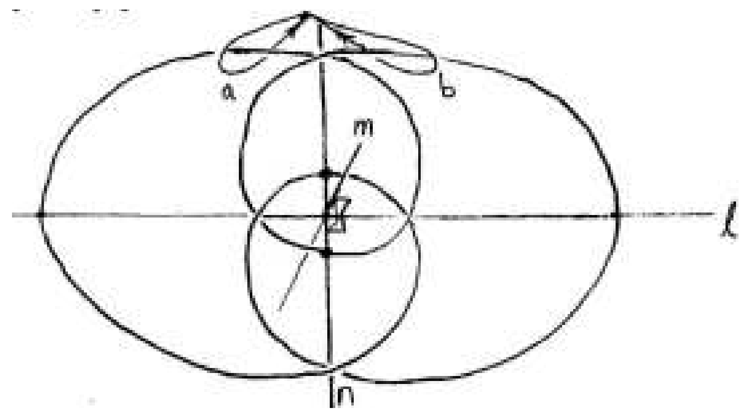
However, $w(a^{-1}, b^{-1}) = I$ only if λ^{-2} is a root of the same polynomial. This is not the case in the current example.

2. The geometric condition that a and b have a common fixed point at infinity implies the algebraic condition that a and b generate a solvable group. (In fact, the commutator subgroup is abelian.)

Geometric Corollary 5.4.2. *Any complete hyperbolic manifold M^3 whose fundamental group is generated by two elements a and b admits an involution s (an isometry of order 2) which takes a to a^{-1} and b to b^{-1}. If the generators are conjugate, there is a $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ action on M generated by s together with an involution t which interchanges a and b unless a and b have a common fixed point at infinity.*

Proof. Apply the rotation of 180° about l to the universal cover H^3. This conjugates the group to itself so it induces an isometry on the quotient space M^3. The same is true for rotation around m and n in the case when a and b are conjugate. It can happen that a and b have a common fixed point x at infinity, but since the 5.18
group is discrete they must both be parabolic. A 180° rotation about any line through x sends a to a^{-1} and b to b^{-1}. There is not generally a symmetry group of order four in this case. $\square$

As an example, the complete hyperbolic structure on the complement of the figure-eight knot has symmetry implied by this corollary. (In fact the group of symmetries extends to S^3 itself, since for homological reasons such a symmetry preserves the meridian direction.)



Here is another illustration of how certain relations in subgroups of $\mathrm{PSL}(2, \mathbb{C})$ can imply others:

PROPOSITION 5.4.3. *Suppose a and b are not elliptic. If $a^n = b^m$ for some $n, m \neq 0$, then a and b commute.*

PROOF. If $a^n = b^m$ is hyperbolic, then so are a and b. In fact they fix the same geodesic, acting as translations (perhaps with twists) so they commute. If $a^n = b^m$ 5.19
is parabolic then so are a and b. They must fix the same point at infinity so they act as Euclidean transformations of any horosphere based there. It follows that a and b commute. □

PROPOSITION 5.4.3. *If a is hyperbolic and a^k is conjugate to a^l then $k = \pm l$.*

PROOF. Since translation distance along the fixed line is a conjugacy invariant and $\rho\left(a^k\right) = \pm k \rho\left(a\right)$ (where $\rho\left(\ \right)$ denotes translation distance), the proposition is easy to see. □

Finally, along the same vein, it is sometimes possible to derive some nontrivial topological information about a hyperbolic three-manifold from its fundamental group.

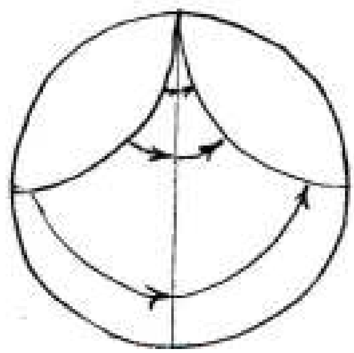
PROPOSITION 5.4.4. *If M^3 is a complete, hyperbolic three-manifold, $a, b \in \pi_1 M^3$ and $[a, b] = 1$, then either*

(i) *a and b belong to an infinite cyclic subgroup generated by x and $x^l = a$, $x^k = b$, or*
(ii) *M has an end, E, homeomorphic to $T^2 \times [0, \infty)$ such that the group generated by a and b is conjugate in $\pi_1 M^3$ to a subgroup of finite index in $\pi_1 E$.*

PROOF. If a and b are hyperbolic then they translate the same geodesic. Since $\pi_1 M^3$ acts as a discrete group on H^3, a and b must act discretely on the fixed geodesic. 5.20

Thus, (i) holds.

If a and b are not both hyperbolic, they must both be parabolic, since they commute. Therefore they can be thought of as Euclidean transformations on a set of horospheres. If the translation vectors are not linearly independent, a and b generate a group of translations of $\mathbb{R}$ and (i) is again true. If the vectors are linearly independent, a and b generate a lattice group $L_{a,b}$ on $\mathbb{R}^2$. Moreover as one approaches the fixed point at infinity, the hyperbolic distance a point x is moved by a and b goes to zero.



Recall that the subgroup $G_\epsilon(X)$ of $\pi_1 M^3$ generated by transformations that moves a point x less than ϵ is abelian. (See pages 4.34-4.35). Therefore all the elements of $G_\epsilon(X)$ commute with a and b and fix the same point p at infinity. By discreteness $G_\epsilon(X)$ acts as a lattice group on the horosphere through x and contains $L_{a,b}$ as a subgroup of finite index. 5.21

Consider a fundamental domain of $G_\epsilon(X)$ acting on the set of horocycles at p which are "contained" in the horocycle H_x through x. It is homeomorphic to the product of a fundamental domain of the lattice group acting on H_x with $[0, \infty)$ and is moved away from itself by all elements in $\pi_1 M^3$ not in $G_\epsilon(X)$. Therefore it is projected down into M^3 as an end homeomorphic to $T^2 \times [0, 1]$. This is case (ii). $\square$

5.22

5.5. The dimension of the deformation space of a hyperbolic three-manifold.

Consider a hyperbolic structure M_0 on $T^2 \times I$. Let α and β be generators for $\mathbb{Z} \oplus \mathbb{Z} = \pi_1(T^2 \times I)$; they satisfy the relation $[\alpha, \beta] = 1$, or equivalently $\alpha\beta = \beta\alpha$. The representation space for $\mathbb{Z} \oplus \mathbb{Z}$ is defined by the equation

$$H(\alpha)\, H(\beta) = H(\beta)\, H(\alpha),$$

where $H(\alpha), H(\beta) \in \mathrm{PSL}(2, \mathbb{C})$. But we have the identity

$$\mathrm{Tr}(H(\alpha)\, H(\beta)) = \mathrm{Tr}(H(\beta)\, H(\alpha)),$$

as well as $\det(H(\alpha)\,H(\beta)) = \det(H(\beta)\,H(\alpha)) = 1$, so this matrix equation is equivalent to two ordinary equations, at least in a neighborhood of a particular non-trivial solution. Consequently, the solution space has a complex dimension four, and the deformation space of M_0 has complex dimension two. This can easily be seen directly: $H(\alpha)$ has one complex degree of freedom to conjugacy, and given $H(\alpha) \neq \mathrm{id}$, there is a one complex-parameter family of transformations $H(\beta)$ commuting with it. This example shows that 5.2.2 is not sharp. More generally, we will improve 5.2.2 for any compact oriented hyperbolic three-manifold M_0 whose boundary contains toruses, under a mild nondegeneracy condition on the holonomy of M_0:

THEOREM 5.6. *Let M_0 be a compact oriented hyperbolic three-manifold whose holonomy satisfies*

(a) *the holonomy around any component of ∂M homeomorphic with T^2 is not trivial, and*
(b) *the holonomy has no fixed point on the sphere at ∞.*

Under these hypotheses, the space of small deformations of M_0 has dimension at least as great as the total dimension of the Teichmüller space of ∂M, that is,

$$\dim_{\mathbb{C}}(\mathrm{Def}(M)) \geq \sum_i \begin{cases} +3\left|\chi((\partial M)_i)\right| & \text{if } \chi((\partial M)_i) < 0, \\ 1 & \text{if } \chi((\partial M)_i) = 0, \\ 0 & \text{if } \chi((\partial M)_i) > 0. \end{cases}$$

REMARK. Condition (b) is equivalent to the statement that the holonomy representation in $\mathrm{PSL}(2, \mathbb{C})$ is irreducible. It is also equivalent to the condition that the holonomy group (the image of the holonomy) be solvable.

EXAMPLES. If N is any surface with nonempty boundary then, by the immersion theorem [Hirsch] there is an immersion ϕ of $N \times S^1$ in $N \times I$ so that ϕ sends $\pi_1(N)$ to $\pi_1(N \times I) = \pi_1(N)$ by the identity map. Any hyperbolic structure on $N \times I$ has a $-6\chi(N)$ complex parameter family of deformations. This induces a $(-6\chi(N))$-parameter family of deformations of hyperbolic structures on $N \times S^1$, showing that the inequality of 5.6 is not sharp in general.

Another example is supplied by the complement M_k of k unknotted unlinked solid tori in S^3. Since $\pi_1(M_k)$ is a free group on k generators, every hyperbolic structure on M_k has at least $3k - 3$ degrees of freedom, while 5.6 guarantees only k degrees of freedom. Other examples are obtained on more interesting manifolds by considering hyperbolic structures whose holonomy factors through a free group.

PROOF OF 5.6. We will actually prove that for any compact oriented manifold M, the complex dimension of the representation space of $\pi_1 M$, near a representation satisfying (a) and (b), is at least 3 greater than the number given in 5.6; this suffices, by 5.1. For this stronger assertion, we need only consider manifolds which have no boundary component homeomorphic to a sphere, since any three-manifold M has the same fundamental group as the manifold $\hat{M}$ obtained by gluing a copy of D^3 to each spherical boundary component of M.

Remark. Actually, it can be shown that when $\partial M \neq 0$, a representation

$$\rho : \pi_1 M \to \mathrm{PSL}(2, \mathbb{C})$$

is the holonomy of some hyperbolic structure for M if and only if it lifts to a representation in $\mathrm{SL}(2, \mathbb{C})$. (The obstruction to lifting is the second Stiefel–Whitney class ω_2 of the associated H^3-bundle over M.) It follows that if H_0 is the holonomy of a hyperbolic structure on M, it is also the holonomy of a hyperbolic structure on $\hat{M}$, provided $\partial \hat{M} \neq \emptyset$. Since we are mainly concerned with structures which have more geometric significance, we will not discuss this further.

Let H_0 denote any representation of $\pi_1 M$ satisfying (a) and (b) of 5.6. Let $T_1, \ldots, T_k$ be the components of ∂M which are toruses.

Lemma 5.6.1. *For each i, $1 \leq i \leq k$, there is an element $\alpha_i \in \pi_1(M)$ such that the group generated by $H_0(\alpha_i)$ and $H_0(\pi_1(T_i))$ has no fixed point at ∞. One can choose α_i so $H_0(\alpha_i)$ is not parabolic.*

Proof of 5.6.1. If $H_0(\pi_1 T_i)$ is parabolic, it has a unique fixed point x at ∞ and the existence of an α_i' not fixing x is immediate from condition (b). If $H_0(\pi_1 T_i)$ has two fixed points x_1 and x_2, there is $H_0(\beta_1)$ not fixing x_1 and $H_0(\beta_2)$ not fixing x_2. If $H_0(\beta_1)$ and $H_0(\beta_2)$ each have common fixed points with $H_0(\pi_1 T_i)$, $\alpha_1' = \beta_1 \beta_2$ does not.

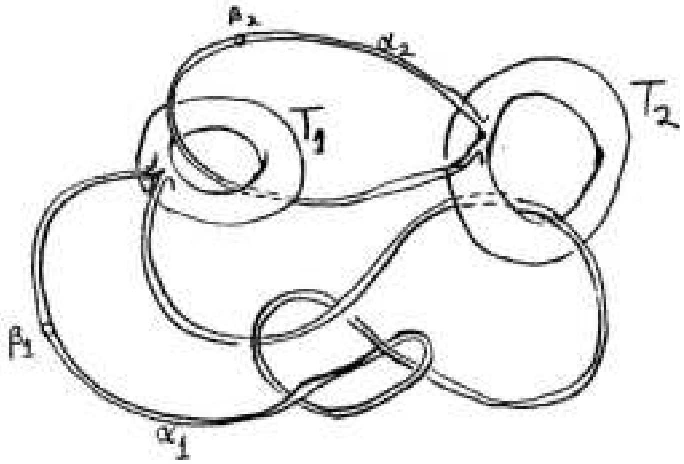
If $H_0(\alpha_i')$ is parabolic, consider the commutators $\gamma_n = [\alpha_i'^n, \beta]$ where $\beta \in \pi_1 T_i$ is some element such that $H_0(\beta) \neq 1$. If $H_0[\alpha_i'^n, \beta]$ has a common fixed point x with $H_0(\beta)$ then also $\alpha_i'^n \beta \alpha_i'^{-n}$ fixes x so β fixes $\alpha_i'^{-n} x$; this happens for at most three values of n. We can, after conjugation, take $H_0(\alpha_i') = \left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$. Write

$$H_0(\beta \alpha_i'^{-1} \beta^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

where $a + d = 2$ and $c \neq 0$ since $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ is not an eigenvector of β. We compute $\mathrm{Tr}(\gamma_n) = 2 + n^2 c$; it follows that γ_n can be parabolic ($\Leftrightarrow \mathrm{Tr}(\gamma_n) = \pm 2$) for at most 3 values of n. This concludes the proof of Lemma 5.6.1. □

Let $\{\alpha_i, 1 \leq i \leq k\}$ be a collection of simple disjoint curves based on T_i and 5.26
representing the homotopy classes of the same names. Let $N \subset M$ be the manifold obtained by hollowing out nice neighborhoods of the α_i. Each boundary component

of N is a surface of genus ≥ 2, and M is obtained by attaching k two-handles along non-separating curves on genus two surfaces $S_1, \ldots, S_k \subset \partial N$.



Let α_i also be represented by a curve of the same name on S_i, and let β_i be a curve on S_i describing the attaching map for the i-th two-handle. Generators γ_i, δ_i can be chosen for $\pi_1 T_i$ so that $\alpha_i, \beta_i, \gamma_i$, and δ_i generate $\pi_1 B_i$ and $[\alpha_i, \beta_i] \cdot [\gamma_i, \delta_i] = 1$. $\pi_1 M$ is obtained from $\pi_1 M$ by adding the relations $\beta_i = 1$.

Lemma 5.6.2. *A representation ρ of $\pi_1 N$ near H_0 gives a representation of $\pi_1 M$ if and only if the equations* 5.27

$$\begin{aligned} \operatorname{Tr}(\rho(\beta_i)) &= 2 \\ \textit{and } \operatorname{Tr}(\rho[\alpha_i, \beta_i]) &= 2 \end{aligned}$$

are satisfied.

Proof of 5.6.2. Certainly if ρ gives a representation of $\pi_1 M$, then $\rho(\beta_i)$ and $\rho[\alpha_i, \beta_i]$ are the identity, so they have trace 2.

To prove the converse, consider the equation

$$\operatorname{Tr}[A, B] = 2$$

in $\mathrm{SL}(2, \mathbb{C})$. If A is diagonalizable, conjugate so that

$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}.$$

Write

$$BA^{-1}B^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

We have the equations

$$a + d = \lambda + \lambda^{-1}$$

$$\mathrm{Tr}\,[A,B] = \lambda a + \lambda^{-1} d = 2$$

which imply that

$$a = \lambda^{-1},\ d = \lambda.$$

Since $ad - bc = 1$ we have $bc = 0$. This means B has at least one common eigenvector $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ or $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ with A; if $[A,B] \neq 1$, this common eigenvector is the unique eigenvector 5.28
of $[A,B]$ (up to scalars). As in the proof of 5.6.1, a similar statement holds if A is parabolic. (Observe that $[A,B] = [-A,B]$, so the sign of $\mathrm{Tr}\,A$ is not important).

It follows that if $\mathrm{Tr}\ \rho[\alpha_i,\beta_i] = 2$, then since $[\gamma_i,\delta_i] = [\alpha_i,\beta_i]$, either $\rho\,(\alpha_i)$, $\rho\,(\beta_i)$, $\rho\,(\gamma_i)$ and $\rho\,(\delta_i)$ all have a common fixed point on the sphere at infinity, or $\rho[\alpha_i,\beta_i] = 1$.

By construction $H_0, \pi_1 S_i$ has no fixed point at infinity, so for ρ near $H_0 \rho\,\pi_1 S_i$ cannot have a fixed point either; hence $\rho[\alpha_i,\beta_i] = 1$.

The equation $\mathrm{Tr}\ \rho\,(\beta_i) = 2$ implies $\rho\,(\beta_i)$ is parabolic; but it commutes with $\rho\,(\beta_i)$, which is hyperbolic for ρ near H_0. Hence $\rho\,(\beta_i) = 1$. This concludes the proof of Lemma 5.6.2. □

To conclude the proof of 5.6, we consider a handle structure for N with one zero-handle, m one-handles, p two-handles and no three-handles (provided $\partial M \neq \emptyset$). This gives a presentation for $\pi_1 N$ with m generators and p relations, where

$$1 - m + p = \chi(N) = \chi(M) - k.$$

The representation space $R \subset \mathrm{PSL}(2,\mathbb{C})^m$ for $\pi_1 M$, in a neighborhood of H_0, is defined by the p matrix equations

$$r_i = 1, \quad (1 \leq i \leq p),$$

where the r_i are products representing the relators, together with $2k$ equations 5.29

$$\mathrm{Tr}\,\rho(\beta_i) = 2$$

$$\mathrm{Tr}\,\rho([\alpha_i,\beta_i]) = 2 \quad [1 \leq i \leq k]$$

The number of equations minus the number of unknowns (where a matrix variable is counted as three complex variables) is

$$3m - 3p - 2k = -3\chi(M) + k + 3.$$

□

Remark. If M is a closed hyperbolic manifold, this proof gives the estimate of 0 for $\dim_{\mathbb{C}} \mathrm{def}(M)$: simply remove a non-trivial solid torus from M, apply 5.6, and fill in the solid torus by an equation $\mathrm{Tr}(\gamma) = 2$.

5.7

There is a remarkable, precise description for the global deformation space of hyperbolic structures on closed manifolds in dimensions bigger than two:

THEOREM 5.7.1 (Mostow's Theorem [algebraic version]). *Suppose Γ_1 and Γ_2 are two discrete subgroups of the group of isometries of H^n, $n \geq 3$ such that H^n/Γ_i has finite volume and suppose $\phi : \Gamma_1 \to \Gamma_2$ is a group isomorphism. Then Γ_1 and Γ_2 are conjugate subgroups.*

This theorem can be restated in terms of hyperbolic manifolds since a hyperbolic manifold has universal cover H^n with fundamental group acting as a discrete group of isometries. 5.30

THEOREM 5.7.2 (Mostow's Theorem [geometric version]). *If M_1^n and M_2^n are complete hyperbolic manifolds with finite total volume, any isomorphism of fundamental groups $\phi : \pi_1 M_1 \to \pi_1 M_2$ is realized by a unique isometry.*

REMARK. Multiplication by an element in either fundamental group induces the identity map on the manifolds themselves so that ϕ needs only to be defined up to composition with inner automorphisms to determine the isometry from M_1 to M_2.

Since the universal cover of a hyperbolic manifold is H^n, it is a $K(\pi, 1)$. Two such manifolds are homotopy equivalent if and only if there is an isomorphism between their fundamental groups.

COROLLARY 5.7.3. *If M_1 and M_2 are hyperbolic manifolds which are complete with finite volume, then they are homeomorphic if and only if they are homotopy equivalent.* (*The case of dimension two is well known.*)

For any manifold M, there is a homomorphism $\operatorname{Diff} M \to \operatorname{Out}(\pi_1 M)$, where $\operatorname{Out}(\pi_1 M) = \operatorname{Aut}(\pi_1 M)/\operatorname{Inn}(\pi_1 M)$ is the group of outer automorphisms. Mostow's Theorem implies this homomorphism splits, if M is a hyperbolic manifold of dimension $n \geq 3$. It is unknown whether the homomorphism splits when M is a surface. When $n = 2$ the kernel $\operatorname{Diff}_0(M)$ is contractible, provided $\chi(M) \leq 0$. If M is a Haken 5.31
three-manifold which is not a Seifert fiber space, Hatcher has shown that $\operatorname{Diff}_0 M$ is contractible.

COROLLARY 5.7.4. *If M^n is hyperbolic (complete, with finite total volume) and $n \geq 3$, then $\operatorname{Out}(\pi_1 M)$ is a finite group, isomorphic to the group of isometries of M^n.*

PROOF. By Mostow's Theorem any automorphism of $\pi_1 M$ induces a unique isometry of M. Since any inner automorphism induces the identity on M, it follows that the group of isometries is isomorphic to $\operatorname{Out}(\pi_1 M)$. That $\operatorname{Out}(\pi_1 M)$ is finite is immediate from the fact that the group of isometries, $\operatorname{Isom}(M^n)$, is finite.

To see that $\mathrm{Isom}(M^n)$ is finite, choose a base point and frame at that point and suppose first that M is compact. Any isometry is completely determined by the image of this frame (essentially by "analytic continuation"). If there were an infinite sequence of isometries there would exist two image frames close to each other. Since M is compact, the isometries, ϕ_1, ϕ_2, corresponding to these frames would be close on all of M. Therefore ϕ, is homotopic to ϕ_2. Since the isometry $\phi_2^{-1}\phi_1$ induces the trivial outer automorphism on $\pi_1 M$, it is the identity; i.e., $\phi_2 = \phi_1$.

If M is not compact, consider the submanifold $M_\epsilon \subset M$ which consists of points which are contained in an embedded hyperbolic disk of radius ϵ. Since M has finite total volume, M_ϵ is compact. Moreover, it is taken to itself under any isometry. The 5.32
argument above applied to M_ϵ implies that the group of isometries of M is finite even in the non-compact case. □

Remark. This result contrasts with the case $n = 2$ where $\mathrm{Out}(\pi_1 M^2)$ is infinite and quite interesting.

The proof of Mostow's Theorem in the case that H^n/Γ is not compact was completed by Prasad. Otherwise, 5.7.1 and 5.7.2 (as well as generalizations to other homogeneous spaces) are proved in Mostow. We shall discuss Mostow's proof of this theorem in 5.10, giving details as far as they can be made geometric. Later, we will give another proof due to Gromov, valid at least for $n = 3$.

5.8. Generalized Dehn surgery and hyperbolic structures.

Let M be a non-compact, hyperbolic three-manifold, and suppose that M has a finite number of ends $E_1, \dots, E_k$, each homeomorphic to $T^2 \times [0, \infty)$ and isometric to the quotient space of the region in H^3 (in the upper half-space model) above an interior Euclidean plane by a group generated by two parabolic transformations which fix the point at infinity. Topologically M is the interior of a compact manifold $\bar{M}$ whose boundary is a union of $T_1, \dots, T_k$ tori.

Recall the operation of generalized Dehn surgery on M (§4.5); it is parametrized 5.33
by an ordered pair of real numbers (a_i, b_i) for each end which describes how to glue a solid torus to each boundary component. If nothing is glued in, this is denoted by ∞ so that the parameters can be thought of as belonging to S^2 (i.e., the one point compactification of $\mathbb{R}^2 \approx H_1(T^2, \mathbb{R})$). The resulting space is denoted by $M_{d_1,\dots,d_k}$ where $d_i = (a_i, b_i)$ or ∞.

In this section we see that the new spaces often admit hyperbolic structures. Since $M_{d_1,\dots,d_k}$ is a closed manifold when $d_i = (a_i, b_i)$ are primitive elements of $H_1(T^2, \mathbb{Z})$, this produces many closed hyperbolic manifolds. First it is necessary to see that small deformations of the complete structure on M induce a hyperbolic structure on *some* space $M_{d_1,\dots,d_k}$.

LEMMA 5.8.1. *Any small deformation of a "standard" hyperbolic structure on $T^2 \times [0,1]$ extends to some $(D^2 \times S^1)_d$. $d = (a,b)$ is determined up to sign by the traces of the matrices representing generators α, β of $\pi_1 T^2$.*

PROOF. A "standard" structure on $T^2 \times [0,1]$ means a structure as described on an end of M truncated by a Euclidean plane. The universal cover of $T^2 \times [0,1]$ is the region between two horizontal Euclidean planes (or horospheres), modulo a group of translations. If the structure is deformed slightly the holonomy determines the new structure and the images of α and β under the holonomy map H are slightly 5.34
perturbed.

If $H(\alpha)$ is still parabolic then so is $H(\beta)$ and the structure is equivalent to the standard one. Otherwise $H(\alpha)$ and $H(\beta)$ have a common axis l in H^3. Moreover since $H(\alpha)$ and $H(\beta)$ are close to the original parabolic elements, the endpoints of l are near the common fixed point of the parabolic elements. If $T^2 \times [0,1]$ is thought to be embedded in the end, $T^2 \times [0,\infty)$, this means that the line lies far out towards ∞ and does not intersect $T^2 \times [0,1]$. Thus the developing image of $T^2 \times [0,1]$ in H^3 for new structure misses l and can be lifted to the universal cover

$$\widetilde{H^3 - l}$$

of $H^3 - l$.

This is the geometric situation necessary for generalized Dehn surgery. The extension to $(D^2 \times S^1)_d$ is just the completion of

$$\widetilde{H^3 - l}/\{\tilde{H}(\alpha), \tilde{H}(\beta)\}$$

where $\tilde{H}$ is the lift of H to the cover

$$\widetilde{H^3 - l}.$$

Recall that the completion depends only on the behavior of $\widetilde{H(\alpha)}$ and $\widetilde{H(\beta)}$ along l. In particular, if $\tilde{H}(\ )$ denotes the complex number determined by the pair (translation distance along l, rotation about l), then the Dehn surgery coefficients $d = (a,b)$ are determined by the formula:

$$a\,\tilde{H}(\alpha) + b\,\tilde{H}(\beta) = \pm 2\pi i.$$

The translation distance and amount of rotation of an isometry along its fixed line is determined by the trace of its matrix in $\mathrm{PSL}(2,\mathbb{C})$. This is easy to see since trace is a conjugacy invariant and the fact is clearly true for a diagonal matrix. In 5.35
particular the complex number corresponding to the holonomy of a matrix acting on H^3 is $\log \lambda$ where $\lambda + \lambda^{-1}$ is its trace. □

The main result concerning deformations of M is

THEOREM 5.8.2. *If $M = M_{\infty,\dots,\infty}$ admits a hyperbolic structure then there is a neighborhood U of $(\infty,\dots,\infty)$ in $S^2 \times S^2 \times \cdots \times S^2$ such that for all $(d_1,\dots,dk) \in U, M_{d_1,\dots,d_k}$ admits a hyperbolic structure.*

PROOF. Consider the compact submanifold $M_0 \subset M$ gotten by truncating each end. M_0 has boundary a union of k tori and is homeomorphic to the manifold $\bar{M}$ such that $M =$ interior $\bar{M}$. By theorem 5.6, M_0 has a k complex parameter family of non-trivial deformations, one for each torus. From the lemma above, each small deformation gives a hyperbolic structure on some $M_{d_1,\dots,d_k}$. It remains to show that the d_i vary over a neighborhood of $(\infty,\dots,\infty)$.

Consider the function

$$\mathrm{Tr} : \mathrm{Def}(M) \to (\mathrm{Tr}(H(\alpha_1)),\dots,\mathrm{Tr}(H(\alpha_k)))$$

which sends a point in the deformation space to the k-tuple of traces of the holonomy of $\alpha_1,\alpha_2,\dots,\alpha_k$, where α_i,β_i generate the fundamental group of the i-th torus. Tr is a holomorphic (in fact, algebraic) function on the algebraic variety 5.36
$\mathrm{Def}(M)$. $\mathrm{Tr}(M_{\infty,\dots,\infty}) = (\pm 2,\dots,\pm 2)$ for some fixed choice of signs. Note that $\mathrm{Tr}(H(\alpha_i)) = \pm 2$ if and only if $H(\alpha_i)$ is parabolic and $H(\alpha_i)$ is parabolic if and only if the i-th surgery coefficient d_i equals ∞. By Mostow's Theorem the hyperbolic structure on $M_{\infty,\dots,\infty}$ is unique. Therefore $d_i = \infty$ for $i = l,\dots,k$ only in the original case and $\mathrm{Tr}^{-1}(\pm 2,\dots,\pm 2)$ consists of exactly one point. Since $\dim(\mathrm{Def}(M)) \geq k$ it follows from [ ] that the image under Tr of a small open neighborhood of $M_{\infty,\dots,\infty}$ is an open neighborhood of $(\pm 2,\dots,\pm 2)$.

Since the surgery coefficients of the i-th torus depend on the trace of both $H(\alpha_i)$ and $H(\beta_i)$, it is necessary to estimate $H(\beta_i)$ in terms of $H(\alpha_i)$ in order to see how the surgery coefficients vary. Restrict attention to one torus T and conjugate the original developing image of $M_{\infty,\dots,\infty}$ so that the parabolic fixed point of the holonomy, $H_0,(\pi_1 T)$, is the point at infinity. By further conjugation it is possible to put the holonomy matrices of the generators α,β of $\pi_1 T$ in the following form:

$$H_0(\alpha) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad H_0(\beta) = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}.$$

Note that since $H_0(\alpha), H_0(\beta)$ act on the horospheres about ∞ as a two-dimensional lattice of Euclidean translations, c and l are linearly independent over $\mathbb{R}$. Since 5.37
$H_0(\alpha), H_0(\beta)$ have $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ as an eigenvector, the perturbed holonomy matrices

$$H(\alpha), H(\beta)$$

will have common eigenvectors near $\begin{bmatrix}1\\0\end{bmatrix}$, say $\begin{bmatrix}1\\\epsilon_1\end{bmatrix}$ and $\begin{bmatrix}1\\\epsilon_2\end{bmatrix}$. Let the eigenvalues of $H(\alpha)$ and $H(\beta)$ be (λ, λ^{-1}) and (μ, μ^{-1}) respectively. Since $H(\alpha)$ is near $H_0(\alpha)$,

$$H(\alpha)\begin{bmatrix}0\\1\end{bmatrix} \approx \begin{bmatrix}1\\1\end{bmatrix}.$$

However

$$H(\alpha)\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{\epsilon_1 - \epsilon_2} H(\alpha)\left(\begin{bmatrix}1\\\epsilon_1\end{bmatrix} - \begin{bmatrix}1\\\epsilon_2\end{bmatrix}\right) = \frac{1}{\epsilon_1 - \epsilon_2}\left(\begin{bmatrix}\lambda\\\lambda\epsilon_1\end{bmatrix} - \begin{bmatrix}\lambda^{-1}\\\lambda^{-1}\epsilon_2\end{bmatrix}\right).$$

Therefore

$$\frac{\lambda - \lambda^{-1}}{\epsilon_1 - \epsilon_2} \approx 1.$$

Similarly,

$$\frac{\mu - \mu^{-1}}{\epsilon_1 - \epsilon_2} \approx c.$$

For λ, μ near l,

$$\frac{\log(\lambda)}{\log(\mu)} \approx \frac{\lambda - 1}{\mu - 1} \approx \frac{\lambda - \lambda^{-1}}{\mu - \mu^{-1}} \approx \frac{1}{c}.$$

Since $\tilde{H}(\alpha) = \log\lambda$ and $\tilde{H}(\beta) = \log\mu$ this is the desired relationship between $\tilde{H}(\alpha)$ and $\tilde{H}(\beta)$.

The surgery coefficients (a, b) are determined by the formula 5.38

$$a\tilde{H}(\alpha) + b\tilde{H}(\beta) = \pm 2\pi i.$$

From the above estimates this implies that

$$(a + bc) \approx \frac{\pm 2\pi i}{\log\lambda}.$$

(Note that the choice of sign corresponds to a choice of λ or λ^{-1}.) Since 1 and c are linearly independent over $\mathbb{R}$, the values of (a, b) vary over an open neighborhood of ∞ as λ varies over a neighborhood of 1. Since $\mathrm{Tr}(H(\alpha)) = \lambda + \lambda^{-1}$ varies over a neighborhood of 2 (up to sign) in the image of $\mathrm{Tr} : \mathrm{Def}(M) \to \mathbb{C}^k$, we have shown that the surgery coefficients for the $M_{d_1,\dots,d_k}$ possessing hyperbolic structures vary over an open neighborhood of ∞ in each component. □

Example. The complement of the Borromean rings has a complete hyperbolic structure. However, if the trivial surgery with coefficients $(1, 0)$ is performed on one component, the others are unlinked. (In other words, $M_{(1,0),\infty,\infty}$ is S^3 minus two unlinked circles.) The manifold $M_{(1,0),x,y}$ (where M is S^3 minus the Borromean rings) is then a connected sum of lens spaces if x, y are primitive elements of $H_1(T_i^2, \mathbb{Z})$ so it cannot have a hyperbolic structure. Thus it may often happen that an infinite number of non-hyperbolic manifolds can be obtained by surgery from a hyperbolic one. However, the theorem does imply that if a finite number of integral pairs of coefficients is excluded from *each boundary component*, then all remaining three-manifolds obtained by Dehn surgery on M are also hyperbolic.

5.39

5.9. A Proof of Mostow's Theorem.

This section is devoted to a proof of Mostow's Theorem for closed hyperbolic n-manifolds, $n \geq 3$. The proof will be sketchy where it seems to require analysis. With a knowledge of the structure of the ends in the noncompact, complete case, this proof extends to the case of a manifold of finite total volume; we omit details. The outline of this proof is Mostow's.

Given two closed hyperbolic manifolds M_1 and M_2, together with an isomorphism of their fundamental groups, there is a homotopy equivalence inducing the isomorphism since M_1 and M_2 are $K(\pi, 1)$'s. In other words, there are maps $f_1 : M_1 \to M_2$ and $f_2 : M_2 \to M_1$ such that $f_1 \circ f_2$ and $f_2 \circ f_1$ are homotopic to the identity. Denote lifts of f_1, f_2 to the universal cover H^n by $\tilde{f}_1, \tilde{f}_2$ and assume $\tilde{f}_1 \circ \tilde{f}_2$ and $\tilde{f}_2 \circ \tilde{f}_1$ are equivariantly homotopic to the identity.

The first step in the proof is to construct a correspondence between the spheres at infinity of H^n which extends $\tilde{f}_1$ and $\tilde{f}_2$.

Definition. A map $g : X \to Y$ between metric spaces is a *pseudo-isometry* if there are constants c_1, c_2 such that $c_1^{-1} d(x_1, x_2) - c_2 \leq d(gx_1, gx_2) \leq c_1 d(x_1, x_2)$ for all $x_1, x_2 \in X$.

Lemma 5.9.1. *$\tilde{f}_1, \tilde{f}_2$ can be chosen to be pseudo-isometries.*

Proof. Make f_1 and f_2 simplicial. Then since M_1 and M_2 are compact, f_1 and 5.40
f_2 are Lipschitz and lift to $\tilde{f}_1$ and $\tilde{f}_2$ which are Lipschitz with the same coefficient. It follows immediately that there is a constant c_1 so that $d(\tilde{f}_i x_1, \tilde{f}_i x_2) \leq c_1 d(x_1, x_2)$ for $i = 1, 2$ and all $x_1, x_2 \in H^n$.

If $x_i = \tilde{f}_1 y_i$, then this inequality implies that

$$d(\tilde{f}_2 \circ \tilde{f}_1(y_1),\ \tilde{f}_2 \circ \tilde{f}_1(y_2)) \leq c_1 d(\tilde{f}_1 y_1, \tilde{f}_1 y_2).$$

However, since M_1 is compact, $\tilde{f}_2 \circ \tilde{f}_1$ is homotopic to the identity by a homotopy that moves every point a distance less than some constant b. It follows that

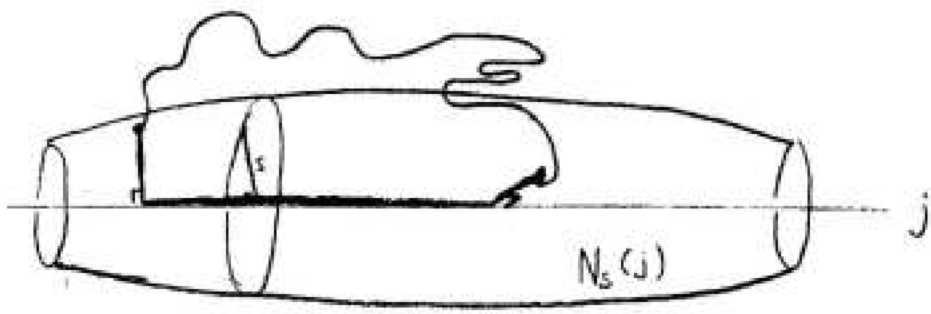
$$d(y_1, y_2) - 2b \leq d(\tilde{f}_2 \circ \tilde{f}_1 y_1, \tilde{f}_2 \circ \tilde{f}_1 y_2),$$

from which the lower bound $c_1^{-1} d(y_1, y_2) - c_2 \leq d(\tilde{f}_1 y_1, \tilde{f}_1 y_2)$ follows. □

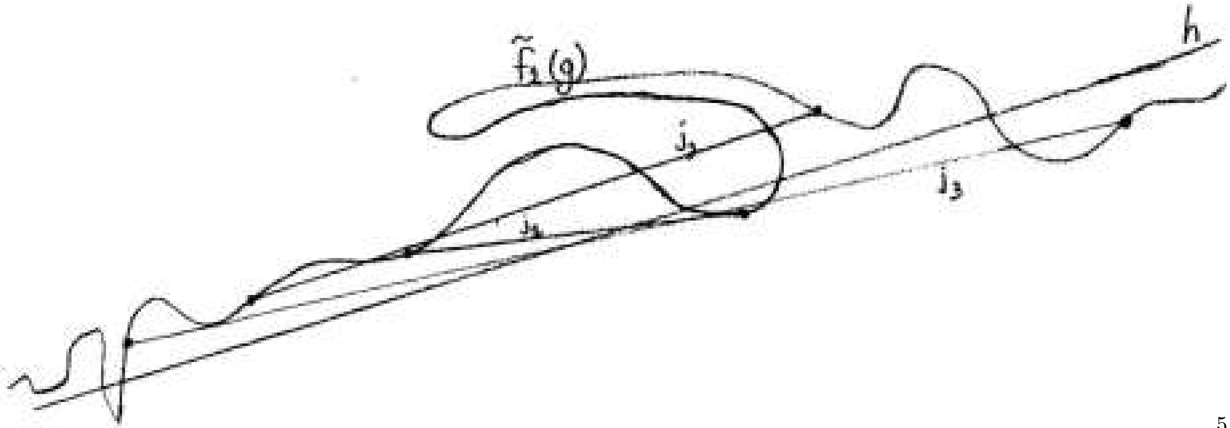
Using this lemma it is possible to associate a unique geodesic with the image of a geodesic.

PROPOSITION 5.9.2. *For any geodesic $g \subset H^n$ there is a unique geodesic h such that $f_1(g)$ stays in a bounded neighborhood of h.*

PROOF. If j is any geodesic in H^n, let $N_s(j)$ be the neighborhood of radius s about j. We will see first that if s is large enough there is an upper bound to the length of any bounded component of $g - \big(\tilde{f}_1^{-1}(N_s(j))\big)$, for any j. In fact, the perpendicular projection from $H^n - N_s(j)$ to j decreases every distance by at least a factor of $1/\cosh s$, so any long path in $H^n - N_s(j)$ with endpoints on $\partial N_s(j)$ can be replaced by a much shorter path consisting of two segments perpendicular to j, together with a segment of j. 5.41



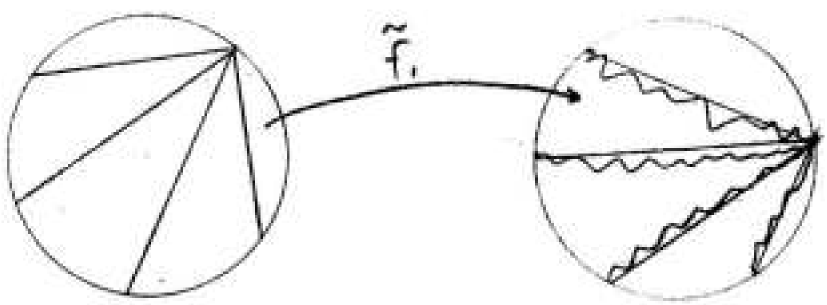
When this fact is applied to a line j joining distant points p_1 and p_2 on $\tilde{f}_1(g)$, it follows that the segment of g between p_1 and p_2 must intersect each plane perpendicular to j a bounded distance from j. It follows immediately that there is a limit line h to such lines j as p_1 and p_2 go to $+\infty$ and $-\infty$ on $\tilde{f}_1(g)$, and that $\tilde{f}_1(g)$ remains a bounded distance from h. Since no two lines in H^n remain a bounded distance apart, h is unique. $\square$

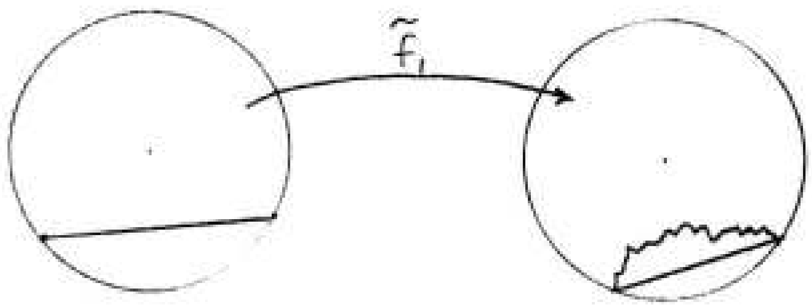


5.42

COROLLARY 5.9.3. $\tilde{f}_1 : H^n \to H^n$ *induces a one-to-one correspondence between the spheres at infinity.*

Proof. There is a one-to-one correspondence between points on the sphere at infinity and equivalence classes of directed geodesics, two geodesics being equivalent if they are parallel, or asymptotic in positive time. The correspondence of 5.9.2 between geodesics in $\tilde{M}_1$ and geodesics in $\tilde{M}_2$ obviously preserves this relation of parallelism, so it induces a map on the sphere at infinity. This map is one-to-one since any two distinct points in the sphere at infinity are joined by a geodesic, hence must be taken to the two ends of a geodesic. $\square$





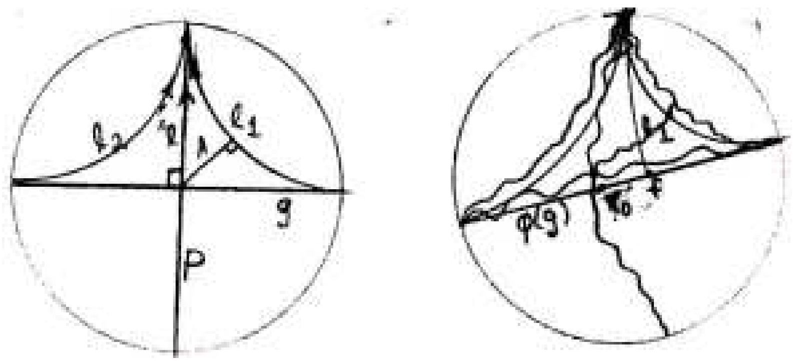
5.43

The next step in the proof of Mostow's Theorem is to show that the extension of $\tilde{f}_1$ to the sphere at infinity S_∞^{n-1} is continuous. One way to prove this is by citing Brouwer's Theorem that every function is continuous. Since this way of thinking is not universally accepted (though it is valid in the current situation), we will give another proof, which will also show that f is quasi-conformal at S_∞^{n-1}. A basis for the neighborhoods of a point $x \in S_\infty^{n-1}$ is the set of disks with center x. The boundaries of the disks are $(n-2)$-spheres which correspond to hyperplanes in H^2 (i.e., to $(n-1)$-spheres perpendicular to S_∞^{n-1} whose intersections with S_∞^{n-1} are the $(n-2)$-spheres).

For any geodesic g in $\tilde{M}_1$, let $\phi(g)$ be the geodesic in $\tilde{M}_2$ which remains a bounded distance from $\tilde{f}_1(g)$.

LEMMA 5.9.4. *There is a constant c such that, for any hyperplane P in H^n and any geodesic g perpendicular to P, the projection of $\tilde{f}_1(P)$ onto $\phi(g)$ has diameter $\leq c$.*

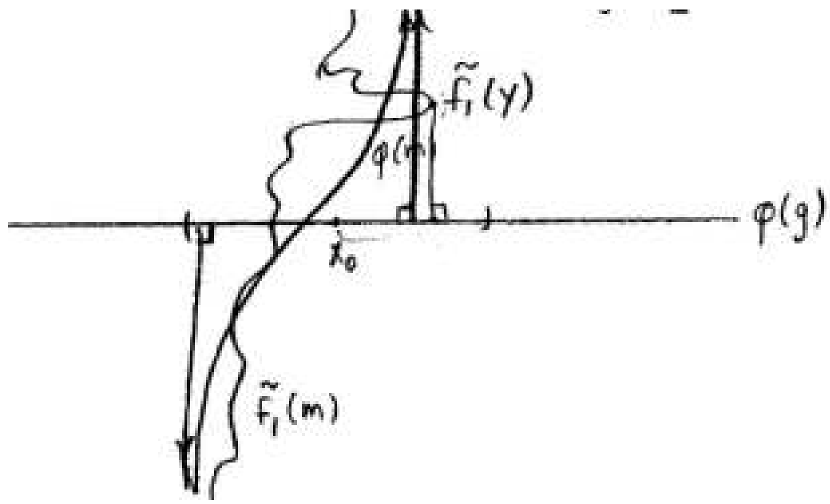
PROOF. Let x be the point of intersection of P and g and let l be a geodesic ray based at x. Then there is a unique geodesic l_1 which is parallel to l in one direction and to a fixed end of g in the other. Let A denote the shortest arc between x and l_1. It has length d, where d is a fixed contrast ($= \operatorname{arc}\cosh\sqrt{2}$). 5.44



The idea of the proof is to consider the image of this picture under $\tilde{f}_l$. Let $\phi(l), \phi(l_1), \phi(g)$ denote the geodesics that remain a bounded distance from l, l_1 and g respectively. Since ϕ preserves parallelism $\phi(l)$ and $\phi(l_1)$ are parallel. Let $l^{\perp}$ denote the geodesic from the endpoint on S^{n-1}_{∞} of $\phi(l)$ which is perpendicular to $\phi(g)$. Also let x_0 be the point on $\phi(g)$ nearest to $\tilde{f}_l(x)$.

Since $\tilde{f}_l(x)$ is a pseudo-isometry the length of $\tilde{f}_l(A)$ is at most $c_1 d$ where c_1 is a fixed constant. Since $\phi(l_1)$ and $\phi(g)$ are less than distance s (for a fixed constant s) from $\tilde{f}_l(l_1)$ and $\tilde{f}_l(g)$ respectively, it follows that x_0 is distance less than $C_1 d + 2s = \bar{d}$ from $\phi(l_1)$. This implies that the foot of $l^{\perp}$ (i.e., $l^{\perp} \cap \phi(g)$) lies distance less than $\bar{d}$ to one side of x_0. By considering the geodesic l_2 which is parallel to l and to the other end of g, it follows that f lies a distance less than $\bar{d}$ from x_0. 5.45

Now consider any point $y \in P$. Let m be any line through y. The endpoints of $\phi(m)$ project to points on $\phi(g)$ within a distance $\bar{d}$ of x_0; since $\tilde{f}_l(y)$ is within a distance s of $\phi(m)$, it follows that y projects to a point not farther than $\bar{d} + s$ from x_0. $\square$



Corollary 5.9.5. *The extension of $\tilde{f}_l$ to S^{n-1}_∞ is continuous.*

Proof. For any point $y \in S^{n-1}_\infty$, consider a directed geodesic g bending toward y, and define $\tilde{f}_l(y)$ to be the endpoint of $\phi(g)$. The half-spaces bounded by hyperplanes perpendicular to $\phi(g)$ form a neighborhood basis for $\tilde{f}_l(y)$. For any such half-space H, there is a point $x \in g$ such that the projection of $\tilde{f}_l(x)$ to $\phi(g)$ is a distance $> C$ from ∂H. Then the neighborhood of y bounded by the hyperplane through x perpendicular to g is mapped within H. ☐

Below it will be necessary to use the concept of quasi-conformality. If f is a homeomorphism of a metric space X to itself, f is K-*quasi-conformal* if and only if for all $z \in X$ 5.46

$$\lim_{r\to 0} \frac{\sup_{x,y\in S_r(z)} d\left(f(x), f(y)\right)}{\inf_{x,y\in S_r(z)} d\left(f(x), f(y)\right)} \le K$$

where $S_r(Z)$ is the sphere of radius r around Z, and x and y are diametrically opposite. K measures the deviation of f from conformality, is equal to 1 if f is conformal, and is unchanged under composition with a conformal map. f is called *quasi-conformal* if it is K-quasi-conformal for some K.

Corollary 5.9.6. *$\tilde{f}_l$ is quasi-conformal at S^{n-1}_∞.*

Proof. Use the upper half-space model for H^n since it is conformally equivalent to the ball model and suppose x and $\tilde{f}_l x$ are the origin since translation to the origin is also conformal. Then consider any hyperplane P perpendicular to the geodesic g from 0 to the point at infinity. By Lemma 5.9.4 there is a bound, depending only on $\tilde{f}_l$, to the diameter of the projection of $\tilde{f}_l(P)$ onto $\phi(g) = g$. Therefore, there are hyperplanes P_1, P_2 perpendicular to g contained in and containing $\tilde{f}_l(P)$ and the distance (along g) between P_1 and P_2 is uniformly bounded for all planes P.

But this distance equals $\log r, r > 1$, where r is the ratio of the radii of the $n-2$ spheres

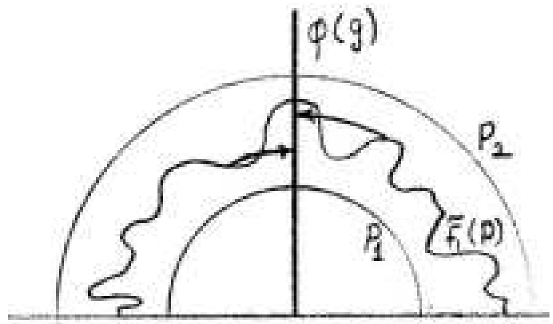
$$S^{n-2}_{p_1}, \quad S^{n-2}_{p_2}$$

in S^{n-1}_∞ corresponding to P_1 and P_2. The image of the $n-2$ sphere S^{n-2}_P corresponding 5.47
to P lies between $S^{n-2}_{p_2}$ and $S^{n-2}_{p_1}$ so that r is an upper bound for the ratio of maximum to minimum distances on

$$\tilde{f}_l(S^{n-2}_p).$$

Since $\log r$ is uniformly bounded above, so is r and $\tilde{f}_l$ is quasi-conformal. □

Corollary 5.9.6 was first proved by Gehring for dimension $n = 3$, and generalized to higher dimensions by Mostow.



At this point, it is necessary to invoke a theorem from analysis (see Bers).

Theorem 5.9.7. *A quasi-conformal map of an $n-1$-manifold, $n > 2$, has a derivative almost everywhere (= a.e.).*

Remark. It is at this stage that the proof of Mostow's Theorem fails for $n = 2$. 5.48
The proof works to show that $\tilde{f}_l$ extends quasi-conformally to the sphere at infinity, S^1_∞, but for a one-manifold this does not imply much.

Consider $\tilde{f}_l : S^{n-1}_\infty \to S^{n-1}_\infty$; by theorem 5.9.7 $d\tilde{f}_l$ exists a.e. At any point x where the derivative exists, the linear map $d\tilde{f}_l(x)$ takes a sphere around the origin to an ellipsoid. Let $\lambda_1, \dots, \lambda_{n-1}$ be the lengths of the axes of the ellipsoid. If we normalize so that $\lambda_1 \cdot \lambda_2 \cdots \lambda_{n-1} = 1$, then the λ_i are conformal invariants. In particular denote the maximum ratio of the λ_i's at x by $e(x)$, the eccentricity of $\tilde{f}_l$ at x. Note that if $\tilde{f}_l$ is K-quasi-conformal, the supremum of $e(x)$, $x \in S^{n-1}_\infty$, is K. Since $\pi_1 M_1$ acts on S^{n-1}_∞ conformally and e is invariant under conformal maps, e is a measurable, $\pi_1 M_1$ invariant function on S^{n-1}_∞. However, such functions are very simple because of the following theorem:

Theorem 5.9.8. *For a closed, hyperbolic n-manifold M, $\pi_1 M$ acts ergodically on S^{n-1}_∞, i.e., every measurable, invariant set has zero measure or full measure.*

Corollary 5.9.9. *e is constant a.e.*

Proof. Any level set of e is a measurable, invariant set so precisely one has full measure. $\square$

In fact more is true: 5.49

Theorem 5.9.10. *$\pi_1(M)$ acts ergodically on $S^{n-1}_\infty \times S^{n-1}_\infty$.*

Remark. This theorem is equivalent to the fact that the geodesic flow of M is ergodic since pairs of distinct points on S^{n-1}_∞ are in a one-to-one correspondence to geodesics in H^n (whose endpoints are those points).

From Corollary 5.9.9 e is equal a.e. to a constant K, and if the derivative of $\tilde{f}_l$ is not conformal, $K \neq 1$.

Consider the case $n = 3$. The direction of maximum "stretch" of df defines a measurable line field l on S^2_∞. Then for any two points $x, y \in S^2_\infty$ it is possible to parallel translate the line $l(x)$ along the geodesic between x and y to y and compute the angle between the translation of $l(x)$ and $l(y)$. This defines a measurable $\pi_1 M$-invariant function on $S^2_\infty \times S^2_\infty$. By theorem 5.9.10 it must be constant a.e. In other words l is determined by its "value" at one point. It is not hard to see that this is impossible.

For example, the line field determined by a line at x agrees with the line field below a.e. However, any line field determined by its "value" at y will have the same form and will be incompatible. 5.50

The precise argument is easy, but slightly more subtle, since l is defined only a.e.

The case $n > 3$ is similar.

Now one must again invoke the theorem, from analysis, that a quasi-conformal map whose derivative is conformal a.e. is conformal in the usual sense; it is a sphere-preserving map of S^{n-1}_∞, so it extends to an isometry I of H^n. The isometry I conjugates the action of $\pi_1 M_1$ to the action of $\pi_1 M_2$, completing the proof of Mostow's Theorem. $\square$

5.10. A decomposition of complete hyperbolic manifolds.

5.51

Let M be any complete hyperbolic manifold (possibly with infinite volume). For $\epsilon > 0$, we will study the decomposition $M = M_{(0,\epsilon]} \cup M_{[\epsilon,\infty)}$, where $M_{(0,\epsilon]}$ consists of those points in M through which there is a non-trivial closed loop of length $\leq \epsilon$, and $M_{[\epsilon,\infty)}$ consists of those points through which every non-trivial loop has length $\geq \epsilon$.

In order to understand the geometry of $M_{(0,\epsilon]}$, we pass to the universal cover $\tilde{M} = H^n$. For any discrete group Γ of isometries of H^n and any $x \in H^n$ let $\Gamma_\epsilon(x)$ be the subgroup generated by all elements of Γ which move x a distance $\leq \epsilon$, and let

$\Gamma'_\epsilon(x) \subset \Gamma_\epsilon(x)$ be the subgroup consisting of elements whose derivative is also ϵ-close to the identity.

LEMMA 5.10.1 (The Margulis Lemma). *For every dimension n there is an $\epsilon > 0$ such that for every discrete group Γ of isometries of H^n and for every $x \in H^n$, $\Gamma'_\epsilon(x)$ is abelian and $\Gamma_\epsilon(x)$ has an abelian subgroup of finite index.*

REMARK. This proposition is much more general than stated; if "abelian" is replaced by "nilpotent," it applies in general to discrete groups of isometries of Riemannian manifolds with bounded curvature. The proof of the general statement is essentially the same.

PROOF. In any Lie group G, since the commutator map $[\ ,\ ] : G \times G \to G$ has derivative 0 at $(1, 1)$, it follows that the size of the commutator of two small elements 5.52
is bounded above by some constant times the product of their sizes. Hence, if Γ'_ϵ is any discrete subgroup of G generated by small elements, it follows immediately that the lower central series $\Gamma'_\epsilon \supset [\Gamma'_\epsilon, \Gamma'_\epsilon] \supset [\Gamma'_\epsilon, [\Gamma'_\epsilon, \Gamma'_\epsilon]], \dots$ is finite (since there is a lower bound to the size of elements of Γ'_ϵ). In other words, Γ'_ϵ is nilpotent. When G is the group of isometries of hyperbolic space, it is not hard to see (by considering, for instance, the geometric classification of isometries) that this implies Γ'_ϵ is actually abelian.

To guarantee that $\Gamma_\epsilon(x)$ has an abelian subgroup of finite index, the idea is first to find an ϵ_1 such that $\Gamma'_{\epsilon_1}(x)$ is always abelian, and then choose ϵ many times smaller than ϵ_1, so the product of generators of $\Gamma_\epsilon(x)$ will lie in $\Gamma'_{\epsilon_1}(x)$. Here is a precise recipe:

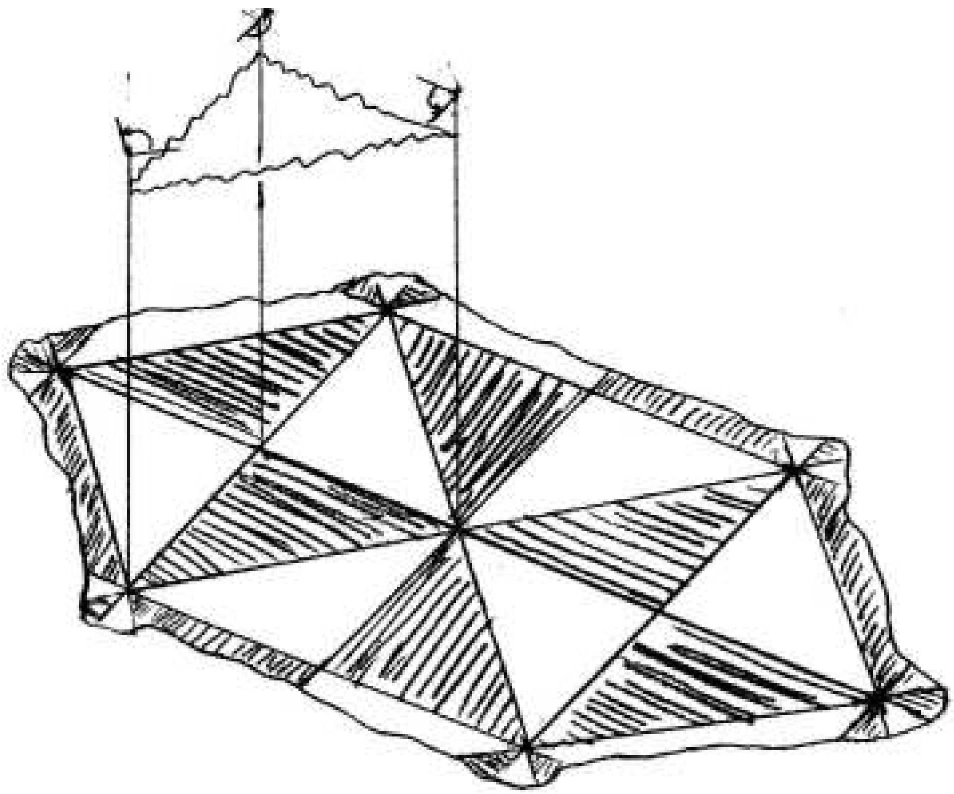
Let N be large enough that any collection of elements of $O(n)$ with cardinality $\geq N$ contains at least one pair separated by a distance not more than $\epsilon_{1/3}$.

Choose $\epsilon_2 \leq \epsilon_{1/3}$ so that for any pair of isometries ϕ_1 and ϕ_2 of H^n which translate a point x a distance $\leq \epsilon_2$, the derivative at x of $\phi_1 \circ \phi_2$ (parallel translated back to x) is estimated within $\epsilon_{1/6}$ by the product of the derivatives at x of ϕ_1 and ϕ_2 (parallel translated back to x).

Now let $\epsilon = \epsilon_{2/2N}$, so that a product of $2N$ isometries, each translating x a distance $\leq \epsilon$, translates x a distance $\leq \epsilon_2$. Let $g_1, \dots, g_k$ be the set of elements of Γ 5.53
which move x a distance $\leq \epsilon$; they generate $\Gamma_\epsilon(x)$. Consider the cosets $\gamma\,\Gamma'_{\epsilon_1}(x)$, where $\gamma \in \Gamma_\epsilon(x)$; the claim is that they are all represented by γ's which are words of length $< N$ in the generators $g_1, \dots, g_k$. In fact, if $\gamma = g_{i_1} \cdot \ldots \cdot g_{i_l}$ is any word of length $\geq N$ in the g_i's, it can be written $\gamma = \alpha \cdot \epsilon' \cdot \beta, (\alpha, \epsilon', \beta \neq 1)$ where $\epsilon' \cdot \beta$ has length $\leq N$, and the derivative of ϵ' is within $\epsilon_{1/3}$ of 1. It follows that $(\alpha\beta)^{-1} \cdot (\alpha\epsilon'\beta) = \beta^{-1}\epsilon'\beta$ is in $\Gamma'_{\epsilon_1}(x)$; hence the coset $\gamma\Gamma'_{\epsilon_1}(x) = (\alpha\beta)\Gamma'_{\epsilon_1}(x)$. By induction, the claim is verified. Thus, the abelian group $\Gamma'_{\epsilon_1}(x)$ has finite index in the group generated by $\Gamma_\epsilon(x)$ and $\Gamma'_{\epsilon_1}(x)$, so $\Gamma'_{\epsilon_1}(x) \cap \Gamma_\epsilon(x)$ with finite index. □

EXAMPLES. When $n = 3$, the only possibilities for discrete abelian groups are $\mathbb{Z}$ (acting hyperbolically or parabolically), $\mathbb{Z} \times \mathbb{Z}$ (acting parabolically, conjugate to a group of Euclidean translations of the upper half-space model), $\mathbb{Z} \times \mathbb{Z}_n$ (acting as a group of translations and rotations of some axis), and $\mathbb{Z}_2 \times \mathbb{Z}_2$ (acting by 180° rotations about three orthogonal axes). The last example of course cannot occur as $\Gamma'_\epsilon(x)$. Similarly, when ϵ is small compared to $1/n$, $\mathbb{Z} \times \mathbb{Z}_n$ cannot occur as $\Gamma'_\epsilon(x)$.

Any discrete group Γ of isometries of Euclidean space E^{n-1} acts as a group of isometries of H^n, via the upper half-space model. 5.54

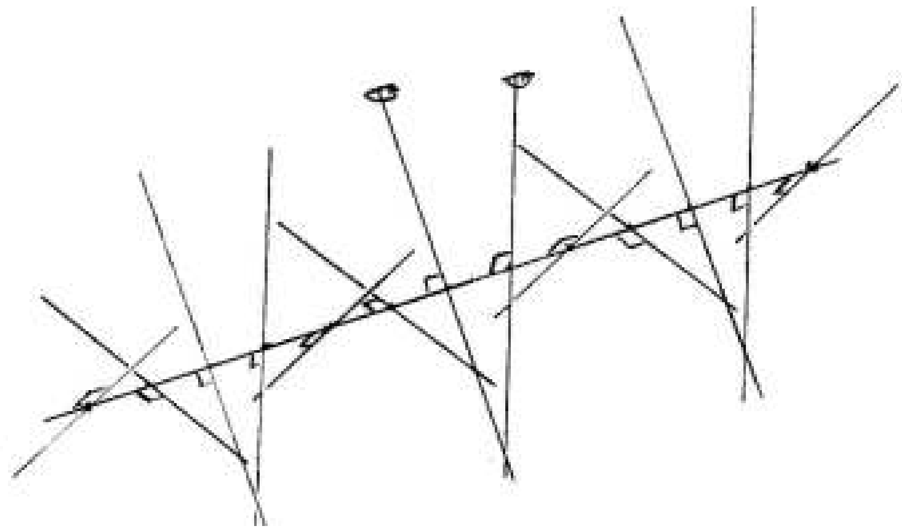


For any x sufficiently high (in the upper half space model), $\Gamma_\epsilon(x) = \Gamma$. Thus, 5.10.1 contains as a special case one of the Bieberbach theorems, that Γ contains an abelian subgroup of finite index. Conversely, when $\Gamma_\epsilon(x) \cap \Gamma'_{\epsilon_1}(x)$ is parabolic, $\Gamma_\epsilon(x)$ must be a Bieberbach group. To see this, note that if $\Gamma_\epsilon(x)$ contained any hyperbolic element γ, no power of γ could lie in $\Gamma'_{\epsilon_1}(x)$, a contradiction. Hence, $\Gamma_\epsilon(x)$ must consist of parabolic and elliptic elements with a common fixed point p at ∞, so it acts as a group of isometries on any horosphere centered at p.

If $\Gamma_\epsilon(x) \cap \Gamma'_{\epsilon_1}(x)$ is not parabolic, it must act as a group of translations and rotations of some axis a. Since it is discrete, it contains $\mathbb{Z}$ with finite index (provided $\Gamma_\epsilon(x)$ is infinite). It easily follows that $\Gamma_\epsilon(x)$ is either the product of some finite

FIGURE 1. The infinite dihedral group acting on H^3.

subgroup F of $O(n-1)$ (acting as rotations about a) with $\mathbb{Z}$, or it is the semidirect 5.55
product of such an F with the infinite dihedral group, $\mathbb{Z}/2 * \mathbb{Z}/2$.



For any set $S \subset H^n$, let $B_r(S) = \{x \in H^n \,|\, d(x, S) \le r\}$.

COROLLARY 5.10.2. *There is an $\epsilon > 0$ such that for any complete oriented hyperbolic three-manifold M, each component of $M_{(0,\epsilon]}$ is either*

(1) *a horoball modulo $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$, or*
(2) *$B_r(g)$ modulo $\mathbb{Z}$, where g is a geodesic.*

The degenerate case $r = 0$ may occur.

PROOF. Suppose $x \in M_{(0,\epsilon]}$. Let $\tilde{x} \in H^3$ be any point which projects to x. There is some covering translation γ which moves x a distance $\le \epsilon$. If γ is hyperbolic, let a be its axis. All rotations around a, translations along a, and uniform contractions of hyperbolic space along orthogonals to a commute with γ. It follows that $\tilde{M}_{(0,\epsilon]}$ 5.56
contains $B_r(a)$, where $r = d(a, x)$, since γ moves any point in $B_r(a)$ a distance $\le \epsilon$. Similarly, if γ is parabolic with fixed point p at ∞, $\tilde{M}_{(0,\epsilon]}$ contains a horoball about p passing through x. Hence $M_{(0,\epsilon]}$ is a union of horoballs and solid cylinders $B_r(a)$. Whenever two of these are not disjoint, they correspond to two covering transformations γ_1 and γ_2 which move some point x a distance $\le \epsilon$; γ_1 and γ_2 must commute (using 5.10.1), so the corresponding horoballs or solid cylinders must be concentric, and 5.10.2 follows. $\square$

5.11. Complete hyperbolic manifolds with bounded volume.

It is easy now to describe the structure of a complete hyperbolic manifold with finite volume; for simplicity we stick to the case $n = 3$.

PROPOSITION 5.11.1. *A complete oriented hyperbolic three-manifold with finite volume is the union of a compact submanifold* (*bounded by tori*) *and a finite collection of horoballs modulo* $\mathbb{Z} \oplus \mathbb{Z}$ *actions.*

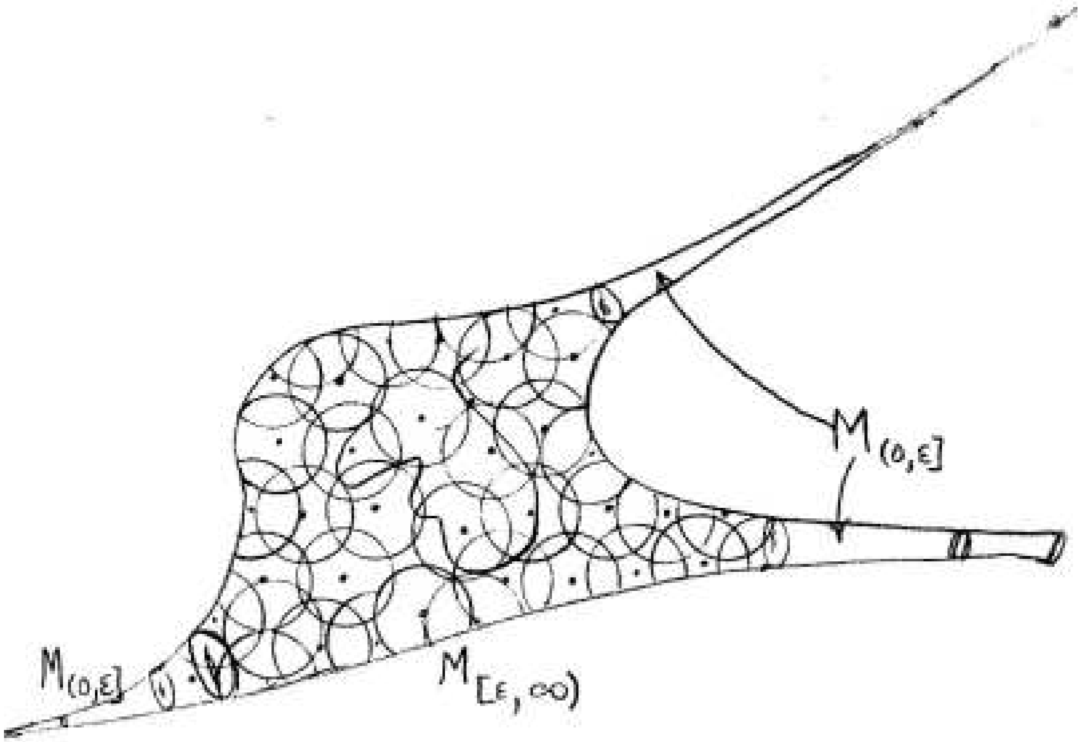
PROOF. $M_{[\epsilon,\infty)}$ must be compact, for otherwise there would be an infinite sequence of points in $M_{[\epsilon,\infty)}$ pairwise separated by at least ϵ. This would give a sequence of hyperbolic $\epsilon/2$ balls disjointly embedded in $M_{[\epsilon,\infty)}$, which has finite volume. $M_{(0,\epsilon]}$ must have finitely many components (since its boundary is compact). The proposition is obtained by lumping all compact components of $M_{(0,\epsilon]}$ with $M_{[\epsilon,\infty)}$. $\square$

5.57

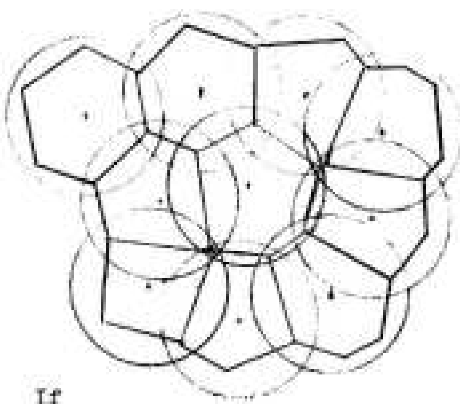
With somewhat more effort, we obtain Jørgensen's theorem, which beautifully describes the structure of the set of all complete hyperbolic three-manifolds with volume bounded by a constant C:

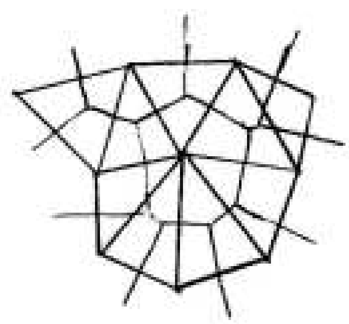
THEOREM 5.11.2 (Jørgensen's theorem [first version]). *Let $C > 0$ be any constant. Among all complete hyperbolic three-manifolds with volume $\leq C$, there are only finitely many homeomorphism types of $M_{[\epsilon,\infty)}$. In other words, there is a link L_c in S^3 such that every complete hyperbolic manifold with volume $\leq C$ is obtained by Dehn surgery along L_C.* (*The limiting case of deleting components of L_C to obtain a non-compact manifold is permitted.*)

PROOF. Let V be any maximal subset of $M_{[\epsilon,\infty)}$ having the property that no two elements of V have distance $\leq \epsilon/2$. The balls of radius $\epsilon/4$ about elements of V are embedded; since their total volume is $\leq C$, this gives an upper bound to the cardinality of V. The maximality of V is equivalent to the property that the balls of radius $\epsilon/2$ about V cover.



The combinatorial pattern of intersections of this set of $\epsilon/2$-balls determines $M_{[\epsilon,\infty)}$ up to diffeomorphism. There are only finitely many possibilities. (Alternatively a triangulation of $M_{[\epsilon,\infty)}$ with vertex set V can be constructed as follows. First, form a cell division of $M_{[\epsilon,\infty)}$ whose cells are indexed by V, associating to each $v \in V$ the subset of $M_{[\epsilon,\infty)}$ consisting of $x \in M_{[\epsilon,\infty)}$ such that $d(x, v) < d(x, v')$ for all $v' \in V$.





If V is in general position, faces of the cells meet at most four at a time. (The dual cell division is a triangulation.)

Any two hyperbolic manifolds M and N such that $M_{[\epsilon,\infty)} = N_{[\epsilon,\infty)}$ can be obtained from one another by Dehn surgery. All manifolds with volume $\leq C$ can therefore be obtained from a fixed finite set of manifolds by Dehn surgery on a fixed link in each manifold. Each member of this set can be obtained by Dehn surgery on some link in S^3, so all manifolds with volume $\leq C$ can be obtained from S^3 by Dehn surgery on the disjoint union of all the relevant links. $\square$ 5.59

The full version of Jørgensen's Theorem involves the geometry as well as the topology of hyperbolic manifolds. The geometry of the manifold $M_{[\epsilon,\infty)}$ completely determines the geometry and topology of M itself, so an interesting statement comparing the geometry of $M_{[\epsilon,\infty)}$'s must involve the approximate geometric structure. Thus, if M and N are complete hyperbolic manifolds of finite volume, Jørgensen defines M to be *geometrically near* N if for some small ϵ, there is a diffeomorphism which is approximately an isometry from the hyperbolic manifold $M_{[\epsilon,\infty)}$ to $N_{[\epsilon,\infty)}$. It would suffice to keep ϵ fixed in this definition, except for the exceptional cases when M and N have closed geodesics with lengths near ϵ. This notion of geometric nearness gives a topology to the set $\mathcal{H}$ of isometry classes of complete hyperbolic manifolds of finite volume. Note that neither coordinate systems nor systems of generators for the fundamental groups have been chosen for these hyperbolic manifolds; the homotopy class of an approximate isometry is arbitrary, in contrast with the definition for Teichmüller space. Mostow's Theorem implies that every closed manifold M in $\mathcal{H}$ is an isolated point, since $M_{[\epsilon,\infty)} = M$ when ϵ is small enough. On the other hand, a manifold in $\mathcal{H}$ with one end or *cusp* is a limit point, by the hyperbolic Dehn surgery theorem 5.9. A manifold with two ends is a limit point of limit points and a manifold with k ends is a k-fold limit point. 5.60

Mostow's Theorem implies more generally that the number of cusps of a geometric limit M of a sequence $\{M_i\}$ of manifolds distinct from M must strictly exceed the lim sup of the number of cusps of M_i. In fact, if ϵ is small enough, $M_{(0,\epsilon]}$ consists only of cusps. The cusps of M_i are contained in $M_{i_{(0,\epsilon]}}$; if all its components are cusps, and if $M_{i_{[\epsilon,\infty)}}$ is diffeomorphic with $M_{[\epsilon,\infty)}$ then M_i is diffeomorphic with M so M_i is isometric with M.

The volume of a hyperbolic manifold gives a function $v : \mathcal{H} \to \mathbb{R}_+$. If two manifolds M and N are geometrically near, then the volumes of $M_{[\epsilon,\infty)}$ and $N_{[\epsilon,\infty)}$ are approximately equal. The volume of a hyperbolic solid torus r_0 centered around a geodesic of length l may be computed as

$$\text{volume (solid torus)} = \int_0^{r_0} \int_0^{2\pi} \int_0^{l} \sinh r \cosh r \, dt \, d\theta \, dr = \pi l \sinh^2 r_0$$

while the area of its boundary is

$$\text{area (torus)} = 2\pi l \sinh r_0 \cosh r_0.$$

Thus we obtain the inequality

$$\frac{\text{area } (\partial \text{ solid torus})}{\text{volume (solid torus)}} = \frac{1}{2}\frac{\sinh r_0}{\cosh r_0} < \frac{1}{2}.$$

The limiting case as $r_0 \to \infty$ can be computed similarly; the ratio is $1/2$. Applying 5.61
this to M, we have

5.11.2. $$\text{volume } (M) \le \text{ volume } (M_{[\epsilon,\infty)}) + \tfrac{1}{2} \text{ area } (\partial M_{[\epsilon,\infty)}).$$

It follows easily that v is a continuous function on $\mathcal{H}$. Changed this label to 5.11.2a.

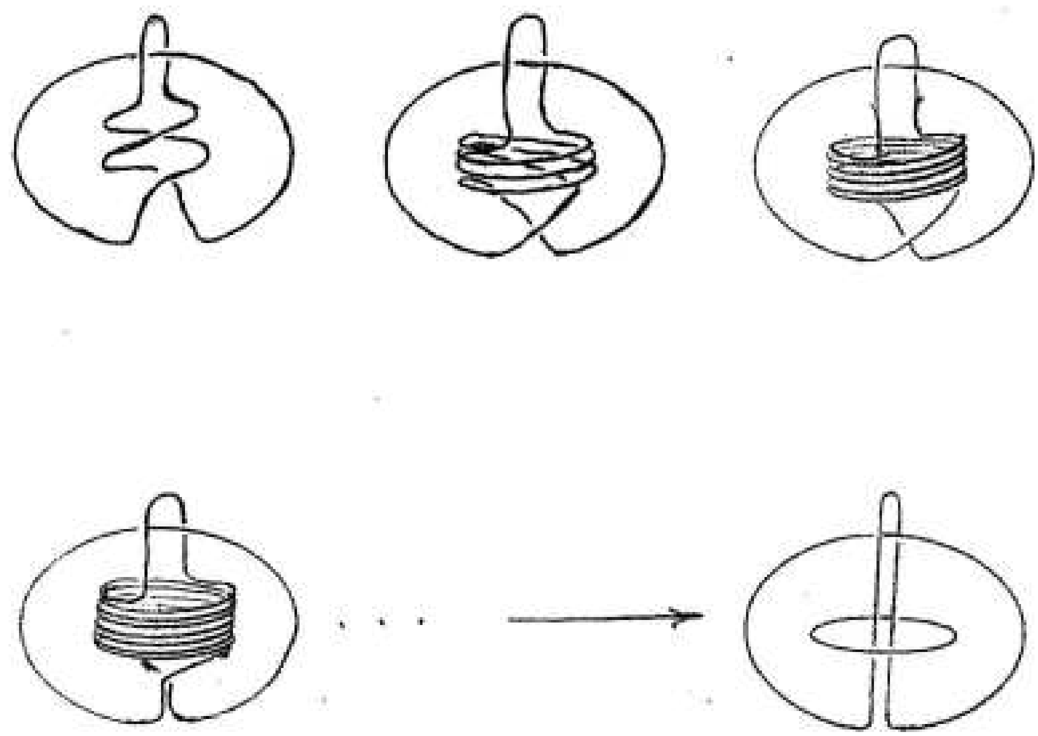
5.12. Jørgensen's Theorem.

Theorem 5.12.1. *The function $v : \mathcal{H} \to \mathbb{R}_+$ is proper. In other words, every sequence in $\mathcal{H}$ with bounded volume has a convergent subsequence. For every C, there is a finite set $M_1, \dots, M_k$ of complete hyperbolic manifolds with volume $\le C$ such that all other complete hyperbolic manifolds with volume $\le C$ are obtained from this set by the process of hyperbolic Dehn surgery* (as in 5.9)*.*

Proof. Consider a maximal subset of V of $M_{[\epsilon,\infty)}$ having the property that no two elements of V have distance $\le \epsilon/2$ (as in 5.11.1). Choose a set of isometries of the $\epsilon/2$ balls centered at elements of V with a standard $\epsilon/2$-ball in hyperbolic space. The set of possible gluing maps ranges over a compact subset of $\text{Isom}(H^3)$, so any sequence of gluing maps (where the underlying sequence of manifolds has volume $\le C$) has a convergent subsequence. It is clear that in the limit, the gluing maps still give a hyperbolic structure on $M_{[\epsilon,\infty)}$, approximately isometric to the limiting $M_{[\epsilon,\infty)}$'s. We must verify that $M_{[\epsilon,\infty)}$ extends to a complete hyperbolic manifold. To see this, note that whenever a complete hyperbolic manifold N has a geodesic which is very short compared to ϵ, the radius of the corresponding solid torus in $N_{(0,\epsilon]}$
becomes large. (Otherwise there would be a short non-trivial curve on $\partial N_{(0,\epsilon]}$—but 5.62
such a curve has length $\ge \epsilon$). Thus, when a sequence $\{M_{i_{[\epsilon,\infty)}}\}$ converges, there are approximate isometries between arbitrarily large balls $B_r(M_{i_{[\epsilon,\infty)}})$ for large i, so in the limit one obtains a complete hyperbolic manifold. This proves that v is a proper function. The rest of §5.12 is merely a restatement of this fact. □

Remark. Our discussion in §5.10, 5.11 and 5.12 has made no attempt to be numerically efficient. For instance, the proof that there is an ϵ such that $\Gamma_\epsilon(x)$ has an abelian subgroup of finite index gives the impression that ϵ is microscopic. In fact, ϵ can be rather large; see Jørgensen, for a more efficient approach. It would be extremely interesting to have a good estimate for the number of distinct $M_{[\epsilon,\infty)}$'s

Figure eight knot



Whitehead Link

where M has volume $\leq C$, and it would be quite exciting to find a practical way of computing them. The development in 5.10, 5.11, and 5.12 is completely inefficient in this regard. Jørgensen's approach is much more explicit and efficient.

Example. The sequence of knot complements below are all obtained by Dehn surgery on the Whitehead link, so 5.8.2 implies that all but a finite number possess complete hyperbolic structures. (A computation similar to that of Theorem 4.7 shows that in fact they all possess hyperbolic structures.) This sequence converges, in $\mathcal{H}$, to the Whitehead link complement: 5.63

Note. Gromov proved that in dimensions $n \neq 3$, there is only a finite number of complete hyperbolic manifolds with volume less than a given constant. He proved this more generally for negatively curved Riemannian manifolds with curvature varying between two negative constants. His basic method of analysis was to study the injectivity radius

$$\begin{aligned}\mathrm{inj}(x) &= \tfrac{1}{2}\inf\{\text{lengths of non-trivial closed loops through } x\} \\ &= \sup\{r \mid \text{the exponential map is injective on the ball of radius } r \text{ in } T(x)\}.\end{aligned}$$

Basically, in dimensions $n \neq 3$, little can happen in the region M^n_ϵ of M^n where $\mathrm{inj}(x)$ is small. This was the motivation for the approach taken in 5.10, 5.11 and 5.12. Gromov also gave a weaker version of hyperbolic Dehn surgery, 5.8.2: he showed that many of the manifolds obtained by Dehn surgery can be given metrics of negative curvature close to -1.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 6

Gromov's invariant and the volume of a hyperbolic manifold

6.1

6.1. Gromov's invariant

Let X be any topological space. Denote the real singular chain complex of X by $C_*(k)$. (Recall that $C_k(X)$ is the vector space with a basis consisting of all continuous maps of the standard simplex Δ^k into X.) Any k-chain c can be written uniquely as a linear combination of the basis elements. Define the norm $\|c\|$ of c to be the sum of the absolute values of its coefficients,

6.1.1. $$\|c\| = \sum |a_i| \qquad \text{where } c = \sum a_i \sigma_i, \quad \sigma_i : \Delta^k \to X.$$

Gromov's norm on the real singular homology (really it is only a pseudo-norm) is obtained from this norm on cycles by passing to homology: if $a \in H_k(X;\mathbb{R})$ is any homology class, then the norm of α is defined to be the infimum of the norms of cycles representing α, Labelled this 6.1.2.def

Definition 6.1.2 (First definition).

$$\|\alpha\| = \inf \{\|z\| \mid z \text{ is a singular cycle representing } \alpha\}.$$

It is immediate that

$$\|\alpha + \beta\| \le \|\alpha\| + \|\beta\|$$

and for $\lambda \in \mathbb{R}$,

$$\|\lambda\alpha\| \le |\lambda| \, \|\alpha\|.$$

If $f : X \to Y$ is any continuous map, it is also immediate that 6.2

6.1.2. $$\|f_*\alpha\| \le \|\alpha\|.$$

In fact, for any cycle $\sum a_i\sigma_i$ representing α, the cycle $\sum a_i f \circ \sigma_i$ represents $f_*\alpha$, and $\| \sum a_i f \circ \sigma_i\| = \sum |a_i| \le \| \sum a_i \sigma_i\|$. (It may happen that $f \circ \sigma_i = f \circ \sigma_j$; even when $\sigma_i \ne \sigma_j$.) Thus $\|f_*\alpha\| \le \inf \|a_i f \circ \sigma_i\| \le \|\alpha\|$. In particular, the norm of the fundamental class of a closed oriented manifold M gives a characteristic number of M, Gromov's invariant of M, satisfying the inequality that for any map $f : M_1 \to M_2$,

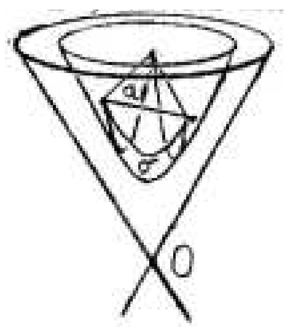
6.1.3. $$\| [M_1] \| \ge |\deg f| \, \| [M_2] \|.$$

What is not immediate from the definition is the existence of any non-trivial examples where $\| [M] \| \ne 0$.

Example. The n-sphere $n \geq 1$ admits maps $f : S^n \to S^n$ of degree 2 (and higher). As a consequence of 6.1.2 $\| [S^n] \| = 0$. More explicitly, one may picture a sequence $\{z_i\}$ representing the fundamental class of S^1, where z_i is $(\frac{l}{i})\sigma_i$ and σ_i wraps a 1-simplex i times around S^1. Since $\|z_i\| = \frac{1}{i}$, $\| [S^1] \| = 0$.

As a trivial example, $\| [S^0] \| = 2$.

Consider now the case of a complete hyperbolic manifold M^n. Any $k+1$ points $v_0, \ldots, v_k$ in $\tilde{M}^n = H^n$ determine a *straight* k-simplex $\sigma_{v_0,\ldots,v_k} : \Delta^k \to H^n$, whose image is the convex hull of $v_0, \ldots, v_k$. There are various ways to define canonical parametrizations for $\sigma_{v_0,\ldots,v_k}$; here is an explicit one. Consider the quadratic form 6.3
model for H^n (§2.5). In this model, $v_0, \ldots, v_k$ become points in $\mathbb{R}^{n+1}$, so they determine an affine simplex α. [In barycentric coordinates, $\alpha(t_0, \ldots, t_k) = \sum t_i v_i$. This parametrization is natural with respect to affine maps of $\mathbb{R}^{n+1}$]. The central projection from O of α back to one sheet of hyperboloid $Q = x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1$ gives a parametrized straight simplex $\sigma_{v_0,\ldots,v_k}$ in H^n, natural with respect to isometries of H^n.



Any singular simplex $\tau : \Delta^k \to M$ can be lifted to a singular simplex $\tilde{\tau}$ in $\tilde{M} = H^n$, since Δ^k is simply connected. Let straight $(\tilde{\tau})$ be the straight simplex with the same vertices as $\tilde{\tau}$ and let straight(τ) be the projection of $\tilde{\tau}$ back to M. Since the straightening operation is natural, straight(τ) does not depend on the lift $\tilde{\tau}$. Straight extends linearly to a chain map

$$\text{straight } : C_*(M) \to C_*(M),$$

chain homotopic to the identity. (The chain homotopy is constructed from a canonical homotopy of each simplex τ to straight(τ).) It is clear that for any chain c, $\|$ straight $(c)\| \leq \|c\|$. Hence, in the computation of the norm of a homology class in M, it suffices to consider only straight simplices. 6.3

Proposition 6.1.4. *There is a finite supremum v_k to the k-dimensional volume of a straight k-simplex in hyperbolic space H^n provided $k \neq 1$.*

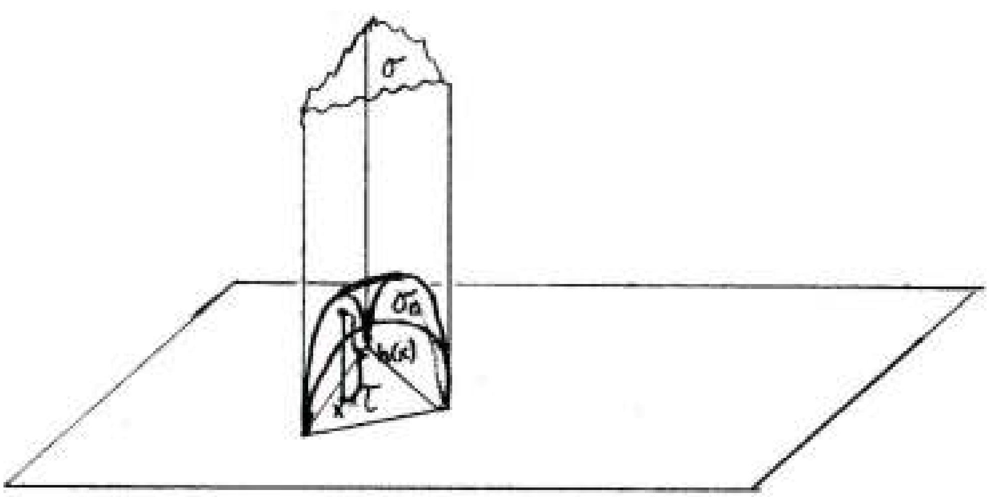
Proof. It suffices to consider ideal simplices with all vertices on S_∞, since any finite simplex fits inside one of these. For $k = 2$, there is only one ideal simplex up to isometry. We have seen that 2 copies of the ideal triangle fit inside a compact surface (§3.9). Thus it has finite volume, which equals π by the Gauss-Bonnet theorem. When $k = 3$, there is an efficient formula for the computation of the volume of an ideal 3-simplex; see Milnor's discussion of volumes in chapter 7. The volume of such simplices attains its unique maximum at the regular ideal simplex, which has all angles equal to 60°. Thus we have the values

6.1.5.
$$v_2 = 3.1415926\ldots = \pi$$
$$v_3 = 1.0149416\ldots$$

It is conjectured that in general, v_k is the volume of the regular ideal k-simplex; if so, Milnor has computations for more values, and a good asymptotic formula as $k \to \infty$. In lieu of a proof of this conjecture, an upper bound can be obtained for v_k from the inductive estimate

6.1.6.
$$v_k < \frac{v_{k-1}}{k-1}.$$

To prove this, consider any ideal k-simplex σ in H^k. Arrange σ so that one of its vertices is the point at ∞ in the upper half-space model, so that σ looks like a triangular chimney lying above a $k-1$ face σ_0 of σ. 6.5



Let dW^k be the Euclidean volume element, so hyperbolic volume is $dV^k = (\frac{1}{x_k})^k dW^k$. Let τ denote the projection of σ_0 to E^{n-1}, and let $h(x)$ denote the Euclidean height of σ_0 above the point $x \in \tau$. The volume of σ is

$$v(\sigma) = \int_\tau \int_h^\infty t^{-k}\, dt\, dW^{k-1}$$

(where dW^{k-1} is the Euclidean $k-1$ volume element for τ). Integrating, we obtain

$$(k-1)\,v(\sigma) = \int_\tau h^{-(k-1)}\,dW^{k-1}.$$

The volume of σ_0 is obtained by a similar integral, where dW^{k-1} is replaced by the Euclidean volume element for σ, which is never smaller than dW^{k-1}. We have $(k-1)v(\sigma) < v(\sigma_0) \le v_{k-1}$. □

We are now ready to find non-trivial examples for Gromov's invariant:

COROLLARY 6.1.7. *Every closed oriented hyperbolic manifold M^n of dimension $n > 1$ satisfies the inequality*

$$\|\,[M]\,\| \ge \frac{v(M)}{v_n}.$$

PROOF. Let Ω be the hyperbolic volume form for M, so that $\int_M \Omega = v(M)$. If $z = \sum z_i\sigma_i$ is any straight cycle representing $[M]$, then 6.6

$$v(M) = \int_M \Omega = \sum z_i \int_{\Delta^n} \sigma_i^*\Omega \le \sum |z_i|\,v_n.$$

Dividing by v_n, we obtain $\|z\| \ge v(M)/v_n$. The infimum over all such z gives 6.1.7 □

A similar proof shows that the norm of element $0 \ne \alpha \in H_k(M,\mathbb{R})$ where $k \ne 1$ is non-zero. Instead of Ω, use an k-form ω representing some multiple $\lambda\alpha$ such that ω has Riemannian norm ≤ 1 at each point of M. (In fact, ω need only satisfy the inequality $\omega(V) \le 1$ where V is a *simple* k-vector of Riemannian norm 1.) Then the inequality $\|\alpha\| \ge \lambda/v_k$ is obtained.

Intuitively, Gromov's norm measures the efficiency with which multiples of a homology class can be represented by simplices. A complicated homology class needs many simplices.

Gromov proved the remarkable theorem that the inequality of 6.1.7 is actually equality. Instead of proving this, we will take the alternate approach to Gromov's theorem developed in [Milnor and Thurston, "Characteristic numbers for three-manifolds"], of changing the definition of $\|\ \|$ to one which is technically easier to work with. It can be shown that past and future definitions are equivalent. However, we have no further use for the first definition, 6.1.2, so henceforth we shall simply abandon it.

For any manifold M, let $C^1(\Delta^k, M)$ denote the space of maps of Δ^k to M, with the C^1 topology. We define a new notion of chains, where a k-chain is a Borel measure μ on $C^1(\Delta^K, M)$ with compact support and bounded total variation. [The total variation of a measure μ is $\|\mu\| = \sup\{\int f\,d\mu \,|\, |f| \le 1\}$. Alternately, μ can be decomposed into a positive and negative part, $\mu = \mu_+ - \mu_-$ where μ_+ and μ_- are 6.7

positive. Then $\|\mu\| = \int d\mu_+ + \int d\mu_-]$. Let the group of k-chains be denoted $\mathcal{C}_k(M)$. There is a map $\partial : \mathcal{C}_k(M) \to \mathcal{C}_{k-1}(M)$, defined in an obvious way. It is not difficult to prove that the homology obtained by using these chains is the standard homology for M; see [Milnor and Thurston, "Characteristic numbers for three-manifolds"] for more details. (Note that integration of a k-form over an element of $\mathcal{C}_k(M)$ is defined; this gives a map from $\mathcal{C}_*(M)$ to currents on M. Some condition such as compact support for μ is necessary; otherwise one would have pathological cycles such as $\sum(\frac{1}{i})^2\sigma_i$, where σ_i wraps Δ^1 i times around S^1. The measure has total variation $\sum(\frac{1}{i})^2 < \infty$, yet the cycle would seem to represent the infinite multiple $\sum(\frac{1}{i})[S^1]$ of $[S^1]$.)

Definition 6.1.8 (Second definition). i Let $\alpha \in H^k(M;\mathbb{R})$, where M is a manifold. Gromov's norm $\|\alpha\|$ is defined to be

$$\|\alpha\| = \inf\{\|u\| \,|\mu \in \mathcal{C}^k(M) \text{ represents } \alpha\}.$$

Theorem 6.2 (Gromov). *Let M^n be any closed oriented hyperbolic manifold. Then*

$$\|\,[M]\,\| = \frac{v(M^n)}{v_n}.$$

Proof. The proof of corollary 6.1.7 works equally well with the new definition as with the old. The point is that the straightening operation is completely uniform, so it works with measure-cycles. What remains is to prove that $\|\,[M]\,\| \le v(M)/v_n$, or in other words, the fundamental cycle of M can be represented efficiently by a cycle using simplices which have (on the average) nearly maximal volume. 6.8

Let σ be any singular k-simplex in H^n. A chain $\mathrm{smear}_M(\sigma) \in \mathcal{C}_k(M)$ can be constructed, which is a measure supported on all isometric maps of σ into M, weighted uniformly. With more notation, let h denote Haar measure on the group of orientation-preserving isometries of H^n, $\mathrm{Isom}_+(H^n)$. Let h be normalized so that the measure of the set of isometries taking a point $x \in H^n$ to a region $R \subset H^n$ is the volume of R. Haar measure on $\mathrm{Isom}_+(H^n)$ is invariant under both right and left multiplication, so it descends to a measure (also denoted h) on the quotient space $P(M) = \pi_1 M \backslash \mathrm{Isom}_+(H^n)$.

There is a map from $P(M)$ to $\mathcal{C}^1(\Delta^k, M)$, which associates to a coset $\pi_1 M\varphi$ the singular simplex $p \circ \varphi \circ \sigma$, where $p : H^n \to M$ is the covering projection. The measure h pushes forward to give a chain $\mathrm{smear}_M(\sigma) \in \mathcal{C}_k(M)$. Since h is invariant on both sides, $\mathrm{smear}_M(\sigma)$ depends only on the isometry class of σ. Smearing extends linearly to $\mathcal{C}_k(H^n)$. Furthermore, $\mathrm{smear}_M \partial c = \partial\, \mathrm{smear}_M\, c$.

Let σ now be any straight simplex in H^n, and σ_- a reflected copy of σ. Then $\frac{1}{2}\,\mathrm{smear}_M(\sigma - \sigma_-))$ is a cycle, since the faces of σ and σ_- cancel out in pairs, up to isometries. We have

$$\|\tfrac{1}{2}\ \mathrm{smear}_M\,(\sigma - \sigma_-)\| = v(M).$$

The homology class of this cycle can be computed by integration of the hyperbolic form Ω from M. The integral over each copy of σ is $v(\sigma)$, so the total integral is $v(M)v(\sigma)$. Thus, the cycle represents

$$[\tfrac{1}{2}\ \mathrm{smear}\,(\sigma - \sigma_-)] = v(\sigma)[M]$$

so that 6.9

$$\|v(\sigma)[M]\,\| \le v(M).$$

Dividing by $v(\sigma)$ and taking the infimum over σ, we obtain 6.2. □

Corollary 6.2.1. *If $f : M_1 \to M_2$ is any map between closed oriented hyperbolic n-manifolds, then*

$$v(M_1) \ge |\deg f| v(M_2).$$

Gromov's theorem can be generalized to any (G, X)-manifold, where G acts transitively on X with compact isotropy groups.

To do this, choose an invariant Riemannian metric for X and normalize Haar measure on G as before. The smearing operation works equally well, so that one has a chain map

$$\mathrm{smear}_M : \mathcal{C}_k(X) \to \mathcal{C}_k(M).$$

In fact, if N is a second (G, X)-manifold, one has a chain map

$$\mathrm{smear}_{N,M} : \mathcal{C}_k(N) \to \mathcal{C}_k(M),$$

defined first on simplices in N via a lift to X, and then extended linearly to all of $\mathcal{C}_k(N)$. If z is any cycle representing $[N]$, then $\mathrm{smear}_{N,M}(z)$ represents

$$(v(N)/v(M))[M].$$

This gives the inequality

$$\frac{\|\,[N]\,\|}{v(N)} \ge \frac{\|\,[M]\,\|}{v(M)}.$$

Interchanging M and N, we obtain the reverse inequality, so we have proved the following result:

Theorem 6.2.2. *For any pair (G, X), where G acts transitively on X with compact isotropy groups and for any invariant volume form on X, there is a constant C such that every closed oriented (G, X)-manifold M satisfies* 6.10

$$\|\,[M]\,\| = Cv(M),$$

(where $v(M)$ is the volume of M). □

This line may be pursued still further. In a hyperbolic manifold a smeared k-cycle is homologically trivial except in dimension $k = 0$ or $k = n$, but this is not generally true for other (G, X)-manifolds when G does not act transitively on the frame bundle of X. The *invariant cohomology* $H^*_G(X)$ is defined to be the cohomology of the cochain complex of differential forms on X invariant by G. If α is any invariant cohomology class for X, it defines a cohomology class α_M on any (G, X)-manifold M. Let $PD(\gamma)$ denote the Poincaré dual of a cohomology class γ.

THEOREM 6.2.3. *There is a norm $\| \ \|$ in $H^*_G(X)$ such that for any closed oriented (G, X)-manifold M,*

$$\big\|\mathrm{PD}(\alpha_m)\| = v(M)\|\alpha\|.$$

PROOF. It is an exercise to show that the map

$$\mathrm{smear}_{M,M} : H_*(M) \to H_*(M)$$

is a retraction of the homology of M to the Poincaré dual of the image in M of $H^*_G(X)$. The rest of the proof is another exercise. □

In these variations, 6.2.2 and 6.2.3, on Gromov's theorem, there does not seem to be any general relation between the proportionality constants and the maximal 6.11
volume of simplices. However, the inequality 6.1.7 readily generalizes to any case when X possesses and invariant Riemannian metric of non-positive curvature.

6.3. Gromov's proof of Mostow's Theorem

Gromov gave a very quick proof of Mostow's theorem for hyperbolic three-manifolds, based on 6.2. The proof would work for hyperbolic n-manifolds if it were known that the regular ideal n-simplex were the unique simplex of maximal volume. The proof goes as follows.

This is now known to be true.

LEMMA 6.3.1. *If M_1 and M_2 are homotopy equivalent, closed, oriented hyperbolic manifolds, then $v(M_1) = v(M_2)$.*

PROOF. This follows immediately by applying 6.2 to the homotopy equivalence $M_1 \leftrightarrow M_2$. □

Let $f_1 : M_1 \to M_2$ be a homotopy equivalence and let $\tilde{f}_1 : \tilde{M}_1 \to \tilde{M}_2$ be a lift of f_1. From 5.9.5 we know that $\tilde{f}_1$ extends continuously to the sphere S^{n-1}_∞.

LEMMA 6.3.2. *If $n = 3$, $\tilde{f}_1$ takes every 4-tuple of vertices of a positively oriented regular ideal simplex to the vertices of a positively oriented regular ideal simplex.*

PROOF. Suppose the contrary. Then there is a regular ideal simplex σ such that the volume of the simplex straight$(\tilde{f}_1\sigma)$ spanned by the image of its vertices is $v_3 - \epsilon$, with $\epsilon > 0$. There are neighborhoods of the vertices of σ in the disk such that for any simplex σ' with vertices in these neighborhoods, $v\big(\text{straight}(\tilde{f}_1\sigma')\big) \leq v_3 - \epsilon/2$. Then for every finite simplex σ'_0 very near to σ, this means that a definite Haar measure 6.12
of the isometric copies σ' of σ'_0 near σ' have $v\big(\text{straight}(\tilde{f}_1\sigma'_0)\big) < v_3 - \epsilon/2$. Such a simplex σ'_0 can be found with volume arbitrarily near v_3. But then the "total volume" of the cycle $z = \frac{1}{2}\,\text{smear}(\sigma'_0 - \sigma'_{0-})$ strictly exceeds the total volume of straight$(f_* z)$, contradicting 6.3.1. □

To complete the proof of Mostow's theorem in dimension 3, consider any ideal regular simplex σ together with all images of σ coming from repeated reflections in the faces of σ. The set of vertices of all these images of σ is a dense subset of S^2_∞. Once $\tilde{f}_1$ is known on three of the vertices of σ, it is determined on this dense set of points by 6.3.2, so $\tilde{f}_1$ must be a fractional linear transformation of S^2_∞, conjugating the action of $\pi_1 M_1$ to the action of $\pi_1 M_2$. This completes Gromov's proof of Mostow's theorem. □

In this proof, the fact that f_1 is a homotopy equivalence was used to show (a) that $v(M_1) = v(M_2)$ and (b) that $\tilde{f}_1$ extends to a map of S^2_∞. With more effort, the proof can be made to work with only assumption (a):

THEOREM 6.4 (Strict version of Gromov's theorem). *Let $f : M_1 \to M_2$ be any map of degree $\neq 0$ between closed oriented hyperbolic three-manifolds such that Gromov's inequality* 6.2.1 *is equality, i.e.,*

$$v(M_1) = |\deg f|\, v(M_2).$$

Then f is homotopic to a map which is a local isometry. If $|\deg f| = 1$, f is a homotopy equivalence and otherwise it is homotopic to a covering map.

PROOF. The first step in the proof is to show that a lift $\tilde{f}$ of f to the universal covering spaces extends to S^2_∞. Since the information in the hypothesis of 6.4 has to do with volume, not topology, we will know at first only that this extension is a measurable map of S^2_∞. Then, the proof of Section 6.3 will be adapted to the current situation.

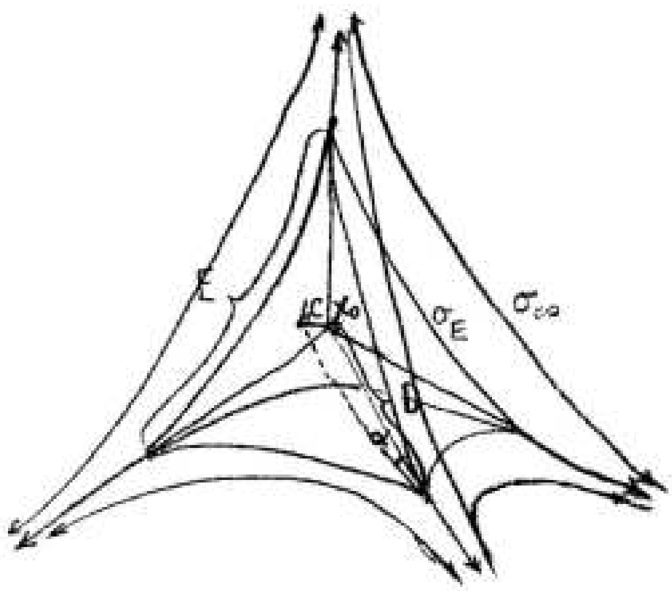
The proof works most smoothly if we have good information about the asymptotic behavior of volumes of simplices. Let σ_E be a regular simplex in H^3 all of whose edge lengths are E.

THEOREM 6.4.1. *The volume of σ_E differs from the maximal volume v_3 by a quantity which decreases exponentially with E.*

Proof. Construct copies of simplices σ_E centered at a point $x_0 \in H^3$ by drawing the four rays from a point x_0 through the vertices of an ideal regular simplex σ_∞ centered at x_0. The simplex whose vertices are on these rays, a distance D from x_0, is isometric to σ_E for some E. Let C be the distance from x_0 to any face of this simplex. The derivative $dv(\sigma_E)/dD$ is less than the area of $\partial\sigma_E$ times the maximal normal velocity of a face of σ_E. If α is the angle between such a face and the ray through x_0, we have

$$\frac{dv(\sigma_E)}{dD} < 2\pi \sin\alpha.$$

From the hyperbolic law of sines (2.6.16) $\sin\alpha = \sinh C/\sinh D$, showing that $dv(\sigma_I)/dD$ decreases exponentially with D (since $\sinh C$ is bounded). The corresponding statement for E follows since asymptotically, $E \sim 2D + \text{constant}$. $\square$ 6.14



Lemma 6.4.2. *Any simplex with volume close to v_3 has all dihedral angles close to 60°.*

Proof. Such a simplex is properly contained in an ideal simplex with any two face planes the same, so with one common dihedral angle. 6.4.2 follows form ??? $\square$

Lemma 6.4.3. *There is some constant C such that for every simplex σ with volume near v_3 and for any angle β on a face of σ,*

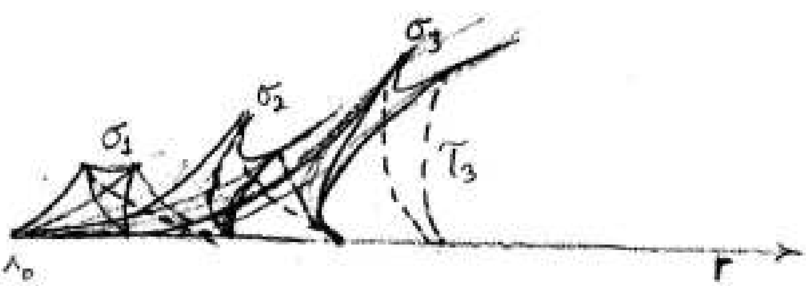
$$v_3 - v(\sigma) \geq C\beta^2.$$

Proof. If the vertex v has a face angle of β, first enlarge σ so that the other three vertices are at ∞, without changing a neighborhood of v. Now prolong one of

the edges through v to S^2_∞, and push v out along this edge. The new spike added to σ beyond v has thickness at v estimated by a linear function of β (from 2.6.12), 6.15
so its volume is estimated by a quadratic function of β. (This uses the fact that a cross-section of the spike is approximately an equilateral triangle.) $\square$

LEMMA 6.4.4. *For every point x_0 in M_1, and almost every ray r through x_0, $f_1(r)$ converges to a point on S^2_∞.*

PROOF. Let $x_0 \in H^3$, and let r be some ray emanating from x_0. Let the simplex σ_i (with all edges having length i) be placed with a vertex at x_0 and with one edge on r, and let τ_i be a simplex agreeing with σ_i in a neighborhood of x_0 but with the edge on r lengthened, to have length $i+1$.



The volume of σ_i and $\tau_i \supset \sigma_i$ deviate from the supremal value by an amount ϵ_i decreasing exponentially with i, so $\text{smear}_{M_1} \tau_i$ and $\text{smear}_{M_1} \sigma_i$ are very efficient cycles representing a multiple of $[M_1]$. Since $v(M_1) = |\deg f|\, v(M_2)$, the cycles straight $f_* \,\text{smear}_{M_1} \sigma_i$ and straight $f_* \,\text{smear}_{M_1} \tau_i$ must also be very efficient. In other words, for all but a set of measure at most $v(M_1)\epsilon_i/v_3$ of simplices σ in $\text{smear}\, \sigma_i$ (or near $\text{smear}\, \tau_i$), the simplex straight $f\sigma$ must have volume $\geq v_3 - \epsilon_i$. 6.16

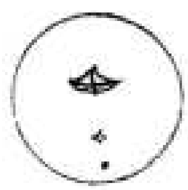
Let B be a ball around x_0 which embeds in M_i. The chains $\text{smear}_B \sigma_i$ and $\text{smear}_B \tau_i$ correspond to the measure for $\text{smear}_M \sigma_i$ and $\text{smear}_M \tau_i$ restricted to those singular simplices with the first vertex in the image of B in M_1. Thus for all but a set of measure at most $\big(2v(M_1)/v_3\big) \sum_{i=i_0}^\infty \epsilon_i$ of isometries I with take x_0 to B, all simplices $I(\sigma_i)$ and $I(\tau_i)$ for all $i > i_0$ are mapped to simplices straight $\tilde f\, \text{smear}_B \sigma$ with volume $\geq v_3 - \epsilon_i$. By 6.4.3, the sum of all face angles of the image simplices is a geometically convergent series. It follows that for all but a set of small measure of rays r emanating from points in B, $f(r)$ converges to a point on S^2_∞; in fact, by letting $i_0 \to \infty$, it follows that for almost every ray r emanating from points in B, $\tilde f(r)$ converges. Then there must be a point x' in B such that for almost every ray r emanating from x', $\tilde f(r)$ converges. Since each ray emanating from a point in H^3 is asymptotic to some ray emanating from x', this holds for rays through all points in H^3. $\square$

Remark. This measurable extension of $\tilde{f}$ to S^2_∞ actually exists under very general circumstances, with no assumption on the volume of M_1 and M_2. The idea is that if g is a geodesic in M_1, $\tilde{f}(g)$ behaves like a random walk on $\tilde{M}_2$. Almost every random walk in hyperbolic space converges to a point on S^{n-1}_∞. (Moral: always carry a map when you are in hyperbolic space!)

Lemma 6.4.5. *The measurable extension of $\tilde{f}$ to S^2_∞ carries the vertices of almost every positively oriented ideal regular simplex to the vertices of another positively oriented ideal regular simplex.*

Proof. Consider a point x_0 in H^3 and a ball B about x_0 which embeds in M, as before. Let σ_i be centered at x_0. As before, for almost all isometries I which take 6.17
x_0 to B, the sequence $\{\text{straight}\, \tilde{f} \circ I \circ \sigma_i\}$ has volume converging to v_3, and all four vertices converging to S^2_∞.

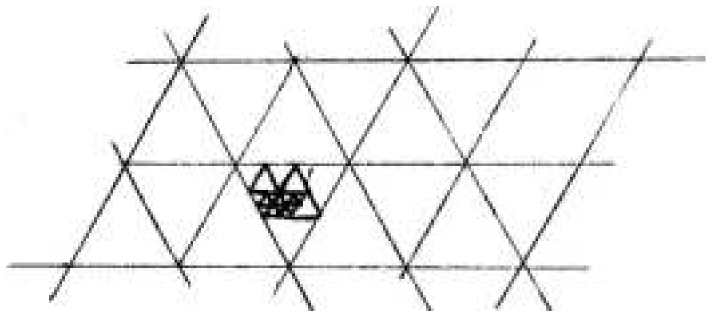
If for almost all I these four vertices converge to distinct points, we are done. Otherwise, there is a set of positive measure of ideal regular simplices such that the image of the vertex set of σ is degenerate: either all four vertices are mapped to the same point, or three are mapped to one point and the fourth to an arbitrary point. We will show this is absurd. If the degenerate cases occur

with positive measure, there is some pair of points v_0 and v_1 with $\tilde{f}(v_0) = \tilde{f}(v_1)$ such that for almost all regular ideal simplices spanned by v_0, v_1, v_2, v_3, either $\tilde{f}(v_2) = \tilde{f}(v_0)$ or $\tilde{f}(v_3) = \tilde{f}(v_0)$. Thus, there is a set A of positive measure with $\tilde{f}(A)$ a single point. Almost every regular ideal simplex with two vertices in A has one other vertex in A. It is easy to conclude that A must be the entire sphere. (One method is to use ergodicity as in the proof of 6.4 which will follow.) The image point $\tilde{f}(A)$ is invariant under covering transformations of M_1. This implies that the image of $\pi_1 M_1$ in $\pi_1 M_2$ has a fixed point on S_∞, which is absurd. □

We resume the proof of 6.4 here. It follows from 6.4.5 that there is a vertex v_0 such that for almost all regular ideal simplices spanned by v_0, v_1, v_2, v_3, the image vertices
span a regular ideal simplex. Arrange v_0 and $\tilde{f}(v_0)$ to be the point at infinity in the 6.18
upper half-space model. Three other points v_1, v_2, v_3 span a regular ideal simplex with v_0 if and only if they span an equilateral triangle in the plane, E^2. By changing coordinates, we may assume that f maps vertices of almost all equilateral triangles parallel to the x-axis to the vertices of an equilateral triangle in the plane. In complex

notation, let $\omega = \sqrt[3]{-1}$, so that $0, 1, \omega$ span an equilateral triangle. For almost all $z \in \mathbb{C}$, the entire countable set of triangles spanned by vertices of the form $z + 2^{-k}n$, $z + 2^{-k}(n+1)$, $z + 2^{-k}(n+\omega)$, for $k, n \in \mathbb{Z}$, are mapped to equilateral triangles.



Then the map $\tilde{f}$ must take the form

$$\tilde{f}\big(z + 2^{-k}(n + m\omega)\big) = g(z) + h(z) \cdot 2^{-k}(n + m\omega), \qquad k, n, m \in \mathbb{Z},$$

for almost all z. The function h is invariant a.e. by the dense group T of translations of the form $z \mapsto z + 2^{-k}(n + m\omega)$. This group is ergodic, so h is constant a.e. Similar reasoning now shows that g is constant a.e., so that f is essentially a fractional linear transformation on the sphere S^2_∞. Since $\tilde{f} \circ T_\alpha = T_{f_*\alpha} \circ \tilde{f}$, this shows that $\pi_1 M_1$ is conjugate, in $\operatorname{Isom}(H^3)$, to a subgroup of $\pi_1 M_2$. $\square$

6.19

6.5. Manifolds with Boundary

There is an obvious way to extend Gromov's invariant to manifolds with boundary, as follows. If M is a manifold and $A \subset M$ a submanifold, the relative chain group $\mathcal{C}_k(M, A)$ is defined to be the quotient $\mathcal{C}_k(M)/\mathcal{C}_k(A)$. The norm on $\mathcal{C}_k(M)$ goes over to a norm on $\mathcal{C}_k(M, A)$: the norm $\|\mu\|$ of an element of $\mathcal{C}_k(M, A)$ is the total variation of μ restricted to the set of singular simplices that do not lie in A. The norm $\|\gamma\|$ of a homology class $\gamma \in H_k(M, A)$ is defined, as before, to be the infimal norm of relative cycles representing γ. Gromov's invariant of a compact, oriented manifold with boundary $(M, \partial M)$ is $\big\|[M, \partial M]\big\|$, where $[M, \partial M]$ denotes the relative fundamental cycle.

There is a second interesting definition which makes sense in an important special case. For concreteness, we shall deal only with the case of three-manifold whose boundary consists of tori. For such a manifold M, define

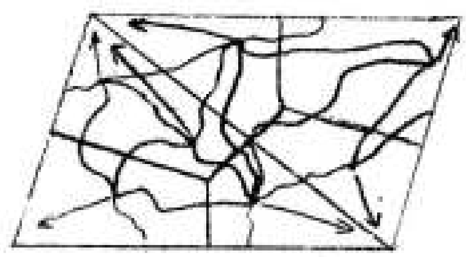
$$\|\,[M, \partial M]\,\|_0 = \lim_{a \to 0} \inf\{\|z\| \,|\, z \,\text{straight}\, [M, \partial M] \text{ and } \|\partial Z\| \le a\}.$$

Observe that ∂z represents the fundamental cycle of ∂M, so that a necessary condition for this definition to make sense is that $\|\,[\partial M]\,\| = 0$. This is true in the present situation that ∂M consists of tori, since the torus admits self-maps of degree > 1.

Then $\|(M, \partial M)\|_0$ is the limit of a non-decreasing sequence, so to insure the existence of the limit we need only find an upper bound. This involves a special property of the torus.

PROPOSITION 6.5.1. *There is a constant K such that z is any homologically trivial cycle in $\mathcal{C}_2(T^2)$, then z bounds a chain c with $\|c\| \leq K\|z\|$.*

PROOF. Triangulate T^2 (say, with two “triangles” and a single vertex). Partition T^2 into disjoint contractible neighborhoods of the vertices. Consider first the case that no simplices in the support of z have large diameter. Then there is a chain homotopy of z to its simplicial approximation $a(z)$.



The chain homotopy has a norm which is a bounded multiple of the norm of z. Since simplicial singular chains form a finite dimensional vector space, $a(z)$ is homologous to zero by a homology whose norm is a bounded multiple of the norm of $a(z)$. This gives the desired result when the simplices of z are not large. In the general case, pass to a very large cover $\tilde{T}^2$ of T^2. For any finite sheeted covering space $p : \tilde{M} \to M$ there is a canonical chain map, transfer: $\mathcal{C}_*(M) \to \mathcal{C}_*(\tilde{M})$. The transfer of a singular simplex is simply the average of its lifts to $\tilde{M}$; this extends in an obvious way to measures on singular simplices. Clearly $p \circ$ transfer $=$ id, and $\|$ transfer $c\| = \|c\|$. If z is any cycle on T^2, then for a sufficiently large finite cover $\tilde{T}^2$ of T^2, the transfer of z to $\tilde{T}^2 = T^2$ has no large 2-simplices in its support. Then transfer z is the boundary of a chain c with $\|c\| \leq K\|z\|$ for some fixed K. The projection of c back to the base space completes the proof. $\square$

6.21

We now have upper bounds for $\|\,[M, \partial M]\,\|_0$. In fact, let z be any cycle representing $[M, \partial M]$, and let ϵ be any cycle representing $[\partial M]$. By piecing together z with a homology from ∂z to ϵ given by 6.5.1, we find a cycle z' representing $[M, \partial M]$ with $\|z'\| \leq \|z\| + K(\|\partial z\| + \|\epsilon\|)$. Passing to the limit as $\|\epsilon\| \to 0$, we find that $\|\,[M, \partial M]\,\| \leq \|z\| + K\|\partial z\|$.

The usefulness of the definition of $\|\,[M, \partial M]\,\|_0$ arises from the easy

PROPOSITION 6.5.2. *Let $(M, \partial M)$ be a compact oriented three-manifold, not necessarily connected, with ∂M consisting of tori. Suppose $(N, \partial N)$ is an oriented manifold obtained by gluing together certain pairs of boundary components of M. Then*

$$\| [N, \partial N] \|_0 \leq \| [M, \partial M] \|_0.$$

COROLLARY 6.5.3. *If $(S, \partial S)$ is any Seifert fiber space, then*

$$\| [S, \partial S] \|_0 = \| [S, \partial S] \| = 0.$$

(The case $\partial S = \phi$ is included.)

PROOF OF COROLLARY. If S is a circle bundle over a connected surface M with non-empty boundary, then S (or a double cover of it, if the fibers are not oriented) is $M \times S^1$. Since it covers itself non-trivially its norm (in either sense) is 0. If S is a circle bundle over a closed surface M, it is obtained by identification of $(M - D^2) \times S^1$ with $D^2 \times S^1$, so its norm is also zero. If S is a Seifert fibration, it is obtained by identifying solid torus neighborhoods of the singular fibers with the complement which is a fibration. □

PROOF OF 6.5.2. A cycle z representing $[M, \partial M]$ with $\|\partial z\| \leq \epsilon$ goes over to a chain on $[N, \partial N]$, which can be corrected to be a cycle z' with $\|z\|' \leq \|z\| + K\epsilon$. □ 6.22

If M is a complete oriented hyperbolic manifold with finite total volume, recall that M is the interior of a compact manifold $\bar{M}$ with boundary consisting of tori. Both $\| [\bar{M}, \partial \bar{M}] \|$ and $\| [\bar{M}, \partial \bar{M}] \|_0$ can be computed in this case:

LEMMA 6.5.4 (Relative version of Gromov's Theorem). *If M is a complete oriented hyperbolic three-manifold with finite volume, then*

$$\| [\bar{M}, \partial \bar{M}] \|_0 = \| [\bar{M}, \partial \bar{M}] \| = \frac{v(M)}{v_3}.$$

PROOF. Let σ be a 3-simplex whose volume is nearly the maximal value, v_3. Then $\mathrm{smear}_M \, \sigma$ is a measure on singular cycles with non-compact support. Restrict this measure to simplices not contained in $M_{(0,\epsilon]}$, and project to $M_{[\epsilon,\infty)}$ by a retraction of M to $M_{[\epsilon,\infty)}$. Since the volume of $M_{(0,\epsilon]}$ is small for small ϵ, this gives a relative fundamental cycle z' for

$$(M_{[\epsilon,\infty)}, \partial M_{[\epsilon,\infty)}) = (\bar{M}, \partial \bar{M})$$

with $\|z'\| \approx \frac{v(M)}{v_3}$ and with $\|\partial z'\|$ small. This proves that

$$\frac{v(M)}{v_3} \geq \| [\bar{M}, \partial \bar{M}] \|_0.$$

There is an immediate inequality

$$\| [\bar{M}, \partial \bar{M}] \|_0 \geq \| [\bar{M}, \partial \bar{M}] \|.$$

To complete the proof, we will show that $\| [\bar{M}, \partial \bar{M}] \| \geq v(M)/v_3$. This is done by a straightening operation, as in 6.1.7. For this, note that if σ is any simplex lying in 6.23
$M_{(0,\epsilon]}$, then straight(σ) also lies in $M_{(0,\epsilon]}$, since $M_{(0,\epsilon]}$ is convex. Hence we obtain a chain map

$$\text{straight} : \mathcal{C}_*(M, M_{(0,\epsilon]}) \to \mathcal{C}_*(M, M_{(0,\epsilon]}),$$

chain homotopic to the identity, and not increasing norms. As in 6.1.7, this gives the inequality

$$\| [M, M_{(0,\epsilon]}] \| \geq \frac{v(M_{[\epsilon,\infty)})}{v_3}.$$

Since for small ϵ there is a chain isomorphism between $\mathcal{C}_k(M, M_{(0,\epsilon]})$ and $\mathcal{C}_k(\bar{M}, \partial \bar{M})$ which is a $\| \ \|$-isometry, this proves 6.5.4. □

Here is an inequality which enables one to compute Gromov's invariant for much more general three-manifolds:

THEOREM 6.5.5. *Suppose M is a closed oriented three-manifold and $H \subset M$ is a three-dimensional submanifold with a complete hyperbolic structure of finite volume. Suppose $\bar{H}$ is embedded in M and that $\partial \bar{H}$ is incompressible. Then*

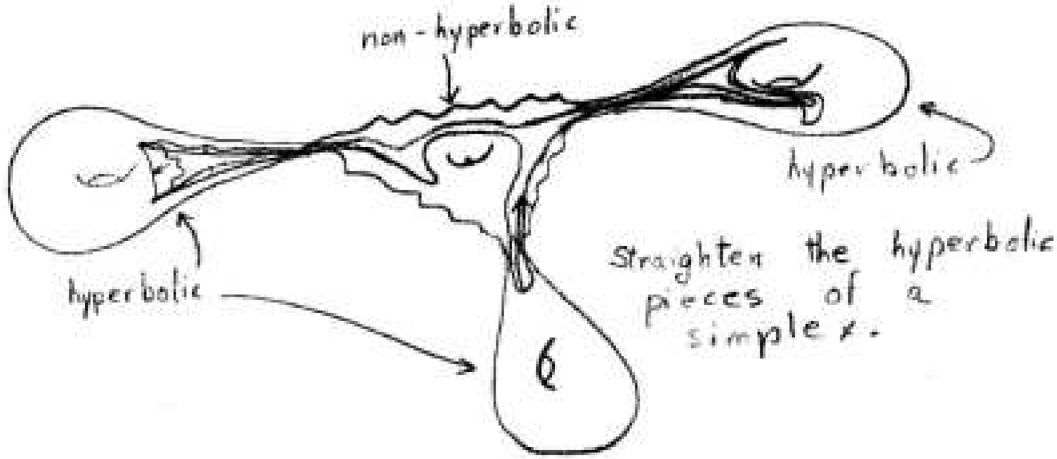
$$\| [M] \| \geq \frac{v(H)}{v_3}.$$

REMARK. Of course, the hypothesis that $\partial \bar{H}$ is incompressible is necessary; otherwise M might be S^3. If H were not hyperbolic, further hypotheses would be needed to obtain an inequality. Consider, for instance, the product $M_g \times I$ where M_g is a surface of genus $g > 1$. Then $\| [M_g] \| = 2\, v(M_g)/\pi = 4\, |\chi(M_g)|$, so

$$\| [M_g \times I, \partial(M_g \times I)] \| \geq \| [M_g] \| \geq 4\, |\chi(M_g)|.$$

On the other hand, one can identify the boundary of this manifold to obtain $M_g \times S^1$, 6.24
which has norm 0. The boundary can also be identified to obtain hyperbolic manifolds (see §4.6, or §). Since finite covers of arbitrarily high degree and with arbitrarily high norm can also be obtained by gluing the boundary of the same manifold, no useful inequality is obtained in either direction.

PROOF. Since this is a digression, we give only a sketch of a proof.



□

With 6.5.5 combined with 6.5.2, one can compute Gromov's invariant for any manifold which is obtained from Seifert fiber spaces and complete hyperbolic manifolds of finite volume by identifying along incompressible tori.

The strict and relative versions of Gromov's theorems may be combined; here is the most interesting case:

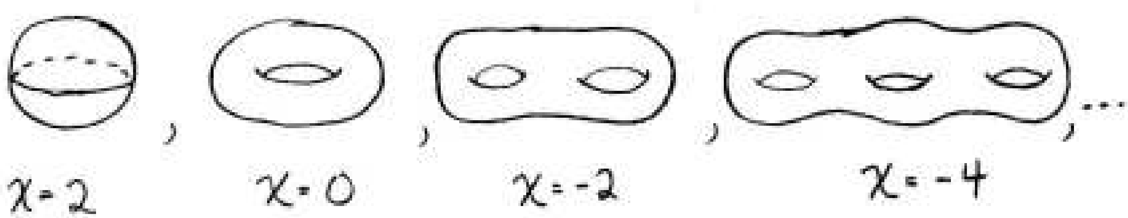
THEOREM 6.5.6. *Suppose M_1 is a complete hyperbolic manifold of finite volume and that $M_2 \neq M_1$ is a complete hyperbolic manifold obtained topologically by replacing certain cusps of M_1 by solid tori. Then $v(M_1) > v(M_2)$.*

6.25

PROOF. No new ideas are needed. Consider some map $f : M_1 \to M_2$ which collpases certain components of $M_{1_{(0,\epsilon]}}$ to short geodesics in M_2. Now apply the proof of 6.4. □

6.6. Ordinals

Closed oriented surfaces can be arranged very neatly in a single sequence,



in terms of their Euler characteristic. What happens when we arrange all hyperbolic three-manifolds in terms of their volume? From Jørgensen's theorem, 5.12 it follows that the set of volumes is a closed subset of $\mathbb{R}_+$. Furthermore, by combining Jørgensen's theorem with the relative version of Gromov's theorem, 6.5.4, we obtain

Corollary 6.6.1. *The set of volumes of hyperbolic three-manifolds is well-ordered.*

Proof. Let $v(M_1) \geq v(M_2) \geq \ldots \geq v(M_k) \geq \ldots$ be any non-ascending sequence of volumes. By Jørgensen's theorem, by passage to a subsequence we may assume that the sequence $\{M_i\}$ converges geometrically to a manifold M, with $v(M) \leq \lim v(M_i)$. By 6.5.2, eventually $\| [M_i] \|_0 \leq \| [M] \|_0$, so 6.5.4 implies that the sequence of volumes is eventually constant. □

6.26

Corollary 6.6.2. *The volume is a finite-to-one function of hyperbolic manifolds.*

Proof. Use the proof of 6.6.1, but apply the strict inequality 6.5.6 in place of 6.5.2, to show that a convergent sequence of manifolds with non-increasing volume must be eventually constant. □

In view of these results, the volumes of complete hyperbolic manifolds are indexed by countable ordinals. In other words, there is a smallest volume v_1, a next smallest volume v_2, and so forth. This sequence $v_1 < v_2 < v_3 < \cdots < v_k < \cdots$ has a limit point v_ω, which is the smallest volume of a complete hyperbolic manifold with one cusp. The next smallest manifold with one cusp has volume $v_{2\omega}$. It is a limit of manifolds with volumes $v_{\omega+1}, v_{\omega+2}, \ldots, v_{\omega+k}, \ldots$. The first volume of a manifold with two cusps is v_{ω^2}, and so forth. (See the discussion on pp. 5.59–5.60, as well as Theorem 6.5.6.) The set of all volumes has order type ω^ω. These volumes are indexed by the ordinals less than ω^ω, which are represented by polynomials in ω. Each volume of a manifold with k cusps is indexed by an ordinal of the form $\alpha \cdot \omega^k$, (where the product $\alpha \cdot \beta$ is the ordinal corresponding to the order type obtained by replacing each element of α with a copy of β). There are examples where α is a limit ordinal. These can be constructed from coverings of link complements. For instance, the Whitehead link complement has two distinct 2-fold covers; one has two cusps and the other has three, so the common volume corresponds to an ordinal divisible by ω^3. I do not know any examples of closed manifolds corresponding to limit ordinals.

It would be very interesting if a computer study could determine some of the low volumes, such as $v_1, v_2, v_\omega, v_{\omega^2}$. It seems plausible that some of these might come from Dehn surgery on the Borromean rings.

There is some constant C such that every manifold with k cusps has volume $\geq C \cdot k$. This follows from the analysis in 5.11.2: the number of boundary components of $M_{[\epsilon,\infty)}$ is bounded by the number of disjoint $\epsilon/2$ balls which can fit in M. It would be interesting to calculate or estimate the best constant C.

COROLLARY 6.6.3. *The set of values of Gromov's invariant $\| [\,] \|_0$ on the class of connected manifolds obtained from Seifert fiber spaces and complete hyperbolic manifolds of finite volume by identifying along incompressible tori is a closed well-ordered subset of $\mathbb{R}^+$, with order type ω^ω.*

We shall see later (§) that this class contains all Haken manifolds with toral boundaries.

PROOF. Extend the volume function to $v(M) = v_3 \cdot \| [M] \|_0$ when M is not hyperbolic. From 6.5.5 and 6.5.2, we know that every value of v is a finite sum of volumes of hyperbolic manifolds. Suppose $\{w_i\}$ is a bounded sequence of values of v. Express each w_i as the sum of volumes of hyperbolic pieces of a manifold M_i with $v(M)_i = w_i$. The number of terms is bounded, since there is a lower bound to the volume of a hyperbolic manifold, so we may pass to an infinite subsequence where the number of terms in this expression is constant. Since every infinite sequence of ordinals has an infinite non-decreasing subsequence, we may pass to a subsequence of w_i's where all terms in these expressions are non-decreasing. This proves that 6.28
the set of values of v is well-ordered. Furthermore, our subsequence has a limit $w = v_{\alpha_1} + \cdots + v_{\alpha_k}$, which is expressed as a sum of limits of non-decreasing sequences of volumes. Each v_{α_j} is the volume of a hyperbolic manifold M_j with at least as many cusps as the limiting number of cusps of the corresponding hyperbolic piece of M_i. Therefore, the $\bar{M}_j$'s may be glued together to obtain a manifold M with $v(M) = w$. This shows the set of values of v is closed. The fact that the order type is ω^ω can be deduced easily by showing that every value of v is not in the k-th derived set, for some integer k; in fact, $k \leq v/C$, where C is the constant just discussed. □

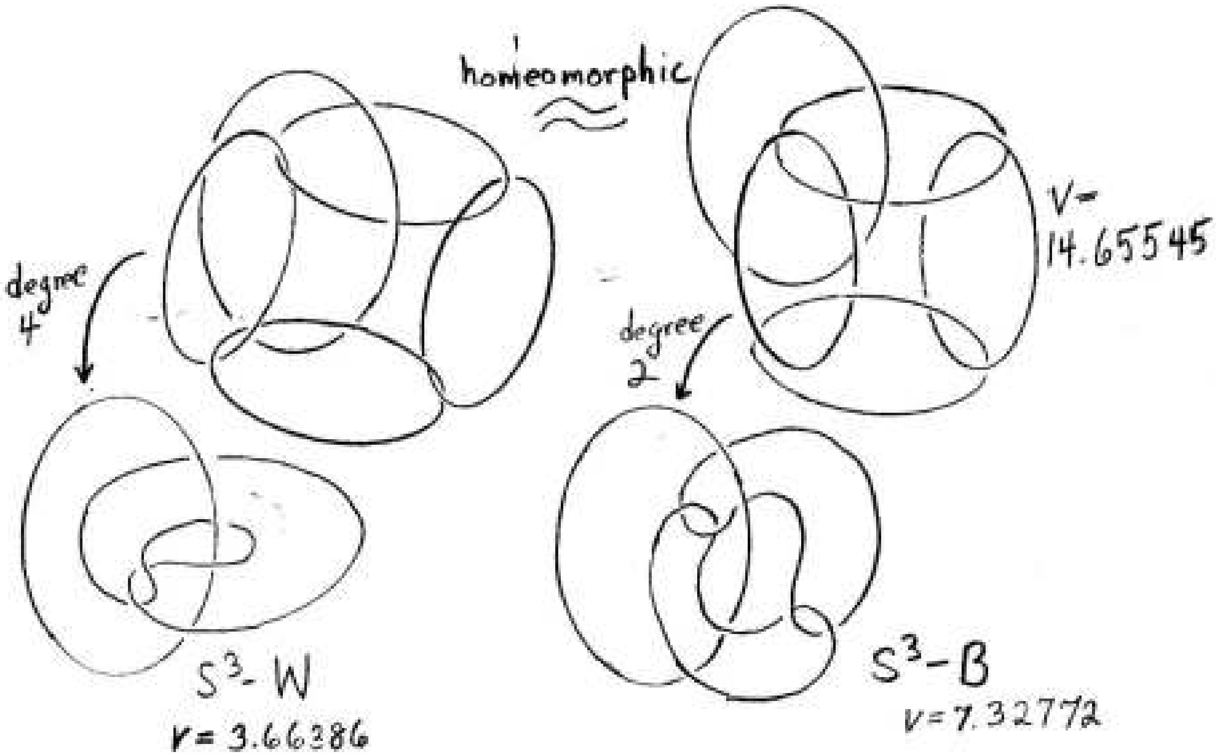
6.7. Commensurability

DEFINITION 6.7.1. If Γ_1 and Γ_2 are two discrete subgroups of isometries of H^n, then Γ_1 is *commensurable* with Γ_2 if Γ_1 is conjugate (in the group of isometries of H^n) to a group Γ_1' such that $\Gamma' \cap \Gamma_2$ has finite index in Γ_1' and in Γ_2.

DEFINITION 6.7.2. Two manifolds M_1 and M_2 are *commensurable* if they have finited sheeted covers $\tilde{M}_1$ and $\tilde{M}_2$ which are homeomorphic.

Commensurability in either sense is an equivalence relation, as the reader may easily verify. 6.29
Labelled 6.7.3.ex

EXAMPLE 6.7.3. If W is the Whitehead link and B is the Borromean rings, then $S^3 - W$ has a four-sheeted cover homeomorphic with a two sheeted cover of $S^3 - B$:



The homeomorphism involves cutting along a disk, twisting 360° and gluing back. Thus $S^3 - W$ and $S^3 - B$ are commensurable. One can see that $\pi_1(S^3 - W)$ and $\pi_1(S^3 - B)$ are commensruable as discrete subgroups of $\mathrm{PSL}(2, \mathbb{C})$ by considering the tiling of H^3 by regular ideal octahedra. Both groups preserve this tiling, so they are contained in the full group of symmetries of the octahedral tiling, with finite index. Therefore, they intersect each other with finite index.

$$\pi_1(S^3 - B) \subset \text{Symmetries (octahedral tiling)} \supset \pi_1(S^3 - W)$$
$$\pi_1(S^3 - B) \supset \pi_1(S^3 - B) \cap \pi_1(S^3 - W) \subset \pi_1(S^3 - W)$$

6.30

Warning. Two groups Γ_1 and Γ_2 can be commensurable, and yet not be conjugate to subgroups of finite index in a single group.

Proposition 6.7.3. *If M_1 is a complete hyperbolic manifold with finite volume and M_2 is commensurable with M_1, then M_2 is homotopy equivalent to a complete hyperbolic manifold.*

Proof. This is a corollary of Mostow's theorem. Under the hypotheses, M_2 has a finite cover M_3 which is hyperbolic. M_3 has a finite cover M_4 which is a regular cover of M_2, so that $\pi_1(M_4)$ is a normal subgroup of $\pi_1(M_2)$. Consider the action of $\pi_1(M_2)$ on $\pi_1(M_4)$ by conjugation. $\pi_1(M_4)$ has a trivial center, so in other words the action of $\pi_1(M_4)$ on itself is effective. Then for every $\alpha \in \pi_1(M_2)$, since some power of α^k is in $\pi_1(M_4)$, α must conjugate $\pi_1(M_4)$ non-trivially. Thus $\pi_1(M_2)$ is isomorphic to a group of automorphisms of $\pi_1(M_4)$, so by Mostow's theorem it is a discrete group of isometries of H^n. □

In the three-dimensional case, it seems likely that M_1 would actually be hyperbolic. Waldhausen proved that two Haken manifolds which are homotopy equivalent are homeomorphic, so this would follow whenever M_1 is Haken. There are some sorts of properties of three-manifolds which do not change under passage to a finite-sheeted cover. For this reason (and for its own sake) it would be interesting to have a better understanding of the commensurability relation among three-manifolds. This is difficult to approach from a purely topological point of view, but there is a great deal of information about commensurability given by a hyperbolic structure. For instance, in the case of a complete non-compact 6.31

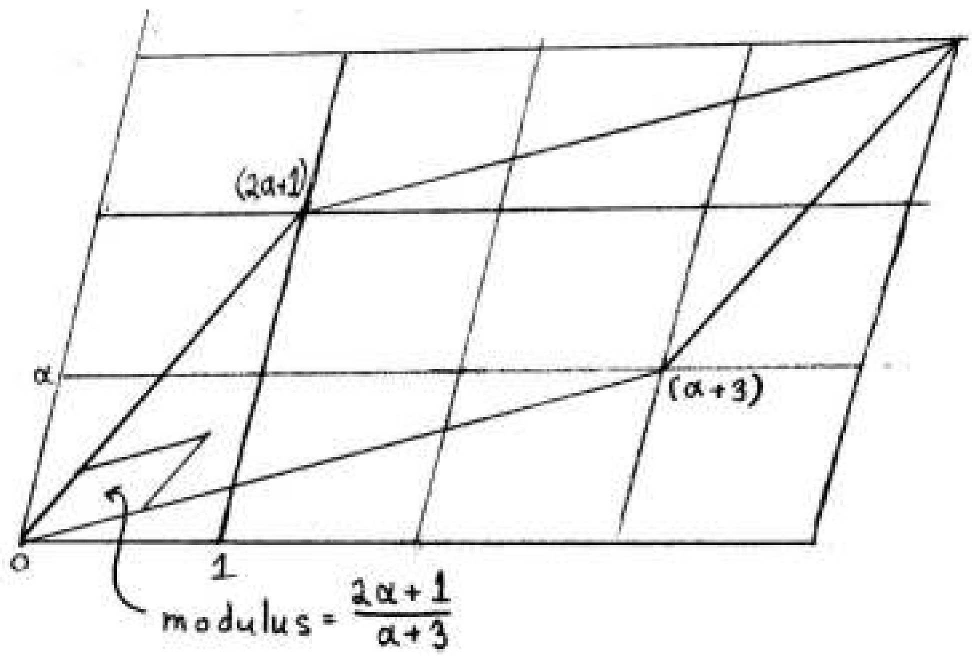
hyperbolic three-manifold M of finite volume, each cusp gives a canonical Euclidean structure on a torus, well-defined up to similarity. A convenient invariant for this structure is obtained by arranging M so that the cusp is the point at ∞ in the upper half space model and one generator of the fundamental group of the cusp is a translation $z \mapsto z + 1$. A second generator is then $z \mapsto z + \alpha$. The set of complex numbers $\alpha_1 \dots \alpha_k$ corresponding to various cusps is an invariant of the commensurability class of M well-defined up to the equivalence relation

$$\alpha_i \sim \frac{n\alpha_i + m}{p\alpha_i + q},$$

where

$$n, m, pq \in \mathbb{Z}, \quad \begin{vmatrix} n & m \\ p & q \end{vmatrix} \neq 0.$$

$(n, m, p$ and q depend on $i)$.



6.32

In particular, if $\alpha \sim \beta$, then they generate the same fields $\mathbb{Q}(\alpha) = \mathbb{Q}(\beta)$.

Note that these invariants α_i are always algebraic numbers, in view of

Proposition 6.7.4. *If Γ is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$ such that H^3/Γ has finite volume, then Γ is conjugate to a group of matrices whose entries are algebraic.*

Proof. This is another easy consequence of Mostow's theorem. Conjugate Γ so that some arbitrary element is a diagonal matrix

$$\begin{bmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{bmatrix}$$

and some other element is upper triangular,

$$\begin{bmatrix} \lambda & x \\ 0 & \lambda^{-1} \end{bmatrix}.$$

The component of Γ in the algebraic variety of representations of Γ having this form is 0-dimensional, by Mostow's theorem, so all entries are algebraic numbers. $\square$

One can ask the more subtle question, whether all entries can be made algebraic integers. Hyman Bass has proved the following remarkable result regarding this question:

Theorem 6.7.5 (Bass). *Let M be a complete hyperbolic three-manifold of finite volume. Then either $\pi_1(M)$ is conjugate to a subgroup of $\mathrm{PSL}(2, \mathcal{O})$, where $\mathcal{O}$ is the ring of algebraic integers, or M contains a closed incompressible surface (not homotopic to a cusp).*

The proof is out of place here, so we omit it. See Bass. As an example, very few knot complements seem to contain non-trivial closed incompressible surfaces. The property that a finitely generated group Γ is conjugate to a subgroup of $\mathrm{PSL}(2, \mathcal{O})$ is equivalent to the property that the *additive* group of matrices generated by Γ 6.33
is finitely generated. It is also equivalent to the property that the trace of every element of Γ is an algebraic integer. It is easy to see from this that every group commensurable with a subgroup of $\mathrm{PSL}(2, \mathcal{O})$ is itself conjugate to a subgroup of $\mathrm{PSL}(2, \mathcal{O})$. (If $\operatorname{Tr} \gamma^n = a$ is an algebraic integer, then an eigenvalue λ of γ satisfies $\lambda^{2n} - a\lambda^n + 1 = 0$. Hence λ, λ^{-1} and $\operatorname{Tr} \gamma = \lambda + \lambda^{-1}$ are algebraic integers).

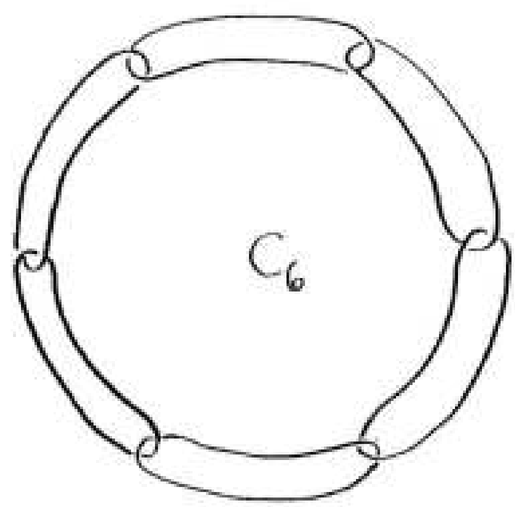
If two manifolds are commensurable, then their volumes have a rational ratio. We shall see examples in the next section of incommensurable manifolds with equal volume.

Questions 6.7.6. Does every commensurability class of discrete subgroups of $\mathrm{PSL}(2, \mathbb{C})$ have a finite collection of maximal groups (up to isomorphism)?

Is the set of volumes of three-manifolds in a given commensurability class a discrete set, consisting of multiples of some number V_0?

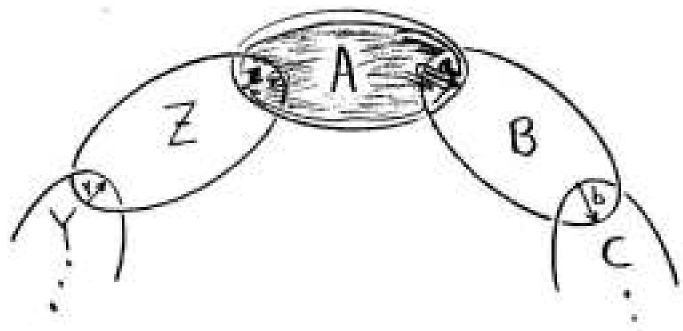
6.8. Some Examples

Example 6.8.1. Consider the k-link chain C_k pictured below:

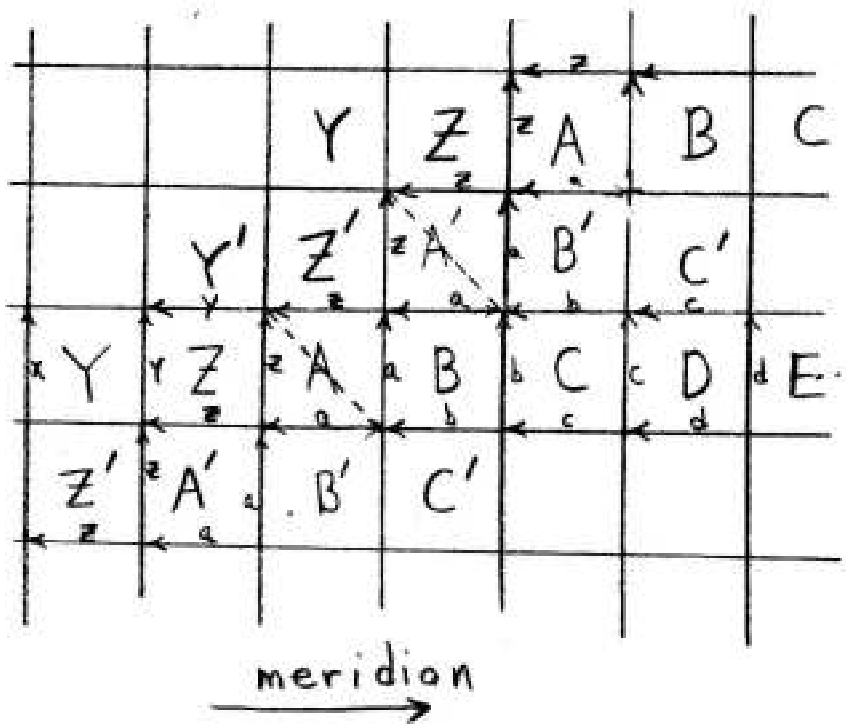


6.34

If each link of the chain is spanned by a disk in the simplest way, the complement of the resulting complex is an open solid torus.

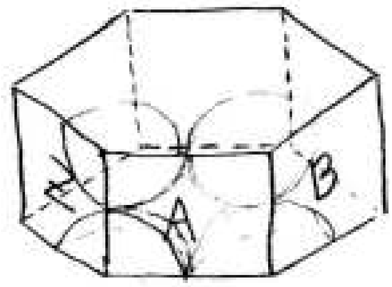


$S^3 - C_k$ is obtained from a solid torus, with the cell division below on its boundary, by deleting the vertices and identifying.

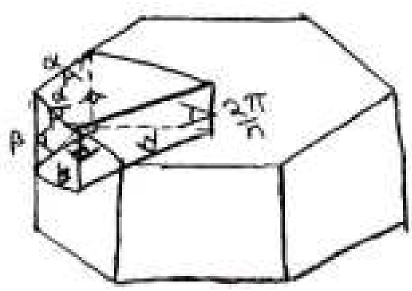


6.35

To construct a hyperbolic structure for $S^3 - C_k$, cut the solid torus into two drums.

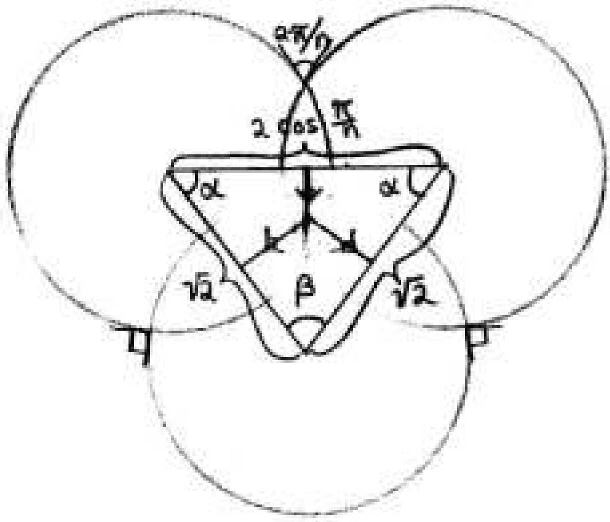


Let P be a regular k-gon in H^3 with all vertices on S^2_∞. If P' is a copy of P obtained by displacing P along the perpendicular to P through its center, then P' and P can be joined to obtain a regular hyperbolic drum. The height of P' must be adjusted so that the reflection through the diagonal of a rectangular side of the drum is an isometry of the drum. If we subdivide the drum into $2k$ pieces as shown,



6.36

the condition is that there are horospheres about the ideal vertices tangent to three faces. Placing the ideal vertex at ∞ in upper half-space, we have a figure bounded by three vertical Euclidean planes and three Euclidean hemispheres of equal radius r. Here is a view from above:



From this figure, we can compute the dihedral angles α and β of the drum to be

$$\alpha = \operatorname{arc\,cos}\left(\frac{\cos \pi/k}{\sqrt{2}}\right), \quad \beta = \pi - 2\alpha.$$

Two copies of the drum with these angles can now be glued together to give a hyperbolic structure on $S^3 - C_k$. (Note that the total angle around an edge is $4\alpha + 2\beta = 2\pi$. Since the horospheres about vertices are matched up by the gluing maps, we obtain a complete hyperbolic manifold).

From Milnor's formula (6), p. 7.15, for the volume, we can compute some values. 6.37

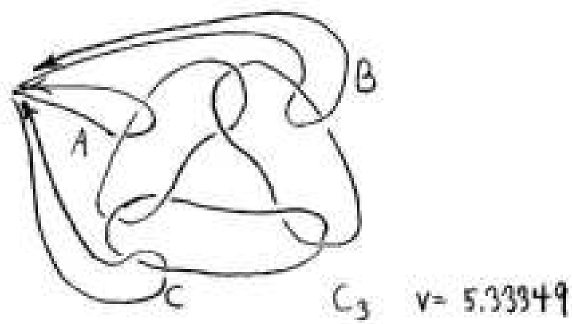
k	$v(S^3 - C_k)$	$v(S^3 - C_k)/k$	
2	0	0	(Seifert fiber space)
3	5.33349	1.77782	$\sim \mathrm{PSL}(2, \mathcal{O}_7)$
4	10.14942	2.53735	$\sim \mathrm{PSL}(2, \mathcal{O}_3)$
5	14.60306	2.92061	
6	18.83169	3.13861	
7	22.91609	3.27373	
10	34.691601	3.4691601	
50	182.579859	3.65159719	
200	732.673784	3.66336892	
1000	3663.84264	3.66384264	
8000	29310.8990	3.66386238	
∞	∞	3.66386238	Whitehead link

Note that the quotient space $(S^3 - C_k)/\mathbb{Z}_k$ by the rotational symmetry of C_k is obtained by generalized Dehn surgery on the White head link W, so the limit of 6.38
$v(C_k)/k$ as $k \to \infty$ is the volume of $S^3 - W$.

Note also that whenever k divides l, then there is a degree $\frac{l}{k}$ map from $S^3 - C_l$ to $S^3 - C_k$. This implies that $v(S^3 - C_l)/l > v(S^3 - C_k)/k$. In fact, from the table it is clear that these numbers are strictly increasing with k.

The cases $k = 3$ and 4 have particular interest.

EXAMPLE 6.8.2. The volume of $S^3 - C_3$ per cusp has a particularly low value (1.7778). The holonomy of the hyperbolic structure can be described by



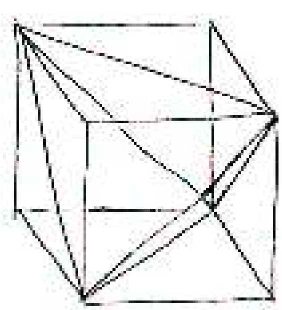
$$H(A) = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}$$

$$H(B) = \begin{bmatrix} 1+\alpha & \alpha \\ -\alpha & 1-\alpha \end{bmatrix}$$

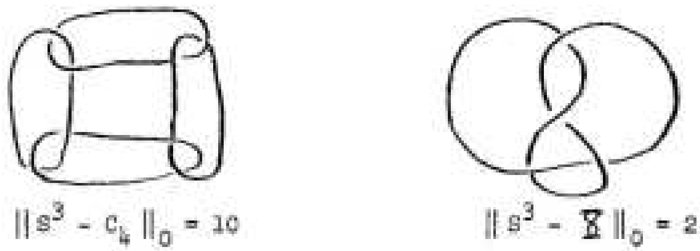
$$H(C) = \begin{bmatrix} 1 & 0 \\ -\alpha & 1 \end{bmatrix}$$

where $\alpha = \frac{-1+\sqrt{-7}}{2}$. Thus $\pi_1(X^3 - C_3)$ is a subgroup of $\mathrm{PSL}(2, \mathcal{O}_7)$ where $\mathcal{O}_d$ is the ring of integers in $\mathbb{Q}\sqrt{-d}$. See §7.4. Referring to Humbert's formula 7.4.1, we find $v(H^3/\,\mathrm{PSL}(2, \mathcal{O}_7) = .8889149\ldots$, so $\pi_1(S^3 - C_3)$ has index 6 in this group.

EXAMPLE 6.8.3. When $k = 4$, the rectangular-sided drum becomes a cube with all dihedral angles $60°$. This cube may be subdivided into five regular ideal tetrahedra: 6.39



Thus $S^3 - C_4$ is commensurable with S^3- figure eight knot, since $\pi_1(S^3 - C_4)$ preserves a tiling of H^3 by regular ideal tetrahedra.



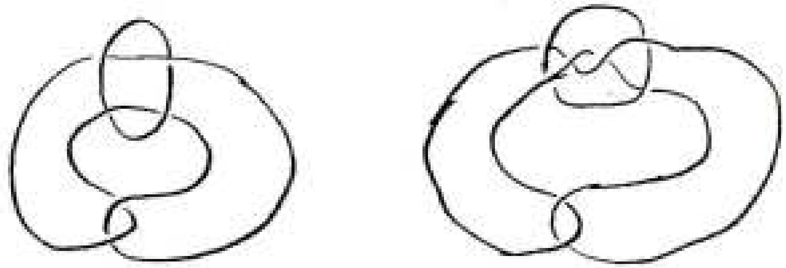
commensurable with $\mathrm{PSL}(2, \mathcal{O}_3)$

$S^3 - C_k$ is homeomorphic to many other link complements, since we can cut along any disk spanning a component of C_k, twist some integer number of times and glue back to obtain a link with a complement homeomorphic to that of C_k. Furthermore, if we glue back with a half-integer twist, we obtain a link whose complement is hyperbolic with the same volume as $S^3 - C_k$. This follows since twice-punctured spanning disks are totally geodesic thrice-punctured spheres in the hyperbolic structure of $S^3 - C_k$. The thrice-punctured sphere has a unique hyperbolic structure, and all six isotopy classes of diffeomorphisms are represented by isometries. 6.40

Using such operations, we obtain these examples for instance:

EXAMPLE 6.8.4.

$v = v(S^3 - C_3)$
$= 5.33349$



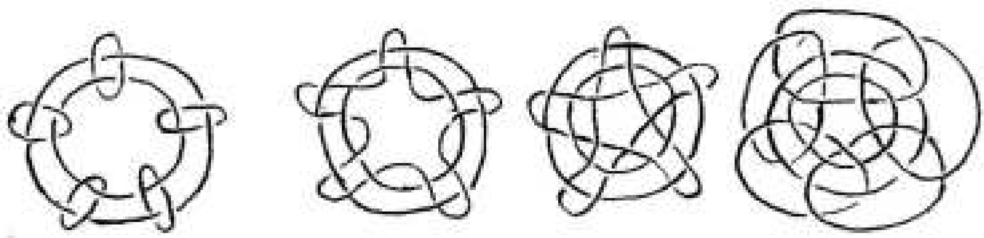
commensurable with C_3

The second link has a map to the figure-eight knot obtained by erasing a component of the link. Thus, by 6.5.6, we have

$$v(S^3 - C_3) = 5.33340\ldots > 2.02988 = v(S^3 - \text{figure eight knot}).$$

These links are commensurable with C_3, since they give rise to identical tilings of H^3 by drums. As another example, the links below are commensurable with C_{10}:

EXAMPLE 6.8.5.



$k = 5$ Commensurable with C_{10} $\quad v = 34.69616$

6.41

The last three links are obtained from the first by cutting along 5-times punctured disks, twisting, and gluing back. Since this gluing map is a diffeomorphism of the surface which extends to the three-manifold, it must come from an isometry of a 6-punctured sphere in the hyperbolic structure. (In fact, this surface comes from the top of a 10-sided drum).

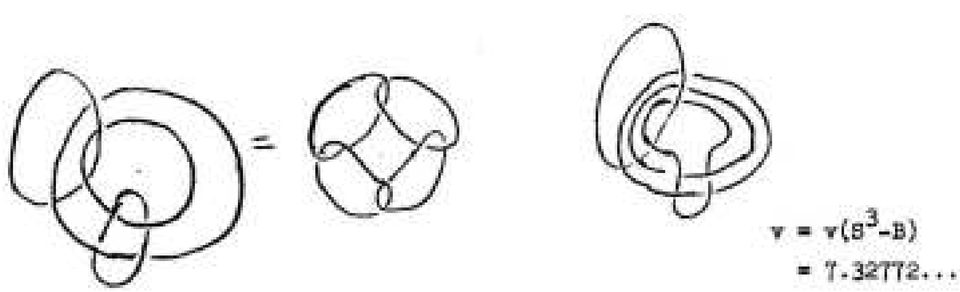
The compex modulus associated with a cusp of C_n is

$$\frac{1}{2} \quad \left(1 + \sqrt{\frac{1+\sin^2\frac{\pi}{n}}{\cos^2\frac{\pi}{n}}}\ i\right).$$

Clearly we have an infinite family of incommensurable examples.

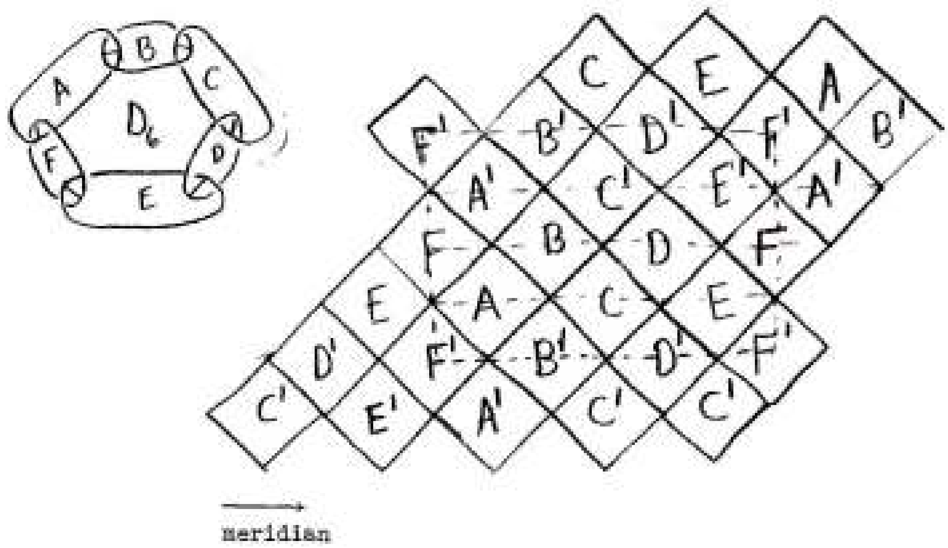
By passing to the limit $k \to \infty$ and dividing by $\mathbb{Z}$, we get these links commensurable with $S^3 - W$ and $S^3 - B$, for instance:

Example 6.8.6.

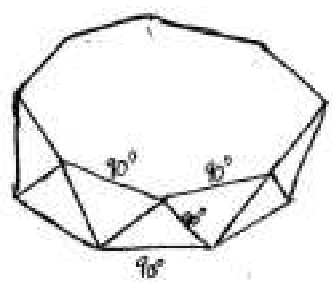


Many other chains, with different amounts of twist, also have hyperbolic structures. They all are obtained, topologically, by identifying faces of a tiling of the boundary of a solid torus by rectangles. Here is another infinite family $D_{2k} (\geq 3)$ which is easy to compute: 6.42

Example 6.8.7.



Hyperbolic structures can be realized by subdividing the solid torus into 4 drums with triangular sides:



6.43

Regular drums with all dihedral angles 90° can be glued together to give $S^3 - D_k$. By methods similar to Milnor's in 7.3, the formula for the volume is computed to be

$$v(S^3 - D_{2k}) = 8k\big(\mu(\tfrac{\pi}{4} + \tfrac{\pi}{2k}) + \mu(\tfrac{\pi}{4} - \tfrac{\pi}{2k})\big).$$

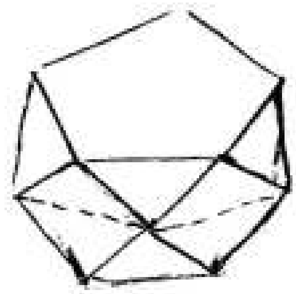
Thus we have the values

k	$v(S^3 - D_{2k})$	$v(S^3 - D_{2k})/(2k)$
3	14.655495	2.44257
4	24.09218	3.01152
5	32.55154	3.25515
6	40.59766	3.38314
100	732.750	3.66288
1000	7327.705	3.66386
∞	∞	3.66386

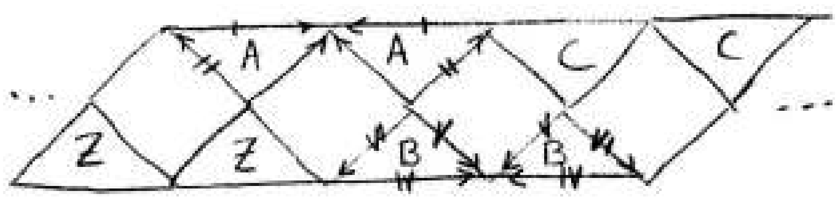
The cases $k = 3$ and $k = 4$ have algebraic significance. They are commensurable with $\mathrm{PSL}(2, \mathcal{O}_1)$ nad $\mathrm{PSL}(2, \mathcal{O}_2)$, respectively. When $k = 3$, the drum is an octahedron and $v(S^3 - D_{2k}) = 4v(S^3 - W)$.

Note that the volume of $(S^3 - D_{12})$ is 20 times the volume of the figure-eight knot complement. 6.44

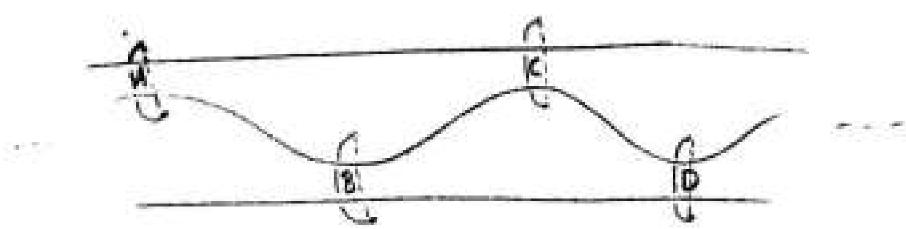
Two copies of the triangular-sided drum form this figure:



The faces may be glued in other patterns to obtain link complements. For instance, if k is even we can first identify



the triangular faces, to obtain a ball minus certain arcs and curves on the boundary.



6.45

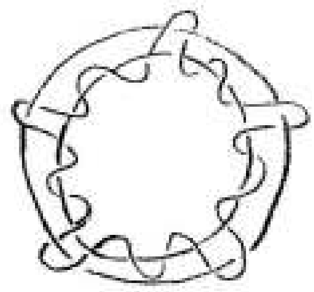
If we double this figure, we obtain a complete hyperbolic structure for the complement of this link, E_l:

EXAMPLE 6.8.8.

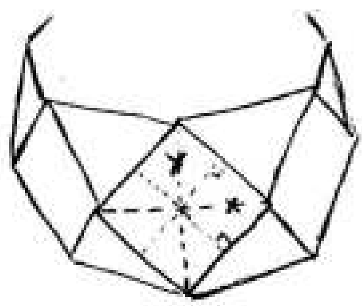
E_5

Alternatively, we can identify the boundary of the ball to obtain

Example 6.8.9.

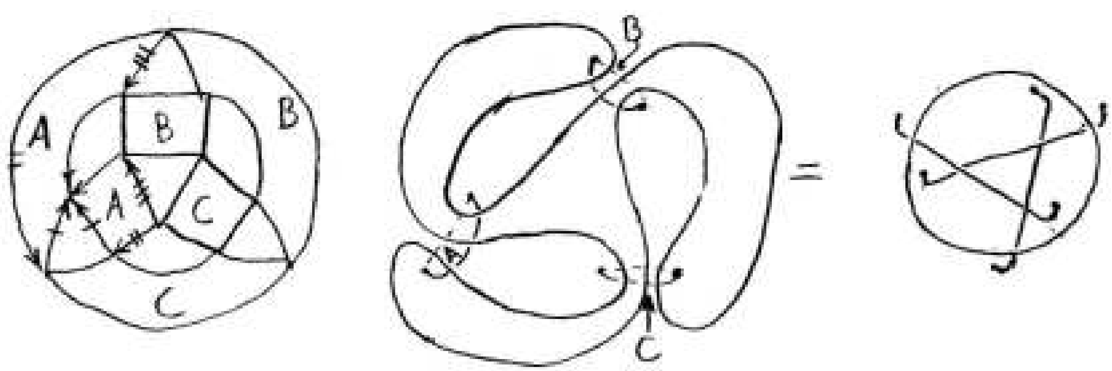


In these examples, note that the rectangular faces of the doubled drums



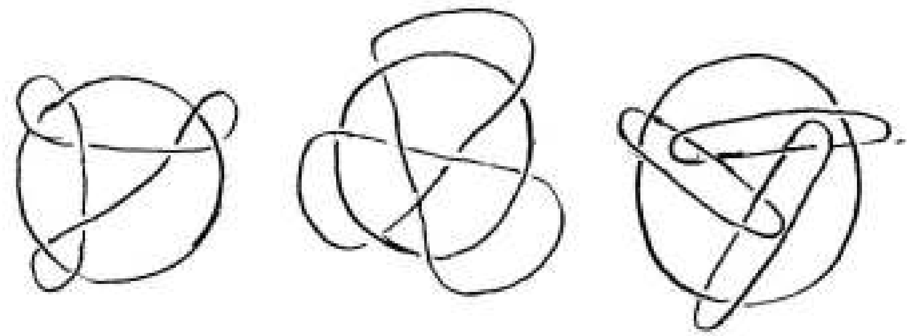
have complete symmetry, and some of the link complements are obtained by gluing maps which interchange the diagonals, while others preserve them. These links are generally commensurable even when they have the same volume; this can be proven by computing the moduli of the cusps. 6.46

There are many variations. Two copies of the drum with 8 triangular faces, glued together, give a cube with its corners chopped off. The 4-sided faces can be glued, to obtain the ball minus these arcs and curves:



The two faces of the ball may be glued together (isometrically) to give any of these link complements: 6.47

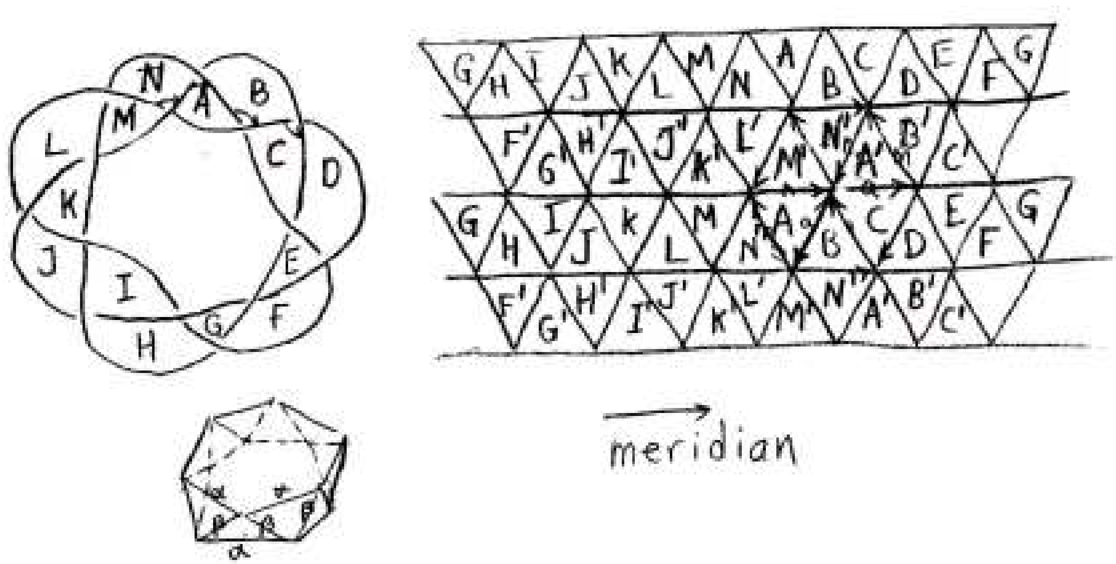
Example 6.8.10.



$v = 12.04692 = \frac{1}{2}v(S^3 - D_8) > v(C^3)$ (commensurable with $\mathrm{PSL}(2, \mathbb{Z}\sqrt{-2})$)

The sequence of link complements, F_n below can also be given hyperbolic structures obtained from a third kind of drum:

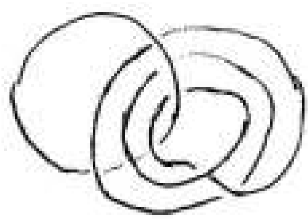
Example 6.8.11.



6.48

The regular drum is determined by its angles α and $\beta = \pi - \alpha$. Any pair of angles works to give a hyperbolic structure; one verifies that when the angle $\alpha = \arccos(\cos \frac{\pi}{2n} - \frac{1}{2})$, the hyperbolic structure is complete. The case $n = 1$ gives a trivial knot. In the case $n = 2$, the drums degenerate into simplices with 60° angles, and we obtain once more the hyperbolic structure on F_2 = figure eight knot. When $n = 3$, the angles are 90°, the drums become octahedra and we obtain $F_3 = B$. Passing to the limit $n = \infty$, and dividing by $\mathbb{Z}$, we obtain the following link, whose complement is commensurable with S^3 − figure eight knot:

EXAMPLE 6.8.12.



$$v = 4 - 05977\ldots$$

With these examples, many maps between link complements may be constructed. The reader should experiment for himself. One gets a feeling that volume is a very good measure of the complexity of a link *complement*, and that the ordinal structure is really inherent in three-manifolds.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 7

Computation of volume

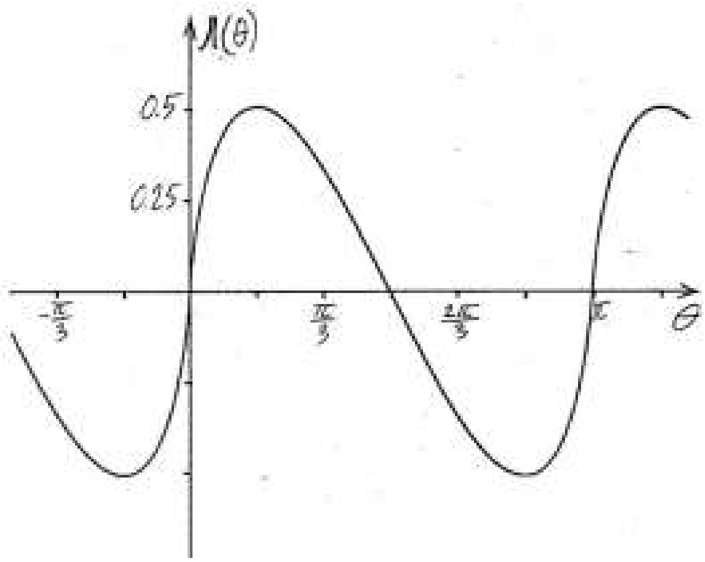
by J. W. Milnor

7.1. The Lobachevsky function $\text{л}(\theta)$.

This preliminary section will decribe analytic properties, and conjecture number theoretic properties, for the function

$$\text{л}(\theta) = -\int_0^\theta \log|2\sin u|\,du.$$

Here is the graph of this function:



Thus the first derivative $\text{л}'(\theta)$ is equal to $-\log|2\sin\theta|$, and the second derivative $\text{л}''(\theta)$ is equal to $-\cot\theta$. I will call $\text{л}(\theta)$ the *Lobachevsky function.* (This name is not quite accurate historically, since Lobachevsky's formulas for hyperbolic volume were expressed rather in terms of the function

$$\int_0^\theta \log(\sec u)\,du = \text{л}(\theta+\pi/2)+\theta\log 2$$

for $|\theta| \leq \pi/2$. However our function $\pi(\theta)$ is clearly a close relative, and is more convenient to work with in practice. Compare Clausen [3]). references are at end of chapter

Another close relative of $\text{л}(\theta)$ is the *dilogarithm function*

$$\psi(z) = \sum_{n=1}^{\infty} z^n/n^2 \qquad \text{for } |z| \leq 1,$$

which has been studied by many authors. (See for example [1], [2], [8], [9], [12], [13].) Writing

$$\psi(z) = -\int_0^z \log(1-w)\, dw/w$$

(where $|w| \leq 1$, the substitution $w = e^{2i\theta}$ yields 7.2

$$\log(1-w)\, dw/w = \big(\pi - 2\theta + 2i\log(2\sin\theta)\big)\, d\theta$$

for $0 < \theta < \pi$, hence

$$\psi(e^{2i\theta}) - \psi(1) = -\theta(\pi - \theta) + 2i\text{л}(\theta)$$

for $0 \leq \theta \leq \pi$. Taking the imaginary part of both sides, this proves the following: 7.3

Lemma 7.1.2. *The Lobachevsky function has uniformly convergent Fourier series expansion*

$$\text{л}(\theta) = \frac{1}{2}\sum_{n=1}^{\infty} \sin(2n\theta)/n^2.$$

Apparently, we have proved this formula only for the case $0 \leq \theta \leq \pi$. However, this suffices to show that $\text{л}(0) = \text{л}(\pi) = 0$. Since the derivative

$$d\text{л}(\theta)/d\theta = -2\log|\sin 2\theta|$$

is periodic of period π, this proves the following.

Lemma 7.1.3. *The function $\text{л}(\theta)$ is itself periodic of period π, and is an odd function, that is, $\text{л}(-\theta) = -\text{л}(\theta)$.*

It follows that the equation in 7.1.2 is actually valid for all values of θ.

The equation $z^n - 1 = \prod_{j=0}^{n-1}(z - e^{-2\pi i j/n})$ for $z = e^{2\pi i u}$ leads to the trigonometric identity

$$2\sin nu = \prod_{j=0}^{n-1} 2\sin(u + j\pi/n).$$

Integrating the logarithm of both sides and multiplying by n, this yields the following for $n \geq 1$, and hence for all n.

Lemma 7.1.4. *The identity*

$$\text{Л}(n\theta) = \sum_{j \bmod n} n\,\text{Л}(\theta + j\pi/n)$$

is valid for any integer $n \neq 0$. (Compare [14].)

Here the sum is to be taken over all residue classes modulo $|n|$. Thus for $n = 2$ 7.4
we get

$$\tfrac{1}{2}\text{Л}(2\theta) = \text{Л}(\theta) + \text{Л}(\theta + \pi/2),$$

or equivalently

$$\tfrac{1}{2}\text{Л}(2\theta) = \text{Л}(\theta) - \text{Л}(\pi/2 - \theta).$$

As an example, for $\theta = \pi/6$:

$$\tfrac{3}{2}\text{Л}(\pi/3) = \text{Л}(\pi/6).$$

(It is interesting to note that the function $\text{Л}(\theta)$ attains its maximum,

$$\text{Л}(\pi/6) = .5074\ldots,$$

at $\theta = \pi/6$.)

It would abe interesting to know whether there are any other such linear relations between various values of $\text{Л}(\theta)$ with rational coefficients. Here is an explicit guess.

Conjecture (A). Restricting attention to angles θ which are rational multiples of π, every rational linear relation between the real numbers $\text{Л}(\theta)$ is a consequence of 7.1.3 and 7.1.4.

(If we consider the larger class consisting of all angles θ for which $e^{i\theta}$ is algebraic then it definitely is possible to give other Q-linear relations. Compare [4].)

A different but completely equivalent formulation is the following. 7.5

Conjecture (B). Fixing some denominator $N \geq 3$, the real numbers $\text{Л}(\pi j/N)$ with j relatively prime to N and $0 < j < N/2$ are linearly independent over the rationals.

These numbers span a rational vector space v_N, conjectured to have dimension $\phi(N)/2$, where it is easy to check that $v_N \subset v_M$ whenever N divides M. Quite likely the elements $\text{Л}(\pi j/N)$ with $1 \leq j \leq \phi(N)/2$ would provide an alternative basis for this vector space.

I have tested these conjectures to the following extent. A brief computer search has failed to discover any other linear relations with small integer coefficients for small values of N.

To conclude this section, here is a remark about computation. The Fourier series 7.1.2 converges rather slowly. In order to get actual numerical values for $\text{л}(\theta)$, it is much better to work with the series

$$\text{л}(\theta) = \theta\left(1 - \log|2\theta| + \sum_{n=1}^{\infty} \frac{B_n}{2n}\,\frac{(2\theta)^{2n}}{(2n+1)!}\right),$$

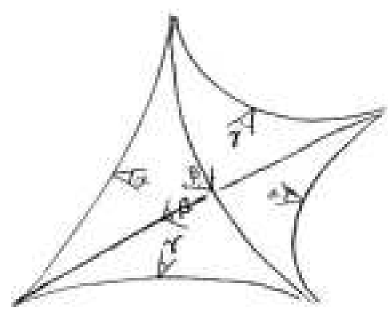
which is obtained by twice integrating the usual Laurent series expansion for the cotangent of θ. Here

$$B_1 = \tfrac{1}{6}, \quad B_2 = \tfrac{1}{30}, \ldots$$

are Bernoulli numbers. This series converges for $|\theta| = \pi$, and hence converges reasonably well for $|\theta| \leq \pi/2$. 7.6

7.2

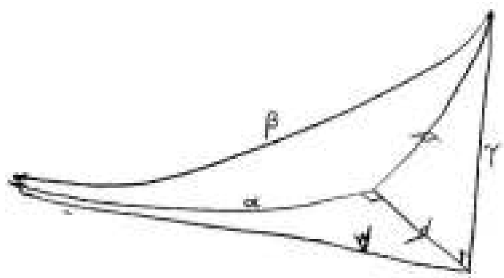
Having discussed the Lobachevsky function, we will see how it arises in the computation of hyperbolic volumes. The first case is the ideal simplex, i.e., a tetrahedron whose vertices are at ∞ and whose edges are geodesics which converge to the vertices at ∞. Such a simplex is determined by the dihedral angles formed between pairs of faces. The simplex intersects any small horosphere based at a vertex in a triangle whose interior angles are precisely the three dihedral angles along the edges meeting at that vertex. Since a horosphere is isometric to a Euclidean plane, the sum of the dihedral angles at an infinite vertex equals 2π. It follows by an easy computation that the dihedral angles of opposite edges are equal.



Call the three dihedral angles determining the simplex α, β, γ and denote the simplex by $\Sigma_{\alpha,\beta,\gamma}$. The main result of this section is:

THEOREM 7.2.1. *The volume of the simplex* $\sum_{\alpha,\beta,\gamma}$ *equals* $\text{л}(\alpha) + \text{л}(\beta) + \text{л}(\gamma)$.

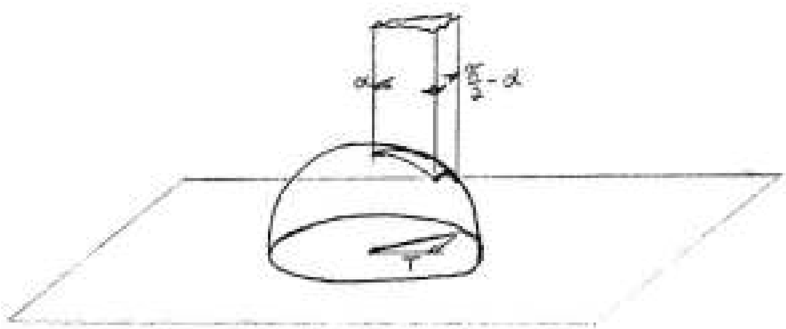
In order to prove this theorem a preliminary computation is necessary. Consider the simplex $S_{\alpha,\beta,\gamma}$ pictured below, with three right dihedral angles and three other 7.7
dihedral angles α, β, γ and suppose that one vertex is at infinity. (Thus $\alpha + \beta = \pi/2$.)



It turns out that any simplex can be divided by barycentric subdivision into simplices with three right angles so this is a natural object to consider. The decomposition of $\sum_{\alpha,\beta,\gamma}$ is demonstrated below, but first a computation, due to Lobachevsky.

LEMMA 7.2.2. *The volume of* $S_{\alpha,\pi/2-\alpha,\gamma}$ *equals* $\frac{1}{4}[\text{л}(\alpha+\gamma)+\text{л}(\alpha-\gamma)+2\text{л}(\pi/2-\alpha)]$.

PROOF. Consider the upper half-space model of H^3, and put the infinite vertex of $S_{\alpha,\pi/2-\alpha,\gamma}$ at ∞. The edges meeting that vertex are just vertical lines. Furthermore, assume that the base triangle lies on the unithemisphere (which is a hyperbolic plane). Recall that the line element for the hyperbolic metric in this model is $ds^2 = \frac{dx^2+dy^2+dz^2}{z^2}$ so that the volume element is $dV = \frac{dz\,dy\,dz}{z^3}$. Projecting the base triangle to the (x, y) plane produces a Euclidean triangle T with angles $\alpha, \pi/2-\alpha, \pi/2$, 7.8
which we may take to be the locus $0 \le x \le \cos\gamma, 0 \le y \le x \tan\alpha$, with γ as above.



REMARK. This projection of the unit hemisphere gives Klein's projective model for H^2. The angles between lines are not their hyperbolic angles; rather, they are the dihedral angles of corresponding planes in H^3.

Now it is necessary to compute

(1). $$V = \int_{x,y\in T}\int\int_{z\ge\sqrt{1-x^2-y^2}} \frac{dx\,dy\,dz}{z^3}.$$

Integrating with respect to z gives

(2). $$V = \int_T\int \frac{dx\,dy}{2(1-x^2-y^2)}.$$

Setting $a = \sqrt{1-x^2}$, we have 7.9

(3).
$$V = \int_0^{\cos\gamma} dx \int_0^{x\tan\alpha} \frac{dy}{2(a^2-y^2)} = \int_0^{\cos\gamma} \left(\frac{x}{4a} \log \frac{a + x\tan\alpha}{a - x\tan\alpha} \right)$$
$$= \int_0^{\cos\gamma} \frac{dx}{4a} \log \frac{2(a\cos\alpha + x\sin\alpha)}{2(a\cos\alpha - x\sin\alpha)}.$$

If we set $x = \cos\theta$, then $a = \sqrt{1-x^2} = \sin\theta$ and $\frac{dx}{a} = -d\theta$. Then (3) becomes

(4).
$$V = \frac{1}{4} \int_{\pi/2}^{\gamma} -d\theta \log\left(\frac{2\sin(\theta+\alpha)}{2\sin(\theta-\alpha)} \right)$$
$$= \frac{1}{4} [л(\gamma+\alpha) - л(\gamma-\alpha) - л(\pi/2+\alpha) + л(\pi/2-\alpha)].$$

Since $л(\gamma) - \alpha) = -л(\alpha-\gamma)$ and $л(\pi/2+\alpha) = -л(\pi/2-\alpha)$ by 7.1.3, this is the desired formula. □

Suppose that two vertices are at infinity in $S_{a,\pi/2-\alpha,\gamma}$. Then $\alpha = \gamma$. The lemma above implies that volume

$$(S_{\alpha,\pi/2-\alpha,a}) = \tfrac{1}{4}[л(2\alpha) + 2л(\pi/2-\alpha)].$$

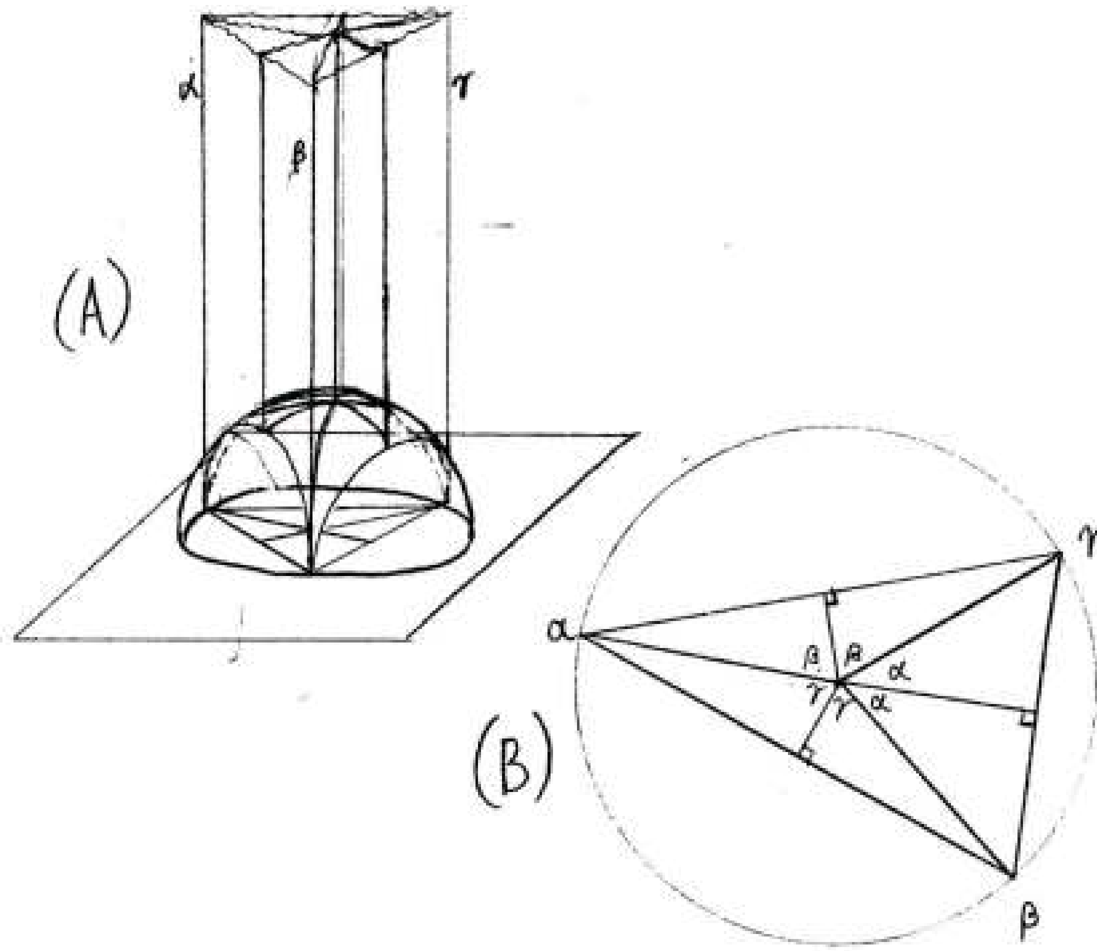
By lemmas 7.1.3 and 7.1.4

$$л(\pi/2-\alpha) = -л(\pi/2+\alpha) \quad \text{and} \quad л(2\alpha) = 2\big(л(\alpha) + л(\alpha+\pi/2)\big)$$

so that

(5).
$$\text{Volume}\,(S_{\alpha,\pi/2-\alpha,\alpha}) = \tfrac{1}{2}л(\alpha).$$

To see how $\sum_{\alpha,\beta,\gamma}$ decomposes into simplices of the above type, consider the upper half-space model of H^3. Put one vertex at the point at infinity and the base 7.10
on the unit sphere. Drop the perpendicular from ∞ to the sphere and draw the perpendiculars from the intersection point x on the base to each of the three edges on the base. Connect x to the remaining three vertices. Taking the infinite cone on the lines in the base gives the decomposition. (See (A) below.) Projecting onto the (x,y) plane gives a triangle inscribed in the unit circle with x projected into its center. Figure (B) describes the case when x is in the interior of the base (which happens when $\alpha, \beta, \gamma < \pi/2$). Not that the pairs of triangles which share a perpendicular are similar triangles. It follows that the angles around x are as described.

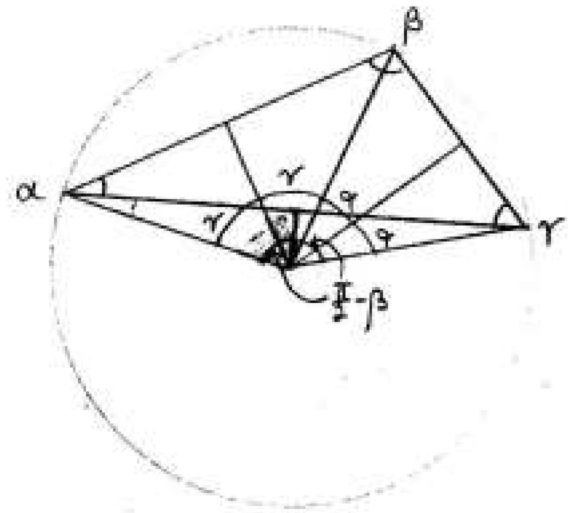


7.11

Each sub-simplex has two infinite vertices and three dihedral angles of $\pi/2$ so that they are of the type considered above. Thus

$$\text{Volume}\Big(\sum_{\alpha,\beta,\gamma}\Big) = 2\big(\tfrac{1}{2}\text{Л}(\gamma) + \tfrac{1}{2}\text{Л}(\beta) + \tfrac{1}{2}\text{Л}(\alpha)\big).$$

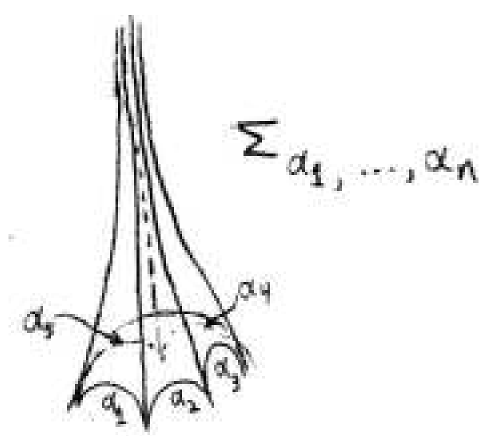
In the case when x is not in the interior of the base triangle, $\Sigma_{\alpha,\beta,\gamma}$ can still be thought of as the sum of six simplices each with three right dihedral angles. However, some of the simplices must be considered to have negative volume. The interested reader may supply the details, using the picture below.



Example. The complement of the figure-eight knot was constructed in 3.1 by gluing two copies of $\Sigma_{\pi/3,\pi/3,\pi/3}$. Thus its volume is $6\text{Л}(\pi/3) = 2.02988\ldots$.

Remark. It is not hard to see that the $(\pi/3, \pi/3, \pi/3)$ simplex has volume greater than any other three-dimensional simplex. A simplex with maximal volume must 7.12
have its vertices at infinity since volume can always be increased by pushing a finite vertex out towards infinity. To maximize $V = \text{Л}(\alpha) + \text{Л}(\beta) + \text{Л}(\gamma)$ subject to the restraint $\alpha+\beta+\gamma = 0$ we must have $\text{Л}'(\alpha) = \text{Л}'(\beta) = \text{Л}'(\gamma)$ which implies easily that $\alpha = \beta = \gamma = \pi/3$. (The non-differentiability of $\text{Л}(\alpha)$ at $\alpha = 0$ causes no trouble, since V tends to zero when α, β or γ tends to zero.)

Theorem 7.2.1 generalizes to a formula for the volume of a figure which is an infinite cone on a planar n-gon with all vertices at infinity. Let the dihedral angles formed by the triangular faces with the base plane be $(\alpha_1, \ldots, \alpha_n)$ and denote the figure with these angles by $\Sigma_{\alpha_1,\ldots,\alpha_n}$.



Theorem 7.2.3. (i) $\sum_{i=1}^{n} \alpha_i = \pi$. (ii) Volume $(\Sigma_{\alpha_1,\ldots,\alpha_n}) = \sum_{i=1}^{n} \text{Л}(\alpha_i)$.

Proof. The proof is by induction. The case $n = 3$ is Theorem 1. Suppose the theorem to be true for $n = k-1$. It suffices to prove it for $n = k$. 7.13

Consider the base k-gon for $\Sigma_{\alpha_1,\ldots,\alpha_k}$ and divide it into a $k-1$-gon and a triangle. Take the infinite cone on each of these two figures. If the new dihedral angle on the triangle side is β, the new angle on the $k-1$-gon side in $\pi - \beta$. By the induction hypothesis

$$\left(\sum_{i=1}^{2} \alpha_i\right) + \beta = \pi \quad \text{and} \left(\sum_{i=3}^{n} \alpha_i\right) + \pi - \beta = \pi.$$

Part (i) follows by adding the two equations. Similarly by the induction hypothesis,

$$\mathrm{Vol}(\Sigma_{\alpha_1,\alpha_2,\beta}) = \left(\sum_{i=1}^{2} \text{Л}(\alpha_i)\right) + \text{Л}(\beta)$$

and

$$\mathrm{Vol}(\Sigma_{\alpha_3,\ldots,\alpha_n,\pi-\beta}) = \left(\sum_{i=3}^{n} \text{Л}(\alpha_i)\right) + \text{Л}(\pi - \beta).$$

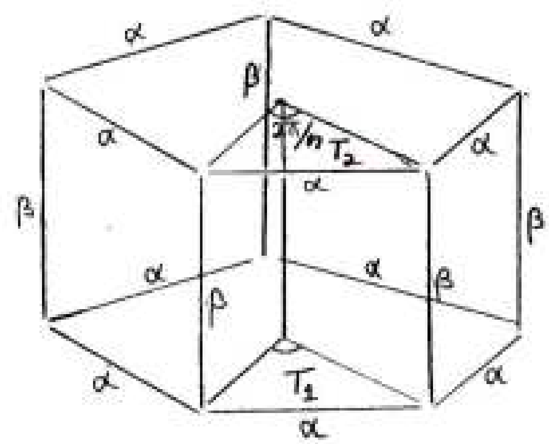
Part (ii) follows easily since $\text{Л}(\pi - \beta) = -\text{Л}(\beta)$. ☐

Example. The complement of the Whitehead link was constructed from a regular ideal octahedron which in turn, is formed by gluing two copies of the infinite cone on a regular planar quadrilateral. Thus its volume equals $8\text{Л}(\pi/4) = 3.66386\ldots$. Similarly, the complement of the Borromean rings has volume $16\text{Л}(\pi/4) = 7.32772\ldots$ since it is obtained by gluing two ideal octahedra together.

7.3

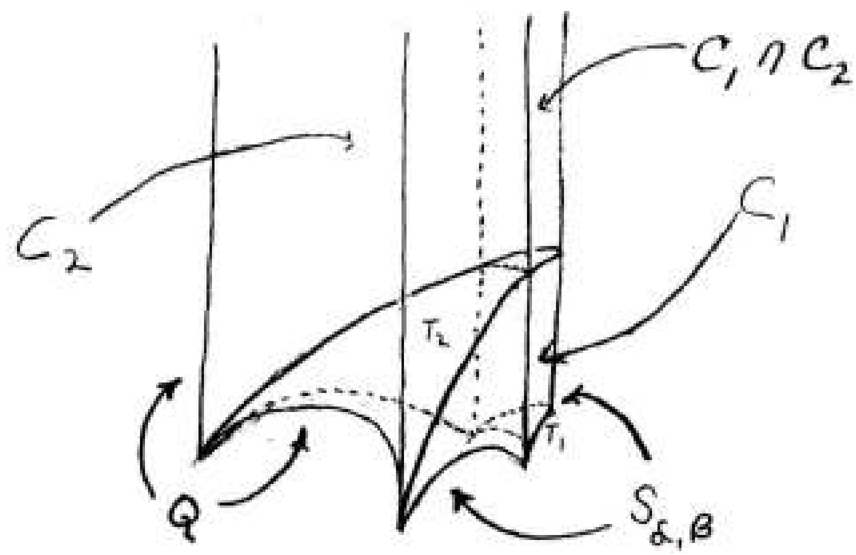
It is difficult to find a general pattern for constructing manifolds by gluing infinite tetrahedra together. A simpler method would be to reflect in the sides of a tetrahedron to form a discrete subgroup of the isometries of H^3. Unfortunately this method yields few examples since the dihedral angles must be of the form π/a, $a \in \mathbb{Z}$ in order that the reflection group be discrete with the tetrahedron as fundamental domain. The only cases when the sum of the angles is π are $\Sigma_{\pi/2,\pi/4,\pi/4}$, $\Sigma_{\pi/3,\pi/3,\pi/3}$ 7.14
and $\Sigma_{\pi/3,\pi/3,\pi/6}$ corresponding to the three Euclidean triangle groups.

Here is a construction for polyhedra in H^3 due to Thurston. Take a planar regular n-gon with vertices at infinity on each of two distinct planes in H^3 and join the corresponding vertices on the two figures by geodesics. If this is done in a symmetric way the sides are planer rectangles meeting each other at angle β and meeting the bases at angle α. Denote the resulting polyhedra by $\mathcal{N}_{\alpha,\beta}$. Note that $2\alpha + \beta = \pi$ since two edges of an n-gon and a vertical edge form a Euclidean triangle at infinity.

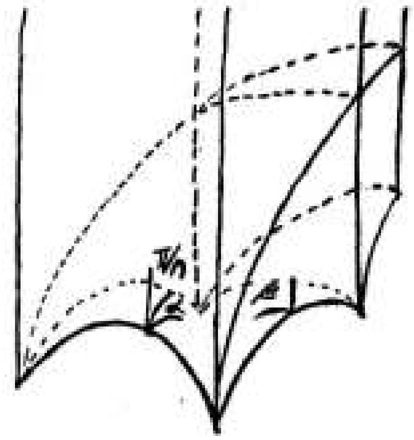


7.15

In order to compute the volume of $\mathcal{N}_{\alpha,\beta}$ consider it in the upper half-space model of H^3. Subdivide $\mathcal{N}_{\alpha,\beta}$ into n congruent sectors $S_{\alpha,\beta}$ by dividing the two n-gons into n congruent triangles and joining them be geodesics. Call the lower and upper triangles of $S_{\alpha,\beta}$, T_1 and T_2 respectively. Consider the infinite cones C_1 and C_2 on T_1 and T_2. They have the same volume since they are isometric by a Euclidean expansion. Hence the volume of $S_{\alpha,\beta}$ is equal to the volume of $Q = (S_{\alpha,\beta} \cup C_2) - C_1$.



Evidently Q is an infinite cone on a quadrilateral. To find its volume it is necessary to compute the dihedral angles at the edges of the base. The angles along the sides are $\frac{\beta}{2}$. The angle at the front face is $\alpha + \gamma$ where γ is the angle between the front
face and the top plane of $\mathcal{N}_{\alpha,\beta}$. Consider the infinite cone on the top n-gon of $\mathcal{N}_{\alpha,\beta}$. 7.16
By (1) of Theorem 7.2.3 the angles along its base are π/n. Thus $\gamma = \pi/n$ and the front angle is $\alpha + \pi/n$. Similarly the back angle is $\alpha - \pi/n$.



By (2) of Theorem 7.2.3 we have

$$\text{(6).}\qquad \mathrm{Vol}(\mathcal{N}_{\alpha,\beta}) = n\,\mathrm{Vol}(Q) = n\big(2\text{л}(\beta/2) + \text{л}(\alpha + \pi/n) + \text{л}(\alpha - \pi/n)\big).$$

If α and β are of the form π/a, $a \in \mathbb{Z}$ then the group generated by the reflections in the sides of $\mathcal{N}_{\alpha,\beta}$ form a discrete group of isometries of H^3. Take a subgroup Γ which is torsion free and orientation preserving. The quotient space H^3/Γ is an oriented, hyperbolic three-manifold with finite volume.

Since $2\alpha + \beta = \pi$ the only choices for (α, β) are $(\pi/3, \pi/3)$ and $(\pi, 4, \pi/2)$. As long as $n > 4$ both of these can be realized since β varies continuously from 0 to $n - 2/n$ as the distance between the two base planes of $\mathcal{N}_{\alpha,\beta}$ varies from 0 to ∞. 7.16
Thus we have the following:

Theorem 7.3.1. *There are an infinite number of oriented three-manifolds whose volume is a finite rational sum of* $\text{л}(\theta)$ *for* θ*'s commensurable with* π*.*

7.4

We will now discuss an arithmetic method for constructing hyperbolic three-manifolds with finite volume. The construction and computation of volume go back to Bianchi and Humbert. (See [5], [7], [10].) The idea is to consider $\mathcal{O}_d$, the ring of integers in an imaginary quadratic field, $\mathbb{Q}(\sqrt{-d})$, where $d \geq 1$ is a square-free integer. Then $\mathrm{PSL}(2, \mathcal{O}_d)$ is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{C})$. Let Γ be a torsion free subgroup of finite index in $\mathrm{PSL}(2, \mathcal{O}_d)$. Since $\mathrm{PSL}(2, \mathbb{C})$ is the group of orientation preserving isometries of H^3, H^3/Γ is an oriented hyperbolic three-manifold. It always has finite volume.

Example. Let $\mathbb{Z}[i]$ be the ring of Gaussian integers. A fundamental domain for the action of $\mathrm{PSL}(2, \mathbb{Z}[i])$ has finite volume. Different choices of Γ give different manifolds; e.g., there is a Γ of index 12 such that H^3/Γ is diffeomorphic to the complement of the Whitehead link; another Γ of index 24 leads to the complement of the Borromean rings. (N. Wielenbert, preprint).

Example. In case $d = 3$, $\mathcal{O}_d$ is $\mathbb{Z}[\omega]$ where $\omega = \frac{-1+\sqrt{-3}}{2}$ and there is a subgroup $\Gamma \subset \mathrm{PSL}(2, \mathbb{Z}[\omega])$ of index 12 such that H^3/Γ is diffeomorphic to the complement of the figure-eight knot. (R. Riley, [11]). In order to calculate the volume of $H^3/\mathrm{PSL}(2, \mathcal{O}_d)$ in general we recall the following definitions. Define the *discriminant*, 7.18
D, of the extension $Q(\sqrt{-d})$ to be

$$D = \begin{cases} d & \text{if } d \equiv 3 (\operatorname{mod} 4), \\ 4d & \text{otherwise.} \end{cases}$$

If $\mathcal{O}_d$ is considered as a lattice in $\mathbb{T}$, then $\sqrt{D}/2$ is the area of $\mathbb{T}/\mathcal{O}_d$. The *Dedekind ζ-function* for a field K is defined to be

$$\zeta_K(S) = \sum_{\mathfrak{a}} 1/N(\mathfrak{a})^S \quad \text{where}$$

$\mathfrak{a}$ runs through all ideals in $\mathcal{O}$ and $N(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$ denotes the norm of $\zeta(S)$ is also equal to

$$\prod_{\mathfrak{P}} \frac{1}{1 - \dfrac{1}{N(\mathfrak{P})^S}}$$

taking all prime ideals of $\mathfrak{P}$.

Theorem 7.4.1 (Essentially due to Humbert).

$$\mathrm{Vol}\big(H^3/\mathrm{PSL}(2, \mathcal{O}_d)\big) = \tfrac{D^{3/2}}{24}\, \zeta_{\mathbb{Q}(\sqrt{-d})}(2)/\zeta_{\mathbb{Q}}(2).$$

This volume can be expressed in terms of Lobachevsky's function using Hecke's formula

$$\zeta_{\mathbb{Q}(\sqrt{-d})}(S)/\zeta_{\mathbb{Q}}(S) = \sum_{n>0} \frac{\left(\frac{-D}{n}\right)}{n^s}.$$

Here $\left(\frac{-D}{n}\right)$ is the *quadratic symbol* where we use the conventions:

(i) If $n = p_1, \dots, p_t, p_i$ prime then $\left(\frac{-D}{n}\right) = \left(\frac{-D}{p_1}\right)\left(\frac{-D}{p_2}\right) \dots \left(\frac{-D}{p_t}\right)$.
(ii) If $p \mid D$ then $\left(\frac{-D}{p}\right) = 0$; $\left(\frac{-D}{1}\right) = +1$.
(iii) for p an odd prime

$$\left(\frac{-D}{p}\right) = \begin{cases} +1 & \text{if } -D \equiv X^2 (\operatorname{mod} p) \text{ for some } X, \\ -1 & \text{if not.} \end{cases}$$

(iv) For $p = 2$

$$\left(\frac{-D}{p}\right) = \begin{cases} +1 & \text{if } -D \equiv 1 \pmod 8, \\ -1 & \text{if } -D \equiv 5 \pmod 8. \end{cases}$$

(Note that $-D \not\equiv 3 \pmod 4$ by definition.)

The function $n \mapsto \left(\frac{-D}{n}\right)$ is equal to $1/\sqrt{-D}$ times its Fourier transform;* i.e.,

(1).
$$\sum_{k \bmod D} \left(\frac{-D}{k}\right) e^{2\pi i k n/D} = \sqrt{-D}\left(\frac{-D}{n}\right).$$

Multiplying by $1/n^2$ and summing over $n > 0$ we get

(2).
$$\sum_{n>0} 1/n^2 \sum_{k=0}^{n-1} \left(\frac{-D}{k}\right) e^{2\pi i k n/D} = \sqrt{-D} \sum_{n>0} \frac{\left(\frac{-D}{n}\right)}{n^2}.$$

For fixed k the imaginary part of the left side is just the Fourier series for $2\text{л}(\pi k/D)$. Since the right side is pure imaginary we have:

(3).
$$2 \sum_{k \bmod D} \left(\frac{-D}{k}\right) \text{л}(\pi k/D) = \sqrt{D} \sum_{n>0} \left(\frac{-D}{n}\right) 1/n^2.$$

Multiplying by $D/24$ and using Hecke's formula leads to

(4).
$$D/12 \sum_{k \bmod D} \left(\frac{-D}{k}\right) \text{л}(\pi k/D) = \operatorname{Vol}\bigl(H^3/\operatorname{PSL}(2, \mathcal{O}_d)\bigr).$$

7.20

Example. In the case $d = 3$, 7.4.4 implies that the volume of $H^3/\bigl(\operatorname{PSL}(2, \mathbb{Z}[\omega]\bigr)$ is $\frac{1}{4}\bigl(\text{л}(\pi/3) - \text{л}(2\pi/3)\bigr) = \frac{1}{2}\text{л}(\pi/3)$. Recall that the complement of the figure-eight knot $S^3 - K$ is diffeomorphic to H^3/Γ where Γ had index 12 in $\operatorname{PSL}(2, \mathbb{Z}[\omega])$. Thus it has volume $6\text{л}(\pi/3)$. This agrees with the volume computed by thinking of $S^3 - K$ as two copies of $\Sigma_{\pi/3,\pi/3,\pi/3}$ tetrahedra glued together.

Similarly the volumes for the complements of the Whitehead link and the Borromean rings can be computed using 7.4.4. The answers agree with those computed geometrically in 7.2.

This algebraic construction also furnishes an infinite number of hyperbolic manifolds with volumes equal to rational, finite linear combinations of л (a rational multiple of π). Note that Conjectures (A) and (B) would imply that any rational relation between the volumes of these manifolds could occur at most as a result of common factors of the integers, d, defining the quadratic fields. In fact, quite likely they would imply that there are no such rational relations.

*Compare Hecke, Vorlesangen über algebr. Zahlen, p. 241. I am grateful to A. Adler for help on this point.

References

[1] L. Euler, *Institutiones calculi integralis*, I, pp. 110-113 (1768).
[2] C. J. Hill, *Journ. reine angew. Math.* (Crelle) 3 (1828), 101-139.
[3] T. Clausen, *Journ. reine angew. Math.* (Crelle) 8 (1932), 298-300.
[4] N. Lobachevsky, *Imaginary Geometry and its Application to Integration* (Russian) Kasan 1836. German translation, Leipzig 1904. (For a modern presentation see B. L. Laptev, Kazan Gos Univ. Uč. Zapiski 114 (1954), 53-77).
[5] L. Bianchi, *Math. Ann.* 38 (1891) 313-333 and 40 (1892), 332-412.
[6] H. Gieseking, Thesis, Münster 1912. (See Magnus, Noneuclidean Tesselations and their Groups, Acad. Press 1974, p. 153.)
[7] G. Humbert, Comptes Rendus 169 (1919) 448-454.
[8] H. S. M. Coxeter, *Quarterly J. Math* 6 (1935), 13-29.
[9] L. Lewin, *Dilogarithms and Associated Functions*, Macdonald (London), 1958.
[10] R. G. Swan, *Bull. AMS* 74 (1968), 576-581.
[11] R. Riley, *Proc. Cambr. Phil. Soc.* 77 (1975) 281-288.
[12] A. M. Gabrielov, I. M. Gel'fand, and M. V. Losik, *Functional Anal. and Appl.* 9 (1975), pp. 49, 193.
[13] S. Bloch, to appear.
[14] D. Kubert and S. Lang, to appear.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002
http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 8

Kleinian groups

Our discussion so far has centered on hyperbolic manifolds which are closed, or at least complete with finite volume. The theory of complete hyperbolic manifolds with infinite volume takes on a somewhat different character. Such manifolds occur very naturally as covering spaces of closed manifolds. They also arise in the study of hyperbolic structures on compact three-manifolds whose boundary has negative Euler characteristic. We will study such manifolds by passing back and forth between the manifold and the action of its fundamental group on the disk. 8.1

8.1. The limit set

Let Γ be any discrete group of orientation-preserving isometries of H^n. If $x \in H^n$ is any point, the *limit set* $L_\Gamma \subset S_\infty^{n-1}$ is defined to be the set of accumulation points of the orbit Γ_x of x. One readily sees that L_Γ is independent of the choice of x by picturing the Poincaré disk model. If $y \in H^n$ is any other point and if $\{\gamma_i\}$ is a sequence of elements of Γ such that $\{\gamma_i x\}$ converges to a point on S_∞^{n-1}, the hyperbolic distance $d(\gamma_i x, \gamma_i y)$ is constant so the Euclidean distance goes to 0; hence $\lim \gamma_i y = \lim \gamma_i x$.

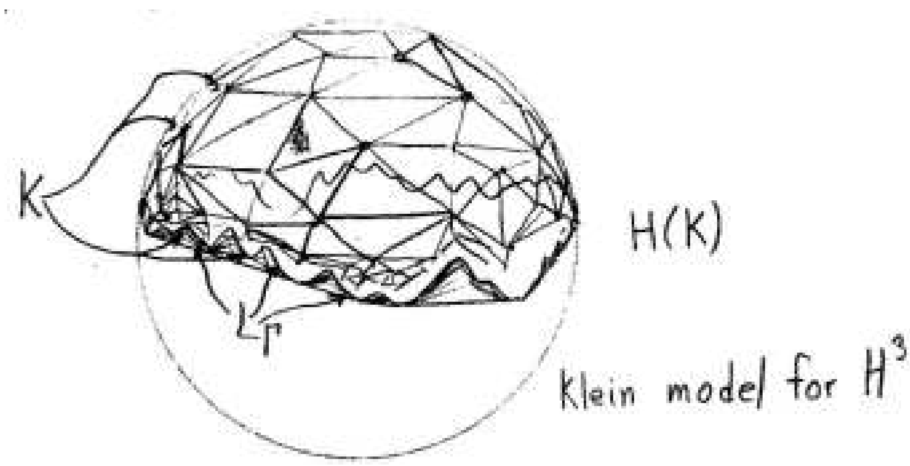
The group Γ is called *elementary* if the limit set consists of $0, 1$ or 2 points.

Proposition 8.1.1. *Γ is elementary if and only if Γ has an abelian subgroup of finite index.* □

When Γ is not elementary, then L_Γ is also the limit set of any orbit on the sphere at infinity. Another way to put it is this:

Proposition 8.1.2. *If Γ is not elementary, then every non-empty closed subset of S_∞ invariant by Γ contains L_Γ.*

Proof. Let $K \subset S_\infty$ be any closed set invariant by Γ. Since Γ is not elementary, K contains more than one element. Consider the projective (Klein) model for H^n, and let $H(K)$ denote the convex hull of K. $H(K)$ may be regarded either as the Euclidean convex hull, or equivalently, as the hyperbolic convex hull in the sense that it is the intersection of all hyperbolic half-spaces whose "intersection" with S_∞ contains K. Clearly $H(K) \cap S_\infty = K$.



Since K is invariant by Γ, $H(K)$ is also invariant by Γ. If x is any point in $H^n \cap H(K)$, the limit set of the orbit Γ_x must be contained in the closed set $H(K)$. Therefore $L_\Gamma \subset K$. $\square$

8.3

A closed set K invariant by a group Γ which contains no smaller closed invariant set is called a minimal set. It is easy to show, by Zorn's lemma, that a closed invariant set always contains at least one minimal set. It is remarkable that in the present situation, L_Γ is the *unique* minimal set for Γ.

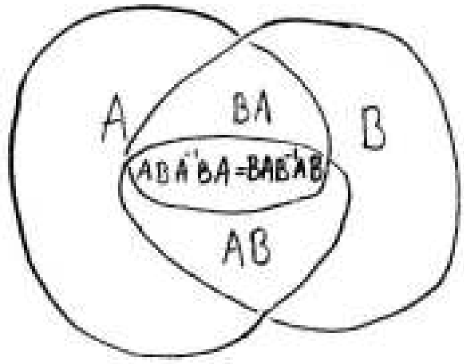
Corollary 8.1.3. *If Γ is a non-elementary group and $1 \neq \Gamma' \triangleleft \Gamma$ is a normal subgroup, then $L_{\Gamma'} = L_\Gamma$.*

Proof. An element of Γ conjugates Γ' to itself, hence it takes $L_{\Gamma'}$ to $L_{\Gamma'}$. Γ' must be infinite, otherwise Γ' would have a fixed point in H^n which would be invariant by Γ so Γ would be finite. It follows from 8.1.2 that $L_{\Gamma'} \supset L_\Gamma$. The opposite inclusion is immediate. $\square$

Examples. If M^2 is a hyperbolic surface, we may regular $\pi_1(M)$ as a group of isometries of a hyperbolic plane in H^3. The limit set is a circle. A group with limit set contained in a geometric circle is called a *Fuchsian group*.

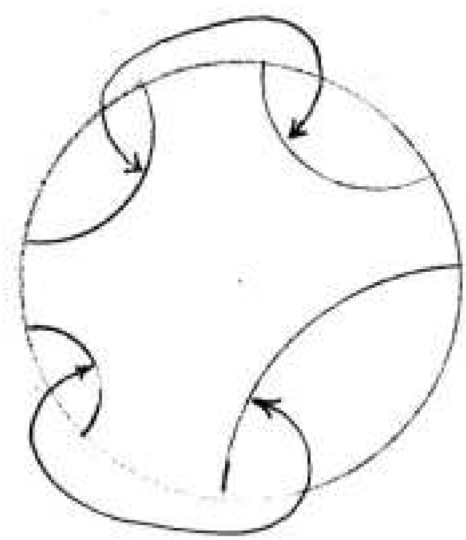
The limit set for a closed hyperbolic manifold is the entire sphere S^{n-1}_∞.

If M^3 is a closed hyperbolic three-manifold which fibers over the circle, then the fundamental group of the fiber is a normal subgroup, hence its limit set is the entire sphere. For instance, the figure eight knot complement has fundamental group $\langle A, B : ABA^{-1}BA = BAB^{-1}AB\rangle$. 8.4



It fibers over S^1 with fiber F a punctured torus. The fundamental group $\pi_1(F)$ is the commutator subgroup, generated by AB^{-1} and $A^{-1}B$. Thus, the limit set of a finitely generated group may be all of S^2 even when the quotient space does not have finite volume.

A more typical example of a free group action is a Schottky group, whose limit set is a Cantor set. Examples of Schottky groups may be obtained by considering H^n minus $2k$ disjoint half-spaces, bounded by hyperplanes. If we choose isometric identifications between pairs of the bounding hyperplanes, we obtain a complete hyperbolic manifold with fundamental group the free group on k generators.



Schottky group
$n = 2$
$k = 2$

8.5

It is easy to see that the limit set for the group of covering transformations is a Cantor set.

8.2. The domain of discontinuity

The domain of discontinuity for a discrete group Γ is defined to be $D_\Gamma = S^{n-1}_\infty - L_\Gamma$. A discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$ whose domain of discontinuity is non-empty is called a *Kleinian group*. (There are actually two ways in which the term Kleinian group is generally used. Some people refer to any discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$ as a Kleinian group, and then distinguish between a type I group, for which $L_\Gamma = S^2_\infty$, and a type II group, where $D_\Gamma \neq \emptyset$. As a field of mathematics, it makes sense for Kleinian groups to cover both cases, but as mathematical objects it seems useful to have a word to distinguish between these cases $D_\Gamma \neq \emptyset$ and $D_\Gamma = \emptyset$.)

We have seen that the action of Γ on L_Γ is minimal—it mixes up L_Γ as much as possible. In contrast, the action of Γ on D_Γ is as discrete as possible.

Definition 8.2.1. If Γ is a group acting on a locally compact space X, the action is *properly discontinuous* if for every compact set $K \subset X$, there are only finitely many $\gamma \in \Gamma$ such that $\gamma K \cap K \neq \emptyset$.

Another way to put this is to say that for any compact set K, the map $\Gamma \times K \to X$ given by the action is a proper map, where Γ has the discrete topology. (Otherwise 8.6
there would be a compact set K' such that the preimage of K' is non-compact. Then infinitely many elements of Γ would carry $K \cup K'$ to itself.)

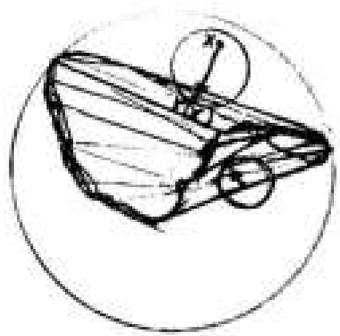
Proposition 8.2.2. *If Γ acts properly discontinuously on the locally compact Hausdorff space X, then the quotient space X is Hausdorff. If the action is free, the quotient map $X \to X/\Gamma$ is a covering projection.*

Proof. Let $x_1, x_2 \in X$ be points on distinct orbits of Γ. Let N_1 be a compact neighborhood of x_1. Finitely many translates of x_2 intersect N_1, so we may assume N_1 is disjoint from the orbit of x_2. Then $\bigcup_{\gamma\in\Gamma} \gamma N_1$ gives an invariant neighborhood of x_1 disjoint from x_2. Similarly, x_2 has an invariant neighborhood N_2 disjoint from N_1; this shows that X/Γ is Hausdorff. If the action of Γ is free, we may find, again by a similar argument, a neighborhood of any point x which is disjoint from all its translates. This neighborhood projects homeomorphically to X/Γ. Since Γ acts transitively on the sheets of X over X/Γ, it is immediate that the projection $X \to X/\Gamma$ is an even covering, hence a covering space. $\square$

Proposition 8.2.3. *If Γ is a discrete group of isometries of H^n, the action of Γ on D_Γ (and in fact on $H^n \cup D^\Gamma$) is properly discontinuous.*

Proof. Consider the convex hull $H(L_\Gamma)$. There is a retraction r of the ball $H^n \cup S_\infty$ to $H(L_\Gamma)$ defined as follows. 8.7

If $x \in H(L_\Gamma)$, $r(x) = x$. Otherwise, map x to the nearest point of $H(L_\Gamma)$. If x is an infinite point in D_Γ, the nearest point is interpreted to be the first point of $H(L_\Gamma)$ where a horosphere "centered" about x touches L_Γ. This point $r(x)$ is always uniquely defined



because $H(L_\Gamma)$ is convex, and spheres or horospheres about a point in the ball are strictly convex. Clearly r is a proper map of $H^n \cup D_\Gamma$ to $H(L_\Gamma) - L_\Gamma$. The action of Γ on $H(L_\Gamma) - L_\Gamma$ is obviously properly discontinuous, since Γ is a discrete group of isometries of $H(L_\Gamma) - L_\Gamma$; the property of $H^n \cup D_\Gamma$ follows immediately. □

Remark. This proof doesn't work for certain elementary groups; we will ignore such technicalities.

It is both easy and common to confuse the definition of properly discontinuous with other similar properties. To give two examples, one might make these definitions:

Definition 8.2.4. The action of Γ is *wandering* if every point has a neighborhood N such that only finitely many translates of N intersect N.

Definition 8.2.5. The action of Γ has *discrete orbits* if every orbit of Γ has an empty limit set.

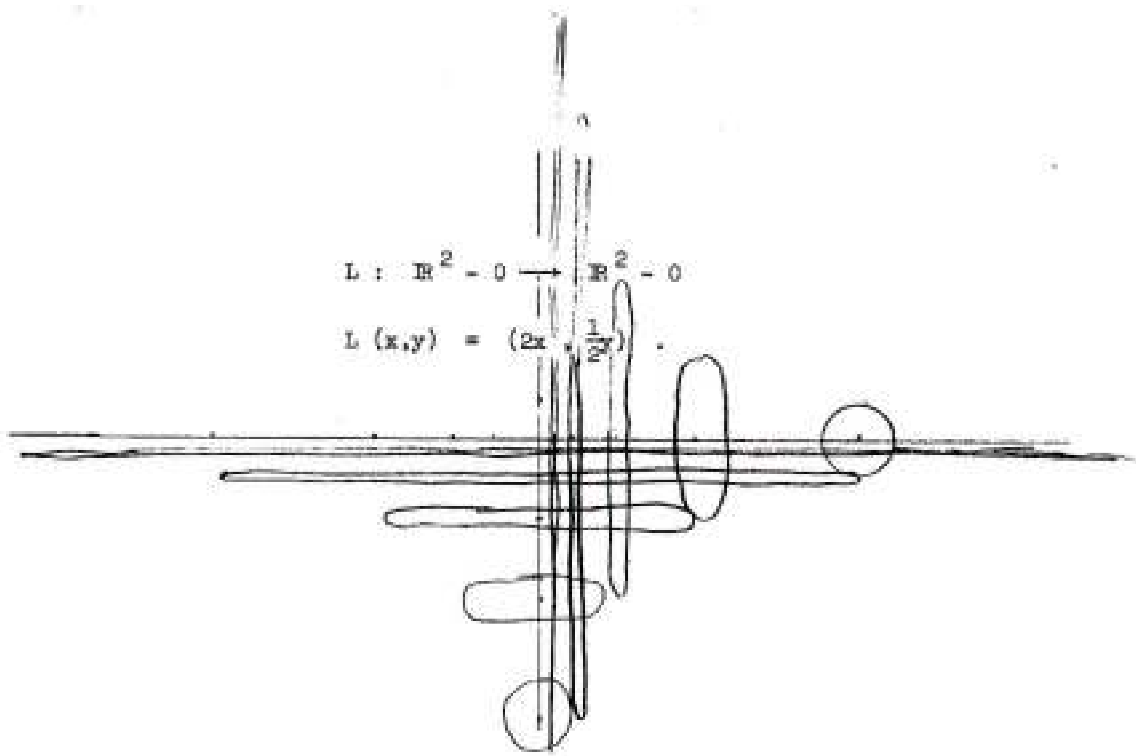
Proposition 8.2.6. *If Γ is a free, wandering action on a Hausdorff space X, the projection $X \to X/\Gamma$ is a covering projection.*

Proof. An exercise. □

Warning. Even when X is a manifold, X/Γ may not be Hausdorff. For instance, consider the map

$$L : \mathbb{R}^2 - 0 \to \mathbb{R}^2 - 0$$

$$L\,(x, y) = (2x, \tfrac{1}{2}y).$$



It is easy to see this is a wandering action. The quotient space is a surface with fundamental group $\mathbb{Z} \oplus \mathbb{Z}$. The surface is non-Hausdorff, however, since points such 8.9
as $(1, 0)$ and $(0, 1)$ do not have disjoint neighborhoods.

Such examples arise commonly and naturally; it is wise to be aware of this phenomenon.

The property that Γ has discrete orbits simply means that for every pair of points x, y in the quotient space X/Γ, x has a neighborhood disjoint from y. This can occur, for instance, in a l-parameter family of Kleinian groups Γ_t, $t \in [0, 1]$. There are examples where $\Gamma_t = \mathbb{Z}$, and the family defines the action of $\mathbb{Z}$ on $[0, 1] \times H^3$ with discrete orbits which is not a wandering action. See § . It is remarkable that the action of a Kleinian group on the set of all points with discrete orbits is properly discontinuous.

8.3. Convex hyperbolic manifolds

The limit set of a group action is determined by a limiting process, so that it is often hard to "know" the limit set directly. The condition that a given group action is discrete involves infinitely many group elements, so it is difficult to verify directly. Thus it is important to have a concrete object, satisfying concrete conditions, corresponding to a discrete group action.

We consider for the present only groups acting freely.

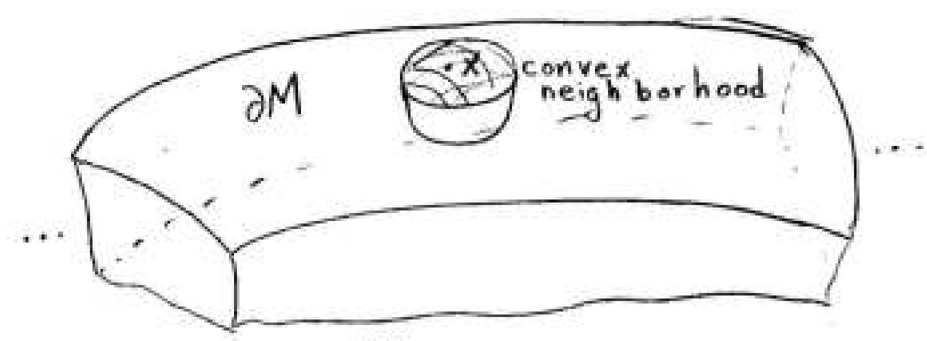
DEFINITION 8.3.1. A complete hyperbolic manifold M with boundary is *convex* if every path in M is homotopic (rel endpoints) to a geodesic arc. (The degenerate 8.10
case of an arc which is a single point may occur.)

PROPOSITION 8.3.2. *A complete hyperbolic manifold M is convex if and only if the developing map $D : \tilde{M} \to H^n$ is a homeomorphism to a convex subset of H^n.*

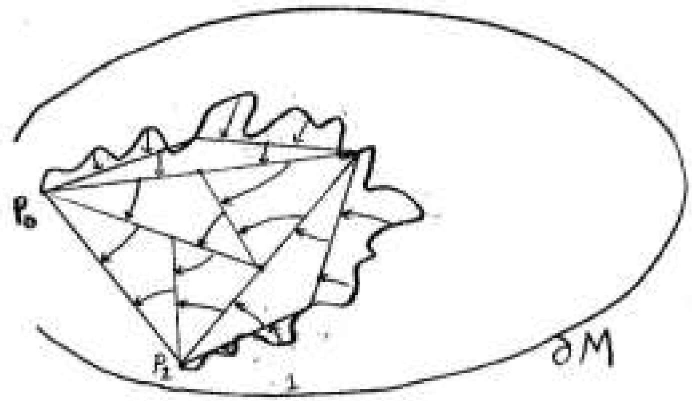
PROOF. If $\tilde{M}$ is a convex subset S of H^n, then it is clear that M is convex, since any path in M lifts to a path in S, which is homotopic to a geodesic arc in S, hence in M.

If M is convex, then D is $1-1$, since any two points in $\tilde{M}$ may be joined by a path, which is homotopic in M and hence in $\tilde{M}$ to a geodesic arc. D must take the endpoints of a geodesic arc to distinct points. $D(\tilde{M})$ is clearly convex. $\square$

We need also a local criterion for M to be convex. We can define M to be *locally convex* if each point



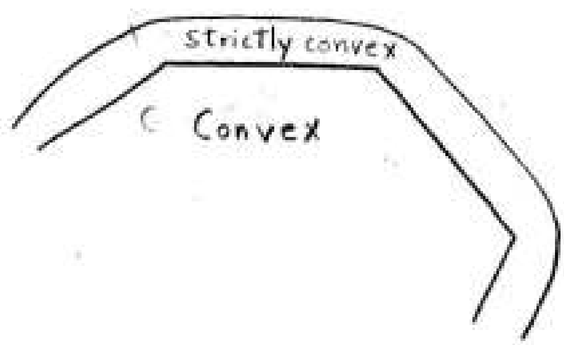
$x \in M$ has a neighborhood isometric to a convex subset of H^n. If $x \in \partial M$, then x will be on the boundary of this set. It is easy to convince oneself that local convexity implies convexity: picture a bath and imagine straightening it out. Because of local 8.11
convexity, one never needs to push it out of ∂M. To make this a rigorous argument, given a path p of length l there is an ϵ such that any path of length $\leq \epsilon$ intersecting $\mathcal{N}_l(p_0)$ is homotopic to a geodesic arc. Subdivide p into subintervals of length between $\epsilon/4$ and $\epsilon/2$. Straighten out adjacent pairs of intervals in turn, putting a new division point in the middle of the resulting arc unless it has length $\leq \epsilon/2$. Any time an interval becomes too small, change the subdivision. This process converges, giving a homotopy of p to a geodesic arc, since any time there are angles not close to π, the homotopy significantly shortens the path.



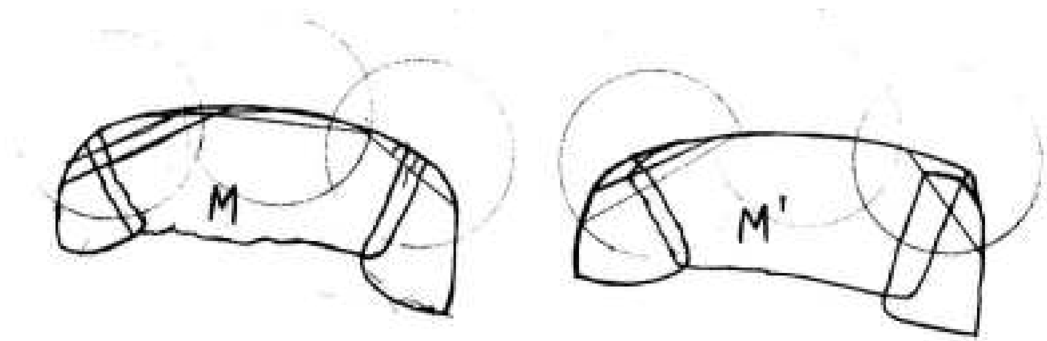
This give us a very concrete object corresponding to a Kleinian group: a complete convex hyperbolic three-manifold M with non-empty boundary. Given a convex manifold M, we can define $H(M)$ to be the intersection of all convex submanifolds M' of M such that $\pi_1 M' \to \pi_1 M$ is an isomorphism. $H(M)$ is clearly the same as $HL_{\pi_1}(M)/\pi_1(M)$. $H(M)$ is a convex manifold, with the same dimension as M except in degenerate cases. 8.12

Proposition 8.3.3. *If M is a compact convex hyperbolic manifold, then any small deformation of the hyperbolic structure on M can be enlarged slightly to give a new convex hyperbolic manifold homeomorphic to M.*

Proof. A convex manifold is *strictly convex* if every geodesic arc in M has interior in the interior of M. If M is not already strictly convex, it can be enlarged slightly to make it strictly convex. (This follows from the fact that a neighborhood of radius ϵ about a hyperplane is strictly convex.)



Thus we may assume that M' is a hyperbolic structure that is a slight deformation of a strictly convex manifold M. We may assume that our deformation M' is small enough that it can be enlarged to a hyperbolic manifold M'' which contains a 2ϵ-neighborhood of M'. Every arc of length l greater than ϵ in M has the middle $(l-\epsilon)$ some uniform distance δ from ∂M; we may take our deformation M' of M small 8.13 enough that such intervals in M' have the middle $l-\epsilon$ still in the interior of M'. This implies that the union of the convex hulls of intersections of balls of radius 3ϵ with M' is locally convex, hence convex. □



The convex hull of a uniformly small deformation of a uniformly convex manifold is locally determined.

Remark. When M is non-compact, the proof of 8.3.3 applies provided that M has a uniformly convex neighborhood and we consider only uniformly small deformations. We will study deformations in more generality in § .

Proposition 8.3.4. *Suppose M_1^n and M_2^n are strictly convex, compact hyperbolic manifolds and suppose $\phi : M_1^n \to M_2^n$ is a homotopy equivalence which is a diffeomorphism on ∂M_1. Then there is a quasi-conformal homeomorphism $f : B^n \to B^n$ of the Poincaré disk to itself conjugating $\pi_1 M_1$ to $\pi_1 M_2$. f is a pseudo-isometry on H^n.*

8.14

Proof. Let $\tilde{\phi}$ be a lift of ϕ to a map from $\tilde{M}_1$ to $\tilde{M}_2$. We may assume that $\tilde{\phi}$ is already a pseudo-isometry between the developing images of M_1 and M_2. Each point p on $\partial\tilde{M}_1$ and $\partial\tilde{M}_2$ has a unique normal ray γ_p; if $x \in \gamma_p$ has distance t from $\partial\tilde{M}_1$ let $f(x)$ be the point on $\gamma_{\tilde{\phi}(p)}$ a distance t from $\partial\tilde{M}_2$. The distance between points at a distance of t along two normal rays γ_{p_1} and γ_{p_2} at nearby points is approximately $\cosh t + \alpha \sinh t$, where d is the distance and θ is the angle between the normals of p_1 and p_2. From this it is evident that f is a pseudo-isometry extending to $\bar{\phi}$. □

Associated with a discrete group Γ of isometries of H^n, there are at least four distinct and interesting quotient spaces (which are manifolds when Γ acts freely). Let us name them:

Definition 8.3.5.

$M_\Gamma = H(L_\Gamma)/\Gamma$, the convex hull quotient.
$N_\Gamma = H^n/\Gamma$, the complete hyperbolic manifold without boundary.
$O_\Gamma = (H^n \cup D_\Gamma)/\Gamma$, the Kleinian manifold.
$P_\Gamma = (H^n \cup D_\Gamma \cup W_\Gamma)/\Gamma$. Here $W_\Gamma \subset \mathbb{P}^n$ is the set of points in the projective model dual to planes in H^n whose intersection with S_∞ is contained in D_Γ.

We have inclusions $H(N_\Gamma) = M_\Gamma \subset N_\Gamma \subset O_\Gamma \subset P_\Gamma$. It is easy to derive the fact that Γ acts properly discontinuously on $H^n \cup D_\Gamma \cup W_\Gamma$ from the proper discontinuity on $H^n \cup D_\Gamma$. M_Γ, N_Γ and O_Γ have the same homotopy type. M_Γ and O_Γ are homeomorphic except in degenerate cases, and $N_\Gamma = \mathrm{int}(O_\Gamma)$ P_Γ is not always connected 8.15
when L_Γ is not connected.

8.4. Geometrically finite groups

Definition 8.4.1. Γ is geometrically finite if $\mathcal{N}_\epsilon(M_\Gamma)$ has finite volume.

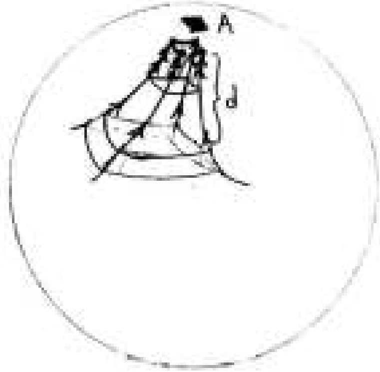
The reason that $\mathcal{N}_\epsilon(M_\Gamma)$ is required to have finite volume, and not just M_Γ, is to rule out the case that Γ is an arbitary discrete group of isometries of $H^{n-1} \subset H^n$. We shall soon prove that geometrically finite *means* geometrically finite (8.4.3).

Theorem 8.4.2 (Ahlfors' Theorem). *If Γ is geometrically finite, then $L_\Gamma \subset S_\infty$ has full measure or 0 measure. If L_Γ has full measure, the action of Γ on S_∞ is ergodic.*

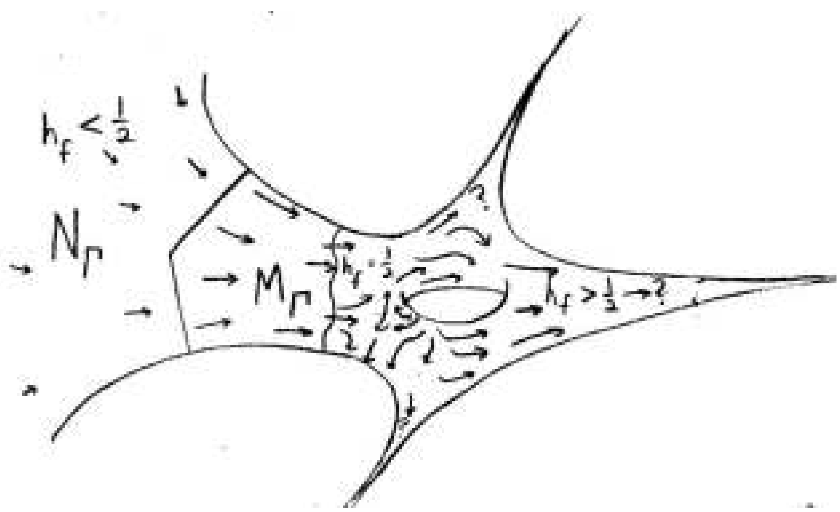
Proof. This statement is equivalent to the assertion that every bounded measurable function f supported on L_Γ and invariant by Γ is constant a.e. (with respect to Lebesque measure on S_∞). Following Ahlfors, we consider the function h_f on H^n determined by f as follows. If $x \in H^n$, the points on S_∞ correspond to rays through x; these rays have a natural "visual" measure V_x. Define $h_f(x)$ to be the average of f with respect to the visual measure V_x. This function h_f is harmonic, i.e., the gradient flow of h_f preserves volume,

$$\operatorname{div}\operatorname{grad} h_f = 0.$$

For this reason, the measure $\frac{1}{V_x(S_\infty)}V_x$ is called *harmonic measure*. To prove this, 8.16
consider the contribution to h_f coming from an infinitesimal area A centered at $p \in S^{n-1}$ (i.e., a Green's function). As x moves a distance d in the direction of p, the visual measure of A goes up exponentially, in proportion to $e^{(n-1)d}$. The gradient of any multiple of the characteristic function of A is in the direction of p, and also proportional in size to $e^{(n-1)d}$. The flow lines of the gradient are orthogonal trajectories to horospheres; this flow contracts linear dimensions along the horosphere in proportion to e^{-d}, so it preserves volume.

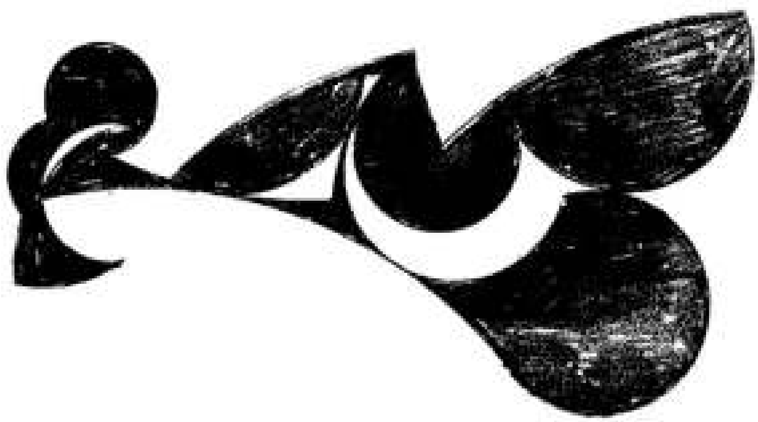


The average h_f of contributions from all the infinitesimal areas is therefore harmonic. We may suppose that f takes only the values of 0 and 1. Since f is invariant by Γ, so is h_f, and h_f goes over to a harmonic function, also h_f, on N_Γ. To complete the proof, observe that $h_f < \frac{1}{2}$ in $N_\Gamma - M_\Gamma$, since each point x in $H^n - H(L_\Gamma)$ lies in a half-space whose intersection with infinity does not meet L_Γ, which means that f is 0 on more than half the sphere, with respect to V_x. The set $\{x \in N_\Gamma | h_f(x) = \frac{1}{2}\}$ 8.17
must be empty, since it bounds the set $\{x \in N_\Gamma | h_f(x) \geq \frac{1}{2}\}$ of finite volume which flows into itself by the volume preserving flow generated by $\operatorname{grad} h_f$. (Observe that $\operatorname{grad} h_f$ has bounded length, so it generates a flow defined everywhere for all time.) But if $\{p | f(p) = 1\}$ has any points of density, then there are $x \in H^{n-1}$ near p with $h_f(x)$ near 1. It follows that f is a.e. 0 or a.e. 1. $\square$

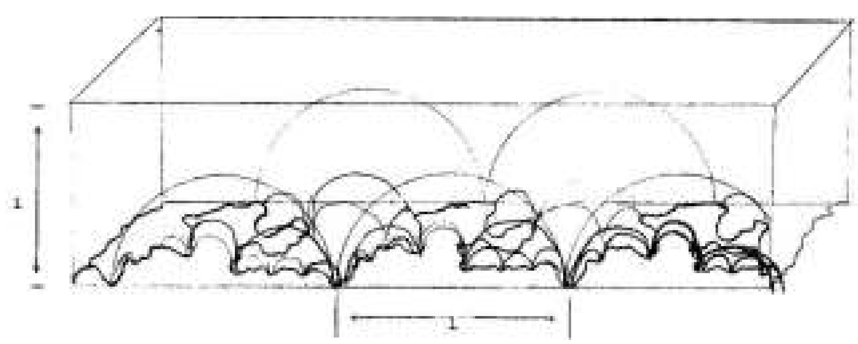


Let us now relate definition 8.4.1 to other possible notions of geometric finiteness. The usual definition is in terms of a fundamental polyhedron for the action of Γ. For concreteness, let us consider only the case $n = 3$. For the present discussion, a finite-sided polyhedron means a region P in H^3 bounded by finitely many planes. P

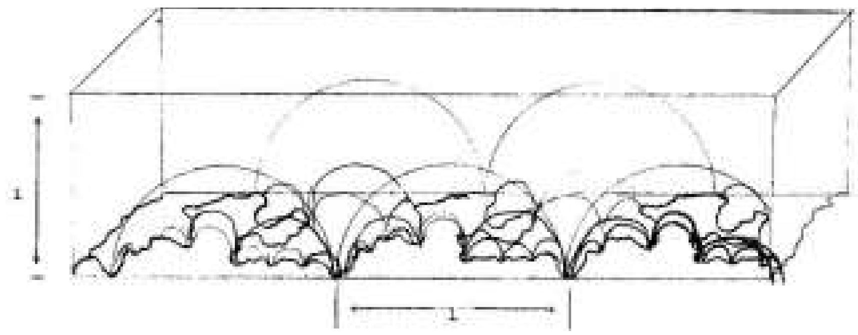
is a fundamental polyhedron for Γ if its translates by Γ cover H^3, and the translates of its interior are pairwise disjoint. P intersects S_∞ in a polygon which unfortunately 8.18
may be somewhat bizarre, since tangencies between sides of $P \cap S_\infty$ may occur.



Sometimes these tangencies are forced by the existence of parabolic fixed points for Γ. Suppose that $p \in S_\infty$ is a parabolic fixed point for some element of Γ, and let π be the subgroup of Γ fixing p. Let B be a horoball centered at p and sufficiently small that the projection of B/P to N_Γ is an embedding. (Compare §5.10.) If $\pi \supset \mathbb{Z} \oplus \mathbb{Z}$, for any point $x \in B \cap H(L_\Gamma)$, the convex hull of πx contains a horoball B', so in particular there is a horoball $B' \subset H(L_\Gamma) \cap B$. Otherwise, $\mathbb{Z}$ is a maximal torsion-free subgroup of π. Coordinates can be chosen so that p is the point at ∞ in the upper half-space model, and $\mathbb{Z}$ acts as translations by real integers. There is some minimal strip $S \subseteq \mathbb{C}$ containing $L_\Gamma \cap \mathbb{C}$; S may interesect the imaginary axis in a finite, half-infinite, or doubly infinite interval. In any case, $H(L_\Gamma)$ is contained in the region R of upper half-space above S, and the part of ∂R of height ≥ 1 lies on ∂H_Γ. 8.19



It may happen that there are wide substrips $S' \subset S$ in the complement of L_Γ. If S' is sufficiently wide, then the plane above its center line intersects $H(L_\Gamma)$ in B, so it gives a half-open annulus in $B/\mathbb{Z}$. If Γ is torsion-free, then maximal, sufficiently wide strips in $S - L_\Gamma$ give disjoint non-parallel half-open annuli in M_Γ; an easy argument shows they must be finite in number if Γ is finitely generated. (This also follows from Ahlfors's finiteness theorem.) Therefore, there is some horoball B' centered at p so that $H(L_\Gamma) \cap B' = R \cap B'$. This holds even if Γ has torsion.



8.20

With an understanding of this picture of the behaviour of M_Γ near a cusp, it is not hard to relate various notions of geometric finiteness. For convenience suppose Γ is torsion-free. (This is not an essential restriction in view of Selberg's theorem—see § .) When the context is clear, we abbreviate $M_\Gamma = M$, $N_\Gamma = N$, etc.

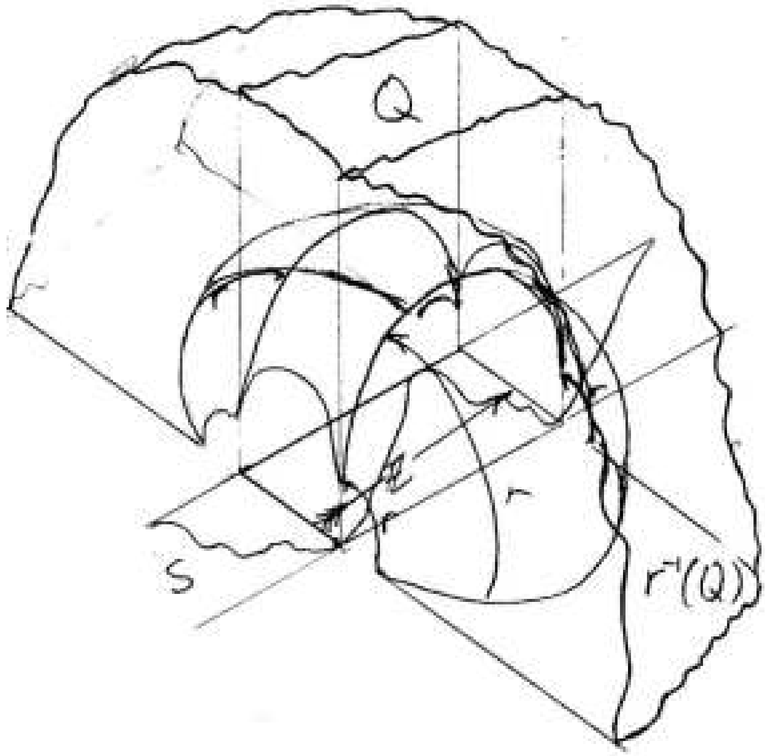
Proposition 8.4.3. *Let* $\Gamma \subset \mathrm{PSL}(2, \mathbb{C})$ *be a discrete, torsion-free group. The following conditions are equivalent:*

(a) Γ *is geometrically finite* (see dfn. 8.4.1).
(b) $M_{[\epsilon,\infty)}$ *is compact.*
(c) Γ *admits a finite-sided fundamental polyhedron.*

Proof. (a) $\Rightarrow$ (b).

Each point in $M_{[\epsilon,\infty)}$ has an embedded $\epsilon/2$ ball in $\mathcal{N}_{\epsilon/2}(M_\Gamma)$, by definition. If (a) holds, $\mathcal{N}_{\epsilon/n}(M_\Gamma)$ has finite volume, so only finitely many of these balls can be disjoint and $M_{\Gamma[\epsilon,\infty)}$ is compact.

(b) $\Rightarrow$ (c). First, find fundamental polyhedra near the non-equivalent parabolic fixed points. To do this, observe that if p is a $\mathbb{Z}$-cusp, then in the upper half-space model, when $p = \infty$, $L_\Gamma \cap \mathbb{C}$ lies in a strip S of finite width. Let R denote the region above S. Let B' be a horoball centered at ∞ such that $R \cap B' = H(L_\Gamma) \cap B'$. Let $r : H^3 \cup D_\Gamma \to H(L_\Gamma)$ be the canonical retraction. If Q is any fundamental polyhedra for the action of $\mathbb{Z}$ in some neighborhood of p in $H(L_\Gamma)$ then $r^{-1}(Q)$ is a fundamental polyhedron in some neighborhood of p in $H^3 \cup D_\Gamma$. 8.21



A fundamental polyhedron near the cusps is easily extended to a global fundamental polyhedron, since O_Γ-(neighborhoods of the cusps) is compact.

(c) $\Rightarrow$ (a). Suppose that Γ has a finite-sided fundamental polyhedron P.

A point $x \in P \cap S_\infty$ is a regular point ($\in D_\Gamma$) if it is in the interior of $P \cap S_\infty$ or of some finite union of translates of P. Thus, the only way x can be a limit point is for x to be a point of tangency of sides of infinitely many translates of P. Since P can have only finitely many points of tangency of sides, infinitely many $\gamma\Gamma$ must identify one of these points to x, so x is a fixed point for some element $\gamma\Gamma$. γ must be parabolic, otherwise the translates of P by powers of γ would limit on the axis of γ. If x is arranged to be ∞ in upper half-space, it is easy to see that $L_\Gamma\mathbb{C}$ must be contained in a strip of finite width. (Finitely many translates of P must form a fundamental domain for $\{\gamma^n\}$, acting on some horoball centered at ∞, since $\{\gamma^n\}$ has finite index in the group fixing ∞. Th faces of these translates of P which do not pass through ∞ lie on hemispheres. Every point in $\mathbb{C}$ outside this finite collection of hemispheres and their translates by $\{\gamma^n\}$ lies in D_Γ.) 8.22

It follows that $v(\mathcal{N}_\epsilon(M)) = v(\mathcal{N}_\epsilon(H(L_\Gamma)) \cap P)$ if finite, since the contribution near any point of $L_\Gamma \cap P$ is finite and the rest of $\mathcal{N}_\epsilon(H(L_\Gamma)) \cap P$ is compact. $\square$

8.5. The geometry of the boundary of the convex hull

Consider a closed curve σ in Euclidean space, and its convex hull $H(\sigma)$. The boundary of a convex body always has non-negative Gaussian curvature. On the other hand, each point p in $\partial H(\sigma) - \sigma$ lies in the interior of some line segment or triangle with vertices on σ. Thus, there is some line segment on $\partial H(\sigma)$ through p, so that $\partial H(\sigma)$ has non-positive curvature at p. It follows that $\partial H(\sigma) - \sigma$ has zero curvature, i.e., it is "developable". If you are not familiar with this idea, you can see it by bending a curve out of a piece of stiff wire (like a coathanger). Now roll the wire around on a big piece of paper, tracing out a curve where the wire touches. Sometimes, the wire may touch at three or more points; this gives alternate ways to roll, and you should carefully follow all of them. Cut out the region in the plane bounded by this curve (piecing if necessary). By taping the paper together, you can envelope the wire in a nice paper model of its convex hull. The physical process of unrolling a developable surface onto the plane is the origin of the notion of the developing map. 8.23

The same physical notion applies in hyperbolic three-space. If K is any closed set on S_∞, then $H(K)$ is convex, yet each point on $\partial H(K)$ lies on a line segment in $\partial H(K)$. Thus, $\partial H(K)$ can be developed to a hyperbolic plane. (In terms of Riemannian geometry, $\partial H(K)$ has extrinsic curvature 0, so its intrinsic curvature is the ambient sectional curvature, -1. Note however that $\partial H(K)$ is not usually differentiable). Thus $\partial H(K)$ has the natural structure of a complete hyperbolic surface.

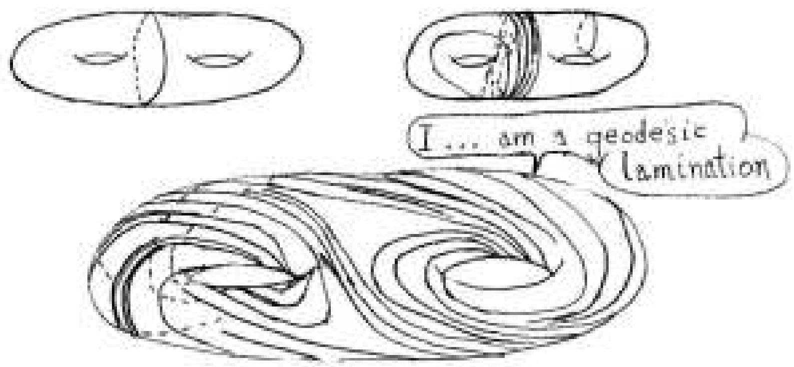
Proposition 8.5.1. *If Γ is a torsion-free Kleinian group, the ∂M_Γ is a hyperbolic surface.* $\square$

The boundary of M_Γ is of course not generally flat—it is bent in some pattern. Let $\gamma \subset \partial M_\Gamma$ consist of those points which are not in the interior of a flat region of 8.24
∂M_Γ. Through each point x in γ, there is a unique geodesic g_x on ∂M_Γ. g_x is also a geodesic in the hyperbolic structure of ∂M_Γ. γ is a closed set. If ∂M_Γ has finite area, then γ is compact, since a neighborhood of each cusp of ∂M_Γ is flat. (See §8.4.)

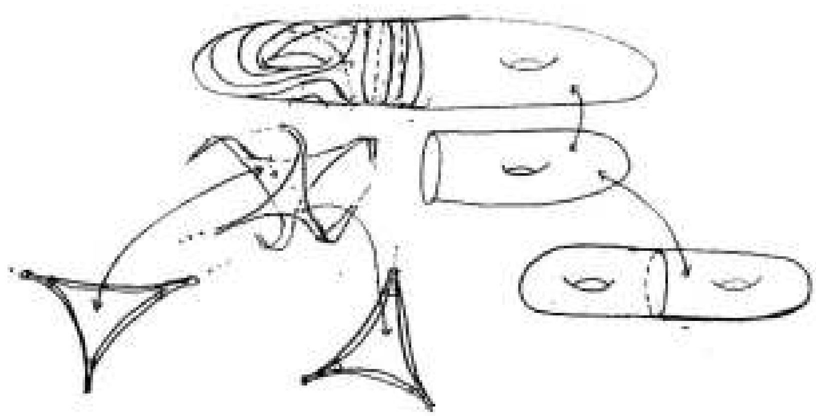
Definition 8.5.2. A *lamination* L on a manifold M^n is a closed subset $A \subset M$ (the *support* of L) with a local product structure for A. More precisely, there is a covering of a neighborhood of A in M with coordinate neighborhoods $U_i \xrightarrow{\phi_i} \mathbb{R}^{n-k} \times \mathbb{R}^k$ so that $\phi_i(A \cap U_i)$ is of the form $\mathbb{R}^{n-k} \times B$, $B \subset \mathbb{R}^k$. The coordinate changes ϕ_{ij} must be of the form $\phi_{ij}(x, y) = (f_{ij}(x, y), g_{ij}(y))$ when $y \in B$. A lamination is like a foliation of a closed subset of M. Leaves of the lamination are defined just as for a foliation.

EXAMPLES. If $\mathcal{F}$ is a foliation of M and $S \subset M$ is any set, the closure of the union of leaves which meet S is a lamination.

Any submanifold of a manifold M is a lamination, with a single leaf. Clearly, the bending locus γ for ∂M_Γ has the structure of a lamination: whenever two points of γ are nearby, the directions of bending must be nearly parallel in order that the lines of bending do not intersect. A lamination whose leaves are geodesics we will call a *geodesic lamination.* 8.25

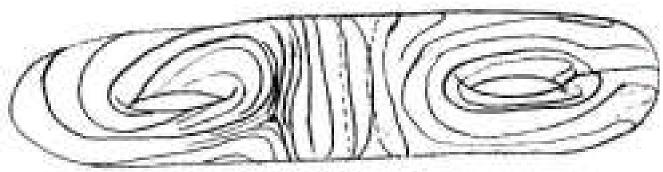


By consideration of Euler characteristic, the lamination γ cannot have all of ∂M as its support, or in other words it cannot be a foliation. The complement $\partial M - \gamma$ consists of regions bounded by closed geodesics and infinite geodesics. Each of these regions can be doubled along its boundary to give a complete hyperbolic surface, which of course has finite area. There

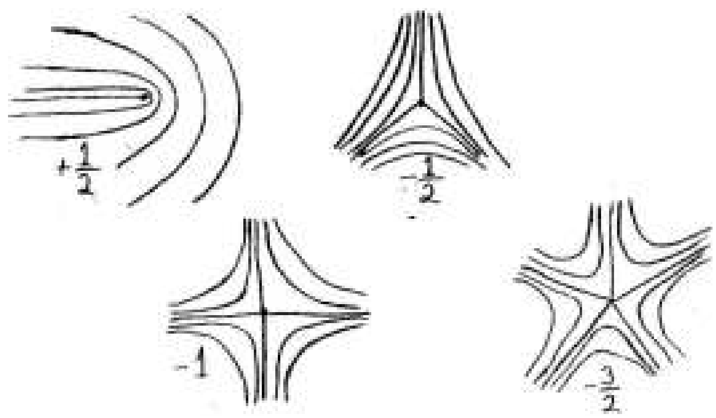


8.26

is a lower bound for π for the area of such a region, hence an upper bound of $2|\chi(\partial M)|$ for the number of components of $\partial M - \gamma$. Every geodesic lamination γ on a hyperbolic surface S can be extended to a foliation with isolated singularities on the complement. There



is an index formula for the Euler characteristic of S in terms of these singularities. Here are some values for the index.



8.27

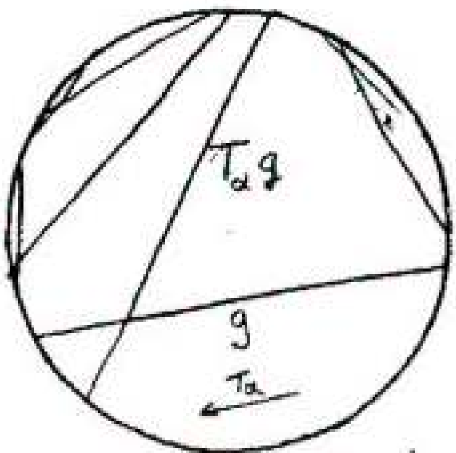
From the *existence* of an index formula, one concludes that the Euler characteristic of S is half the Euler characteristic of the double of $S - \gamma$. By the Gauss-Bonnet theorem,

$$\mathrm{Area}(S - \gamma) = \mathrm{Area}(S)$$

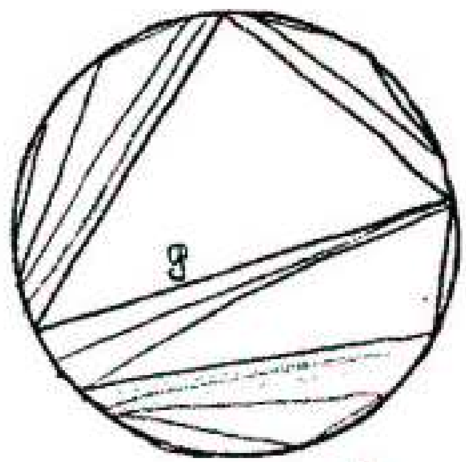
or in other words, γ has measure 0. To give an idea of the range of possibilities for geodesic laminations, one can consider an arbitrary sequence $\{\gamma_i\}$ of geodesic laminations: simple closed curves, for instance. Let us say that $\{\gamma_i\}$ converges geometrically to γ if for each $x \in \operatorname{support} \gamma$, and for each ϵ, for all great enough i the support of γ_i intersects $\mathcal{N}_\epsilon(x)$ and the leaves of $\gamma_i \cap \mathcal{N}_\epsilon(x)$ are within ϵ of the direction of the leaf of γ through x. Note that the support of γ may be smaller than the limiting support of γ_i, so the limit of a sequence may not be unique. See §8.10. An easy diagonal argument shows that every sequence $\{\gamma_i\}$ has a subsequence which converges geometrically. From limits of sequences of simple closed geodesics, uncountably many geodesic laminations are obtained.

Geodesic laminations on two homeomorphic hyperbolic surfaces may be compared by passing to the circle at ∞. A directed geodesic is determined by a pair of points $(x_1, x_2) \in S^1_\infty \times S^1_\infty - \Delta$, where Δ is the diagonal $\{(x,x)\}$. A geodesic without direction is a point on $J = (S^1_\infty \times S^1_\infty - \Delta/\mathbb{Z}_2)$, where $\mathbb{Z}_2$ acts by interchanging coordinates. Topologically, J is an open Moebius band. It is geometrically realized in the Klein (projective) model for H^2 as the region outside H^2. A geodesic g projects 8.28

to a simple geodesic on the surface S if and only if the covering translates of its pairs of end points never strictly separate each other.

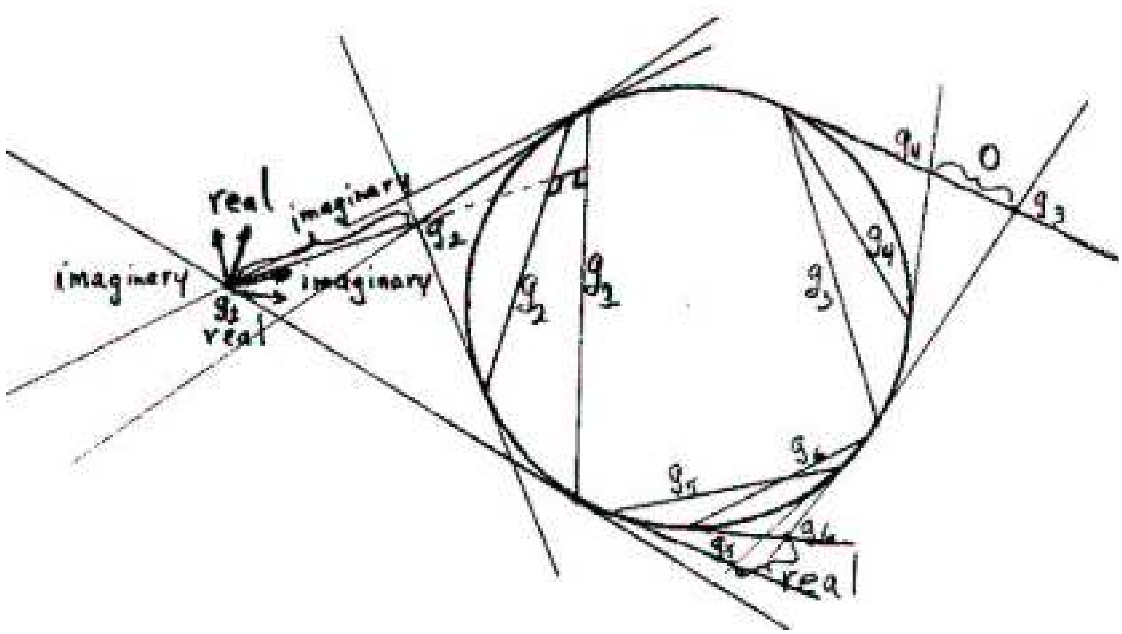


g is not simple



g is simple

Geometrically, J has an indefinite metric of type $(1, 1)$, invariant by covering translates. (See §2.6.) The light-like geodesics, of zero length, are lines tangent to S^1_∞; lines which meet H^2 when extended have imaginary arc length. A point $g \in J$ projects to a simple geodesic in S if and only if no covering translate $T_\alpha(g)$ has a positive real distance from g.



8.29

Let $\mathcal{S} \subset J$ consist of all elements g projecting to simple geodesics on S. Any geodesic $\subset H^2$ which has a translate intersecting itself has a neighborhood with the same property, hence $\mathcal{S}$ is closed.

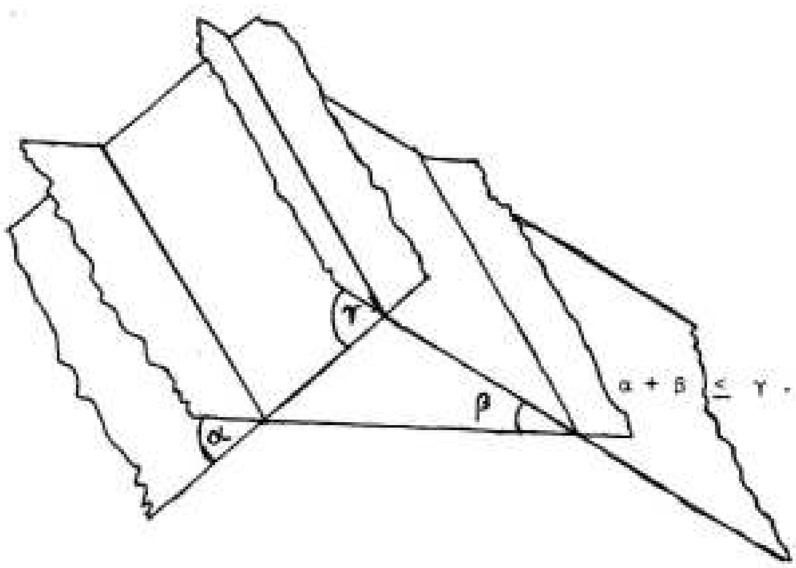
If γ is any geodesic lamination on S, Let $\mathcal{S}_\gamma \subset J$ be the set of lifts of leaves of γ to H^2. $\mathcal{S}_\gamma$ is a closed invariant subset of $\mathcal{S}$. A closed invariant subset of $C \subset J$ gives rise to a geodesic lamination if and only if all pairs of points of C are separated by an imaginary (or 0) distance. If $g \in \mathcal{S}$, then the closure of its orbit, $\overline{\pi_1(S)g}$ is such a set, corresponding to the geodesic lamination $\bar{g}$ of S. Every homeomorphism between surfaces when lifted to H^2 extends to S^1_∞ (by 5.9.5). This determines an extension to J. Geodesic laminations are transferred from one surface to another via this correspondence.

8.6. Measuring laminations

Let L be a lamination, so that it has local homeomorphisms $\phi_i : L \cap U_i \approx \mathbb{R}^{n-k} \times B_i$. A *transverse measure* μ for L means a measure μ_i defined on each local leaf space B_i, in such a way that the coordinate changes are measure preserving. Alternatively one may think of μ as a measure defined on every k-dimensional submanifold transverse to L, supported on $T^k \cap L$ and invariant under local projections along leaves of L. We will always suppose that μ is finite on compact transversals. The simplest example of a transverse measure arises when L is a closed submanifold; in this case, one can take μ to count the number of intersections of a transversal with L. 8.30

We know that for a torsion-free Kleinian group Γ, ∂M_Γ is a hyperbolic surface bent along some geodesic lamination γ. In order to complete the picture of ∂M_Γ, we need a quantitative description of the bending. When two planes in H^3 meet along a line, the angle they form is constant along that line. The flat pieces of ∂M_Γ meet each other along the geodesic lamination γ; the angle of meeting of two planes generalizes to a transverse "bending" measure, β, for γ. The measure β applied to an arc α on ∂M_Γ transverse to γ is the total angle of turning of the normal to ∂M_Γ along α (appropriately interpreted when γ has isolated geodesics with sharp bending). In order to prove that β is well-defined, and that it determines the local isometric embedding in H^3, one can use local polyhedral approximations to ∂M_Γ. Local outer approximations to ∂M_Γ can be obtained by extending the planes of local flat regions. Observe that when three planes have pairwise intersections in H^3 but no triple intersection, the dihedral angles satisfy the inequality

$$\alpha + \beta \leq \gamma.$$



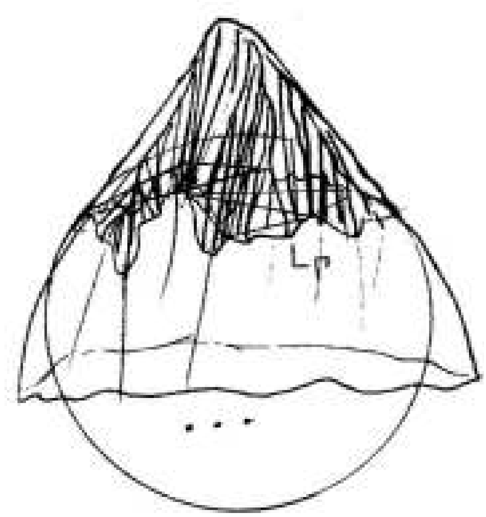
8.31

(The difference $\gamma - (\alpha + \beta)$ is the area of a triangle on the common perpendicular plane.) From this it follows that as outer polyhedral approximations shrink toward M_Γ, the angle sum corresponding to some path α on ∂M_Γ is a monotone sequence, converging to a value $\beta(\alpha)$. Also from the monotonicity, it is easy to see that for short paths α_t, $[0 \le t \le 1]$, $\beta(\alpha)$ is a close approximation to the angle between the tangent planes at α_0 and α_1. This implies that the hyperbolic structure on ∂M_Γ, together with the geodesic lamination γ and the transverse measure β, completely determines the hyperbolic structure of N_Γ in a neighborhood of ∂M_Γ.

The bending measure β has for its support all of γ. This puts a restriction on the structure of γ: every isolated leaf L of γ must be a closed geodesic on ∂M_Γ. (Otherwise, a transverse arc through any limit point of L would have infinite measure.) This limits the possibilities for the intersection of a transverse arc with γ to a Cantor set and/or a finite set of points.

When γ contains more than one closed geodesic, there is obviously a whole family of possibilities for transverse measures. There are (probably atypical) examples of families of distinct transverse measures which are not multiples of each other even for certain geodesic laminations such that every leaf is dense. There are many other examples which possess unique transverse measures, up to constant multiples. Compare Katok. 8.32

Here is a geometric interpretation for the bending measure β in the Klein model. Let P_0 be the component of P_Γ containing N_Γ (recall definition 8.3.5). Each point in $\tilde{P}_0$ outside S_∞ is dual to a plane which bounds a half-space whose intersection with S_∞ is contained in D_Γ. $\partial\tilde{P}_0$ consists of points dual to planes which meet L_Γ in at least one point. In particular, each plane meeting $\tilde{M}_\Gamma$ in a line or flat of $\partial\tilde{M}_\Gamma$ is dual to a point on $\partial\tilde{P}_0$. If $\bar{\pi} \in \partial\tilde{P}_0$ is dual to a plane π touching L_Γ at x, then one of the line segments $\overline{\bar{\pi}x}$ is also on $\partial\tilde{P}_0$. This line segments consists of points dual to planes touching L_Γ at x and contained in a half-space bounded by π. The reader may check that $\tilde{P}_0$ is convex. The natural metric of type $(2,1)$ in the exterior of S_∞ is degenerate on $\partial\tilde{P}_0$, since it vanishes on all line segments corresponding to a family of planes tangent at S_∞. Given a path α on $\partial\tilde{M}_\Gamma$, there is a dual path $\bar{\alpha}$ consisting of points dual to planes just skimming M_Γ along α. The length of $\bar{\alpha}$ is the same as $\beta(\alpha)$.



8.33

Remark. The interested reader may verify that when N is a component of ∂M_Γ such that every leaf of $\gamma \cap N$ is dense in $\gamma \cap N$, then the action of $\pi_1 n$ on the appropriate component of $\partial\tilde{P}_0 - L_\Gamma$ is minimal (i.e., every orbit is dense). This action is approximated by actions of $\pi_1 N$ as covering transformations on surfaces just inside $\partial\tilde{P}_0$.

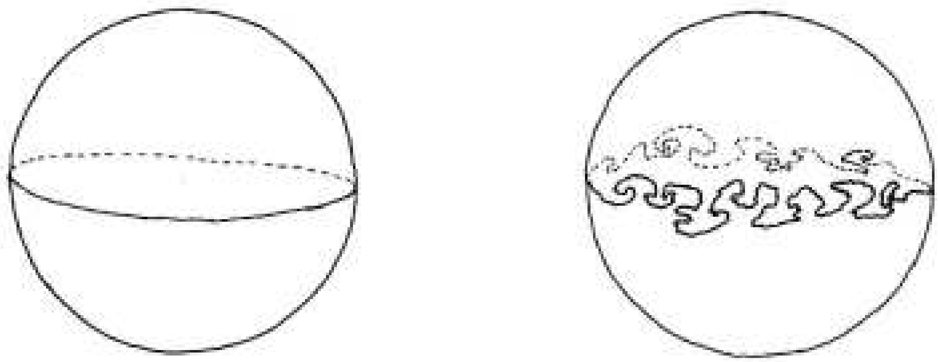
8.7. Quasi-Fuchsian groups

Recall that a Fuchsian group (of type I) is a Kleinian group Γ whose limit set L_Γ is a geometric circle. Examples are the fundamental groups of closed, hyperbolic surfaces. In fact, if the Fuchsian group Γ is torsion-free and has no parabolic elements, then Γ is the group of covering transformations of a hyperbolic surface. Furthermore, the Kleinian manifold $O_\Gamma = (H^3 \cup D_\Gamma)/\Gamma$ has a totally geodesic surface as a spine.

Note. The type of a Fuchsian group should not be confused with its type as a Kleinian group. To say that Γ is a Fuchsian group of type I means that $L_\Gamma = S^1$, but it is a Kleinian group of type II since $D_\Gamma \neq \emptyset$.

Suppose $M = N^2 \times I$ is a convex hyperbolic manifold, where N^2 is a closed surface. Let Γ' be the group of covering transformations of M, and let Γ be a Fuchsian group coming from a hyperbolic structure on N. Γ and Γ' are isomorphic as groups; we want to show that their actions on the closed ball B^3 are topologically conjugate.

Let M_Γ and $M_{\Gamma'}$ be the convex hull quotients ($M_\Gamma \approx N^2$ and $M_{\Gamma'} \approx N^2 \times I$). Thicken M_Γ and $M_{\Gamma'}$ to strictly convex manifolds. The thickened manifolds are 8.34
diffeomorphic, so by Proposition 8.3.4 there is a quasi-conformal homeomorphism of B^3 conjugating Γ to Γ'. In particular, $L_{\Gamma'}$ is homeomorphic to a circle. Γ', which has convex hull manifold homeomorphic to $N^2 \times I$ and limit set $\approx S^1$, is an example of a quasi-Fuchsian group.



Definition 8.7.1. The Kleinian group Γ is called a *quasi-Fuchsian group* if L_Γ is topologically S^1.

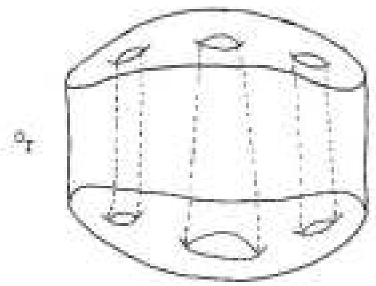
Proposition 8.7.2 (Marden). *For a torsion-free Kleinian group Γ, the following conditions are equivalent.*

(i) Γ *is quasi-Fuchsian.*
(ii) D_Γ *has precisely two components.*
(iii) Γ *is quasi-conformally conjugate to a Fuchsian group.*

Proof. Clearly (iii) $\Longrightarrow$ (i) $\Longrightarrow$ (ii). To show (ii) $\Longrightarrow$ (iii), consider

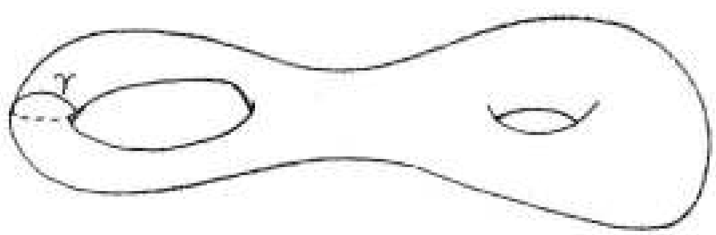
$$O_\Gamma = (H^3 \cup D_\Gamma)/\Gamma.$$

Suppose that no element of Γ interchanges the two components of D_Γ. Then O_Γ is a three-manifold with two boundary components (labelled, for example, N_1 and N_2), and 8.35



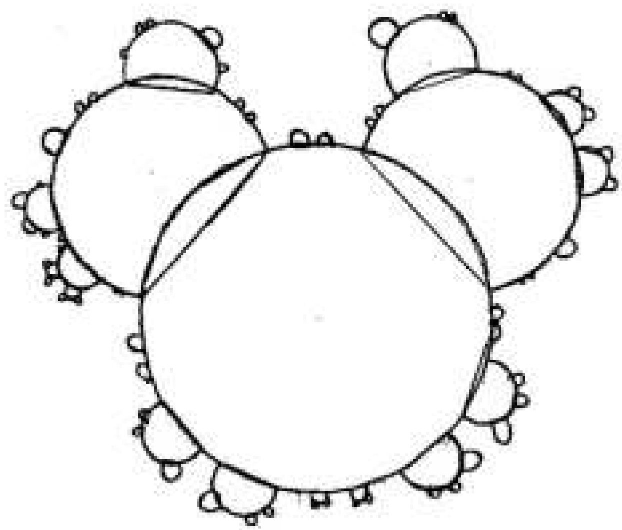
$\Gamma = \pi_1(O_\Gamma) \approx \pi_1(N_1) \approx \pi_1(N_2)$. By a well-known theorem about three-manifolds (see Hempel for a proof), this implies that O_Γ is homeomorphic to $N_1 \times I$. By the above discussion, this implies that Γ' is quasi-conformally conjugate to a Fuchsian group. A similar argument applies if O_Γ has one boundary component; in that case, O_Γ is the orientable interval bundle over a non-orientable surface. The reverse implication is clear. □

Example 8.7.3 (Mickey mouse). Consider a hyperbolic structure on a surface of genus two. Let us construct a deformation of the corresponding Fuchsian group by bending along a single closed geodesci γ by an angle of $\pi/2$. This



8.36

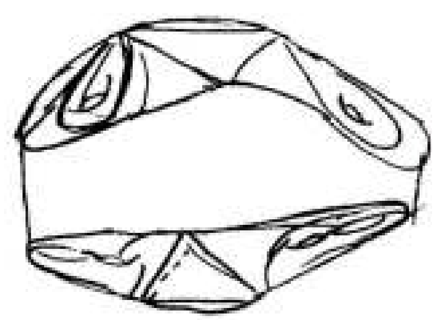
will give rise to a quasi-Fuchsian group if the geodesic is short enough. We may visualize the limit set by imagining bending a hyperbolic plane along the lifts of γ, one by one.



We want to understand how the geometry changes as we deform quasi-Fuchsian groups. Even though the topology doesn't change, geometrically things can become very complicated. For example, given any $\epsilon > 0$, there is a quasi-Fuchsian group Γ whose limit set L_Γ is ϵ-dense in S^2, and there are limits of quasi-Fuchsian groups with $L_\Gamma = S^2$.

Our goal here is to try to get a grasp of the geometry of the convex hull quotient $M = M_\Gamma$ of a quasi-Fuchsian group Γ. M_Γ is a convex hyperbolic manifold which is homeomorphic to $N^2 \times I$, and the two boundary components are hyperbolic surfaces bent along geodesic laminations. 8.37

M_Γ

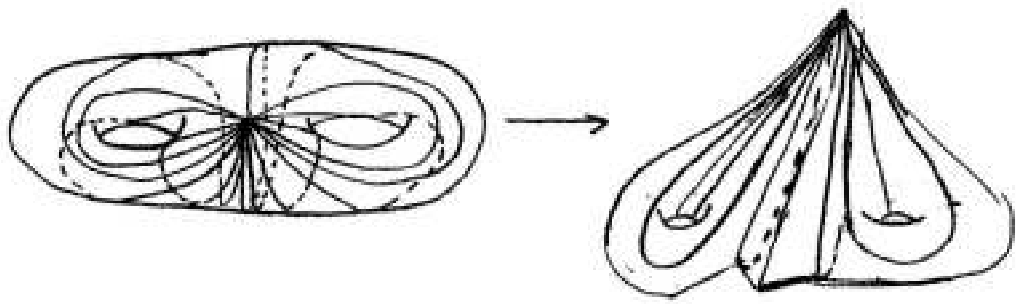


We also need to analyze intermediate surfaces in M_Γ. For example, what kinds of nice surfaces are embedded (or immersed) in M_Γ? Are there isometrically embedded cross sections? Are there cross sections of bounded area near any point in M_Γ?

Here are some ways to map in surfaces.

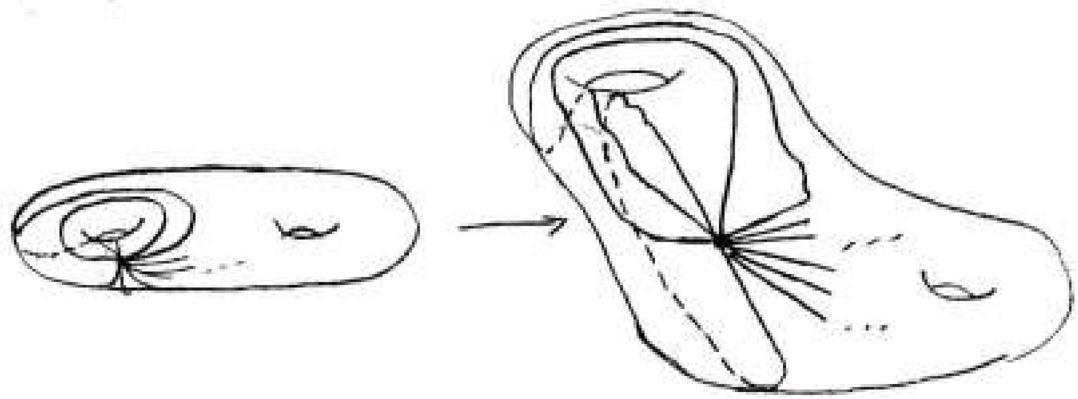
(a) Take the abstract surface N^2, and choose a "triangulation" of N with one vertex. Choose an arbitrary map of N into M. Then straighten the map (see §6.1).

This is a fairly good way to map in a surface, since the surface is hyperbolic away from the vertex. There may be positive curvature concentrated at the vertex, however, since the sum of the angles around the vertex may be quite small. This map can be changed by moving the image of the vertex in M or by changing the triangulation on N.



8.38

(b) Here is another method, which insures that the map is not too bad near the vertex. First pick a closed loop in N, and then choose a vertex on the loop. Now extend this to a triangulation of N with one vertex. To map in N, first map

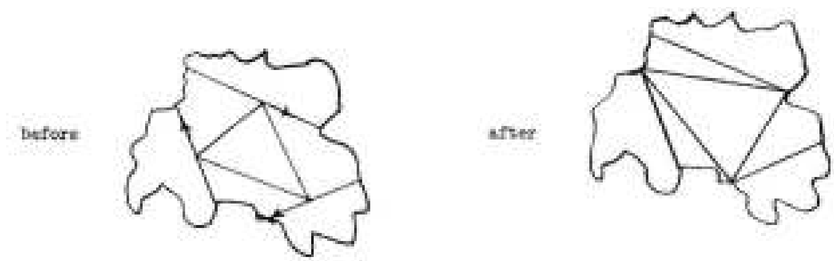


in the loop to the unique geodesic in M in its free homotopy class (this uses a homeomorphism of M to $N \times I$). Now extend this as in (a) to a piecewise straight map $f : N \to M$. The sum of the angles around the vertex is at least 2π, since there is a straight line segment going through the vertex (so the vertex cannot be spiked). It is possible to have the sum of the angles $> 2\pi$, in which case there is negative curvature concentrated near the vertex.

(c) Here is a way to map in a surface with constant negative curvature. Pick an example, as in (b), of a triangulation of N coming from a closed geodesic, and map 8.39
N as in (b). Consider the isotopy obtained by moving the vertex around the loop more and more. The loop stays the same, but the other line segments start spiraling

around the loop, more and more, converging, in the limit, to a geodesic laminated set. The surface N maps into M at each finite stage, and this carries over in the limit to an isometric embedding of a hyperbolic surface. The triangles with an edge on the fixed loop have disappeared in the limit. Compare 3.9.

One can picture what is going on by looking upstairs at the convex hull $H(L_\Gamma)$. The lift $\tilde{f} : \tilde{N} \to H(L_\Gamma)$ of the map from the original triangulation (before isotoping the vertex) is defined as follows. First the geodesic (coming from the loop) and its conjugates are mapped in (these are in the convex hull since their



8.40

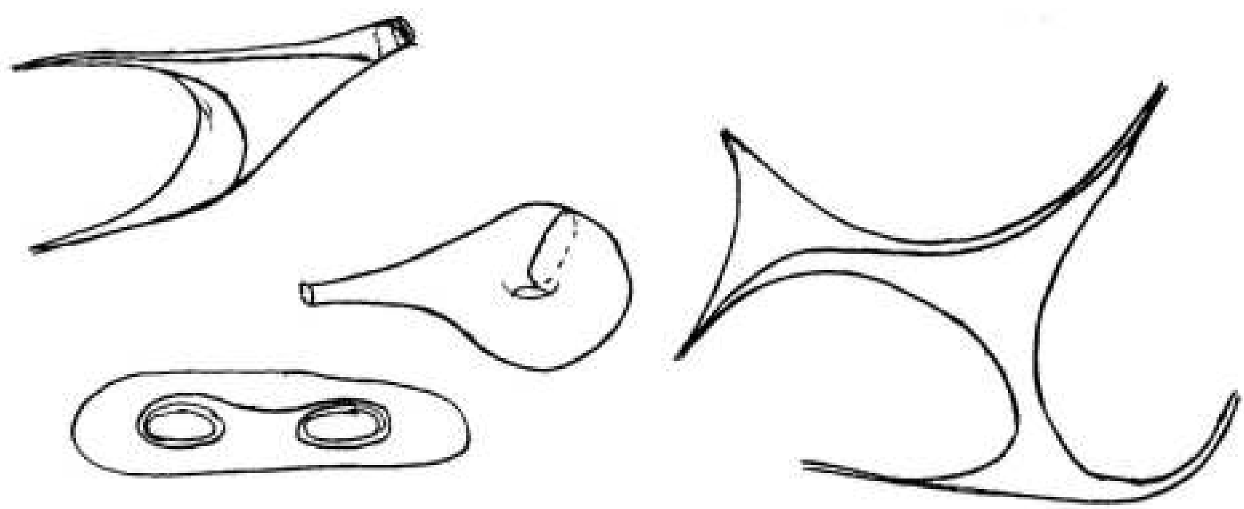
endpoints are in L_Γ). The line segments connect different conjugates of the geodesic, and the triangles either connect three distinct conjugates or two conjugates (when the original loop is an edge of the triangle). As we isotope the vertex around the loop, the image vertices slide along the geodesic (and its conjugates), and in the limit the triangles become asymptotic (and the triangles connecting only two conjugates disappear).

The above method works because the complement of the geodesic lamination (obtained by spinning the triangulation) consists solely of asymptotic triangles. Here is a more general method of mapping in a surface N by using geodesic laminations.

Definition 8.7.5. A geodesic lamination γ on hyperbolic surface S is *complete* if the complementary regions in $S - \gamma$ are all asymptotic triangles.

Proposition 8.7.6. *Any geodesic lamination γ on a hyperbolic surface S can be completed, i.e., γ can be extended to a complete geodesic lamination $\gamma' \supset \gamma$ on S.*

Proof. Suppose γ is not complete, and pick a complementary region A which is not an asymptotic triangle. If A is simply connected, then it is a finite-sided asymptotic polygon, and it is easy to divide A into asymptotic triangles by adding simple geodesics. If A is not simply connected, extend γ to a larger geodesic lamination by adding a simple geodesic α in A 8.41

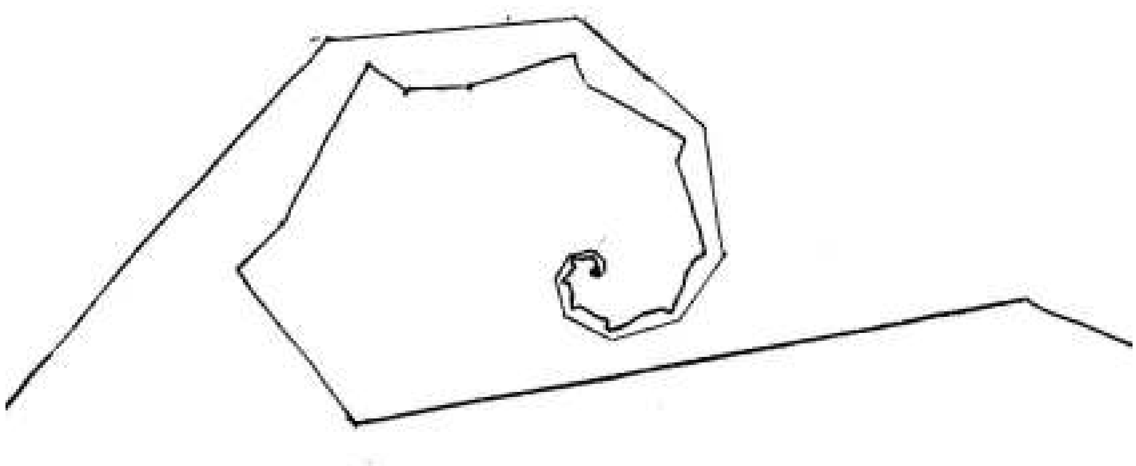


(being careful to add a *simple* geodesic). Either α separates A into two pieces (each of which has less area) or α does not separate A (in which case, cutting along α reduces the rank of the homology. Continuing inductively, after a finite number of steps A separates into asymptotic triangles. □

Completeness is exactly the property we need to map in surfaces by using geodesic laminations.

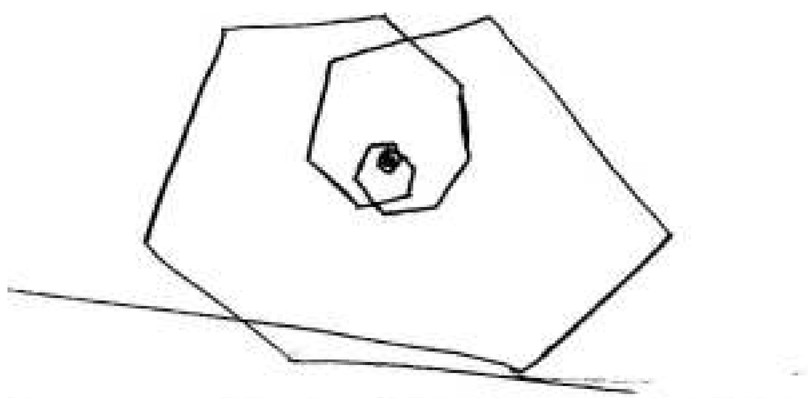
PROPOSITION 8.7.7. *Let S be an oriented hyperbolic surface, and Γ a quasi-Fuchsian group isomorphic to $\pi_1 S$. For every complete geodesic lamination γ on S, there is a unique hyperbolic surface $S' \approx S$ and an isometric map $f : S' \to M_\Gamma$ which is straight* (*totally geodesic*) *in the complement of γ.* (γ here denotes the corresponding geodesic lamination on any hyperbolic surface homeomorphic to S.)

REMARK. By an isometric map $f : M_1 \to M_2$ from one Riemannian manifold to another, we mean that for every rectifiable path α_t in M_1, $f \circ \alpha_t$ is rectifiable and 8.42
has the same length as α_t. When f is differentiable, this means that df preserves lengths of tangent vectors. We shall be dealing with maps which are not usually differentiable, however. Our maps are likely not even to be local embeddings. A cross-section of the image of a surface mapped in by method (c) has two polygonal spiral branches, if the closed geodesic corresponds to a covering transformation which is not a pure translation:



(This picture is obtained by considering triangles in H^3 asymptotic to a loxodromic axis, together with their translates.)

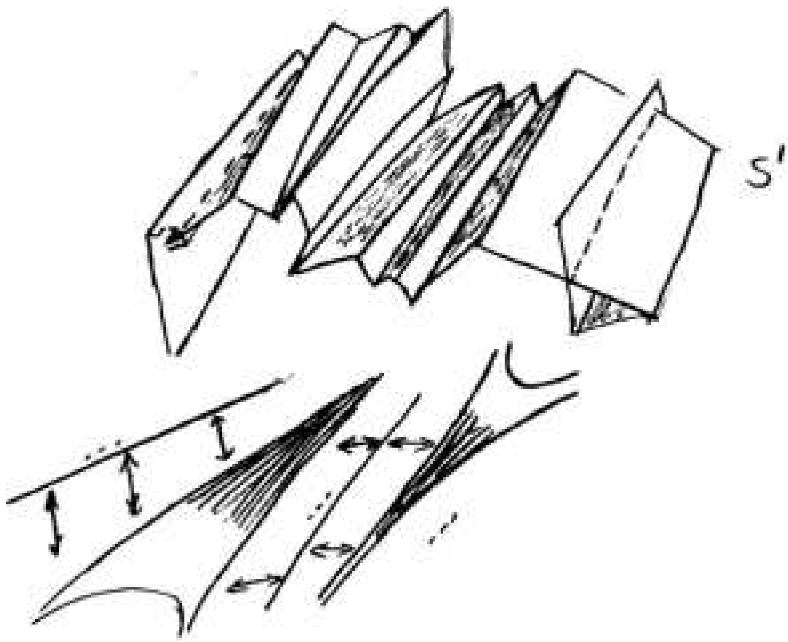
If the triangulation is spun in opposite directions on opposite sides of the geodesic, the polygonal spiral have opposite senses, so there are infinitely many self-intersections.



8.43

Proof. The hyperbolic surface $\tilde{S}'$ is constructed out of pieces. The asymptotic triangles in $\tilde{S} - \tilde{\gamma}$ are determined by triples of points on S^1_∞. We have a canonical identification of S^1_∞ with L_Γ; the corresponding triple of points in L_Γ spans a triangle in H^3, which will be a piece of $\tilde{S}'$. Similarly, corresponding to each leaf of $\tilde{\gamma}$ there is a canonical line in H^3. These triangles and lines fit together just as on $\tilde{S}$; from this the picture of $\tilde{S}'$ should be clear. Here is a formal definition. Let P_γ be the set of all "pieces" of $\tilde{\gamma}$, i.e., P_γ consists of all leaves of $\tilde{\gamma}$, together with all components of $\tilde{S} - \tilde{\gamma}$. Let P_γ have the (non-Hausdorff) quotient topology. The universal cover $\tilde{S}'$

is defined first, to consist of ordered pairs (x, p), where $p \in P_\gamma$ and x is an element of the piece of H^3 corresponding to p. Γ acts on this space $\tilde{S}'$ in an obvious way; the quotient space is defined to be S'. It is not hard to find local coordinates for S', showing that it is a (Hausdorff) surface.



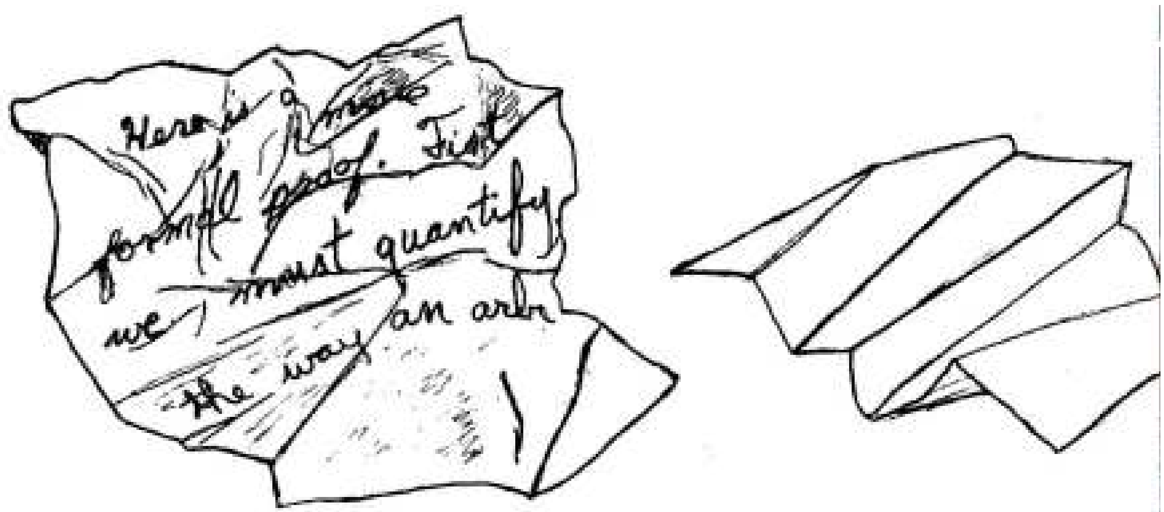
8.44

An appeal to geometric intuition demonstrates that S' is a hyperbolic surface, mapped isometrically to M_Γ, straight in the complement of γ. Uniqueness is evident from consideration of the circle at ∞. □

Remark. There are two approaches which a reader who prefers more formal proofs may wish to check. The first approach is to verify 8.7.7 first for laminations all of whose leaves are either isolated or simple limits of other leaves (as in (c)), and then extend to all laminations by passing to limits, using compactness properties of uncrumpled surfaces (§8.8). Alternatively, he can construct the hyperbolic structure on S' directly by describing the local developing map, as a limit of maps obtained by considering only finitely many local flat pieces. Convergence is a consequence of the finite total area of the flat pieces of S'.

8.8. Uncrumpled surfaces

There is a large qualitative difference between a crumpled sheet of paper and one which is only wrinkled or crinkled. Crumpled paper has fold lines or bending lines going any which way, often converging in bad points.



8.45

Definition 8.8.1. An *uncrumpled surface* in a hyperbolic three-manifold N is a complete hyperbolic surface S of finite area, together with an isometric map $f : S \to N$ such that every $x \in S$ is in the interior of some straight line segment which is mapped by f to a straight line segment. Also, f must take every cusp of S to a cusp of N.

The set of uncrumpled surfaces in N has a well-behaved topology, in which two surfaces $f_1 : S_1 \to N$ and $f_2 : S_2 \to N$ are close if there is an approximate isometry $\phi : S_1 \to S_2$ making f_1 uniformly close to $f_2 \circ \phi$. Note that the surfaces have no preferred coordinate systems.

Let $\gamma \subset S$ consist of those points in the uncrumpled surfaces which are in the interior of unique line segments mapped to line segments.

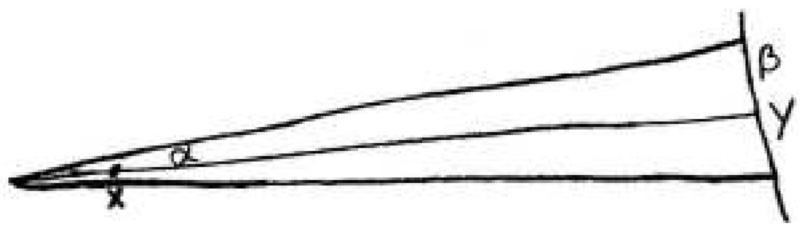
Proposition 8.8.2. *γ is a geodesic lamination. The map f is totally geodesic in the complement of γ.*

Proof. If $x \in S - \gamma$, then there are two transverse line segments through x mapped to line segments. Consider any quadrilateral about x with vertices on these segments; since f does not increase distances, the quadrilateral must be mapped to a plane. Hence, a neighborhood of x is mapped to a plane.

8.46

Consider now any point $x \in \gamma$, and let α be the unique line segment through x which is mapped straight. Let α be extended indefinitely on S. Suppose there were some point y on α in the interior of some line segment $\beta \not\subset \alpha$ which is mapped straight. One may assume that the segment $\overline{xy}$ of α is mapped straight. Then, by considering long skinny triangles with two vertices on β and one vertex on α, it would follow that a neighborhood of x is mapped to a plane—a contradiction.

Thus, the line segments in γ can be extended indefinitely without crossings, so γ must be a geodesic lamination. $\square$



If $U = S \xrightarrow{f} N$ is an uncrumpled surface, then this geodesic lamination $\gamma \subset S$ (which consists of points where U is not locally flat) is the *wrinkling locus* $\omega(U)$.

The *modular space* $\mathcal{M}(S)$ of a surface S of negative Euler characteristic is the space of hyperbolic surfaces with finite area which are homeomorphic to S. In other words, $\mathcal{M}(S)$ is the Teichmüller space $\mathcal{T}(S)$ modulo the action of the group of home- 8.47
omorphisms of S.

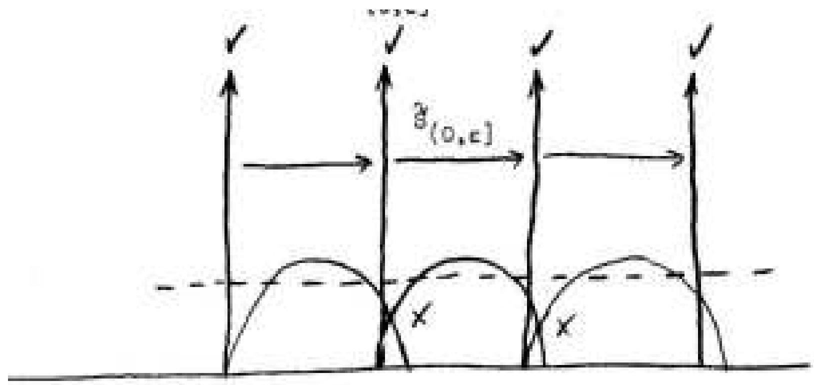
Proposition 8.8.3 (Mumford). *For a surface S, the set $A_\epsilon \subset \mathcal{M}(S)$ consisting of surfaces with no geodesic shorter than ϵ is compact.*

Proof. By the Gauss–Bonnet theorem, all surfaces in $\mathcal{M}(S)$ have the same area. Every non-compact component of $S_{(0,\epsilon]}$ is isometric to a standard model, so the result follows as the two-dimensional version of a part of 5.12. (It is also not hard to give a more direct specifically two-dimensional geometric argument.) $\square$

Denote by $\mathcal{U}(S, N)$ the space of uncrumpled surfaces in N homeomorphic to S with $\pi_1(S) \to \pi_1(N)$ injective. There is a continuous map $\mathcal{U}(S, N) \to \mathcal{M}(S)$ which forgets the isometric map to N.

The behavior of an uncrumpled surface near a cusp is completely determined by its behavior on some compact subset. To see this, first let us prove

Proposition 8.8.4. *There is some ϵ such that for every hyperbolic surface S and every geodesic lamination γ on S, the intersection of γ with every non-compact component of $S_{(0,\epsilon]}$ consists of lines tending toward that cusp.*



8.48

Proof. Thus there are uniform horoball neighborhoods of the cusps of uncrumpled surfaces which are always mapped as cones to the cusp point. Uniform convergence of a sequence of uncrumpled surfaces away from the cusp points implies uniform convergence elsewhere. □

Proposition 8.8.5. *Let $K \subset N$ be a compact subset of a complete hyperbolic manifold N. For any surface S_0, let $W \subset \mathcal{U}(S_0, N)$ be the subset of uncrumpled surfaces $S \xrightarrow{f} N$ such that $f(S)$ intersects K, and satisfying the condition*

(np) $\pi_1(f)$ *takes non-parabolic elements of* $\pi_1 S$ *to non-parabolic elements of* $\pi_1 N$.

Then W is compact.

Proof. The first step is to bound the image of an uncrumpled surface, away from its cusps.

Let ϵ be small enough that for every complete hyperbolic three-manifold M, components of $M_{(0,\epsilon]}$ are separated by a distance of at least (say) 1. Since the area of surfaces in $\mathcal{U}(S_0, N)$ is constant, there is some number d such that any two points in an uncrumpled surface S can be connected (on S) by a path p such that $p \cap S_{[\epsilon,\infty)}$ has length $\leq d$.

If neither point lies in a non-compact component of $S_{(0,\epsilon]}$, one can assume, furthermore, that p does not intersect these components. Let $K' \subset N$ be the set of points which are connected to K by paths whose total length outside compact components of $N_{(0,\epsilon]}$ is bounded by d. Clearly K' is compact and an uncrumpled surface of W must have image in K', except for horoball neighborhoods of its cusps. 8.49

Consider now any sequence $S_1, S_2, \ldots$ in W. Since each short closed geodesic in S_i is mapped into K', there is a lower bound ϵ' to the length of such a geodesic, so by 8.8.3 we can pass to a subsequence such that the underlying hyperbolic surfaces converge in $\mathcal{M}(S)$. There are approximate isometries $\phi_i : S \to S_i$. Then the compositions $f_i \circ \phi_i : S \to N$ are equicontinuous, hence there is a subsequence converging uniformly on $S_{[\epsilon,\infty)}$. The limit is obviously an uncrumpled surface. [To make the picture clear, one can always pass to a further subsequence to make sure that the wrinkling laminations γ_i of S_i converge geometrically.] □

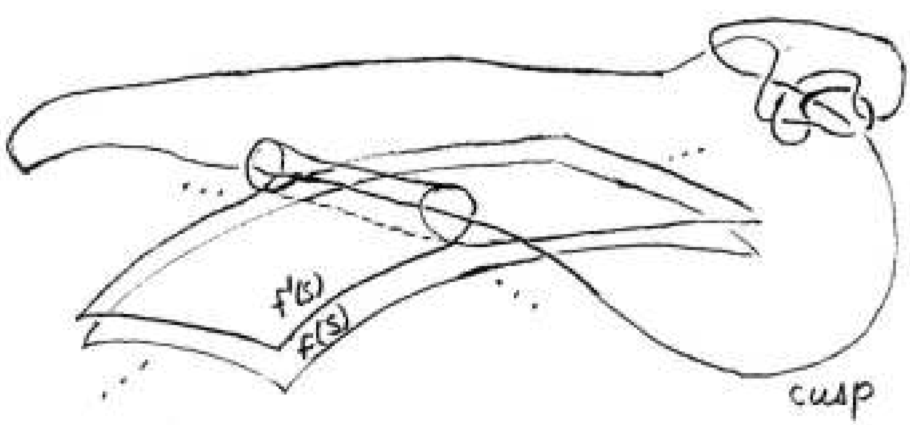
COROLLARY 8.8.6. (a) *Let S be any closed hyperbolic surface, and N any closed hyperbolic manifold. There are only finitely many conjugacy classes of subgroups $G \subset \pi_1 N$ isomorphic to $\pi_1 S$.*

(b) *Let S be any surface of finite area and N any geometrically finite hyperbolic three-manifold. There are only finitely many conjugacy classes of subgroups $G \subset \pi_1 N$ isomorphic to $\pi_1 S$ by an isomorphism which preserves parabolicity* (*in both directions*).

PROOF. Statement (a) is contained in statement (b). The conjugacy class of every subgroup G is represented by a homotopy class of maps of S into N, which is homotopic to an uncrumpled surface (say, by method (c) of §8.7). Nearby uncrumpled surfaces represent the same conjugacy class of subgroups. Thus we have an open cover of the space W by surfaces with conjugate subgroups; by 8.8.5, this is a finite subcover. □ 8.50

REMARK. If non-parabolic elements of $\pi_1 S$ are allowed to correspond to parabolic elements of $\pi_1 N$, then this statement is no longer true.

In fact, if $S \xrightarrow{f} N$ is any surface mapped into a hyperbolic manifold N of finite volume such that a non-peripheral simple closed curve γ in S is homotopic to a cusp of N, one can modify f in a small neighborhood of γ to wrap this annulus a number of times around the cusp. This is likely to give infinitely many homotopy classes of surfaces in N.



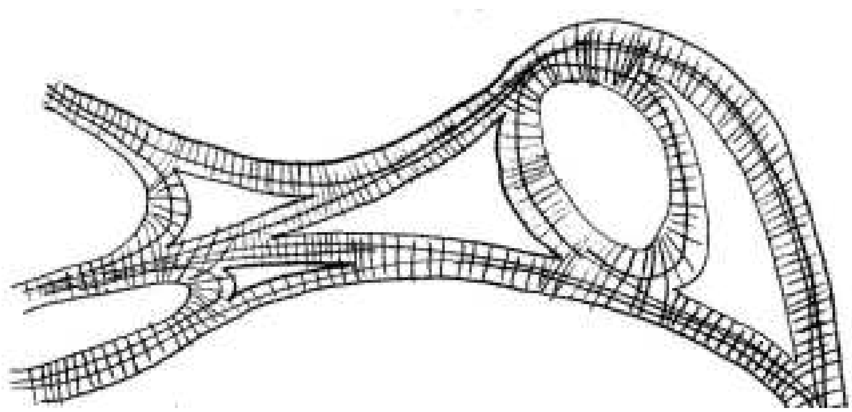
In place of 8.8.5, there is a compactness statement in the topology of geometric convergence provided each component of $S_{[\epsilon,\infty)}$ is required to intersect K. One would allow S to converge to a surface where a simple closed geodesic is pinched to yield a pair of cusps. From this, one deduces that there are finitely many classes of groups G

isomorphic to S up to the operations of conjugacy, and wrapping a surface carrying G around cusps.

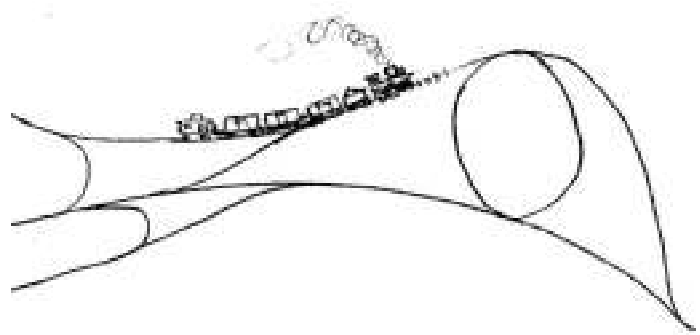
Haken proved a finiteness statement analogous to 8.8.6 for embedded incompressible surfaces in atoroidal Haken manifolds. 8.51

8.9. The structure of geodesic laminations: train tracks

Since a geodesic lamination γ on a hyperbolic surface S has measure zero, one can picture γ as consisting of many parallel strands in thin, branching corridors of S which have small total area.

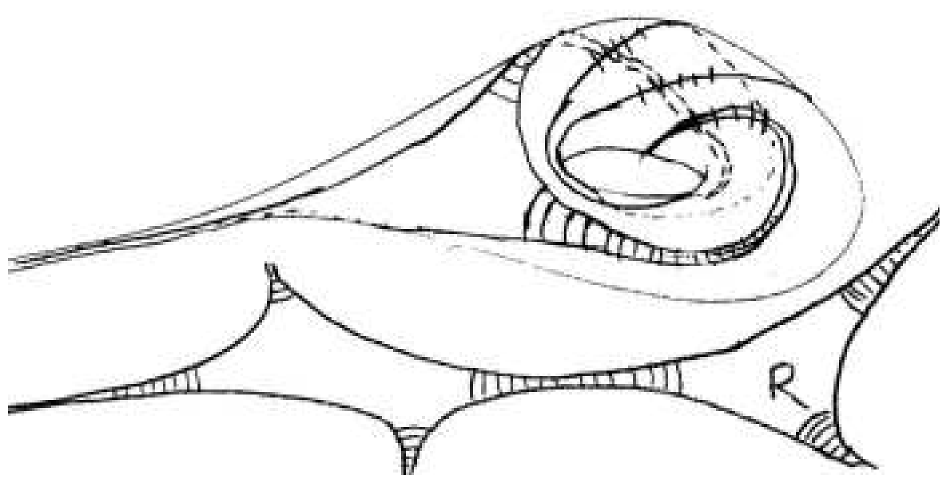


Imagine squeezing the nearly parallel strands of γ in each corridor to a single strand. One obtains a train track τ (with switches) which approximates γ. Each leaf of γ may be imagined as the path of a train running around along τ.



8.52

Here is a construction which gives a precise and nice sequence of train track approximations of γ. Consider a complementary region R in $S-\gamma$. The double dR is a hyperbolic surface of finite area, so $(dR)_{(0,2\epsilon]}$ has a simple structure: it consists of neighborhoods of geodesics shorter than 2ϵ and of cusps. In each such neighborhood there is a canonical foliation by curves of constant curvature: horocycles about a cusp or equidistant curves about a short geodesic. Transfer this foliation to R, and then to S. This yields a foliation $\mathcal{F}$ in the subset of S where leaves of γ are not farther than 2ϵ apart. (A local vector field tangent to $\mathcal{F}$ is Lipschitz, so it is integrable; this is why $\mathcal{F}$ exists. If γ has no leaves tending toward a cusp, then we can make all the leaves of $\mathcal{F}$ be arbitrarily short arcs by making ϵ sufficiently small. If γ has leaves tending toward a cusp, then there can be only finitely many such leaves, since there is an upper bound to the total number of cusps of the complementary regions. Erase all parts of $\mathcal{F}$ in a cusp of a region tending toward a cusp of S; again, when ϵ is sufficiently small all leaves of $\mathcal{F}$ will be short arcs. The space obtained by collapsing all arcs of $\mathcal{F}$ to a point is a surface S' homeomorphic to S, and the image of γ is a train track τ_ϵ on S'. Observe that each switch of τ_ϵ comes from a boundary component of some $dR_{(0,2\epsilon]}$. In particular, there is a uniform bound to the number of switches. From this it is easy to see that there are only finitely many possible types of τ_ϵ, up to homeomorphisms of S' (not necessarily homotopic to the identity). 8.53



In working with actual geodesic laminations, it is better to use more arbitrary train track approximations, and simply sketch pictures; the train tracks are analogous to decimal approximations of real numbers.

Here is a definition of a useful class of train tracks.

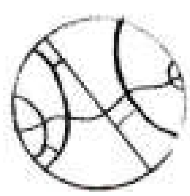
Definitions 8.9.1. A *train track* on a differentiable surface S is an embedded graph τ on S. The edges (branch lines) of τ must be C^1, and all edges at a given vertex (switch) must be tangent. If S has "cusps", τ may have open edges tending toward the cusps. Dead ends are not permitted. (Each vertex v must be in the interior of a C^1 interval on τ through v.) Furthermore, for each component R of $S - \tau$, the double dR of R along the interiors of edges of ∂R must have negative Euler characteristic. A lamination γ on S is *carried* by τ if there is a differentiable map $f : S \to S$ homotopic to the identity taking γ to τ and non-singular on the tangent spaces of the leaves of γ. (In other words, the leaves of γ are trains running around on τ.) The lamination γ is *compatible* with τ if τ can be enlarged to a train track τ' which carries γ. 8.54

Proposition 8.9.2. *Let S be a hyperbolic surface, and let $\delta > 0$ be arbitrary. There is some $\epsilon > 0$ such that for all geodesic laminations γ of S, the train track approximation τ_ϵ can be realized on S in such a way that all branch lines τ_ϵ are C^2 curves with curvature $< \delta$.*

Proof. Note first that by making ϵ sufficiently small, one can make the leaves of the foliation $\mathcal{F}$ very short, uniformly for all γ: otherwise there would be a sequence of γ's converging to a geodesic lamination containing an open set. [One can also see this directly from area considerations.] When all branches of τ_ϵ are reasonably long, one can simply choose the tangent vectors to the switches to be tangent to any geodesic of γ where it crosses the corresponding leaf of $\mathcal{F}$; the branches can be filled in by curves of small curvature. When some of the branch lines are short, group each set of switches connected by very short branch lines together. First map each of these sets into S, then extend over the reasonably long branches. □

Corollary 8.9.3. *Every geodesic lamination which is carried by a close train track approximation τ_ϵ to a geodesic lamination γ has all leaves close to leaves of γ.*

Proof. This follows from the elementary geometrical fact that a curve in hyperbolic space with uniformly small curvature is uniformly close to a unique geodesic. (One way to see this is by considering the planes perpendicular to the curve—they always advance at a uniform rate, so in particular the curve crosses each one only once.)



8.55

□

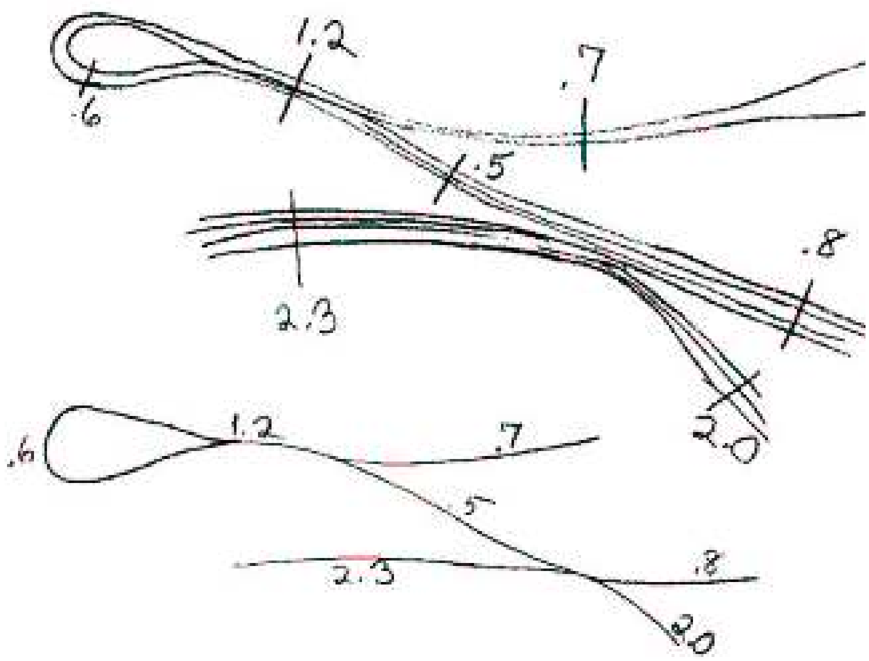
Proposition 8.9.4. *A lamination λ of a surface S is isotopic to a geodesic lamination if and only if*

(a) *λ is carried by some train track τ, and*
(b) *no two leaves of λ take the same (bi-infinite) path on τ.*

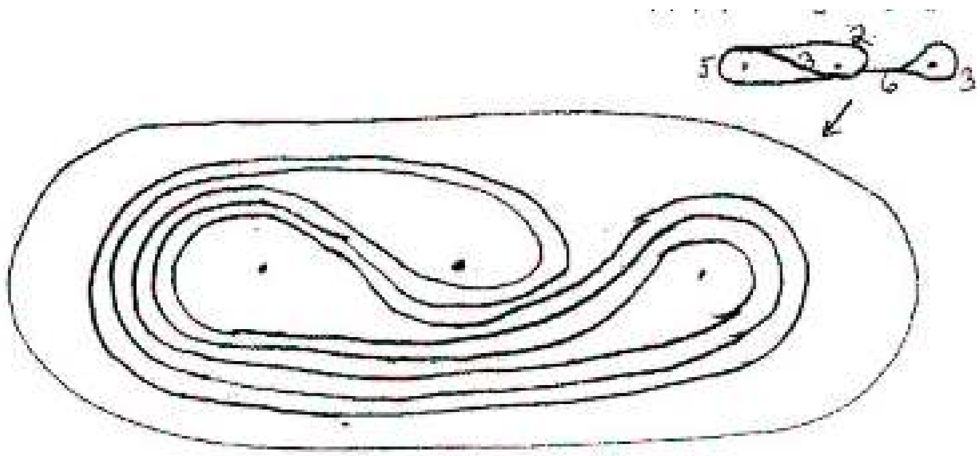
PROOF. Given an arbitrary train track τ, it is easy to construct *some* hyperbolic structure for S on which τ is realized by lines with small curvature. The leaves of λ then correspond to a set of geodesics on S, near τ. These geodesics do not cross, since the leaves of λ do not. Condition (b) means that distinct leaves of λ determine distinct geodesics. When leaves of λ are close, they must follow the same path for a long finite interval, which implies the corresponding geodesics are close. Thus, we obtain a geodesic lamination γ which is isotopic to λ. (To have an isotopy, it suffices to construct a homeomorphism homotopic to the identity. This homeomorphism is constructed first in a neighborhood of τ, then on the rest of S.) $\square$

REMARK. From this, one sees that as the hyperbolic structure on S varies, the corresponding geodesic laminations are all isotopic. This issue was quietly skirted in §8.5.

When a lamination λ has an invariant measure μ, this gives a way to associate a number $\mu(b)$ to each branch line b of any train track which dominates γ: $\mu(b)$ is just 8.56
the transverse measure of the leaves of λ collapsed to a point on b. At a switch, the sum of the "entering" numbers equals the sum of the "exiting" numbers.



Conversely, any assignment of numbers satisfying the switch condition determines a unique geodesic lamination with transverse measure: first widen each branch line b of τ to a corridor of constant width $\mu(b)$, and give it a foliation $\mathcal{G}$ by equally spaced lines.



8.57

As in 8.9.4, this determines a lamination γ; possibly there are many leaves of $\mathcal{G}$ collapsed to a single leaf of γ, if these leaves of $\mathcal{G}$ all have the same infinite path. $\mathcal{G}$ has a transverse measure, defined by the distance between leaves; this goes over to a transverse measure for γ.

8.10. Realizing laminations in three-manifolds

For a quasi-Fuchsian group Γ, it was relatively easy to "realize" a geodesic lamination of the corresponding surface in M_Γ, by using the circle at infinity. However, not every complete hyperbolic three-manifold whose fundamental group is isomorphic to a surface group is quasi-Fuchsian, so we must make a more systematic study of realizability of geodesic laminations.

Definition 8.10.1. Let $f : S \to N$ be a map of a hyperbolic surface to a hyperbolic three-manifold which sends cusps to cusps. A geodesic lamination γ on S is *realizable* in the homotopy class of f if f is homotopic (by a cusp-preserving homotopy) to a map sending each leaf of γ to a geodesic.

Proposition 8.10.2. *If γ is realizable in the homotopy class of f, the realization is* (*essentially*) *unique: that is, the image of each leaf of γ is uniquely determined.*

Proof. Consider a lift $\tilde{h}$ of a homotopy connecting two maps of S into N. If S is closed, $\tilde{h}$ moves every point a bounded distance, so it can't move a geodesic to a different geodesic. If S has cusps, the homotopy can be modified near the cusps of S so $\tilde{h}$ again is bounded. □

8.58

In Section 8.5, we touched on the notion of geometric convergence of geodesic laminations. The *geometric topology* on geodesic laminations is the topology of geometric convergence, that is, a neighborhood of γ consists of laminations γ' which have leaves near every point of γ, and nearly parallel to the leaves of γ. If γ_1 and γ_2 are disjoint simple closed curves, then $\gamma_1 \cup \gamma_2$ is in every neighborhood of γ_1 as well as in every neighborhood of γ_2. The space of geodesic laminations on S with the geometric topology we shall denote $\mathcal{GL}$. The geodesic laminations compatible with train track approximations of γ give a neighborhood basis for γ.

The *measure topology* on geodesic laminations with transverse measures (of full support) is the topology induced from the weak topology on measures in the Möbius band J outside S_∞ in the Klein model. That is, a neighborhood of (γ, μ) consists of (γ', μ') such that for a finite set $f_1, \dots, f_k$ of continuous functions with compact support in J,

$$\left| \int f_i \, d\mu - \int f_i \, d\mu' \right| < \epsilon.$$

This can also be interpreted in terms of integrating finitely many continuous functions on finitely many *transverse* arcs. Let $\mathcal{ML}(S)$ be the space of (γ, μ) on S with the measure topology. Let $\mathcal{PL}(S)$ be $\mathcal{ML}(S)$ modulo the relation $(\gamma, \mu) \sim (\gamma, a\mu)$ where $a > 0$ is a real number.

PROPOSITION 8.10.3. *The natural map $\mathcal{ML} \to \mathcal{GL}$ is continuous.*

8.59

PROPOSITION 8.10.4. *The map $w : \mathcal{U}(S, N) \to \mathcal{GL}(S)$ which assigns to each uncrumpled surface its wrinkling locus is continuous.*

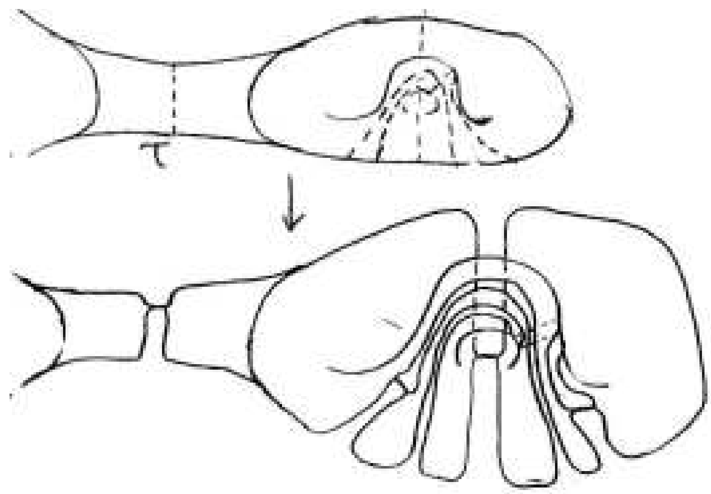
PROOF OF 8.10.3. For any point x in the support of a measure μ and any neighborhood UU of x, the support of a measure close enough to μ must intersect U. $\square$

PROOF OF 8.10.4. An interval which is bent cannot suddenly become straight. Away from any cusps, there is a positive infimum to the "amount" of bending of an interval of length ϵ which intersects the wrinkling locus $w(S)$ in its middle third, and makes an angle of at least ϵ with $w(S)$. (The "amount" of bending can be measure, say, by the different between the length of α and the distance between the image endpoints.) All such arcs must still cross $w(S')$ for any nearby uncrumpled surface S'. $\square$

When S has cusps, we are also interested in measures supported on compact geodesic laminations. We denote this space by $\mathcal{ML}_0(S)$. If (τ, μ) is a train track *description* for (γ, μ), *where $\mu(b) \neq 0$ for any branch of τ*, then neighborhoods for (γ, μ) are described by $\big\{(\tau', \mu')\big\}$, where $\tau \subset \tau'$ and $\big|\mu(b) - \mu'(b)\big| < \epsilon$. (If b is a branch of τ' not in τ, then $\mu(b) = 0$ by definition.)

In fact, one can always choose a hyperbolic structure on S so that τ is a good *approximation* to γ. If S is closed, it is always possible to squeeze branches of τ

together along non-trivial arcs in the complementary regions to obtain a new train track which cannot be enlarged. 8.60



This implies that a neighborhood of (γ,μ) is parametrized by a finite number of real parameters. Thus, $\mathcal{ML}(S)$ is a manifold. Similarly, when S has cusps, $\mathcal{ML}(S)$ is a manifold with boundary $\mathcal{ML}_0(S)$.

PROPOSITION 8.10.5. *$\mathcal{GL}(S)$ is compact, and $\mathcal{PL}(S)$ is a compact manifold with boundary $\mathcal{PL}_0(S)$ if S is not compact.*

PROOF. There is a finite set of train tracks $\tau_1,\dots,\tau_k$ carrying every possible geodesic lamination. (There is an upper bound to the length of a compact branch of τ_ϵ, when S and ϵ are fixed.) The set of projective classes of measures on any particular τ is obviously compact, so this implies $\mathcal{PL}(S)$ is compact. That $\mathcal{PL}(S)$ is a manifold follows from the preceding remarks. Later we shall see that in fact it is the simplest of possible manifolds.

In 8.5, we indicated one proof of the compactness of $\mathcal{GL}(S)$. Another proof goes as follows. First, note that 8.61

PROPOSITION 8.10.6. *Every geodesic lamination γ admits some transverse measure μ (possibly with smaller support).*

PROOF. Choose a finite set of transversals $\alpha_1,\dots,\alpha_k$ which meet every leaf of γ. Suppose there is a sequence $\{l_i\}$ of intervals on leaves of γ such that the total number N_1 of intersection of l_i with the α_j's goes to infinity. Let μ_i be the measure on $\bigcup \alpha_j$ which is $1/N_i$ times the counting measure on $l_i \cap \alpha_j$. The sequence $\{\mu_i\}$ has a subsequence converging (in the weak topology) to a measure μ. It is easy to see that μ is invariant under local projections along leaves of γ, so that it determines a transverse measure.

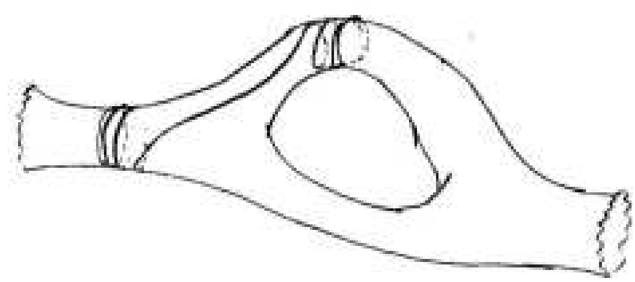
If there is no such sequence $\{l_i\}$ then every leaf is proper, so the counting measure for any leaf will do. $\square$

We continue with the proof of 8.10.5. Because of the preceding result, the image $\mathcal{I}$ of $\mathcal{PL}(S)$ in $\mathcal{GL}(S)$ intersects the closure of every point of $\mathcal{GL}(S)$. Any collection of open sets which covers $\mathcal{GL}(S)$ has a finite subcollection which covers the compact set $\mathcal{I}$; therefore, it covers all of $\mathcal{GL}(S)$. $\square$

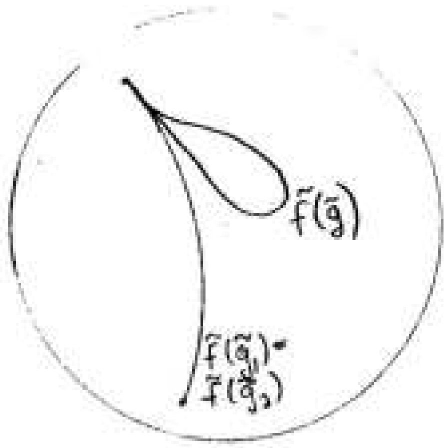
Armed with topology, we return to the question of realizing geodesic laminations. Let $\mathcal{R}_f \subset \mathcal{GL}(S)$ consist of the laminations realizable in the homotopy class of f.

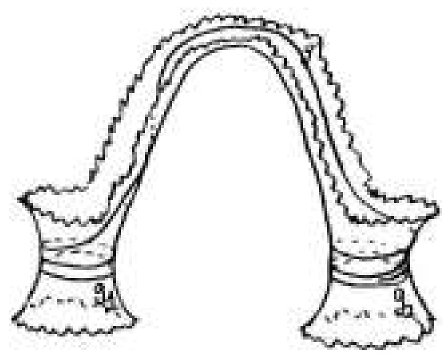
First, if γ consists of finitely many simple closed geodesics, then γ is realizable 8.62
provided $\pi_1(f)$ maps each of these simple closed curves to non-trivial, non-parabolic elements.

If we add finitely many geodesics whose ends spiral around these closed geodesics or converge toward cusps the resulting lamination is also realizable except in the degenerate case that f restricted to an appropriate non-trivial pair of pants on S factors through a map to S^1.



To see this, consider for instance the case of a geodesic g on S whose ends spiral around closed geodesics g_1 and g_2. Lifting f to H^3, we see that the two ends of $\tilde{f}(\tilde{g})$ are asymptotic to geodesics $\tilde{f}(\tilde{g}_1)$ and $\tilde{f}(\tilde{g}_2)$. Then f is homotopic to a map taking g to a geodesic unless $\tilde{f}(\tilde{g}_1)$ and $\tilde{f}(\tilde{g}_2)$ converge to the same point on S_∞, which can only happen if $\tilde{f}(\tilde{g}_1) = \tilde{f}(\tilde{g}_2)$ (by 5.3.2). In this case, f is homotopic to a map taking a neighborhood of $g \cup g_1 \cup g_2$ to $f(g_1) = f(g_2)$. 8.63





The situation is similar when the ends of g tend toward cusps.

These realizations of laminations with finitely many leaves take on significance in view of the next result:

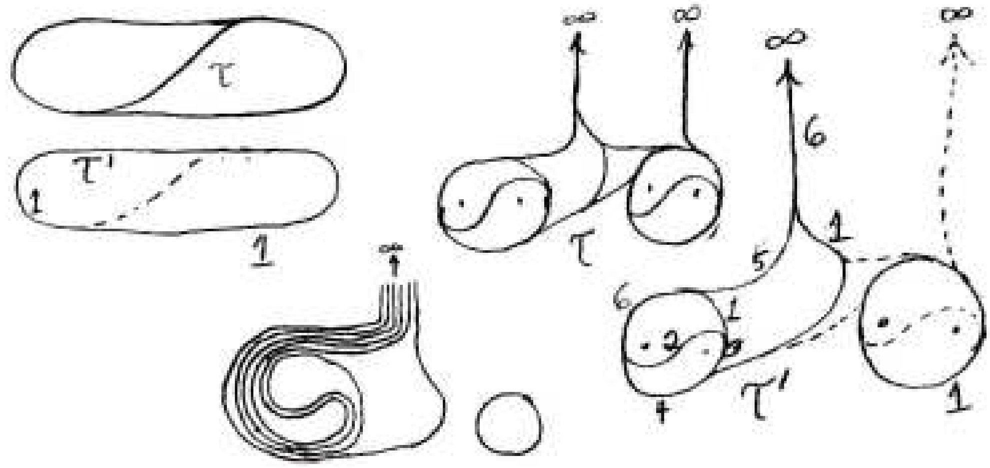
Proposition 8.10.7. (a) *Measures supported on finitely many compact or proper geodesics are dense in $\mathcal{ML}$.*

(b) *Geodesic laminations with finitely many leaves are dense in $\mathcal{GL}$.*

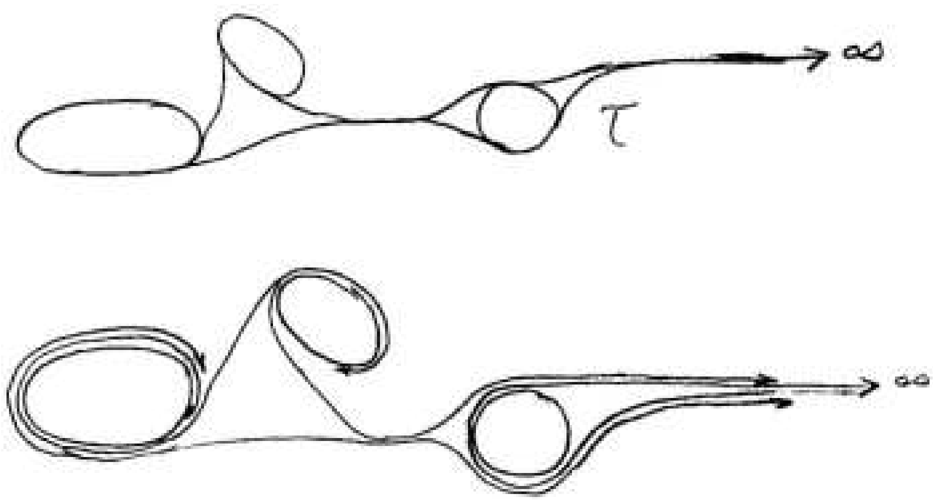
(c) *Each end of a non-compact leaf of a geodesic lamination with only finitely many leaves spirals around some closed geodesic, or tends toward a cusp.*

Proof. If τ is any train track and μ is any measure which is positive on each branch, μ can be approximated by measures μ' which are rational on each branch, since μ is subject only to linear equations with integer coefficients. μ' gives rise to geodesic laminations with only finitely many leaves, all compact or proper. This 8.64
proves (a).

If γ is an arbitrary geodesic lamination, let τ be a close train track approximation of γ and proceed as follows. Let $\tau' \subset \tau$ consist of all branches b of τ such that there exists either a cyclic (repeating) train route or a proper train route through b.

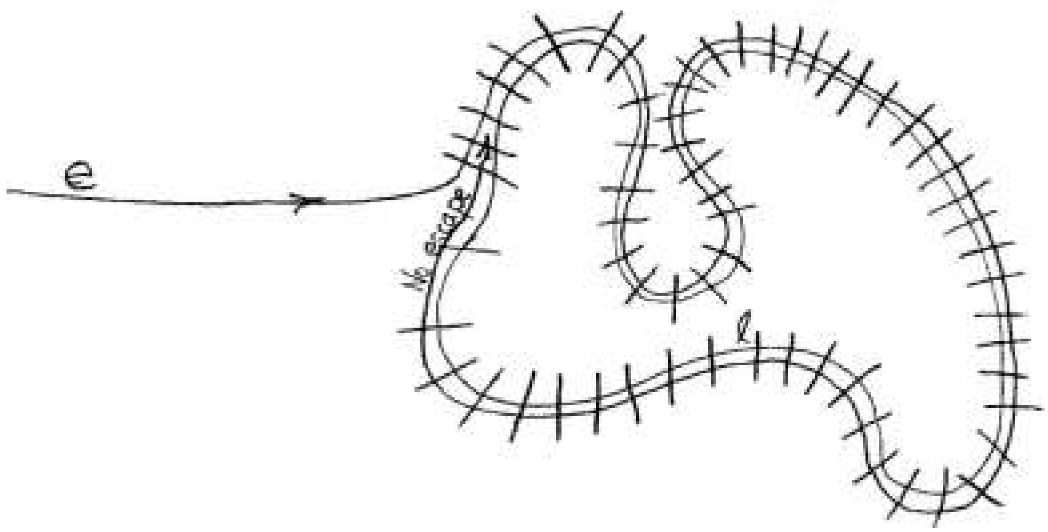


(The reader experienced with toy trains is aware of the subtlety of this question.) There is a measure supported on τ', obtained by choosing a finite set of cyclic and proper paths covering τ' and assigning to a branch b the total number of times these paths traverse. Thus there is a lamination λ' consisting of finitely many compact or proper leaves supported in a narrow corridor about τ'. Now let b be any branch of $\tau - \tau'$. A train starting on b can continue indefinitely, so it must eventually come to τ', in each direction. Add a leaf to λ' representing a shortest path from b to τ' in each direction; if the two ends meet, make them run along side by side (to avoid 8.65
crossings). When the ends approach τ, make them "merge"—either spiral around a closed leaf, or follow along close to a proper leaf. Continue inductively in this way, adding leaves one by one until you obtain a lamination λ dominated by τ and filling out all the branches. This proves (b).



If γ is any geodesic lamination with finitely many (or even countably many) leaves, then the only possible minimal sets are closed leaves; thus each end e of a non-compact must either be a proper end or come arbitrarily near some compact leaf l. By tracing the leaves near l once around l, it is easy to see that this means e spirals around l. $\square$

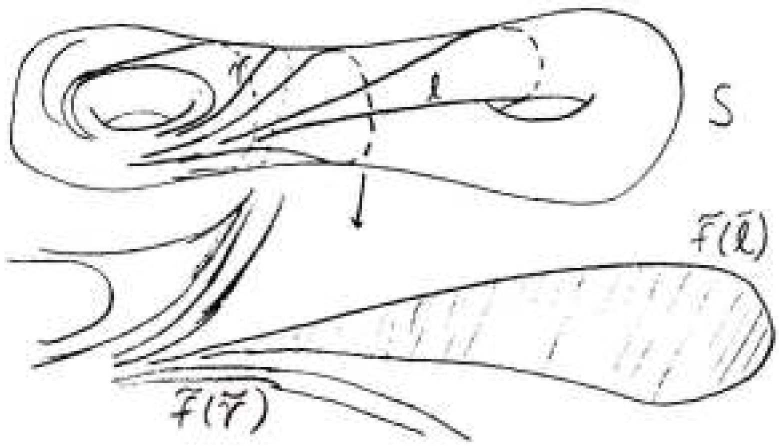
8.66



Thus, if f is non-degenerate, $\mathcal{R}_f$ is dense. Furthermore,

THEOREM 8.10.8. *If* $\pi_1 f$ *is injective, and* f *satisfies* (np) (*that is, if* $\pi_1 f$ *preserves non-parabolicity*), *then* $\mathcal{R}_f$ *is an open dense subset of* $\mathcal{GL}(S)$.

PROOF. If γ is any complete geodesic lamination which is realizable, then a train track approximation τ can be constructed for the image of γ in N^3, in such a way that all branch lines have curvature close to 0. Then all laminations carried by τ are also realizable; they form a neighborhood of γ. Next we will show that any enlargement $\gamma' \supset \gamma$ of a realizable geodesic lamination γ is also realizable. First note that if γ' is obtained by adding a single leaf l to γ, then γ' is also realizable. This is proved in the same way as in the case of a lamination with finitely many leaves: note that each end of l is asymptotic to a leaf of γ. (You can see this by considering $S - \gamma$.) If f is 8.67
homotoped so that $f(\gamma)$ consists of geodesics, then both ends of $\tilde{f}(l)$ are asymptotic to geodesics in $\tilde{f}(\gamma)$. If the two endpoints were not distinct on S_∞, this would imply the existence of some non-trivial identification of γ by f so that $\pi_1 f$ could not be injective.



By adding finitely many leaves to any geodesic lamination γ' we can complete it. This implies that γ' is contained in the wrinkling locus of some uncrumpled surface. By 8.8.5 and 8.10.1, the set of uncrumpled surfaces whose wrinkling locus contains γ is compact. Since the wrinkling locus depends continuously on an uncrumpled surface, the set of $\gamma' \in \mathcal{R}_f$ which contains γ is compact. But any $\gamma' \supset \gamma$ can be approximated by laminations such that $\gamma' - \gamma$ consists of a finite number of leaves. This actually follows from 8.10.7, applied to $d(S - \gamma)$. Therefore, every enlargement $\gamma' \supset \gamma$ is in $\mathcal{R}_f$. 8.68

Since the set of uncrumpled surfaces whose wrinkling locus contains γ is compact, there is a finite set of train tracks $\tau_1, \ldots, \tau_k$ such that for any such surface, $w(S)$ is closely approximated by one of $\tau_1, \ldots, \tau_k$. The set of all laminations carried by at least one of the τ_i is a neighborhood of γ contained in $\mathcal{R}_f$. $\square$

Corollary 8.10.9. *Let Γ be a geometrically finite group, and let $f : S \to N_\Gamma$ be a map as in* 8.10.8. *Then either $\mathcal{R}_f = \mathcal{GL}(S)$ (that is, all geodesic laminations are realizable in the homotopy class of f), or Γ has a subgroup Γ' of finite index such that $N_{\Gamma'}$ is a three-manifold with finite volume which fibers over the circle.*

Conjecture 8.10.10. If $f : S \to N$ is any map from a hyperbolic surface to a complete hyperbolic three-manifold taking cusps to cusps, then the image $\pi_1(f)(\pi_1(S))$ is quasi-Fuchsian if and only if $\mathcal{R}_f = \mathcal{GL}(S)$.

Proof of 8.10.9. Under the hypotheses, the set of uncrumpled surfaces homotopic to $f(S)$ is compact. If each such surface has an essentially unique homotopy to $f(S)$, so that the wrinkling locus on S is well-defined, then the set of wrinkling loci of uncrumpled surfaces homotopic to f is compact, so by 8.10.8 it is all of $\mathcal{GL}(S)$.

Otherwise, there is some non-trivial $h_t : S \to M$ such that $h_1 = h_0 \circ \phi$, where $\phi : S \to S$ is a homotopically non-trivial diffeomorphism. It may happen that ϕ has finite order up to isotopy, as when S is a finite regular covering of another surface in M. The set of all isotopy classes of diffeomorphisms ϕ which arise in this way form a group. If the group is finite, then as in the previous case, $\mathcal{R}_F = \mathcal{GL}(S)$. Otherwise, there is a torsion-free subgroup of finite index (see), so there is an element ϕ of infinite order. The maps f and $\phi \circ f$ are conjugate in Γ, by some element $\beta \in \Gamma$. The group generated by β and $f(\pi_1 S)$ is the fundamental group of a three-manifold which fibers over S^1. $\square$ 8.69

We shall see some justification for the conjecture in the remaining sections of chapter 8 and in chapter 9: we will prove it under certain circumstances.

8.11. The structure of cusps

Consider a hyperbolic manifold N which admits a map $f : S \to N$, taking cusps to cusps such that $\pi_1(f)$ is an isomorphism, where S is a hyperbolic surface. Let $B \subset N$ be the union of the components of $N_{(0,\epsilon]}$ corresponding to cusps of S. f is a relative homotopy equivalence from $(S, S_{(0,\epsilon)})$ to (N, B), so there is a homotopy inverse $g : (N, B) \to (S, S_{(0,\epsilon)})$. If $X \in S_{(\epsilon,\infty)}$ is a regular value for g, then $g^{-1}(x)$ is a one-manifold having intersection number one with $f(S)$, so it has at least one component homeomorphic to R, going out toward infinity in $N - B$ on opposite sides of $f(S)$. Therefore there is a proper function $h : (N - B) \to \mathbb{R}$ with h restricted to $g^{-1}(x)$ a surjective map. One can modify h so that $h^{-1}(0)$ is an incompressible surface. Since g restricted to $h^{-1}(0)$ is a degree one map to S, it must map the fundamental group surjectively as well as injectively, so $h^{-1}(0)$ is homeomorphic to S. $h^{-1}(0)$ divides $N - B$ into two components N_+ and N_- with $\pi_1 N = \pi_1 N_+ = \pi_1 N_- = \pi_1 S$. We can assume that $h^{-1}(0)$ does not intersect $N_{(0,\epsilon]}$ except in B (say, by shrinking ϵ).

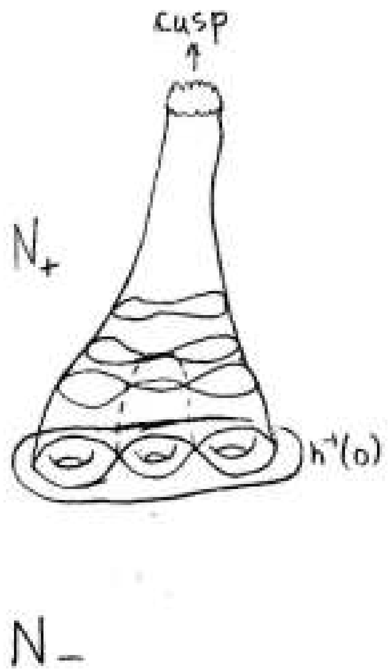
Suppose that N has parabolic elements that are not parabolic on S. The structure 8.71
of the extra cusps of N is described by the following:

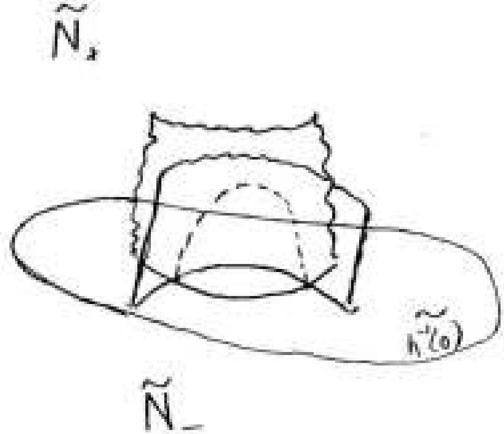
Proposition 8.11.1. *There are geodesic laminations γ_+ and γ_- on S with all leaves compact (i.e., they are finite collections of disjoint simple closed curves) such that the extra cusps in N_e correspond one-to-one with leaves of $\gamma_e (e = +, -)$. In particular, for any element $\alpha \in \pi_1(S)$, $\pi_1(f)(\alpha)$ is parabolic if and only if α is freely homotopic to a cusp of S or to a simple closed curve in γ_+ or γ_-.*

Proof. We need consider only one half, say N_+. For each extra cusp of N_+, there is a half-open cylinder mapped into N_+, with one end on $h^{-1}(0)$ and the other end tending toward the cusp. Furthermore, we can assume that the union of these cylinders is embedded outside a compact set, since we understand the picture in a neighborhood of the cusps. Homotope the ends of the cylinders which lie on $h^{-1}(0)$

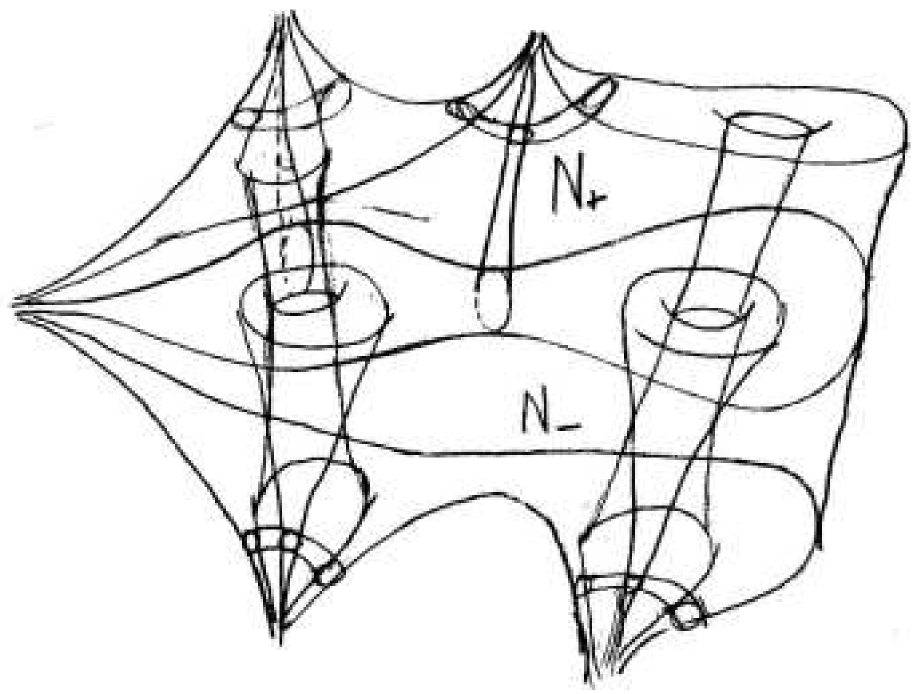
so they are geodesics in some hyperbolic structure on $h^{-1}(0)$. One can assume the cylinders are immersed (since maps of surfaces into three-manifolds are appoximable by immersions) and that they are transverse to themselves and to one another. If there are any self-intersections of the cylinders on $h^{-1}(0)$, there must be a double line which begins and ends on $h^{-1}(0)$. Consider the picture in $\tilde{N}$: there are two translates of universal covers of cylinders which meet in a double line, so that in particular their bounding lines meet twice on $h^{-1}(0)$. This contradicts the fact that they are geodesics in some hyperbolic structure. □

8.71





It actually follows that the collection of cylinders joining simple closed curves to the cusps can be embedded: we can modify g so that it takes each of the extra cusps to a neighborhood of the appropriate simple closed curve $\alpha \subset \gamma_\epsilon$, and then do surgery to make $g^{-1}(\alpha)$ incompressible.



8.72

To study N, we can replace S by various surfaces obtained by cutting S along curves in γ_+ or γ_-. Let P be the union of open horoball neighborhoods of all the cusps of N. Let $\{S_i\}$ be the set of all components of S cut by γ_+ together with those of S cut by γ_-. The union of the S_i can be embedded in $N - P$, with boundary on ∂P, within the convex hull M of N, so that they cut off a compact piece $N_0 \subset N - P$ homotopy equivalent to N, and non-compact ends E_i of $N - P$, with $\partial E_i \subset P \cup S_i$.

Let N now be an arbitrary hyperbolic manifold, and let P be the union of open horoball neighborhoods of its cusps. The picture of the structure of the cusps readily generalizes provided $N - P$ is homotopy equivalent to a compact submanifold N_O, obtained by cutting $N - P$ along finitely many incompressible surfaces $\{S_i\}$ with boundary ∂P.

Applying 8.11.1 to covering spaces of N corresponding to the S_i (or applying its proof directly), one can modify the S_i until no non-peripheral element of one of the S_i is homotopic, outside N_O, to a cusp. When this is done, the ends $\{E_i\}$ of $N - P$ are in one-to-one correspondence with the S_i.

According to a theorem of Peter Scott, every three-manifold with finitely generated fundamental group is homotopy equivalent to a compact submanifold. In general, such a submanifold will not have incompressible boundary, so it is not as well behaved. We will leave this case for future consideration. 8.73

Definition 8.11.2. Let N be a complete hyperbolic manifold, P the union of open horoball neighborhoods of its cusps, and M the convex hull of N. Suppose

E is an end of $N - P$, with $\partial E - \partial P$ an incompressible surface $S \subset M$ homotopy equivalent to E. Then E is a *geometrically tame end* if either

(a) $E \cap M$ is compact, or
(b) the set of uncrumpled surfaces S' homotopic to S and with $S'_{[\epsilon,\infty)}$ contained in E is not compact.

If N has a compact submanifold N_O of $N - P$ homotopy equivalent to N such that $N - P - N_O$ is a disjoint union of geometrically tame ends, then N and $\pi_1 N$ are *geometrically tame*. (These definitions will be extended in §). We shall justify this definition by showing geometric tameness implies that N is analytically, topologically and geometrically well-behaved.

8.12. Harmonic functions and ergodicity

Let N be a complete Riemannian manifold, and h a positive function on N. Let ϕ_t be the flow generated by $-(\operatorname{grad} h)$. The integral of the velocity of ϕ_t is bounded along any flow line:

$$\begin{aligned}\int_x^{\phi_T(x)} \|\operatorname{grad} h\|\, ds &= h(x) - h(\phi_T(x)) \\ &\le h(x) \quad (\text{ for } T > 0).\end{aligned}$$

If A is a subset of a flow line $\{\phi_t(x)\}_{t\ge 0}$ of finite length $l(A)$, then by the Schwarz inequality 8.74

Labelled this 8.12.1.eq

8.12.1. $$T(A) = \int_A \frac{1}{\|\operatorname{grad} h\|}\, ds \ge \frac{l(A)^2}{\int_A \|\operatorname{grad} h\|\, ds} \ge \frac{l(A)^2}{h(x)}$$

where $T(A)$ is the total time the flow line spends in A. Note in particular that this implies $\phi_t(x)$ is defined for all positive time t (although ϕ_t may not be surjective). The flow lines of ϕ_t are moving very slowly for most of their length. If h is harmonic, then the flow ϕ_t preserves volume: this means that if it is not altogether stagnant, it must flow along a channel that grows very wide. A river, with elevation h, is a good image. It is scaled so $\operatorname{grad} h$ is small.

Suppose that N is a hyperbolic manifold, and $S \xrightarrow{f} N$ is an uncrumpled surface in N, so that it has area $-2\pi\chi(S)$. Let a be a fixed constant, suppose also that S has no loops of length $\le a$ which are null-homotopic in N.

PROPOSITION 8.12.1. *There is a constant C depending only on a such that the volume of $\mathcal{N}_1(f(S))$ is not greater than $-C \cdot \chi(S)$. ($\mathcal{N}_1$ denotes the neighborhood of radius* 1.)

PROOF. For each point $x \in S$, let c_x be the "characteristic function" of an immersed hyperbolic ball of radius $1 + a/2$ centered at $f(x)$. In other words, $c_x(y)$ is the number of distinct homotopy classes of paths from x to y of length $\leq 1 + a/2$. Let g be defined by integrating c_x over S; in other words, for $y \in N$, 8.75

$$g(y) = \int_S c_x(y)\, dA.$$

If $v(B_r)$ is the volume of a ball of radius r in H^3, then

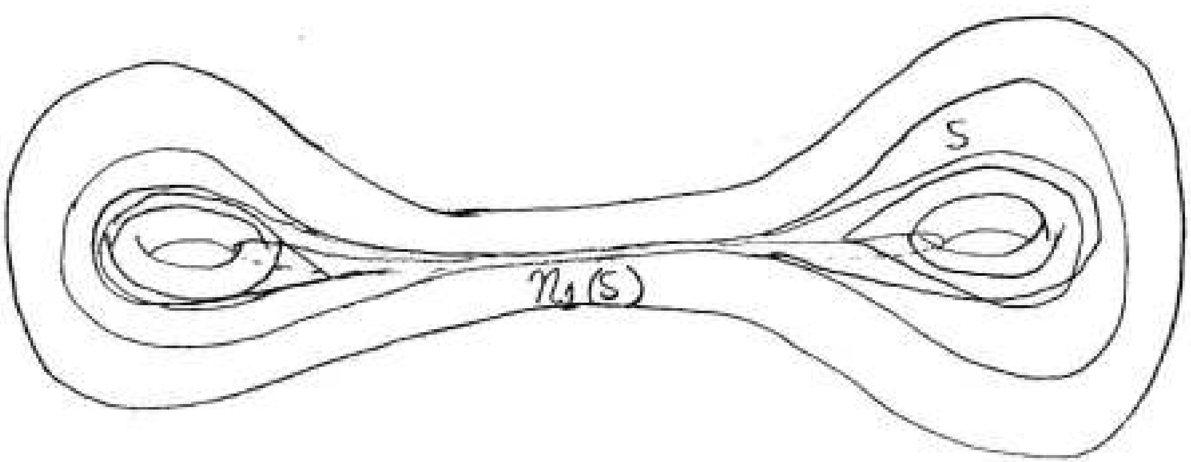
$$\int_N g\, dV = -2\pi\chi(S)\, v(B_{1+a/2}).$$

For each point $y \in \mathcal{N}_1(f(S))$, there is a point x with $d(fx, y) \leq 1$, so that there is a contribution to $g(y)$ for every point z on S with $d(z, y) \leq a/2$, and for each homotopy class of paths on S between z and x of length $\leq a/2$. Thus $g(y)$ is at least as great as the area $A(B_{a/2})$ of a ball in H^2 of radius $a/2$, so that

$$v(\mathcal{N}_1(f(S))) \leq \frac{1}{A(B_{a/2})} \int_N g\, dV \leq -C \cdot \chi(S).$$

□

As $a \to 0$, the best constant C goes to ∞, since one can construct uncrumpled surfaces with long thin waists, whose neighborhoods have very large volume.



8.76

THEOREM 8.12.3. *If N is geometrically tame, then for every non-constant positive harmonic function h on its convex hull M,*

$$\inf_M h = \inf_{\partial M} h.$$

This inequality still holds if h is only a positive superharmonic function, i.e., if $\Delta h = \operatorname{div} \operatorname{grad} h \leq 0$.

COROLLARY 8.12.4. *If $\Gamma = \pi_1 N$, where N is geometrically tame, then L_Γ has measure* 0 *or* 1. *In the latter case, Γ acts ergodically on S^2.*

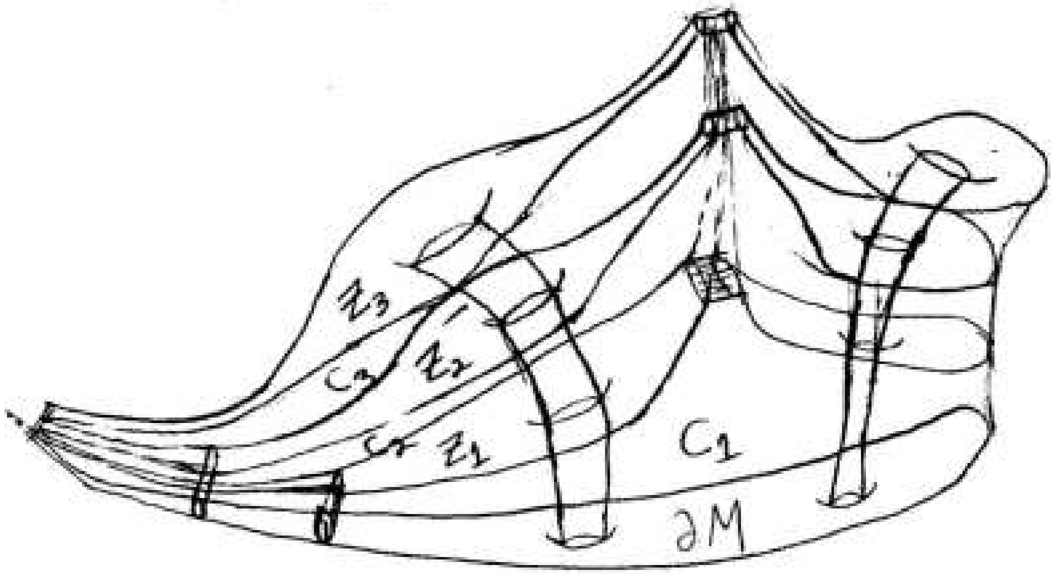
PROOF OF COROLLARY FROM THEOREM. This is similar to 8.4.2. Consider any invariant measurable set $A \subset L_\Gamma$, and let h be the harmonic extension of the characteristic function of A. Since A is invariant, h defines a harmonic function, also h, on N. If $L_\Gamma = S^2$, then by 8.12.3 h is constant, so A has measure 0 to 1. If $L_\Gamma \neq S^2$ then the infimum of $(1 - h)$ is the infimum on ∂M, so it is $\geq \frac{1}{2}$. This implies A has measure 0. This completes the proof of 8.12.4. ☐

Theorem 8.12.3 also implies that when $L_\Gamma = S^2$, the geodesic flow for N is ergodic. We shall give this proof in § , since the ergodicity of the geodesic flow is useful for the proof of Mostow's theorem and generalizations.

PROOF OF 8.12.3. The idea is that all the uncrumpled surfaces in M are narrows, which allow a high flow rate only at high velocities. In view of 8.12.1, most of 8.77
the water is forced off M—in other words, ∂M is low.

Let P be the union of horoball neighborhoods of the cusps of N, and $\{S_i\}$ incompressible surfaces cutting $N - P$ into a compact piece N_0 and ends $\{E_i\}$. Observe that each component of P has two boundary components of $\cup S_i$. In each end E_i which does not have a compact intersection with M, there is a sequence of uncrumpled maps $f_{i,j} : S_i \to E_i \cup P$ moving out of all compact sets in $E_i \cup P$, by 8.8.5. Combine these maps into one sequence of maps $f_j : \cup S_i \to M$. Note that f_j maps $\sum [S_i]$ to a cycle which bounds a (unique) chain C_j of finite volume, and that the supports of the C_j's eventually exhaust M.

If there are no cusps, then there is a subsequence of the f_i whose images are disjoint, separated by distances of at least 2. If there are cusps, modify the cycles $f_j(\sum [S_i])$ by cutting them along horospherical cylinders in the cusps, and replacing the cusps of surfaces by cycles on these horospherical cylinders.



8.78

If the horospherical cylinders are sufficiently close to ∞, the resulting cycle Z_j will have area close to that of $f_j \sum [S_i]$, less than, say, $2\pi \sum |\chi(S_i)| + 1$. Z_j bounds a chain C_j with compact support. We may assume that the support of Z_{j+1} does not intersect $\mathcal{N}_2$ (support C_j). From 8.3.2, it follows that there is a constant K such that for all j,

$$v(\mathcal{N}_1(\text{support}\, Z_j)) \leq K.$$

If $x \in M$ is any regular point for h, then a small enough ball B about x is disjoint from $\phi_1(B)$. To prove the theorem, it suffices to show that almost every flow line through B eventually leaves M. Note that all the images $\{\phi_i(B)\}_{i \in N}$ are disjoint. Since ϕ_t does not decrease volume, almost all flow lines through B eventually leave the supports of all the C_j. If such a flow line does not cross ∂M, it must cross Z_j, hence it intersects $\mathcal{N}_1(\text{support}\, Z_j)$ with length at least two. By 8.12.1, the total length of time such a flow line spends in

$$\bigcup_{j=1}^{J} \mathcal{N}_1(\text{support}\, Z_j)$$

grows as J^2. Since the volume of

$$\bigcup_{j=1}^{J} \mathcal{N}_1(\text{support}\, Z_j)$$

grows only as $K \cdot J$, no set of positive measure of flow lines through B will fit—most have to run off the edge of M. $\square$

Remark. The fact that the area of Z_j is constant is stronger than necessary to obtain the conclusion of 8.3.3. It would suffice for the sum of reciprocals of the areas to form a divergent series. Thus, $\mathbb{R}^2$ has no non-constant positive superharmonic function, although $\mathbb{R}^3$ has.

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 9

Algebraic convergence

9.1. Limits of discrete groups

It is important for us to develop an understanding of the geometry of deformations of a given discrete group. A qualitative understanding can be attained most concretely by considering limits of sequences of groups. The situation is complicated by the fact that there is more than one reasonable sense in which a group can be the limit of a sequence of discrete groups.

DEFINITION 9.1.1. A sequence $\{\Gamma_i\}$ of closed subgroups of a Lie group G *converges geometrically* to a group Γ if

(i) each $\gamma \in \Gamma$ is the limit of a sequence $\{\gamma_i\}$, with $\gamma_i \in \Gamma_i$, and
(ii) the limit of every convergent sequence $\{\gamma_{i_j}\}$, with $\gamma_{i_j} \in \Gamma_{i_j}$, is in Γ.

Note that the geometric limit Γ is automatically closed. The definition means that Γ_i's look more and more like Γ, at least through a microscope with limited resolution. We shall be mainly interested in the case that the Γ_i's and Γ are discrete. The *geometric topology* on closed subgroups of G is the topology of geometric convergence.

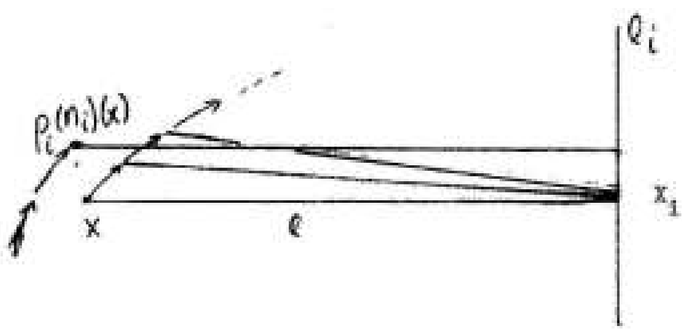
The notion of geometric convergence of a sequence of discrete groups is closely related to geometric convergence of a sequence of complete hyperbolic manifolds of bounded volume, as discussed in 5.11. A hyperbolic three-manifold M determines a subgroup of $\mathrm{PSL}(2, \mathbb{C})$ well-defined *up to conjugacy.* A specific representative of this conjugacy class of discrete groups corresponds to a choice of a base frame: a base 9.2
point p in M together with an orthogonal frame for the tangent space of M at p. This gives a specific way to identify $\tilde{M}$ with H^3. Let $O(\mathcal{H}_{[\epsilon,\infty)})$ consist of all base frames contained in $M_{[\epsilon,\infty)}$, where M ranges over $\mathcal{H}$ (the space of hyperbolic three-manifolds with finite volume). $O(\mathcal{H}_{[\epsilon,\infty)})$ has a topology defined by geometric convergence of groups. The topology on $\mathcal{H}$ is the quotient topology by the equivalence relation of conjugacy of subgroups of $\mathrm{PSL}(2, \mathbb{C})$. This quotient topology is not well-behaved for groups which are not geometrically finite.

DEFINITION 9.1.2. Let Γ be an abstract group, and $\rho_i : \Gamma \to G$ be a sequence of representations of Γ into G. The sequence $\{\rho_i\}$ *converges algebraically* if for every $\gamma \in \Gamma$, $\{\rho_i(\gamma)\}$ converges. The limit $\rho : \Gamma \to G$ is called the algebraic limit of $\{\rho_i\}$.

Definition 9.1.3. Let Γ be a countable group, $\{\rho_i\}$ a sequence of representations of Γ in G with $\rho_i(\Gamma)$ discrete. $\{\rho_i\}$ *converges strongly* to a representation ρ if ρ is the algebraic limit of $\{\rho_i\}$ *and* $\rho\Gamma$ is the geometric limit of $\{\rho_i\Gamma\}$.

Example 9.1.4 (Basic example). There is often a tremendous difference between algebraic limits and geometric limits, growing from the following phenomenon in a sequence of cyclic groups.

Pick a point x in H^3, a "horizontal" geodesic ray l starting at x, and a "vertical" plane through x containing the geodesic ray. Define a sequence of representations 9.3
$\rho_i : Z \to \mathrm{PSL}(2, \mathbb{C})$ as follows. Let x_i be



the point on l at distance i from x, and let l_i be the "vertical" geodesic through x_i: perpendicular to l and in the chosen plane. Now define ρ_i on the generator 1 by letting $\rho_i(1)$ be a screw motion around l_i with fine pitched thread so that $\rho_i(1)$ takes x to a point at approximately a horizontal distance of 1 from x and some high power $\rho_i(n_i)$ takes x to a point in the vertical plane a distance of 1 from x. The sequence $\{\rho_i\}$ converges algebraically to a parabolic representation $\rho : \mathbb{Z} \to \mathrm{PSL}(2, \mathbb{C})$, while $\{\rho_i\mathbb{Z}\}$ converges geometrically to a parabolic subgroup of rank 2, generated by $\rho(\mathbb{Z})$ plus an additional generator which moves x a distance of 1 in the vertical plane. 9.4

This example can be described in matrix form as follows. We make use of one-complex parameter subgroups of $\mathrm{PSL}(2, \mathbb{C})$ of the form

$$\begin{bmatrix} \exp w & a \sinh w \\ 0 & \exp - w \end{bmatrix},$$

with $w \in \mathbb{C}$. Define ρ_n by

$$\rho_n(1) = \begin{bmatrix} \exp w_n & n \sinh w_n \\ 0 & \exp - w_n \end{bmatrix}$$

where $w_n = 1/n^2 + \pi i/n$.

Thus $\{\rho_n(1)\}$ converges to

$$\begin{bmatrix} 1 & \pi i \\ 0 & 1 \end{bmatrix}$$

while $\{\rho_n(n)\}$ converges to

$$\begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

This example can be easily modified without changing the algebraic limit so that $\{\rho_i(\mathbb{Z})\}$ has no geometric limit, or so that its geometric limit is a one-complex-parameter parabolic subgroup, or so that the geometric limit is isomorphic to $\mathbb{Z} \times \mathbb{R}$. 9.5

This example can also be combined with more general groups: here is a simple case. Let Γ be a Fuchsian group, with M_Γ a punctured torus. Thus Γ is a free group on generators a and b, such that $[a, b]$ is parabolic. Let $\rho : \Gamma \to \mathrm{PSL}(2, \mathbb{C})$ be the identity representation. It is easy to see that $\operatorname{Tr} \rho'[a, b]$ ranges over a neighborhood of 2 as ρ' ranges over a neighborhood of ρ. Any nearby representation determines a nearby hyperbolic structure for $M_{[\epsilon,\infty)}$, which can be thickened to be locally convex except near $M_{(0,\epsilon]}$. Consider representations ρ_n with an eigenvalue for

$$\rho_n[a, b] \sim 1 + C/n^2 + \pi i/n.$$

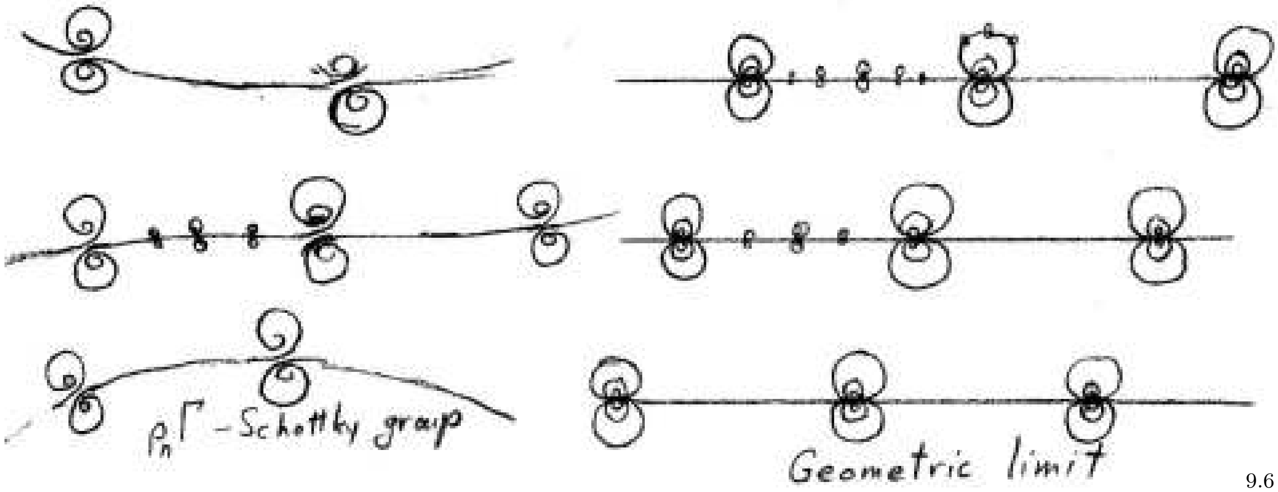
$\rho_n[a, b]$ translates along its axis a distance of approximately $2\operatorname{Re}(C)/n^2$, while rotating an angle of approximately

$$\frac{2\pi}{n} + \frac{2\operatorname{Im}(C)}{n^2}.$$

Thus the n-th power translates by a distance of approximately $2\operatorname{Re}(C)/n$, and rotates approximately

$$2\pi + \frac{2\operatorname{Im}(C)}{n}.$$

The axis moves out toward infinity as $n \to \infty$. For C sufficiently large, the image of ρ_n will be a geometrically finite group (a Schottky group); a compact convex manifold with $\pi_1 = \rho_n(\Gamma)$ can be constructed by piecing together a neighborhood of $M_{[\epsilon,\infty)}$ with (the convex hull of a helix)/$\mathbb{Z}$. The algebraic limit of $\{\rho_n\}$ is ρ, while the geometric limit is the group generated by $\rho(\Gamma) = \Gamma$ together with an extra parabolic generator commuting with $[a, b]$.



9.6

Troels Jørgensen was the first to analyze and understand this phenomenon. He showed that it is possible to iterate this construction and produce examples as above where the algebraic limit is the fundamental group of a punctured torus, but the geometric limit is not even finitely generated. See § .

Here are some basic properties of convergence of sequences of discrete groups.

Proposition 9.1.5. *If $\{\rho_i\}$ converges algebraically to ρ and $\{\rho_i\Gamma\}$ converges geometrically to Γ', then $\Gamma' \supset \rho\Gamma$.*

Proof. Obvious. $\square$

Proposition 9.1.6. *For any Lie group G, the space of closed subgroups of G (with the geometric topology) is compact.*

Proof. Let $\{\Gamma_i\}$ be any sequence of closed subgroups. First consider the case that there is a lower bound to the "size" $d(e, \gamma)$ of elements of $\gamma \in \Gamma_i$. Then there is an upper bound to the number of elements of Γ_i in the ball of radius γ about e, for every γ. The Tychonoff product theorem easily implies the existence of a subsequence converging geometrically to a discrete group.

Now let S be a maximal subspace of $T_e(G)$, the tangent space of G at the identity element e, with the property that for any $\epsilon > 0$ there is a Γ_i whose ϵ-small elements fill out all directions in S, within an angle of ϵ. It is easy to see that S is closed under Lie brackets. Furthermore, a subsequence $\{\Gamma_{i_j}\}$ whose small elements fill out S has 9.7
the property that all small elements are in directions near S. It follows, just as in the previous case, that there is a subsequence converging to a closed subgroup whose tangent space at e is S.

Corollary 9.1.7. *The set of complete hyperbolic manifolds N together with base frames in $N_{[\epsilon,\infty)}$ is compact in the geometric topology.*

□

Corollary 9.1.8. *Let Γ be any countable group and $\{\rho_i\}$ a sequence of discrete representations of Γ in $\mathrm{PSL}(2,\mathbb{C})$ converging algebraically to a representation ρ. If $\rho\Gamma$ does not have an abelian subgroup of finite index then $\{\rho_i\}$ has a subsequence converging geometrically to a discrete group $\Gamma' \supset \circ\Gamma$. In particular, $\rho\Gamma$ is discrete.*

Proof. By 9.1.7, there is a subsequence converging geometrically to *some* closed group Γ'. By 5.10.1, the identity component of Γ' must be abelian; since $\rho\Gamma \subset \Gamma'$, the identity component is trivial. □

Note that if the ρ_i are all faithful, then their algebraic limit is also faithful, since there is a lower bound to $d(\rho_i\gamma x, x)$. These basic facts were first proved in ????

Here is a simple example negating the converse of 9.1.8. Consider any discrete group $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ which admits an automorphism ϕ of infinite order: for instance, Γ might be a fundamental group of a surface. The sequence of representations ϕ^i has no algebraically convergent subsequence, yet $\{\phi^i\Gamma\}$ converges geometrically to Γ. 9.8

There are some simple statements about the behavior of limit sets when passing to a limit. First, if Γ is the geometric limit of a sequence $\{\Gamma_i\}$, then each point $x \in L_\Gamma$ is the limit of a sequence $x_i \in L_{\Gamma_i}$. In fact, fixed points x (eigenvectors) of non-trivial elements of $\gamma \in \Gamma$ are dense in L_Γ; for high i, Γ_i must have an element near γ, with a fixed point near x. A similar statement follows for the algebraic limit ρ of a sequence of representations ρ_i. Thus, the limit set cannot suddenly increase in the limit. It may suddenly decrease, however. For instance, let $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ be any finitely generated group. Γ is residually finite (see §), or in other words, it has a sequence $\{\Gamma_i\}$ of subgroups of finite index converging geometrically to the trivial group (e). $L_{\Gamma_i} = L_\Gamma$ is constant, but $L_{(e)}$ is empty. It is plausible that every finitely generated discrete group $\Gamma \subset \mathrm{PSL}(2,\mathbb{C})$ be a geometric limit of groups with compact quotient.

We have already seen (in 9.1.4) examples where the limit set suddenly decreases in an algebraic limit.

Let Γ be the fundamental group of a surface S with finite area and $\{\rho_i\}$ a sequence of faithful quasi-Fuchsian representations of Γ, preserving parabolicity. Suppose $\{\rho_i\}$ converges algebraically to a representation ρ as a group without any additional parabolic elements. Let N denote $N_{\rho(\Gamma)}$, N_i denote $N_{\rho_i(\Gamma)}$, etc. 9.9

Theorem 9.2. *N is geometrically tame, and $\{\rho_i\}$ converges strongly to ρ.*

Proof. If the set of uncrumpled maps of S into N homotopic to the standard map is compact, then using a finite cover of $\mathcal{GL}(S)$ carried by nearly straight train tracks, one sees that for any discrete representation ρ' near ρ, every geodesic lamination γ of S is realizable in N' near its realizations in N. (Logically, one can think of uncrumpled surfaces as equivariant uncrumpled maps of M^2 into H^3, with the compact-open topology, so that "nearness" makes sense.) Choose any subsequence of the ρ_i's so that the bending loci for the two boundary components of M_i converge in $\mathcal{GL}(S)$. Then the two boundary components must converge to locally convex disjoint embeddings of S in N (unless the limit is Fuchsian). These two surfaces are homotopic, hence they bound a convex submanifold M of N, so $\rho(\Gamma)$ is geometrically finite.

Since $M_{[\epsilon,\infty)}$ is compact, strong convergence of $\{\rho_i\}$ follows form 8.3.3: no unexpected identifications of N can be created by a small perturbation of ρ which preserves parabolicity.

If the set of uncrumpled maps of S homotopic to the standard map is not compact, then it follows immediately from the definition that N has at least one geometrically infinite tame end. We must show that both ends are geometrically tame. The possible phenomenon to be wary of is that the bending loci β_i^+ and β_i^- of the two boundary components of M_i might converge, for instance, to a single point λ in $\mathcal{GL}(S)$. (This would be conceivable if the "simplest" homotopy of one of the two boundary compo- 9.10
nents to a reference surface which persisted in the limit first carried it to the vicinity of the other boundary component.) To help in understanding the picture, we will first find a restriction for the way in which a hyperbolic manifold with a geometrically tame end can be a covering space.

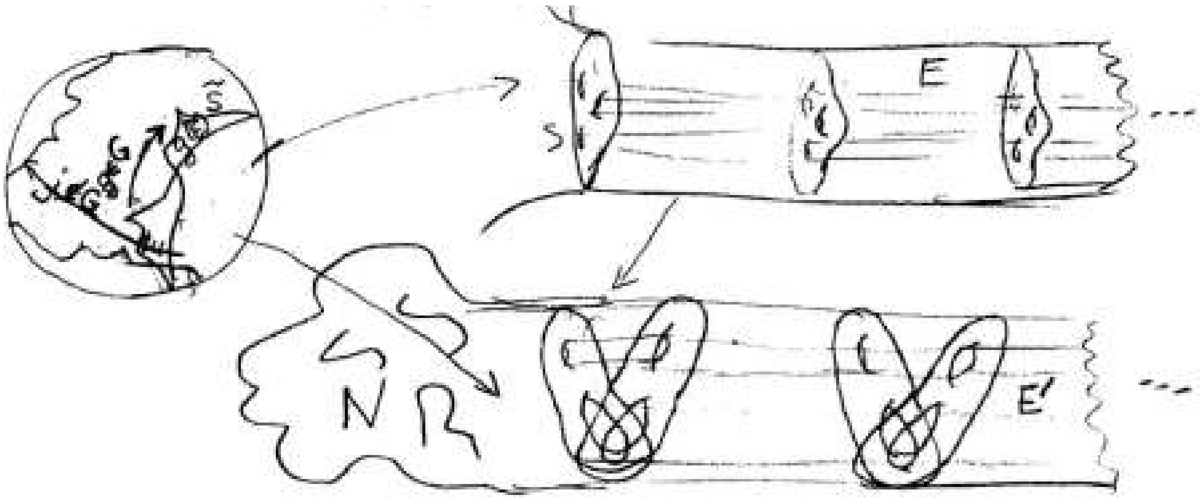
Definition 9.2.1. Let N be a hyperbolic manifold, P a union of horoball neighborhoods of its cusps, E' an end of $N-P$. E' is *almost geometrically tame* if some finite-sheeted cover of E' is (up to a compact set) a geometrically tame end. (Later we shall prove that if E is almost geometrically tame it is geometrically tame.)

Theorem 9.2.2. *Let N be a hyperbolic manifold, and $\tilde N$ a covering space of N such that $\tilde N - \tilde P$ has a geometrically infinite tame end E bounded by a surface $S_{[\epsilon,\infty)}$. Then either N has finite volume and some finite cover of N fibers over S^1 with fiber S, or the image of E in $N-P$, up to a compact set, is an almost geometrically tame end of N.*

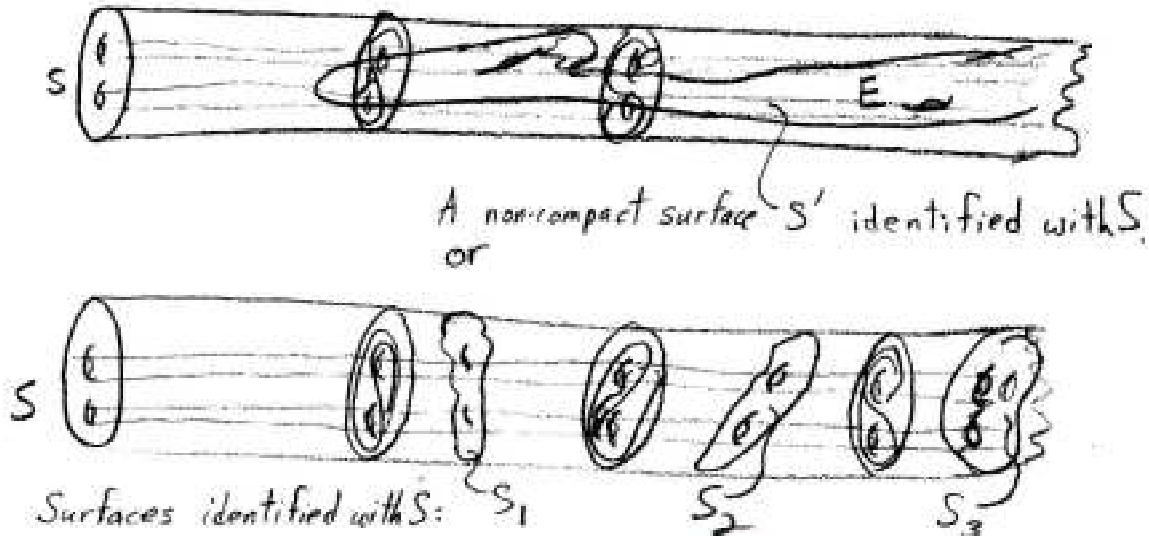
Proof. Consider first the case that all points of E identified with $S_{[\epsilon,\infty)}$ in the projection to N lie in a compact subset of E. Then the local degree of the projection of E to N is finite in a neighborhood of the image of S. Since the local degree is constant except at the image of S, it is everywhere finite.

Let $G \subset \pi_1 N$ be the set of covering transformations of H^3 over N consisting of 9.11
elements g such that $g\tilde E \cap \tilde E$ is all of $\tilde E$ except for a bounded neighborhood of $\tilde S$. G

is obviously a group, and it contains $\pi_1 S$ with finite index. Thus the image of E, up to compact sets, is an almost geometrically tame end of N.

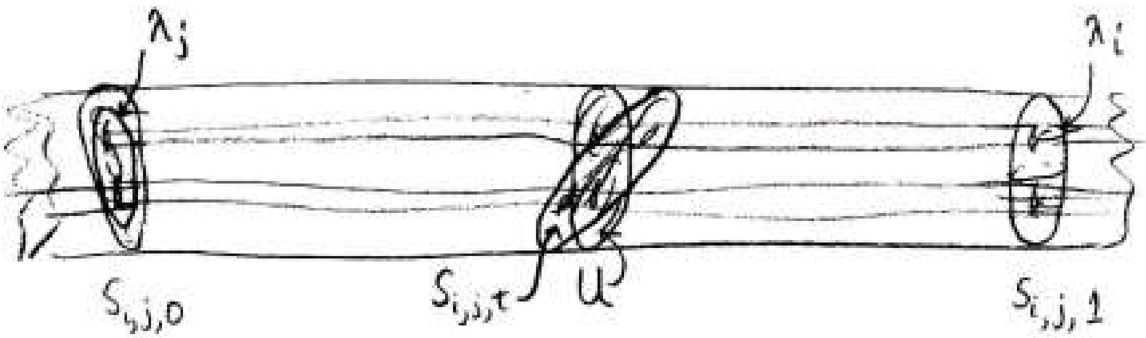


The other case is that $S_{[\epsilon,\infty)}$ is identified with a non-compact subset of E by projection to N. Consider the set I of all uncrumpled surfaces in E whose images intersect the image of $S_{[\epsilon,\infty)}$. Any short closed geodesic on an uncrumpled surface of E is homotopic to a short geodesic of E (not a cusp), since E contains no cusps other than the cusps of S. Therefore, by the proof of 8.8.5, the set of images of I in N is precompact (has a compact closure). If I itself is not compact, then N has a finite cover which fibers over S^1, by the proof of 8.10.9. If I is compact, then (since uncrumpled surfaces cut E into compact pieces), infinitely many components of the set of points identified with $S_{[\epsilon,\infty)}$ are compact and disjoint from S. 9.12



These components consist of immersions of k-sheeted covering spaces of S injective on π_1, which must be homologous to $\pm k\,[S]$. Pick two disjoint immersions with the same sign, homologous say to $-k\,[S]$ and $-l\,[S]$. Appropriate multiples of these cycles are homologous by a compactly supported three-chain which maps to a three-cycle in $N - P$, hence N has finite volume. Theorem 9.2.2 now follows from 8.10.9. $\square$

We continue the proof of Theorem 9.2. We may, without loss of generality, pass to a subsequence of representations ρ_i such that the sequences of bending loci $\{\beta_i^+\}$ and $\{\beta_i^-\}$ converge, in $\mathcal{PL}_0(S)$, to laminations β^+ and β^-. If β^+, say, is realizable for the limit representation ρ, then any uncrumpled surface whose wrinkling locus contains β^+ is embedded and locally convex—hence it gives a geometrically finite end of N. The only missing case for which we must prove geometric tameness is that neither 9.13
β^+ nor β^- is realizable. Let $\lambda_i^\epsilon \in \mathcal{PL}_0(S)$ (where $\epsilon = +, -$) be a sequence of geodesic laminations with finitely many leaves and with transverse measures approximating β_i^ϵ closely enough that the realization of λ_i^ϵ in N_i is near the realization of β_i^ϵ. Also suppose that $\lim \lambda_i^\epsilon = \beta^\epsilon$ in $\mathcal{PL}_0(S)$. The laminations λ_i^ϵ are all realized in N. They must tend toward ∞ in N, since their limit is not realized. We will show that they tend toward ∞ in the ϵ-direction. Imagine the contrary—for definiteness, suppose that the realizations of $\{\lambda_i^+\}$ in N go to ∞ in the $-$ direction. The realization of each λ_i^+ in N_j must be near the realization in N, for high enough j. Connect λ_j^+ to λ_i^+ by a short path $\lambda_{i,j,t}$ in $\mathcal{PL}_0(S)$. A family of uncrumpled surfaces $S_{i,j,t}$ realizing the $\lambda_{i,j,t}$ is not continuous, but has the property that for t near t_0, $S_{i,j,t}$ and S_{i,j,t_0} have points away from their cusps which are close in N. Therefore, for every uncrumpled surface U between $S_{i,j,0}$ and $S_{i,j,1}$ (in a homological sense), there is some t such that $S_{i,j,t} \cap U \cap (N - P)$ is non-void.



9.14

Let γ be any lamination realized in N, and U_j be a sequence of uncrumpled surfaces realizing γ in N_j, and converging to a surface in N. There is a sequence $S_{i(j),j,t(j)}$ of uncrumpled surfaces in N_j intersecting U_j whose wrinkling loci tend toward β^+.

Without loss of generality we may pass to a geometrically convergent subsequence, with geometric limit Q. Q is covered by N. It cannot have finite volume (from the analysis in Chapter 5, for instance), so by 8.14.2, it has an almost geometrically tame end E which is the image of the $-$ end E_- of N. Each element α of $\pi_1 E$ has a finite power $\alpha^k \in \pi_1 E_-$. Then a sequence $\{\alpha_i\}$ approximating α in $\pi_1(N_i)$ has the property that the α_i^k have bounded length in the generators of $\pi_1 S$, this implies that the α_i have bounded length, so α is in fact in $\pi_1 E_-$, and $E_- = E$ (up to compact sets). Using this, we may pass to a subsequence of $S_{i(j),j,t}$'s which converge to an uncrumpled surface R in E. R is incompressible, so it is in the standard homotopy class. It realizes β^+, which is absurd.

there is no 8.14.2

We may conclude that N has two geometrically tame ends, each of which is mapped homeomorphically to the geometric limit Q. (This holds whether or not they are geometrically infinite.) This implies the local degree of $N \to Q$ is finite one or two (in case the two ends are identified in Q). But any covering transformation α of N over Q has a power (its square) in $\pi_1 N$, which implies, as before, that $\alpha \in \pi_1 N$, so that $N = Q$. This concludes the proof of 9.2. □

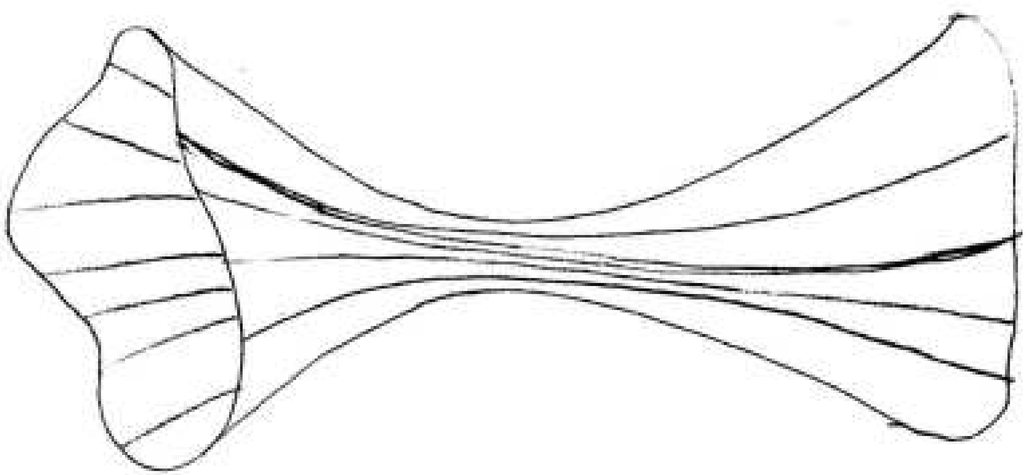
9.3. The ending of an end

In the interest of avoiding circumlocution, as well as developing our image of a geometrically tame end, we will analyze the possibilities for non-realizable laminations in a geometrically tame end.

We will need an estimate for the area of a cylinder in a hyperbolic three-manifold. Given any map $f : S^1 \times [0,1] \to N$, where N is a convex hyperbolic manifold, we may straighten each line $\theta \times [0,1]$ to a geodesic, obtaining a ruled cylinder with the same boundary.

Theorem 9.3.1. *The area of a ruled cylinder (as above) is less than the length of its boundary.*

Proof. The cylinder can be C^0-approximated by a union of small quadrilaterals each subdivided into two triangles. The area of a triangle is less than the minimum of the lengths of its sides (see p. 6.5). ☐



9.16

If the two boundary components of the cylinder C are far apart, then most of the area is concentrated near its boundary. Let γ_1 and γ_2 denote the two components of ∂C.

Theorem 9.3.2. *Area* $(C - \mathcal{N}_r\gamma_1) \leq e^{-r}\, l(\gamma_1) + l(\gamma_2)$ *where* $r \geq 0$ *and* l *denotes length.*

This is derived by integrating the area of a triangle in polar coordinates from any vertex:

$$A = \int \int_0^{T(\theta)} \sinh t\, dt\, d\theta = \int (\cosh T(\theta) - 1)\, d\theta$$

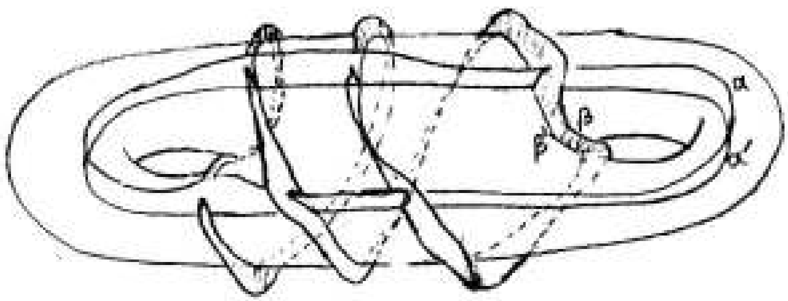
The area outside a neighborhood of radius r of its far edge α is

$$\int \cosh\,(T(\theta) - r) - 1\, d\theta < e^{-r} \int \sinh T(\theta)\, d\theta < e^{-r}\, l(\alpha).$$

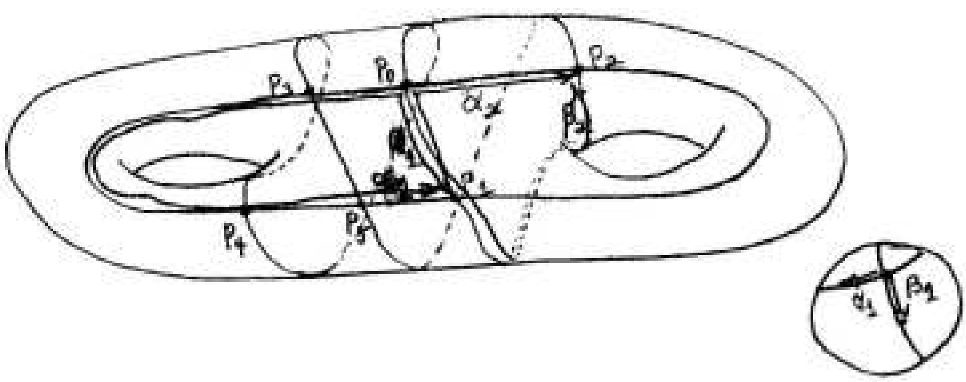
This easily implies 9.3.2

Let E be a geometrically tame end, cut off by a surface $S_{[\epsilon,\infty)}$ in $N - P$, as usual. A curve α in E homotopic to a simple closed curve α' on S gives rise to a ruled cylinder $C_\alpha : S^1 \times [0,1] \to N$.

Now consider two curves α and β homotopic to simple closed curves α' and β' 9.17
on S. One would expect that if α' and β' are forced to intersect, then either α must intersect C_β or β must intersect C_α, as in 8.11.1



We will make this more precise by attaching an invariant to each intersection. Let us assume, for simplicity, that α' and β' are geodesics with respect to some hyperbolic structure on S. Choose one of the intersection points, p_0, of α' and β' as a base point for N. For each other intersection point p_i, let α_i and β_i be paths on α' and β' from p_0 to p_i. Then $\alpha_i * \beta_i^{-1}$ is a closed loop, which is non-trivial in $\pi_1(S)$ when $i \neq 0$ since two geodesics in $\tilde{S}$ have at most one intersection.



There is some ambiguity, since there is more than one path from α_0 to α_i on 9.18
α'; in fact, α_i is well-defined up to a power of α'. Let $\langle g \rangle$ denote the cyclic group generated by an element g. Then $\alpha_i \cdot \beta_i^{-1}$ gives a well-defined element of the double coset space $\langle \alpha' \rangle \backslash \pi_1(S) / \langle \beta' \rangle$. [The double coset $H_1 g H_2 \in H_1 \backslash G / H_2$ of an element $g \in G$ is the set of all elements $h_1 g h_2$, where $h_i \in H_i$.] The double cosets associated to two different intersections p_i and p_j are distinct: if $\langle \alpha' \rangle \alpha_i \beta_i^{-1} \langle \beta' \rangle = \langle \alpha' \rangle a_j \beta_j^{-1} \langle \beta' \rangle$, then there is some loop $\alpha_j^{-1} \alpha'^k \alpha_i \beta_i^{-1} \beta'^l \beta_j$ made up of a path on α' and a path on

β' which is homotopically trivial—a contradiction. In the same way, a double coset $D_{x,y}$ is attached to each intersection of the cylinders C_α and C_β. Formally, these intersection points should be parametrized by the domain: thus, an intersection point means a pair $(x, y) \in (S^1 \times I) \times (S' \times I)$ such that $C_\alpha x = C_\beta y$.

Let $i(\gamma, \delta)$ denote the number of intersections of any two simple geodesics γ and δ on S. Let $D(\gamma, \delta)$ be the set of double cosets attached to intersection points of γ and δ (including p_0). Thus $i(\gamma, \delta) = |D(\gamma, \delta)|$. $D(\alpha, C_\beta)$ and $D(C_\alpha, \beta)$ are defined similarly.

Proposition 9.3.3. $|\alpha \cap C_\beta| + |C_\alpha \cap \beta| \geq i(\alpha', \beta')$. *In fact*

$$D(a, C_\beta) \cup D(C_\alpha, \beta) \supset D(\alpha', \beta').$$

9.19

Proof. First consider cylinders C'_α and C'_β which are contained in E, and which are nicely collared near S. Make C'_α and C'_β transverse to each other, so that the double locus $L \subset (S^1 \times I) \times (S^1 \times I)$ is a one-manifold, with boundary mapped to $\alpha \cup \beta \cup \alpha' \cup \beta'$. The invariant $D_{(x,y)}$ is locally constant on L, so each invariant occurring for $\alpha' \cap \beta'$ occurs for the entire length of interval in L, which must end on α or β. In fact, each element of $D(\alpha', \beta')$ occurs as an invariant of an odd number of points $\alpha \cup \beta$.

Now consider a homotopy h_t of C'_β to C_β, fixing $\beta \cup \beta'$. The homotopy can be perturbed slightly to make it transverse to α, although this may necessitate a slight movement of C_β to a cylinder C''_β. Any invariant which occurs an odd number of times for $a \cap C'_\beta$ occurs also an odd number of times for $\alpha \cap C''_\beta$. This implies that the invariant must also occur for $a \cap C_\beta$. □

Remark. By choosing orientations, we could of course associate signs to intersection points, thereby obtaining an algebraic invariant $\mathcal{D}(\alpha', \beta') \in \mathbb{Z}^{\langle \alpha' \rangle \backslash \pi_1 S / \langle \beta' \rangle}$. Then 9.3.3 would become an equation,

$$\mathcal{D}(\alpha', \beta') = \mathcal{D}(\alpha, C_\beta) + \mathcal{D}(C_\alpha, \beta).$$

Since $\pi_1(S)$ is a discrete group, there is a restriction on how closely intersection points can be clustered, hence a restriction on $|D(\alpha, c_\beta)|$ in terms of the length of α times the area of C_β.

9.20

Proposition 9.3.4. *There is a constant K such that for every curve α in E with distance R from S homotopic to a simple closed curve α' on S and every curve β in E not intersecting C_α and homotopic to a simple curve β' on S,*

$$i(\alpha', \beta') \leq K\big[l(\alpha) + \big(l(\alpha) + 1\big)\big(l(\beta) + e^{-R} + l(\beta')\big)\big].$$

PROOF. Consider intersection points $(x, y) \in S^1 \times (S^1 \times I)$ of α and C_β. Whenever two of them, (x, y) and (x', y'), are close in the product of the metrics induced from N, there is a short loop in N which is non-trivial if $D_{(x,y)} \neq D_{(x',y')}$.

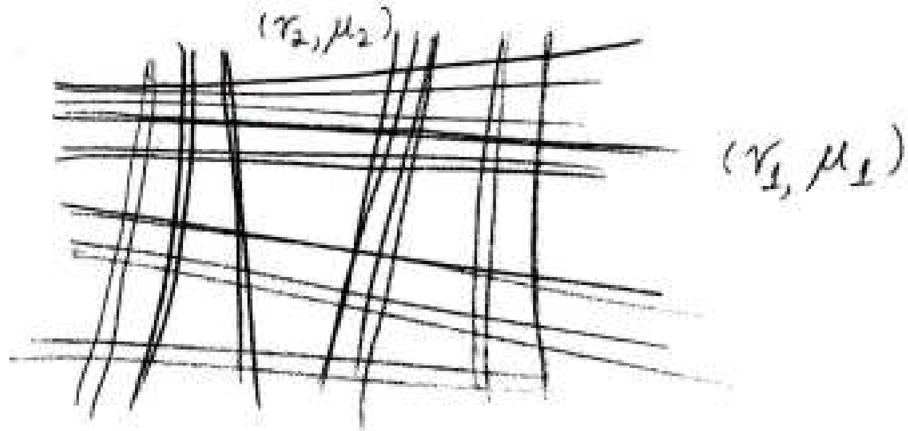
Case (i). α is a short loop. Then there can be no short non-trivial loop on C_β near an intersection point with α. The disks of radius ϵ on C_β about intersection points with α have area greater than some constant, except in special cases when they are near ∂C_β. If necessary, extend the edges of C_β slightly, without substantially changing the area. The disks of radius ϵ must be disjoint, so this case follows from 9.3.2 and 9.3.3.

Case (ii). α is not short. Let $E \subset C_\beta$ consist of points through which there is a short loop homotopic to β. If (x, y) and (x', y') are intersection points with $D_{x,y} \neq D_{x',y'}$ and with y, y' in E, then x and x' cannot be close together—otherwise two distinct conjugates of β would be represented by short loops through the same point. The number of such intersections is thus estimated by some constant times $l(\alpha)$. 9.21

Three intersections of α with $C_\beta - E$ cannot occur close together. $S^1 \times (C_\beta - E)$ contains the balls of radius ϵ, with multiplicity at most 2, and each ball has a definite volume. This yields 9.3.4. $\square$

Let us generalize 9.3.4 to a statement about measured geodesic laminations. Such a lamination (γ, μ) on a hyperbolic surface S has a well-defined "average length" $l_S(\gamma, \mu)$. This can be defined as the total mass of the measure which is locally the product of the transverse measure μ with one-dimensional Lebesgue measure on the leaves of γ. Similarly, a realization of γ in a homotopy class $f : S \to N$ has a length $l_f(\gamma, \mu)$. The length $l_S(\gamma, \mu)$ is a continuous function on $\mathcal{ML}_0(S)$, and $l_f(\gamma)$ is a continuous function where defined. If γ is realized a distance of R from an uncrumpled surface S, then $l_f(\gamma, \mu) \leq (1/\cosh R) l_S(\gamma, \mu)$. This implies that if f preserves non-parabolicity, l_f extends continuously over all of $\mathcal{ML}_0$ so that its zero set is the set of non-realizable laminations.

The intersection number $i\big((\gamma_1, \mu_1), (\gamma_2, \mu_2)\big)$ of two measured geodesic laminations is defined similarly, as the total mass of the measure $\mu_1 \times \mu_2$ which is locally the product of μ_1 and μ_2. (This measure $\mu_1 \times \mu_2$ is interpreted to be zero on any common leaves of γ_1 and γ_2.)



9.22

Given a geodesic lamination γ realized in E, let d_γ be the miniaml distance of an uncrumpled surface through γ from $S_{[\epsilon,\infty)}$.

THEOREM 9.3.5. *There is a constant K such that for any two measured geodesic laminations (γ_1, μ_1) and $(\gamma_2, \mu_2) \in \mathcal{ML}_0(S)$ realized in E,*

$$i\big((\gamma_1, \mu_1), (\gamma_2, \mu_2)\big) \leq K \cdot e^{-2R} l_S(\gamma_1, \mu_1) \cdot l_S(\gamma_2, \mu_2)$$

where $R = \inf(d_{\gamma_1}, d_{\gamma_2})$.

PROOF. First consider the case that γ_1 and γ_2 are simple closed geodesics which are not short. Apply the proof of 9.3.4 first to intersections of γ_1 with C_{γ_2}, then to intersections of C_{γ_1} with γ_2. Note that $l_S(\gamma_i)$ is estimated from below by $e^R l(\gamma_i)$, so the terms involving $l(\gamma_i)$ can be replaced by $Ce^{-R} l(\gamma_i)$. Since γ_1 and γ_2 are not short, one obtains

$$i(\gamma_1, \gamma_2) \leq K \cdot e^{-2R}\, l_S(\gamma_1)\, l_S(\gamma_2),$$

for some constant K. Since both sides of the inequality are homogeneous of degree one in γ_1 and γ_2, it extends by continuity to all of $\mathcal{ML}_0(S)$. □

Consider any sequence $\{(\gamma_i, \mu_i)\}$ of measured geodesic laminations in $\mathcal{ML}_0(S)$ whose realizations go to ∞ in E. If (λ_1, μ_1) and (λ_2, μ_2) are any two limit points of this sequence, 9.3.5 implies that $i(\lambda_1, \lambda_2) = 0$: in other words, the leaves do not cross. The union $\lambda_1 \cup \lambda_2$ is still a lamination. 9.23

DEFINITION 9.3.6. The *ending* lamination $\epsilon(E) \in \mathcal{GL}(S)$ is the union of all limit points λ_i, as above.

Clearly, $\epsilon(E)$ is compactly supported and it admits a measure with full support. The set $\Delta(E) \subset \mathcal{PL}_0(S)$ of all such measures on $\epsilon(E)$ is closed under convex combinations, hence its intersection with a local coordinate system (see p. 8.59) is convex.

In fact, a maximal train track carrying $\epsilon(E)$ defines a single coordinate system containing $\Delta(E)$.

The idea that the realization of a lamination depends continuously on the lamination can be generalized to the ending lamination $\epsilon(E)$, which can be regarded as being realized at ∞.

PROPOSITION 9.3.7. *For every compact subset K of E, there is a neighborhood U of $\Delta(E)$ in $\mathcal{PL}_0(S)$ such that every lamination in $U - \Delta(E)$ is realized in $E - K$.*

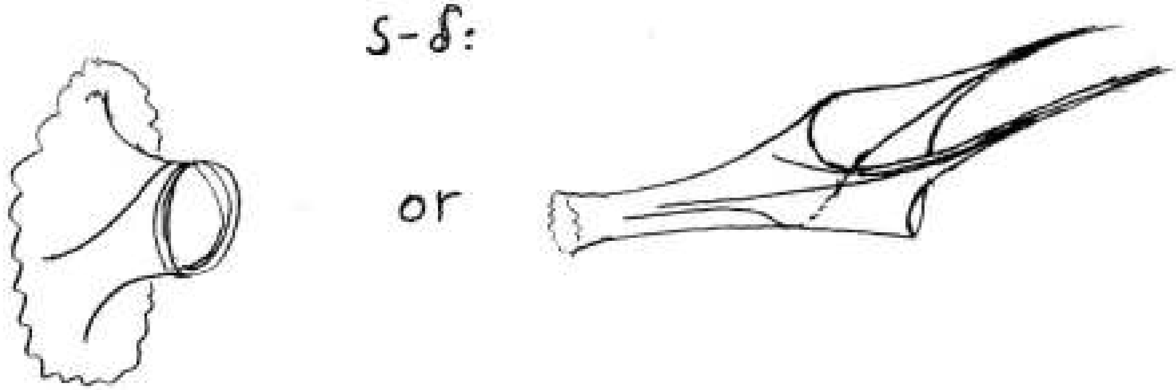
PROOF. It is convenient to pass to the covering of N corresponding to $\pi_1 S$. Let S' be an uncrumpled surface such that K is "below" S' (in a homological sense). Let $\{V_i\}$ be a neighborhood basis for $\Delta(E)$ such that $V_i - \Delta(E)$ is path-connected, and let $\lambda_i \in V_i - \Delta(E)$ be a sequence whose realizations go to ∞ in E. If there is any point $\pi_i \in V_i - \Delta(E)$ which is a non-realizable lamination or whose realization is not "above" S', connect λ_i to π_i by a path in V_i. There must be some element of this path whose realization intersects $S'_{[\epsilon,\infty)}$ (since the realizations cannot go to ∞ while in E.) Even if certain non-peripheral elements of S are parabolic, excess pinching of non-peripheral curves on uncrumpled surfaces intersecting S' can be avoided if S' is far from S, since there are no extra cusps in E. Therefore, only finitely many such π_i's can occur, or else there would be a limiting uncrumpled surface through S realizing the unrealizable. □

PROPOSITION 9.3.8. *Every leaf of $\epsilon(E)$ is dense in $\epsilon(E)$, and every non-trivial simple curve in the complement of $\epsilon(E)$ is peripheral.*

PROOF. The second statement follows easily from 8.10.8, suitably modified if there are extra cusps. The first statement then follows from the next result:

PROPOSITION 9.3.9. *If γ is a geodesic lamination of compact support which admits a nowhere zero transverse measure, then either every leaf of γ is dense, or there is a non-peripheral non-trivial simple closed curve in $S - \gamma$.*

PROOF. Suppose $\delta \subset \gamma$ is the closure of any leaf. Then δ is also an open subset of γ: all leaves of γ near δ are trapped forever in a neighborhood of δ. This is seen by considering the surface $S - \delta$.



9.25

An arc transverse to these leaves would have positive measure, which would imply that a transverse arc intersecting these leaves infinitely often would have infinite measure. (In general, a closed union of leaves $\delta \subset \gamma$ in a general geodesic lamination has only a finite set of leaves of γ intersecting a small neighborhood.)

If $\delta \neq \gamma$, then δ has two components, which are separated by some homotopically non-trivial curve in $S - \gamma$. □

□

Corollary 9.3.10. *For any homotopy class of injective maps $f : S \to N$ from a hyperbolic surface of finite area to a complete hyperbolic manifold, if f preserves parabolicity and non-parabolicity, there are $n = 0$, 1 or 2 non-realizable laminations ϵ_i $[1 \le i \le n]$ such that a general lamination γ on S is non-realizable if and only if the union of its non-isolated leaves is an ϵ_i.*

9.4. Taming the topology of an end

We will develop further our image of a geometrically tame end, once again to avoid circumlocution.

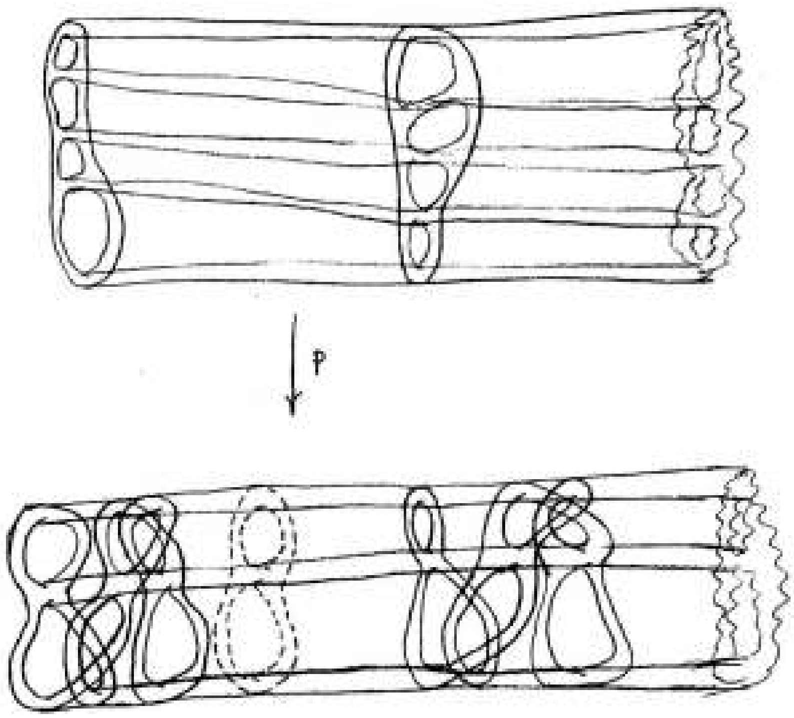
Theorem 9.4.1. *A geometrically tame end $E \subset N - P$ is topologically tame. In other words, E is homeomorphic to the product $S_{[\epsilon,\infty)} \times [0, \infty)$.*

Theorem 9.4.1 will be proved in §§9.4 and 9.5.

Corollary 9.4.2. *Almost geometrically tame ends are geometrically tame.*

Proof that 9.4.1 implies 9.4.2. Let E' be an almost geometrically tame end, finitely covered (up to compact sets) by a geometrically tame end $E = S_{[\epsilon,\infty)} \times [0, \epsilon)$, 9.26
with projection $p : E \to E'$. Let $f : E' \to [0, \epsilon)$ be a proper map. The first step is to find an incompressible surface $S' \subset E'$ which cuts it off (except for compact sets).

Choose t_0 high enough that $p : E \to E'$ is defined on $S_{[\epsilon,\infty)} \times [t_0, \infty)$, and choose $t_1 > t_0$ so that $p\big(S_{[\epsilon,\infty)} \times [t_1, \infty)\big)$ does not intersect $p(S_{[\epsilon,\infty)} \times t_0)$.



9.27

Let $r \in [0, \infty)$ be any regular value for f greater than the supremum of $f \circ p$ on $S_{[\epsilon,\infty)} \times [0, t_1)$. Perform surgery (that is, cut along circles and add pairs of disks) to $f^{-1}(r)$, to obtain a not necessarily connected surface S' in the same homology class which is incompressible in

$$E' - p\big(S_{[\epsilon,\infty)} \times [0, t_0)\big).$$

The fundamental group of S' is still generated by loops on the level set $f = r$. S' is covered by a surface $\tilde{S}'$ in E. $\tilde{S}'$ must be incompressible in $E-$ otherwise there would be a non-trivial disk D mapped into $S_{[\epsilon,\infty)} \times [t_1, \infty)$ with boundary on $\tilde{S}$; $p \circ D$ would be contained in

$$E' - p(S_{[\epsilon,\infty)} \times [0, t_0])$$

so S' would not be incompressible (by the loop theorem). One deduces that $\tilde{S}'$ is homotopic to $S_{[\epsilon,\infty)}$ and S' is incompressible in $N - P$.

If E is geometrically finite, there is essentially nothing to prove—E corresponds to a component of $\partial\tilde{M}$, which gives a convex embedded surface in E'. If E is geometrically infinite, then pass to a finite sheeted cover E'' of E which is a regular cover of E'. The ending lamination $\epsilon(E'')$ is invariant under all diffeomorphisms (up to compact sets) of E''. Therefore it projects to a non-realizable geodesic lamination $\epsilon(E')$ on S'. $\square$

Proof of 9.4.1. We have made use of one-parameter families of uncrumpled surfaces in the last two sections. Unfortunately, these surfaces do not vary continuously. To prove 9.4.1, we will show, in §9.5, how to interpolate with more general surfaces, to obtain a (continuous) proper map $F\colon S_{[\epsilon,\infty)} \times [0,\infty) \to E$. The theorem will follow fairly easily once F is constructed: 9.28

Proposition 9.4.3. *Suppose there is a proper map $F : S_{[\epsilon,\infty)} \times [0,\infty) \to E$ with $F(S_{[\epsilon,\infty)} \times 0)$ standard and with $F\big(\partial S_{[\epsilon,\infty)} \times [0,\infty)\big) \subset \partial(N - P)$. Then E is homeomorphic to $S_{[\epsilon,\infty)} \times [0,\infty)$.*

Proof of 9.4.3. This is similar to 9.4.2. Let $f : E \to [0,\infty)$ be a proper map. For any compact set $K \subset E$, we can find a $t_1 > 0$ so that $F\big(S_{[\epsilon,\infty)} \times [t_1,\infty)\big)$ is disjoint from K. Let r be a regular value for f greater than the supremum of $f \circ F$ on $S_{[\epsilon,\infty)} \times [0,t_1]$. Let $S' = f^{-1}(r)$ and $S'' = (f \circ F)^{-1}(r)$. $F : S'' \to S'$ is a map of degree one, so it is surjective on π_1 (or else it would factor through a non-trivial covering space on S', hence have higher degree). Perform surgery on S' to make it incompressible in the complement of K, without changing the homology class. Now S' must be incompressible in E; otherwise there would be some element α of $\pi_1 S'$ which is null-homotopic in E. But α comes from an element β on S'' which is null-homotopic in $S_{[\epsilon,\infty)} \times [t_1,\infty)$, so its image α is null-homotopic in the complement of K. It follows that S' is homotopic to $S_{[\epsilon,\infty)}$, and that the compact region of E cut off by S' is homeomorphic to $S_{[\epsilon,\infty)} \times I$. By constructing a sequence of such disjoint surfaces going outside of every compact set, we obtain a homeomorphism with $S_{[\epsilon,\infty)} \times [0,\infty)$. $\square$

$\square$

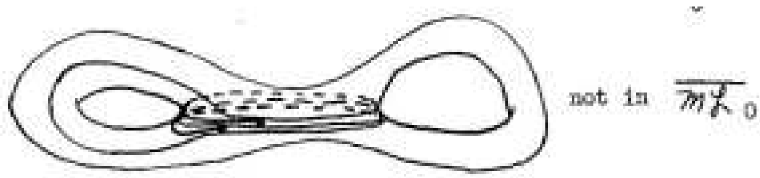
9.5. Interpolating negatively curved surfaces

Now we turn to the task of constructing a continuous family of surfaces moving out 9.29
to a geometrically infinite tame end. The existence of this family, besides completing the proof of 9.4.1, will show that a geometrically tame end has uniform geometry, and it will lead us to a better understanding of $\mathcal{ML}_0(S)$.

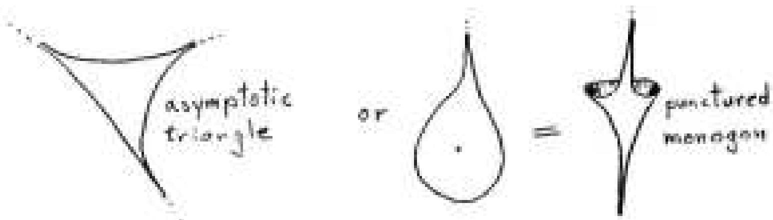
We will work with surfaces which are totally geodesic near their cusps, on esthetic grounds. Our basic parameter will be a family of compactly supported geodesic laminations in $\mathcal{ML}_0(S)$. The first step is to understand when a family of uncrumpled surfaces realizing these laminations is continuous and when discontinuous.

DEFINITION 9.5.1. For a lamination $\gamma \in \mathcal{ML}_0(S)$, let $\mathcal{T}_\gamma$ be the limit set in $\mathcal{GL}(S)$ of a neighborhood system for γ in $\mathcal{ML}_0(S)$. ($\mathcal{T}_\gamma$ is the "qualitative tangent space" of $\mathcal{ML}_0(S)$ at γ).

Let $\overline{\mathcal{ML}}_0(S)$ denote the closure of the image of $\mathcal{ML}_0(S)$ in $\mathcal{GL}(S)$. Clearly $\overline{\mathcal{ML}}_0(S)$ consists of laminations with compact support, but not every lamination with compact support is in $\overline{\mathcal{ML}}_0(S)$:



Every element of $\overline{\mathcal{ML}}_0$ is in $\mathcal{T}_\gamma$ for some $\gamma \in \mathcal{ML}_0$. Let us say that an element $\gamma \in \overline{\mathcal{ML}}_0$ is *essentially complete* if γ is a maximal element of $\overline{\mathcal{ML}}_0$. If $\gamma \in \mathcal{ML}_0$, then γ is essentially complete if and only if $\mathcal{T}_\gamma = \gamma$. A lamination γ is maximal among all
compactly supported laminations if and only if each region of $S - \gamma$ is an asymptotic 9.30
triangle or a neighborhood of a cusp of S with one cusp on its boundary—a punctured monogon.

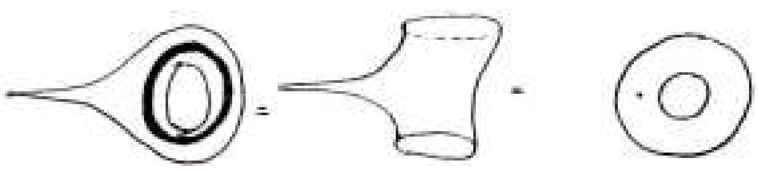


(These are the only possible regions with area π which are simply connected or whose fundamental group is peripheral.) Clearly, if $S - \gamma$ consists of such regions, then γ is essentially complete. There is one special case when essentially complete laminations are not of this form; we shall analyze this case first.

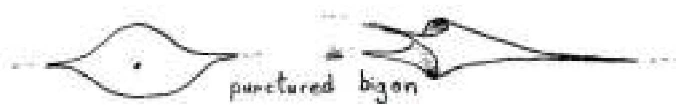
PROPOSITION 9.5.2. *Let $T - p$ denote the punctured torus. An element*

$$\gamma \in \overline{\mathcal{ML}}_0(T - p)$$

is essentially complete if and only if $(T - p) - \gamma$ is a punctured bigon.

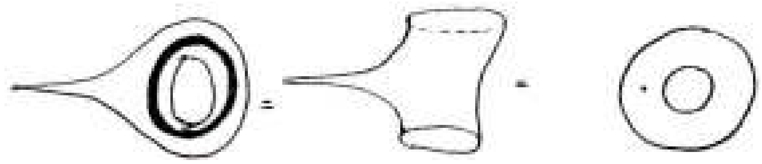


If $\gamma \in \mathcal{ML}_0(T - p)$, then either γ has a single leaf (which is closed), or every leaf of γ is non-compact and dense, in which case γ is essentially complete. If γ has a single closed leaf, then $\mathcal{T}_\gamma$ consists of γ and two other laminations:



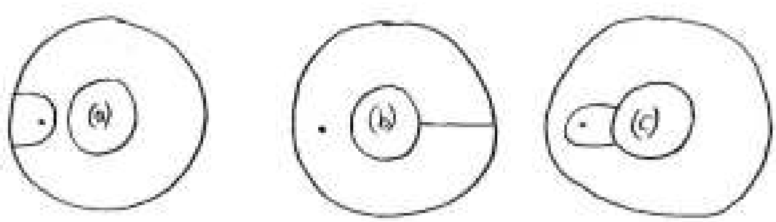
9.31

PROOF. Let $g \in \mathcal{ML}_0(T-p)$ be a compactly supported measured lamination. First, note that the complement of a simple closed geodesic on $T-p$ is a punctured annulus,

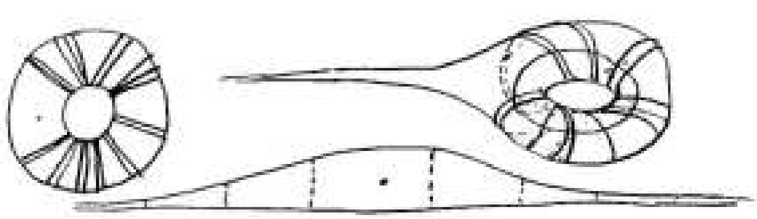


which admits no simple closed geodesics and consequently no geodesic laminations in its interior. Hence if γ contains a closed leaf, then γ consists only of this leaf, and otherwise (by 9.3.9) every leaf is dense.

Now let α be any simple closed geodesic on $T-p$, and consider γ cut apart by α. No end of a leaf of γ can remain forever in a punctured annulus, or else its limit set would be a geodesic lamination. Thus α cuts leaves of γ into arcs, and these arcs have only three possible homotopy classes:

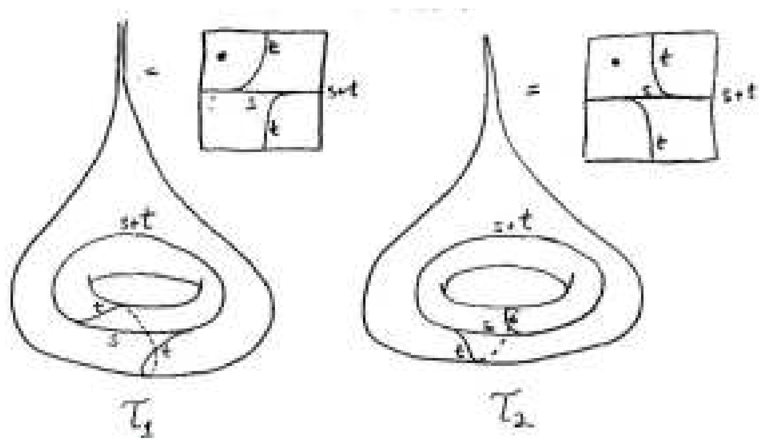


If the measure of the set of arcs of type (a) is m_a, etc., then (since the two boundary components match up) we have $2m_a + m_b = 2m_c + m_b$. But cases (a) and (c) 9.32
are incompatible with each other, so it must be that $m_a = m_c = 0$. Note that γ is orientable: it admits a continuous tangent vector field. By inspection we see a complementary region which is a punctured bigon.



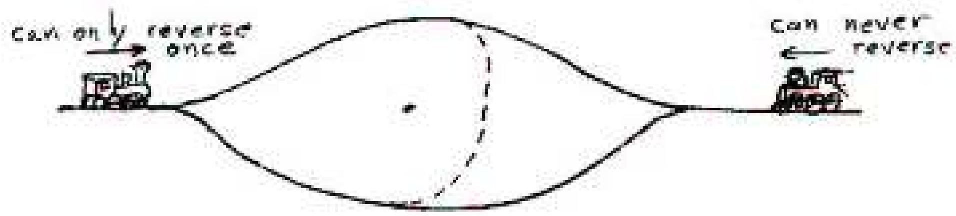
Since the area of a punctured bigon is 2π, which is the same as the area of $T-p$, this is the only complementary region.

It is now clear that a compactly supported measured lamination on $T-p$ with every leaf dense is essentially complete—there is nowhere to add new leaves under a small perturbation. If γ has a single closed leaf, then consider the families of measures on train tracks:

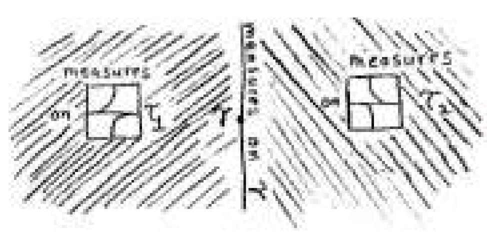


9.33

These train tracks cannot be enlarged to train tracks carrying measures. This can be deduced from the preceding argument, or seen as follows. At most one new branch could be added (by area considerations), and it would have to cut the punctured bigon into a punctured monogon and a triangle.



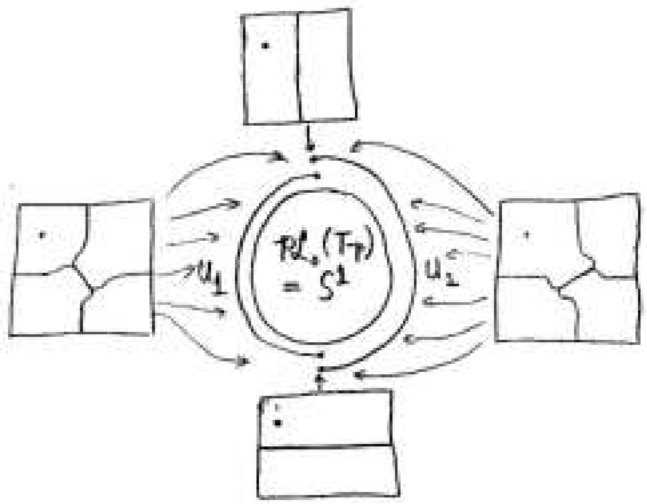
The train track is then orientable in the complement of the new branch, so a train can traverse this branch at most once. This is incompatible with the existence of a positive measure. Therefore $\mathcal{ML}_0(T-p)$ is two-dimensional, so τ_1 and τ_2 carry a neighborhood of γ.



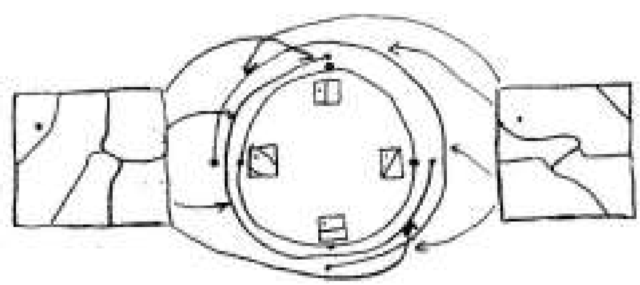
It follows that τ_γ is as shown. □

PROPOSITION 9.5.3. *$\mathcal{PL}_0(T-p)$ is a circle.*

Proof. The only closed one-manifold is S^1. That $\mathcal{PL}_0(T-p)$ is one-dimensional follows from the proof of 9.5.2. Perhaps it is instructive in any case to give a covering of $\mathcal{PL}_0(T-p)$ by train track neighborhoods: 9.34



or, to get open overlaps,



□

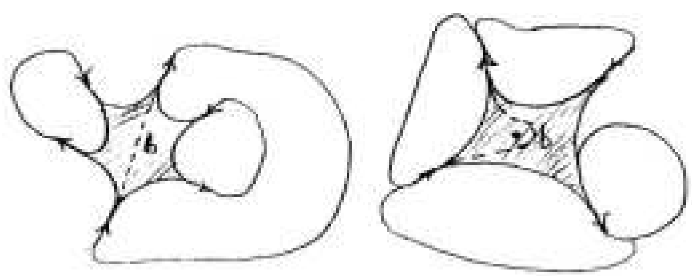
Proposition 9.5.4. *On any hyperbolic surface S which is not a punctured torus, an element $\gamma \in \overline{\mathcal{ML}}_0(S)$ is essentially complete if and only if $S - \gamma$ is a union of triangles and punctured monogons.*

Proof. Let γ be an arbitrary lamination in $\mathcal{ML}_0(S)$, and let τ be any train track approximation close enough that the regions of $S-\tau$ correspond to those of $S-\gamma$. If some of these regions are not punctured monogons or triangles, we will add extra 9.35
branches in a way compatible with a measure.

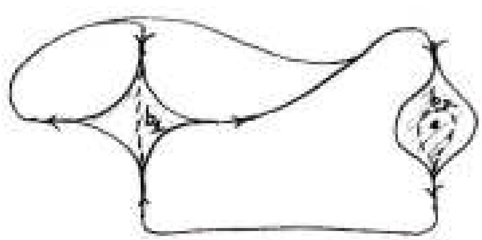
First consider the case that each region of $S-\gamma$ is either simply connected or a simple neighborhood of a cusp of S with fundamental group $\mathbb{Z}$. Then τ is connected. Because of the existence of an invariant measure, a train can get from any part of τ to any other. (The set of points accessible by a given oriented train is a "sink," which can only be a connected component.) If τ is not orientable, then every oriented train can get to any position with any orientation. (Otherwise, the oriented double "cover" of τ would have a non-trivial sink.)

In this case, add an arbitrary branch b to τ, cutting a non-atomic region (of area $> \pi$). Clearly there is some cyclic train path through b, so $\tau \cup b$ admits a positive measure.

If τ is oriented, then each region of $S - \tau$ has an even number of cusps on its boundary. The area of S must be 4π or greater (since the only complete oriented surfaces of finite area having $\chi = -1$ are the thrice punctured sphere, for which $\mathcal{ML}_0$ is empty, and the punctured torus). If there is a polygon with more than four sides, it can be subdivided using a branch which preserves orientation, hence admits a cyclic train path. The case of a punctured polygon with more than two sides is similar. 9.36

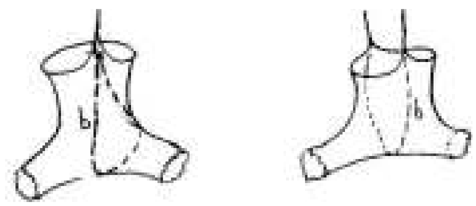


Otherwise, $S - \gamma$ has at least two components. Add one branch b_1 which reverses positively oriented trains, in one region, and another branch b_2 which reverses negatively oriented trains in another.

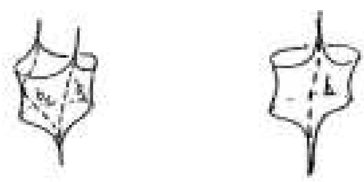


There is a cyclic train path through b_1 and b_2 in $\tau \cup b_1 \cup b_2$, hence an invariant measure.

Now consider the case when $S - \tau$ has more complexly connected regions. If 9.37
a boundary component of such a region R has one or more vertices, then a train pointing away from R can return to at least one vertex pinting toward R. If R is not an annulus, hook a new branch around a non-trivial homotopy class of arcs in R with ends on such a pair of vertices.

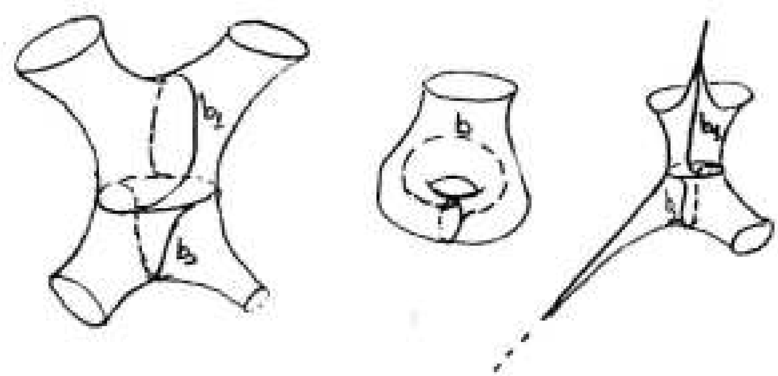


If R is an annulus and each boundary component has at least one vertex, then add one or two branches running across R which admit a cyclic train path.



If R is not topologically a thrice punctured disk or annulus, we can add an interior closed curve to R.

Any boundary component of R which is a geodesic α has another region R' (which may equal R) on the other side. In this case, we can add one or more branches in
R and R' tangent to α in opposite directions on opposite sides, and hooking in ways 9.38
similar to those previously mentioned.



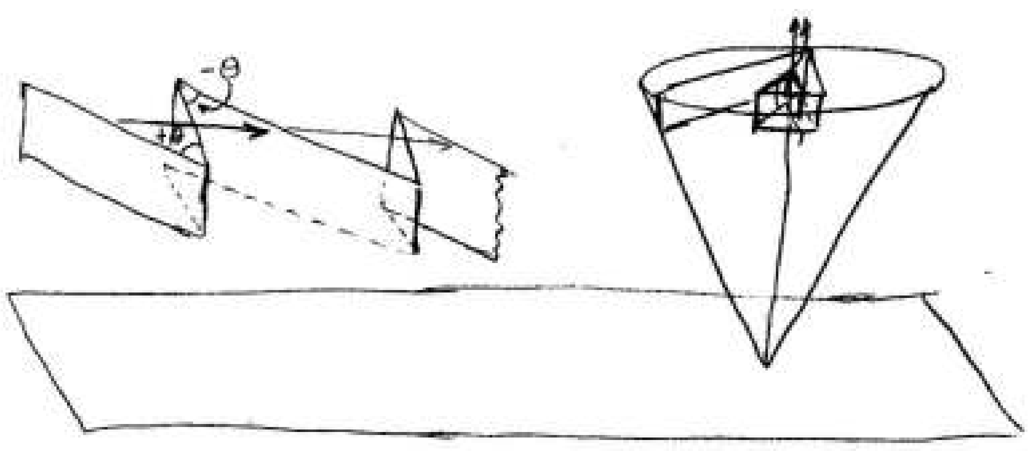
From the existence of these extensions of the original train track, it follows that an element $\gamma \in \mathcal{ML}_0$ is essentially complete if and only if $S - \gamma$ consists of triangles and punctured monogons. Furthermore, every $\gamma \in \overline{\mathcal{ML}_0}$ can be approximated by essentially complete elements $\gamma' \in \mathcal{ML}_0$. In fact, an open dense set has the property that the ϵ-train track approximation τ_ϵ has only triangles and punctured monogons as complementary regions, so generically *every* τ_ϵ has this property. The characterization of essential completeness then holds for $\overline{\mathcal{ML}_0}$ as well. $\square$

Here is some useful geometric information about uncrumpled surfaces.

PROPOSITION 9.5.5. (i) *The sum of the dihedral angles along all edges of the wrinkling locus $w(S)$ tending toward a cusp of an uncrumpled surface S is 0.* (*The sum is taken in the group $S^1 = \mathbb{R} \bmod 2\pi$.*)

(ii) *The sum of the dihedral angles along all edges of $w(S)$ tending toward any* 9.39
side of a closed geodesic γ of $w(S)$ is $\pm\alpha$, where α is the angle of rotation of parallel translation around γ. (*The sign depends on the sense of the spiralling of nearby geodesics toward γ.*)

PROOF. Consider the upper half-space model, with either the cusp or the end of $\tilde{\gamma}$ toward which the geodesics in $w(S)$ are spiralling at ∞. Above some level (in case (a)) or inside some cone (in case (b)), S consists of vertical planes bent along vertical lines. The proposition merely says that the total angle of bending in some fundamental domain is the sum of the parts.

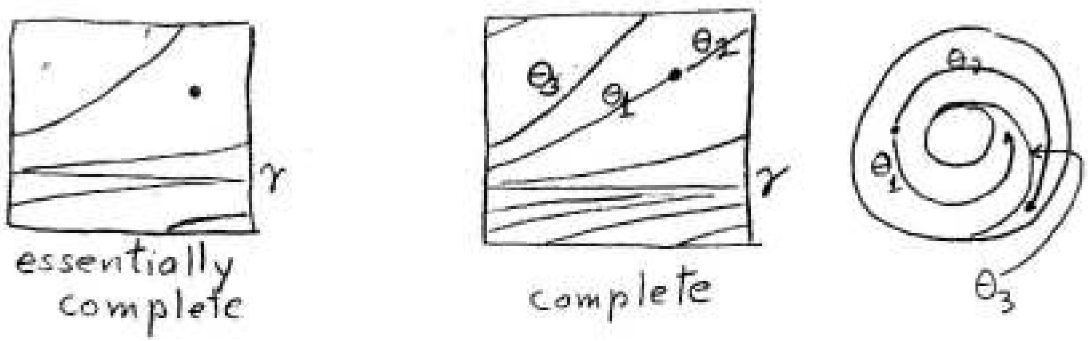


$\square$

COROLLARY 9.5.6. *An uncrumpled surface realizing an essentially complete lamination in $\overline{\mathcal{ML}_0}$ in a given homotopy class is unique. Such an uncrumpled surface is totally geodesic near its cusps.*

PROOF. If the surface S is not a punctured torus, then it has a unique completion obtained by adding a single geodesic tending toward each cusp. By 9.5.5, an uncrumpled surface cannot be bent along any of these added geodesics, so we obtain 9.40
9.5.6.

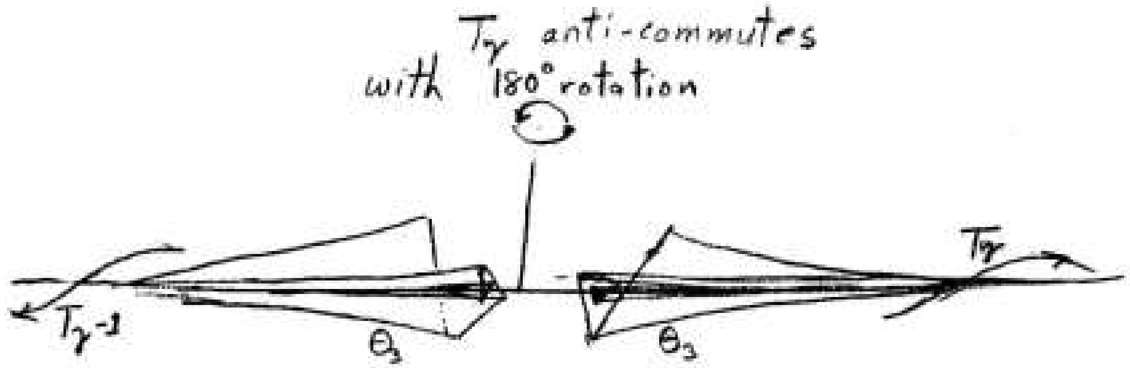
If S is the punctured torus $T - p$, then we consider first the case of a lamination γ which is an essential completion of a single closed geodesic. Complete γ by adding two closed geodesics going from the vertices of the punctured bigon to the puncture.



If the dihedral angles along the infinite geodesics are θ_1, θ_2 and θ_3, as shown, then by 9.5.5 we have

$$\theta_1 + \theta_2 = 0, \quad \theta_1 + \theta_3 = \alpha, \quad \theta_2 + \theta_3 = \alpha,$$

where α is some angle. (The signs are the same for the last two equations because any hyperbolic transformation anti-commutes with a 180° rotation around any perpendicular line.)



9.41

Thus $\theta_1 = \theta_2 = 0$, so an uncrumpled surface is totally geodesic in the punctured bigon. Since simple closed curves are dense in $\mathcal{ML}_0$, every element $g \in \mathcal{ML}_0$ realizable in a given homotopy class has a realization by an uncrumpled surface which is totally geodesic on a punctured bigon. If γ is essentially complete, this means its realizing surface is unique. □

Proposition 9.5.7. *If γ is an essentially complete geodesic lamination, realized by an uncrumpled surface U, then any uncrumpled surface U' realizing a lamination γ' near γ is near U.*

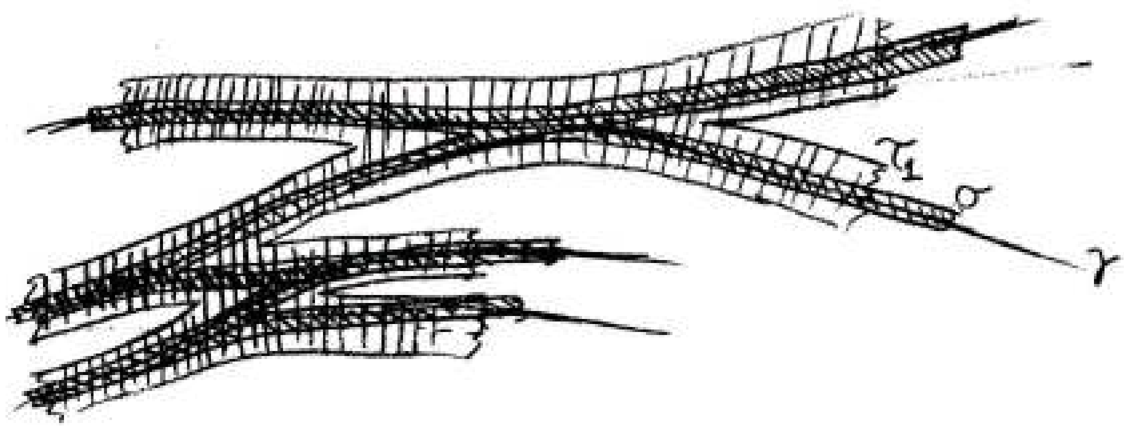
PROOF. You can see this from train track approximations. This also follows from the uniqueness of the realization of γ on an uncrumpled surface, since uncrumpled surfaces realizing laminations converging to γ must converge to a surface realizing γ. □

Consider now a typical path $\gamma_t \in \mathcal{ML}_0$. The path γ_t is likely to consist mostly of essentially complete laminations, so that a family of uncrumpled surfaces U_t realizing γ_t would be usually (with respect to t) continuous. At a countable set of values of t, γ_t is likely to be essentially incomplete, perhaps having a single complementary quadrilateral. Then the left and right hand limits U_{t-} and U_{t+} would probably exist, 9.42
and give uncrumpled surfaces realizing the two essential completions of γ_t. In fact, we will show that any path γ_t can be perturbed slightly to give a "generic" path in which the only essentially incomplete laminations are ones with precisely two distinct completions. In order to speak of generic paths, we need more than the topological structure of $\mathcal{ML}_0$.

PROPOSITION 9.5.8. $\mathcal{ML}$ *and* $\mathcal{ML}_0$ *have canonical PL (piecewise linear) structures.*

PROOF. We must check that changes of the natural coordinates coming from maximal train tracks (pp. 8.59-8.60) are piecewise linear. We will give the proof for $\mathcal{ML}_0$; the proof for $\mathcal{ML}$ is obtained by appropriate modifications.

Let γ be any measured geodesic lamination in $\mathcal{ML}_0(S)$. Let τ_1 and τ_2 be maximal compactly supported train tracks carrying γ, defining coordinate systems ϕ_1 and ϕ_2 from neighborhoods of γ to convex subsets of R^n (consisting of measures on τ_1 and τ_2). A close enough train track approximation σ of γ is carried by τ_1 and τ_2.



9.43

The set of measures on σ go linearly to measures on τ_1 and τ_2. If σ is a maximal compact train track supporting a measure, we are done—the change of coordinates $\phi_2 \circ \phi_2^{-1}$ is linear near γ. (In particular, note that if γ is essentially complete, change of coordinates is *always* linear at γ). Otherwise, we can find a finite set of enlargements of σ, $\sigma_1, \dots, \sigma_k$, so that every element of a neighborhood of γ is closely approximated by one of the σ_i. Since every element of a neighborhood of γ is carried by τ_1 and τ_2, it follows that (if the approximations are good enough) each of the σ_i is carried by τ_1 and τ_2. Each σ_i defines a convex polyhedron which is mapped linearly by ϕ_1 and ϕ_2, so $\phi_2 \circ \phi_1^{-1}$ must be PL in a neighborhood of γ. $\square$

REMARK 9.5.9. It is immediate that change of coordinates involves only rational coefficients. In fact, with more care $\mathcal{ML}$ and $\mathcal{ML}_0$ can be given a piecewise integral linear structure. To do this, we can make use of the set $\mathcal{D}$ of integer-valued measures supported on finite collections of simple closed curves (in the case of $\mathcal{ML}_0$); $\mathcal{D}$ is analogous to the integral lattice in $\mathbb{R}^n$. $\mathrm{GL}_n\,\mathbb{Z}$ consists of linear transformations of $\mathbb{R}^n$ which preserve the integral lattice. The set V_τ of measures supported on a given train track τ is the subset of some linear subspace $V \subset \mathbb{R}^n$ which satisfies a finite number of linear inequalities $\mu(b_i) > 0$. Thus V_τ is the convex hull of a finite number of lines, each passing through an integral point. The integral points in U are closed under integral linear combinations (when such a combination is in U), so they determine an 9.44
integral linear structure which is preserved whenever U is mapped linearly to another coordinate system.

Note in particular that the natural transformations of $\mathcal{ML}_0$ are volume-preserving.

The structure on $\mathcal{PL}$ and $\mathcal{PL}_0$ is a piecewise integral *projective* structure. We will use the abbreviations PIL and PIP for piecewise integral linear and piecewise integral projective.

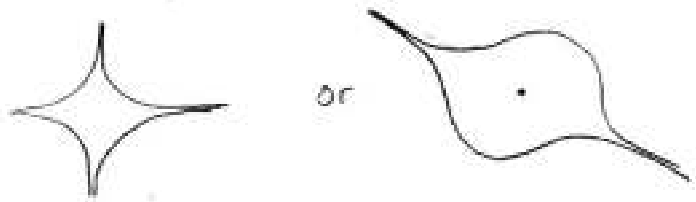
DEFINITION 9.5.10. The *rational depth* of an element $\gamma \in \mathcal{ML}_0$ is the dimension of the space of rational linear functions vanishing on γ, with respect to any natural local coordinate system. From 9.5.8 and 9.5.9, it is clear that the rational depth is independent of coordinates.

PROPOSITION 9.5.11. *If γ has rational depth 0, then γ is essentially complete.*

PROOF. For any $\gamma \in \mathcal{ML}_0$ which is not essentially complete we must construct a rational linear function vanishing on γ. Let τ be some train track approximation of γ which can be enlarged and still admit a positive measure. It is clear that the set of measures on τ spans a proper rational subspace in any natural coordinate system coming from a train track which carries τ. (Note that measures on τ consist of positive linear combinations of integral measures, and that every lamination carried by τ is approximable by one *not* carried by τ.) $\square$ 9.45

PROPOSITION 9.5.12. *If $\gamma \in \mathcal{ML}_0$ has rational depth 1, then either γ is essentially complete or γ has precisely two essential completions. In this case either*

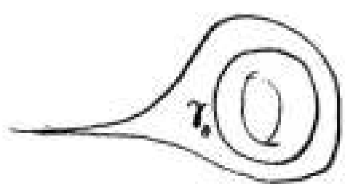
A. *γ has no closed leaves, and all complementary regions have area π or 2π. There is only one region with area 2π unless γ is oriented and* area$(S) = 4\pi$ *in which case there are two. Such a region is either a quadrilateral or a punctured bigon.*



or

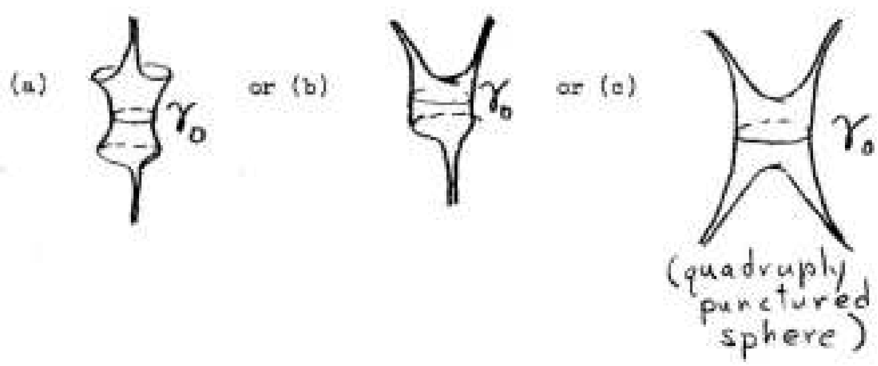
B. *γ has precisely one closed leaf γ_0. Each region touching γ_0 has area 2π. Either*

1. *S is a punctured torus*

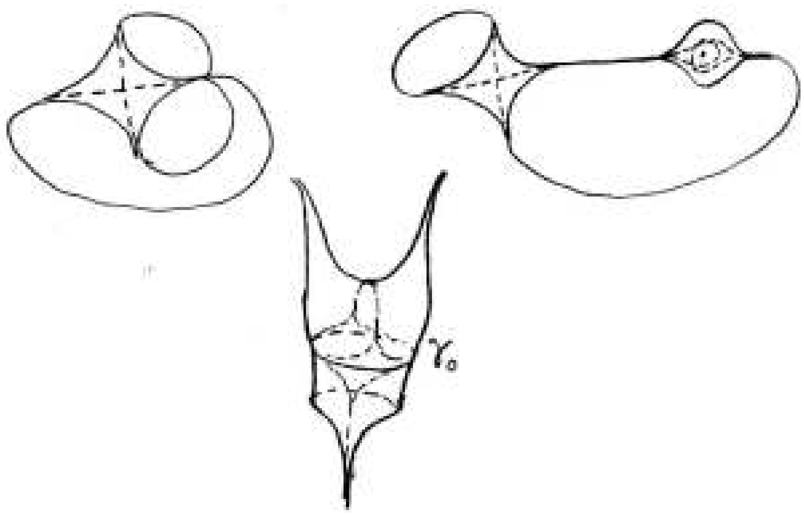


or

2. *γ_0 touches two regions, each a one-pointed crown or a devils cap.*



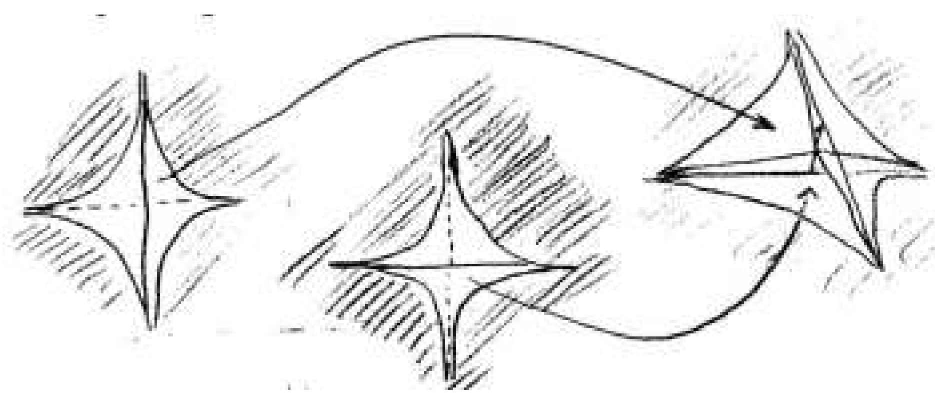
Proof. Suppose γ has rational depth 1 and is not essentially complete. Let τ be a close train track approximation of γ. There is some finite set $\tau_1, \ldots, \tau_k$ of essentially complete enlargements of τ which closely approximate every γ' in a neighborhood of γ. Let σ carry all the τ_i's and let V_σ be its coordinate system. The set of γ corresponding to measures carried by a given proper subtrack of a τ_i is a proper rational subspace of V_σ. Since γ is in a unique proper rational subspace, V_τ, the set of measures V_{τ_i} carried on any τ_i must consist of one side of V_τ. (If V_{τ_i} intersected both sides, by convexity γ would come from a measure positive on all branches of τ_i). Since this works for any degree of approximation of nearby laminations, γ has precisely two essential completions. A review of the proof of 9.5.4 gives the list of possibilities for $\gamma \in \mathcal{ML}_0$ with precisely two essential completions. The ambiguity in the essential completions comes from the manner of dividing a quadrilateral or other region, and the direction of spiralling around a geodesic.



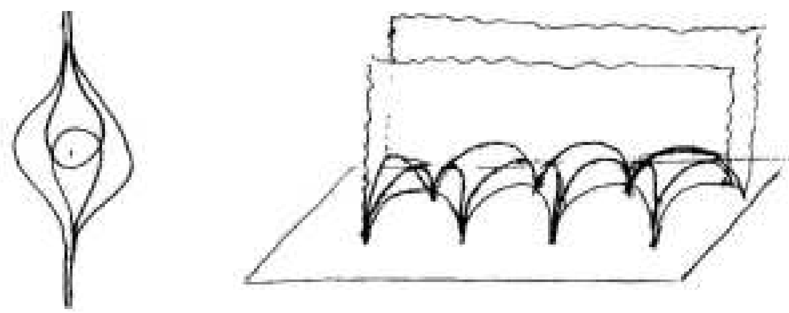
$\square$ 9.47

Remark. There are good examples of $\gamma \in \mathcal{ML}_0$ which have large rational depth but are essentially complete. The construction will occur naturally in another context.

We return to the construction of continuous families of surfaces in a hyperbolic three-manifold. To each essentially incomplete $\gamma \in \mathcal{ML}_0$ of rational depth 1, we associate a one-parameter family of surfaces U_s, with U_0 and U_1 being the two uncrumpled surfaces realizing γ. U_s is constant where U_0 and U_1 agree, including the union of all triangles and punctured monogons in the complement of γ. The two images of any quadrilateral in $S - \gamma$ form an ideal tetrahedron. Draw the common perpendicular p to the two edges not in $U_0 \cap U_1$, triangulate the quadrilateral with 4 triangles by adding a vertex in the middle, and let this vertex run linearly along p, from U_0 to U_1. This extends to a homotopy of S straight on the triangles.

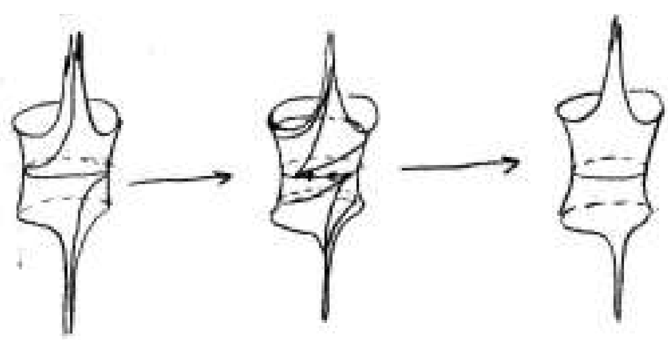


The two images of any punctured bigon in $S - \gamma$ form a solid torus, with the generating curve parabolic. The union of the two essential completions in this punctured bigon gives a triangulation except in a neighborhood of the puncture, with two new vertices at intersection points of added leaves. 9.48



Draw the common perpendiculars to edges of the realizations corresponding to these intersection points, and homotope U_0 to U_1 by moving the added vertices linearly along the common perpendiculars.

When γ has a closed leaf γ_0, the two essential completions of γ have added leaves spiralling around γ_0 in opposite directions. U_0 can be homotoped to U_1 through surfaces with added vertices on γ_0.



9.49

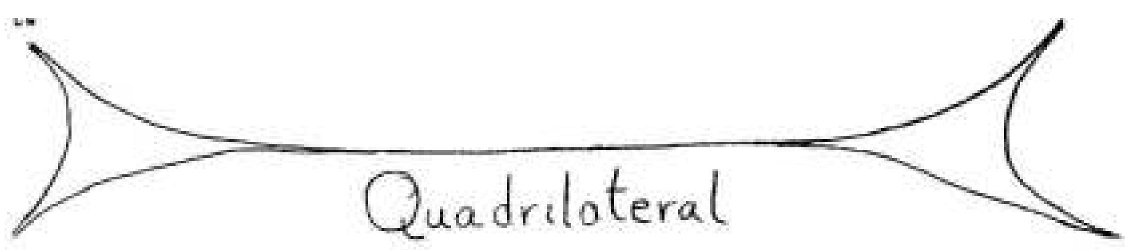
Note that all the surfaces U_s constructed above have the property that any point on U_s is in the convex hull of a small circle about it on U_s. In particular, it has curvature ≤ -1; curvature -1 everywhere except singular vertices, where negative curvature is concentrated.

THEOREM 9.5.13. *Given any complete hyperbolic three-manifold N with geometrically tame end E cut off by a hyperbolic surface $S_{[\epsilon,\infty)}$, there is a proper homotopy $F : S_{[\epsilon,\infty)} \times [0,\infty) \to N$ of S to ∞ in E.*

PROOF. Let V_τ be the natural coordinate system for a neighborhood of $\epsilon(E)$ in $\mathcal{ML}_0(S)$, and choose a sequence $\gamma_i \in V_\tau$ limiting on $\epsilon(E)$. Perturb the γ_i slightly so that the path γ_t $[0 \leq t \leq \infty]$ which is linear on each segment $t \in [i,\, i+1]$ consists of elements of rational depth 0 or 1. Let U_t be the unique uncrumpled surface realizing γ_t when γ_t is essentially complete. When t is not essentially complete, the left and right hand limits U_{t+} and U_{t-} exist. It should now be clear that F exists, since one can cover the closed set $\{U_{t\pm}\}$ by a locally finite cover consisting of surfaces homotopic by small homotopies, and fill in larger gaps between U_{t+} and U_{t-} by the homotopies constructed above. Since all interpolated surfaces have curvature ≤ -1, and they all realize a γ_t, they must move out to ∞. An explicit homotopy can actually be defined, using a new parameter r which is obtained by "blowing up" all parameter values of t with rational depth 1 into small intervals. Explicitly, these parameter values can be enumerated in some order $\{t_j\}$, and an interval of length 2^{-j} inserted 9.50
in the r-parameter in place of t_j. Thus, a parameter value t corresponds to the point or interval

$$r(t) = \left[t + \sum_{\{j|t_j<t\}} 2^- j,\ t + \sum_{\{j|t_j\leq t\}} 2^{-j}\right].$$

Now insert homotopies as constructed above in *each* blown up interval. It is not so obvious that the family of surfaces is still continuous when an infinite family of homotopies is inserted. Usually, however, these homotopies move a very small distance—for instance, γ_t may have a quadrilateral in $S - \gamma_t$, but for all but a locally small number of t's, this quadrilateral looks like two asymptotic triangles to the naked eye, and the homotopy is imperceptible.



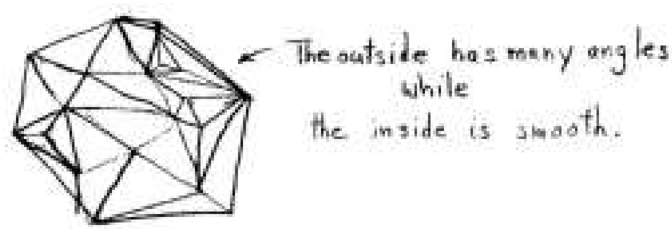
Formally, the proof of continuity is a straightforward generalization of the proof of 9.5.7. The remark which is needed is that if S is a surface of curvature ≤ -1 with a (pathwise) isometric map to a hyperbolic surface homotopic to a homeomorphism, then S is actually hyperbolic and the map is isometric—indeed, the area of S is not greater than the area of the hyperbolic surface. $\square$

REMARKS. 1. There is actually a canonical line of hyperbolic structures on S joining those of U_{t+} and U_{t-}, but it is not so obvious how to map them into E nicely. 9.51

2. An alternative approach to the construction of F is to make use of a sequence of triangulations of S. Any two triangulations with the same number of vertices can be joined by a sequence of elementary moves, as shown.

Although such an approach involves more familiar methods, the author brutally chose to develop extra structure.

3. There should be a good analytic method of constructing F by using harmonic mappings of hyperbolic surfaces. Realizations of geodesic laminations of a surface are analogous to harmonic mappings coming from points at ∞ in Teichmüller space. The harmonic mappings corresponding to a family of hyperbolic structures on S moving along a Teichmüller geodesic to $\epsilon(E)$ ought to move nicely out to ∞ in E. A rigorous proof might involve good estimates of the energy of a map, analogous to §9.3.

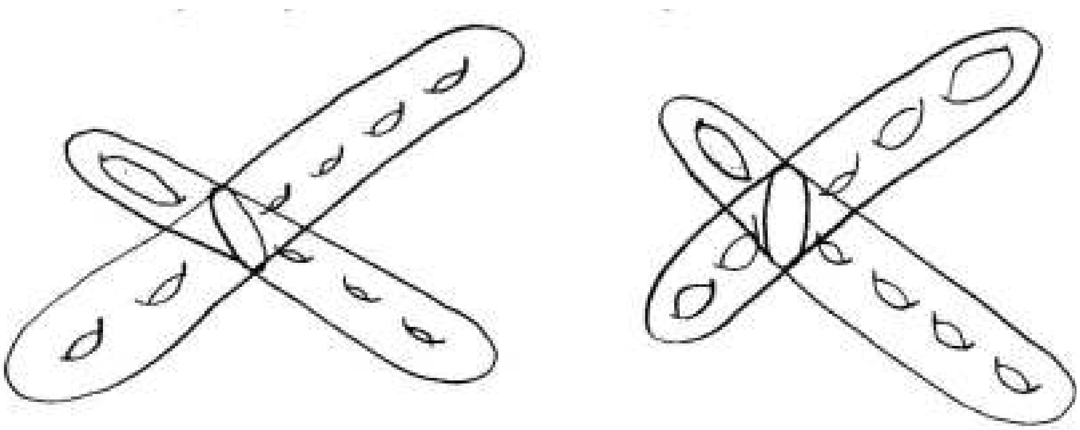


9.52

9.6. Strong convergence from algebraic convergence

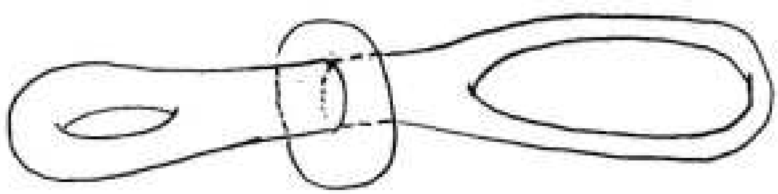
We will take another step in our study of algebraic limits. Consider the space of discrete faithful representations ρ of a fixed torsion free group Γ in $\mathrm{PSL}_2(\mathbb{C})$. The set $\Pi_\rho \subset \Gamma$ of parabolics—i.e., elements $\gamma \in \Gamma$ such that $\rho(\gamma)$ is parabolic—is an important part of the picture; we shall assume that $\Pi_\rho = \Pi$ is constant. When a sequence ρ_i converges algebraically to a representation ρ where $\Pi = \Pi_{\rho_i}$ is constant by $\Pi_\rho \supset \Pi$ is strictly bigger, then elements $\gamma \in \Pi_\rho - \Pi$ are called *accidental parabolics.* The incidence of accidental parabolics can create many interesting phenomena, which we will study later.

One complication is that the quotient manifolds $N_{\rho_i \Gamma}$ need not be homeomorphic; and even when they are, the homotopy equivalence given by the isomorphism of fundamental groups need not be homotopic to a homeomorphism. For instance, consider three-manifolds obtained by gluing several surfaces with boundary, of varying genus, in a neighborhood of their boundary. If every component has negative Euler characteristic, the result can easily be given a complete hyperbolic structure. The homotopy type depends only on the identifications of the boundary components of the original surfaces, but the homeomorphism type depends on the order of arrangement around each image boundary curve.



9.53

As another example, consider a thickened surface of genus 2 union a torus as shown.



It is also easy to give this a complete hyperbolic structure. The fundamental group has a presentation

$$\langle a_1, b_1, a_2, b_2, c : [a_1, b_1] = [a_2, b_2], \big[[a_1, b_1] = c, \big] = 1\rangle.$$

This group has an automorphism

$$a_1 \mapsto a_1,\ b_1 \mapsto b_1,\ c \mapsto c,\ a_2 \mapsto ca_2c^{-1},\ b_2 \mapsto cb_2c^{-1}$$

which wraps the surface of genus two around the torus. No non-trivial power of this automorphism is homotopic to a homeomorphism. From an algebraic standpoint there are infinitely many distinct candidates for the peripheral subgroups.

One more potential complication is that even when a given homotopy equivalence is homotopic to a homeomorphism, and even when the parabolic elements correspond,

there might not be a homeomorphism which preserves cusps. This is easy to picture for a closed surface group Γ: when Π is the set of conjugates of powers of a collection of simple closed curves on the surface, there is not enough information in Π to say which curves must correspond to cusps on which side of S. Another example is when Γ is a free group, and Π corresponds to a collection of simple closed curves on the boundary of a handlebody with fundamental group Γ. The homotopy class of a 9.54
simple closed curve is a very weak invariant here.

Rather than entangle ourselves in cusps and handlebodies, we shall confine ourselves to the case of real interest, when the quotient spaces admit cusp-preserving homeomorphisms.

We shall consider, then, geometrically tame hyperbolic manifolds which have a common model, (N_0, P_0). N_0 should be a compact manifold with boundary, and P_0 (to be interpreted as the "parabolic locus") should be a disjoint union of regular neighborhoods of tori and annuli on ∂N_0, with fundamental groups injecting into $\pi_1 N_0$. Each component of $\partial N_0 - P_0$ should be incompressible.

THEOREM 9.6.1. *Let (N_0, P_0) be as above. Suppose that $\rho_i : \pi_1 N \to \mathrm{PSL}(2, \mathbb{C})$ is a sequence of discrete, faithful representations whose quotient manifolds N_i are geometrically tame and admit homeomorphisms (in the correct homotopy class) to N_0 taking horoball neighborhoods of cusps to P_0. If $\{\rho_i\}$ converges algebraically to a representation ρ, then the limit manifold N is geometrically tame, and admits a homeomorphism (in the correct homotopy class) to N_0 which takes horoball neighborhoods of cusps to P_0.*

We shall prove this first with an additional hypothesis:

9.6.1a. Suppose also that no non-trivial non-peripheral simple curve of a component of $\partial N_0 - P_0$ *is homotopic (in N_0) to P_0.* 9.55

REMARKS. The proof of 9.6.1 (without the added hypothesis) will be given in §9.8. There is no §9.8.

The main case is really when all N_i are geometrically finite. One of the main corollaries, from 8.12.4, is that $\rho(\pi_1 N_0)$ satisfies the property of Ahlfors: its limit set has measure 0 or measure 1.

PROOF OF 9.6.1a. It will suffice to prove that every sequence $\{\rho_i\}$ converging algebraically to ρ has a subsequence converging strongly to ρ. Thus, we will pass to subsequences whenever it is convenient.

Let $S_1, \dots, S_k$ be the components of $\partial N_0 - P_0$, each equipped with a complete hyperbolic metric of finite area. (In other words, their boundary components are made into punctures.) For each i, let P_i denote a union of horoball neighborhoods of cusps of N_i, and let $E_{i,1}, \ldots, E_{i,k}$ denote the ends of $N_i - P_i$ corresponding to $S_1, \ldots, S_k$.

Some of the $E_{i,j}$ may be geometrically finite, others geometrically infinite. We can pass (for peace of mind) to a subsequence so that for each i, the $E_{i,j}$ are all geometrically finite or all geometrically infinite. We pass to a further subsequence so the sequences of bending or ending laminations $\{\beta_{i,j}\}_i$ or $\{\epsilon_{i,j}\}_i$ converge in $\mathcal{GL}_{(S_j)}$. Let χ_j be the limit.

If χ_j is realizable in N, then all nearby laminations have realizations for all representations near ρ, and the $E_{i,j}$ must have been geometrically finite. An uncrumpled surface U realizing χ_j is in the convex hull M of N and approximable by boundary 9.56
components of the convex hulls M_i Since the limit set cannot suddenly increase in the algebraic limit (p. 9.8), U must be a boundary component.

If χ_j is not realizable in N, then it must be the ending lamination for some geometrically infinite tame end E of the covering space of N corresponding to $\pi_1 S_j$, since we have hypothesized away the possibility that it represents a cusp. In view of 9.2.2 and 9.4.2, the image E_j of E in $N - P$ is a geometrically tame end of $N - P$, and $\pi_1 E = \pi_1 S_j$ has finite index in $\pi_1 E_j$.

In either case, we obtain embeddings in $N - P$ of oriented surfaces S_j' finitely covered by $S_{j[\epsilon,\infty)}$. We may assume (after an isotopy) that these embeddings are disjoint, and each surface cuts off (at least) one piece of $N-P$ which is homeomorphic to the product $S_j' \times [0,\infty)$. Since (N,P) is homotopy equivalent to (N_0, P_0), the image of the cycle $\sum \left[S_{j[\epsilon,\infty)}, \partial S_{j[\epsilon,\infty)}\right]$ in (N,P) bounds a chain C with compact support. Except in a special case to be treated later, the S_j' are pairwise non-homotopic and the fundamental group of each S_j' maps isomorphically to a unique side in $N - P$. C has degree 0 "outside" each S_j' and degree some constant l elsewhere. Let N' be the region of $N - P$ where C has degree l. We see that N is geometrically tame, and homotopy equivalent to N'.

The Euler characteristic is a homotopy invariant, so $\chi(N) = \chi(N') = \chi(N_0)$. This imples $\chi(\partial N') = \chi(\partial N_0)$ (by the formula $\chi(\partial M^3) = 2\chi(M^3)$) so in fact the 9.57
finite sheeted covering $S_{j[\epsilon,\infty)} \to S_j'$ has only one sheet—it is a homeomorphism.

Let Q be the geometric limit of any subsequence of the N_i. N is a covering space of Q. Every boundary component of the convex hull M of N is the geometric limit of boundary components of the M_i; consequently, M covers the convex hull of Q. This covering can have only finitely many sheets, since $M - P$ is made of a compact part together with geometrically infinite tame ends. Any element $\alpha \in \pi_1 Q$ has some finite power $\alpha^k \in \pi_1 N$ $[k \geq 1]$. In any torsion-free subgroup of $\mathrm{PSL}(2,\mathbb{C})$, an element has at most one k-th root (by consideration of axes). If we write α as the limit of elements $\rho_i(g_i)$, $g_i \in \pi_1 N_0$, by this remark, g_i must be eventually constant so α is actually in the algebraic limit $\pi_1 N$. $Q = N$, and ρ_i converges strongly to ρ.

A cusp-preserving homeomorphism from N to some N_i, hence to N_0, can be constructed by using an approximate isometry of N' with a submanifold of $N_i - P_i$, for high enough i. The image of N' is homotopy equivalent to N_i, so the fundamental group of each boundary component of N' must map surjectively, as well as injectively, to the fundamental group of the neighboring component of $(N_i, P_i) - N'$. This implies that the map of N' into N_i extends to a homeomorphism from N to N_i.

There is a special case remaining. If any pair of the surfaces S_i' constructed in $N - P$ is homotopic, perform all such homotopies. Unless $N - P$ is homotopy equivalent to a product, the argument continues as before—there is no reason the cover of S_i' must be a connected component of $\partial N_0 - P_0$.

When $N - P$ is homotopy equivalent to the oriented surface S_1' in it, then by a standard argument $N_0 - P_0$ must be homeomorphic to $S_1' \times I$. This is the case essentially dealt with in 9.2. The difficulty is to control both ends of $N - P$—but the argument of 9.2 shows that the ending or bending laminations of the two ends of $N_i - P_i$ cannot converge to the same lamination, otherwise the limit of some intermediate surface would realize χ_i. This concludes the proof of 9.6.1a. □

9.7. Realizations of geodesic laminations for surface groups with extra cusps, with a digression on stereographic coordinates

In order to analyze geometric convergence, and algebraic convergence in more general cases, we need to clarify our understanding of realizations of geodesic laminations for a discrete faithful representation ρ of a surface group $\pi_1(S)$ when certain non-peripheral elements of $\pi_1(S)$ are parabolic. Let $N = N_{\rho\pi_1 S}$ be the quotient three-manifold. Equip S with a complete hyperbolic structure with finite area. As in §8.11, we may embed S in N, cutting it in two pieces the "top" N_+ and the "bottom" N_-. Let γ_+ and γ_- be the (possibly empty) cusp loci for N_+ and N_-, and denote by $S_{1+}, \dots, S_{j+}$ and $S_{1-}, \dots, S_{k-}$ the components of $S - \gamma_+$ and $S - \gamma_-$ (endowed with complete hyperbolic structures with finite area). Let $E_{1+}, \dots, E_{j+}$ and $E_{1-}, \dots, E_{k-}$ denote the ends of $N - P$, where P is the union of horoball neighborhoods of all cusps.

A compactly supported lamination on S_{i+} or S_{i-} defines a lamination on S. In particular, $\epsilon(E_{i\pm})$ may be thought of as a lamination on S for each geometrically infinite tame end of $E_{i\pm}$.

Proposition 9.7.1. *A lamination $\gamma \in \mathcal{GL}_0(S)$ is realizable in N if and only if γ contains no component of γ_+, no component of γ_-, and no $\epsilon(E_{i+})$ or $\epsilon(E_{i-})$.*

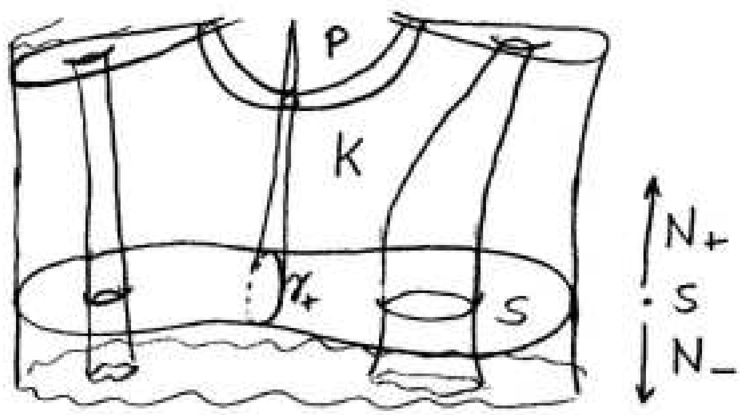
Proof. If γ contains any unrealizable lamination, it is unrealizable, so the necessity of the condition is immediate.

Let $\gamma \in \mathcal{ML}_0(S)$ be any unrealizable compactly supported measured lamination. If γ is not connected, at least one of its components is unrealizable, so we need only consider the case that γ is connected. If γ has zero intersection number with any components of γ_+ or γ_-, we may cut S along this component, obtaining a simpler surface S'. Unless γ is the component of γ_+ or γ_- in question, S' supports γ, so we pass to the covering space of N corresponding to $\pi_1 S'$. The new boundary components of S' are parabolic, so we have made an inductive reduction of this case.

We may now suppose that γ has positive intersection number with each component of γ_+ and γ_-. Let $\{\beta_i\}$ be a sequence of measures, supported on simple closed curves non-parabolic in N which converges to γ. Let $\{U_i\}$ be a sequence of uncrumpled surfaces realizing the β_i. If U_i penetrates far into a component of P 9.60
corresponding to an element α in γ_+ or γ_-, then it has a large ball mapped into P; by area considerations, this ball on U_i must have a short closed loop, which can only be in the homotopy class of α. Then the ratio

$$l_S(\beta_i)/i(\beta_i, \alpha) \geq l_{U_i}(\beta_i)/i(\beta_i, \alpha)$$

is large. Therefore (since $i(\gamma, \alpha)$ is positive and $l_S(\gamma)$ is finite) the U_i, away from their cusps, remain in a bounded neighborhood of $N - P$ in N. If γ_+ (say) is non-empty, one can now find a compact subset K of N so that any U_i intersecting N_+ must intersect K.



By the proof of 8.8.5, if infinitely many U_i intersected K, there would be a convergent subsequence, contradicting the non-realizability of γ. The only remaining possibility is that we have reached, by induction, the case that either N_+ or N_- has no extra cusps, and γ is an ending lamination. 9.61

A general lamination $\gamma \in \mathcal{GL}(S)$ is obtained from a possibly empty lamination which admits a compactly supported measure by the addition of finitely many non-compact leaves. (Let $\delta \subset \gamma$ be the maximal lamination supporting a positive transverse measure. If l is any leaf in $\gamma - \delta$, each end must come close to δ or go to ∞

in S, otherwise one could enlarge δ. By area considerations, such leaves are finite in number.) From §8.10, γ is realizable if and only if δ is. $\square$

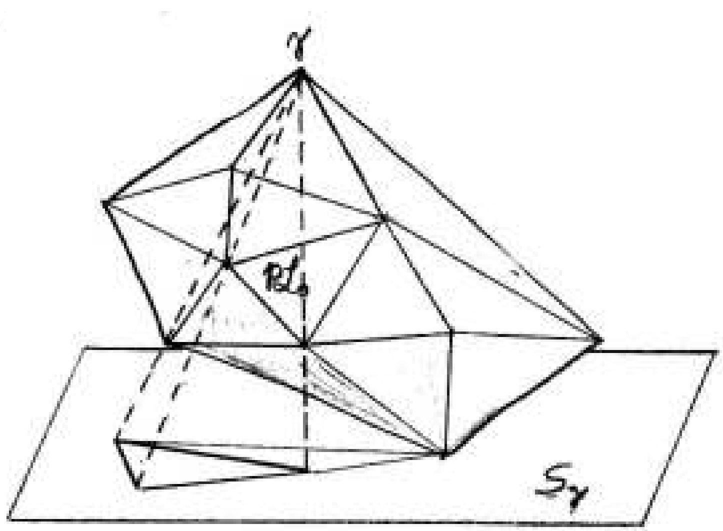
The picture of unrealizable laminations in $\mathcal{PL}_0(S)$ is the following. Let Δ_+ consist of all projective classes of transverse measures (allowing degenerate non-trivial cases) on $\chi_+ = \gamma_+ \cup U_i \epsilon(E_{i+})$. Δ_+ is convex in a coordinate system V_τ coming from any train track τ carrying χ_+.

To see a larger, complete picture, we must find a larger natural coordinate system. This requires a little stretching of our train tracks and imaginations. In fact, it is possible to find coordinate systems which are quite large. For any $\gamma \in \mathcal{PL}_0$, let $\Delta_\gamma \subset \mathcal{PL}_0$ denote the set of projective classes of measures on γ.

Proposition 9.7.2. *Let γ be essentially complete. There is a sequence of train tracks τ_i, where τ_i is carried by τ_{i+1}, such that the union of natural coordinate systems $S_\gamma = U_i V_{\tau_i}$ contains all of $\mathcal{PL}_0 - \Delta_\gamma$.*

The proof will be given presently.

Since τ_i is carried by τ_{i+1}, the inclusion $V_{\tau_i} \subset V_{\tau+1}$ is a projective map (in $\mathcal{ML}_0$, 9.62
the inclusion is linear). Thus S_γ comes naturally equipped with a projective structure. We have not made this analysis, but the typical case is that $\gamma = \Delta_\gamma$. We think of S_γ as a stereographic coordinate system, based on projection from γ. (You may imagine $\mathcal{PL}_0$ as a convex polyhedron in $\mathbb{R}^n$, so that changes of stereographic coordinates are piecewise projective, although this finite-dimensional picture cannot be strictly correct, since there is no fixed subdivision sufficient to make all coordinate changes.)



Corollary 9.7.3. *$\mathcal{PL}_0(S)$ is homeomorphic to a sphere.*

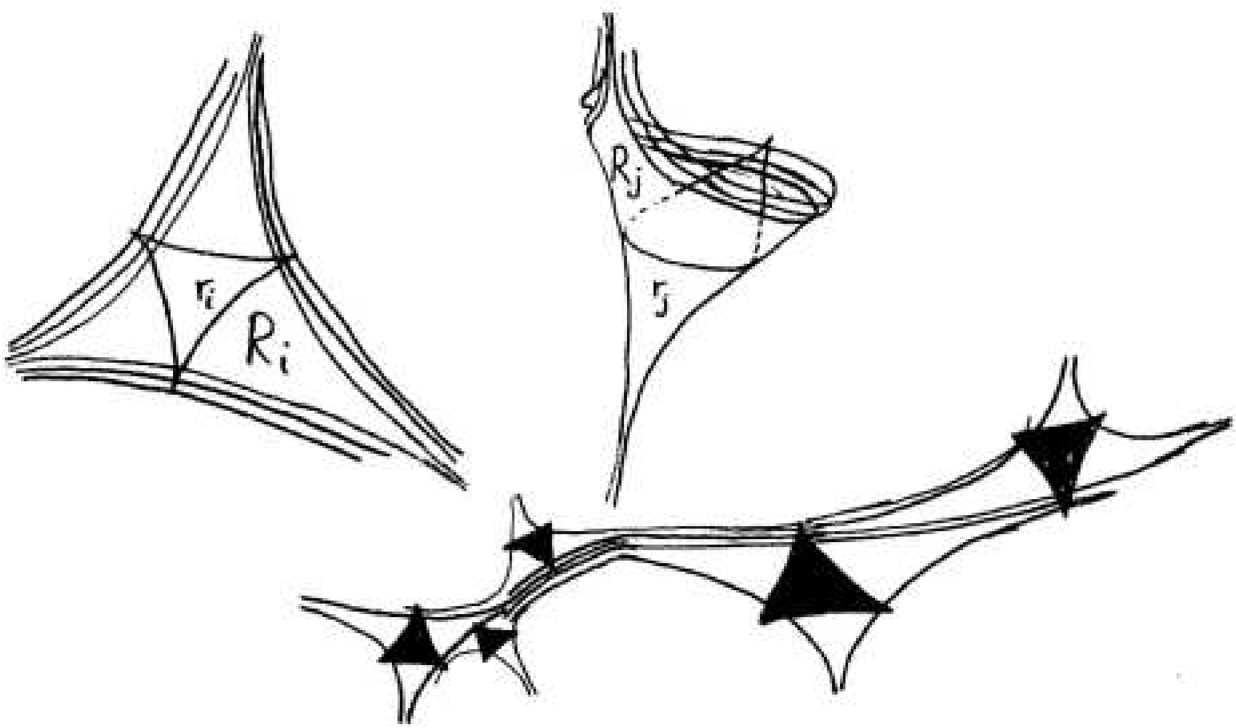
Proof that 9.7.2 implies 9.7.3. Let $\gamma \in \mathcal{PL}_0(S)$ be any essentially complete lamination. Let τ be any train track carrying γ. Then $\mathcal{PL}_0(S)$ is the union of two coordinate systems $V_\tau \cup S_\tau$, which are mapped to convex sets in Euclidean space. 9.63
If $\Delta_\gamma \neq \gamma$, nonetheless the complement of Δ_γ in V_τ is homeomorphic to $V_\tau - \gamma$, so $\mathcal{PL}_0(S)$ is homeomorphic to the one-point compactification of S_γ. □

Corollary 9.7.4. *When $\mathcal{PL}_0(S)$ has dimension greater than 1, it does not have a projective structure.* (In other words, the *pieces* in changes of coordinates have not been eliminated.)

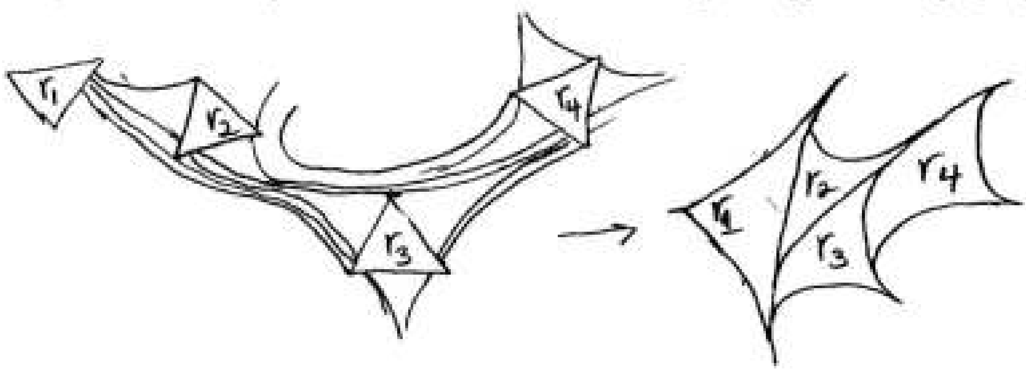
Proof that 9.7.3 implies 9.7.4. The only projective structure on S^n, when $n > 1$, is the standard one, since S^n is simply connected. The binary relation of antipodality is natural in this structure. What would be the antipodal lamination for a simple closed curve α? It is easy to construct a diffeomorphism fixing α but moving any other given lamination. (If $i(\gamma, \alpha) \neq 0$, the Dehn twist around α will do.) □

Remark. When $\mathcal{PL}_0(S)$ is one-dimensional (that is, when S is the punctured torus or the quadruply punctured sphere), the PIP structure *does* come from a projective structure, equivalent to $\mathbb{R}P^1$. The natural transformations of $\mathcal{PL}_0(S)$ are necessarily integral—in $\mathrm{PSL}_2(\mathbb{Z})$.

Proof of 9.7.2. Don't blink. Let γ be essentially complete. For each region R_i of $S - \gamma$, consider a smaller region r_i of the same shape but with finite points, rotated so its points alternate with cusps of R_i and pierce very slightly through the sides of R_i, ending on a leaf of γ. 9.64

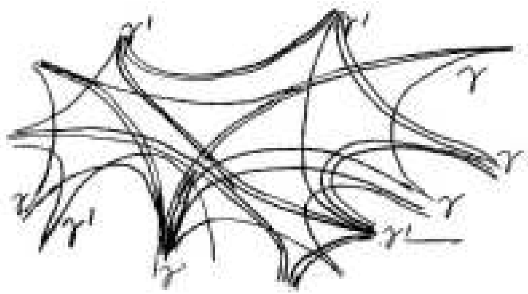


By 9.5.4, 9.5.2 and 9.3.9, both ends of each leaf of γ are dense in γ, so the regions r_i separate leaves of γ into arcs. Each region of $S - \gamma - U_i r_i$ must be a rectangle with two edges on ∂r_i and two on γ, since r_i covers the "interesting" part of R_i. (Or, prove this by area, χ). Collapse all rectangles, identifying the r_i edges with each other, and obtain a surface S' homotopy-equivalent to S, made of $U_i r_i$, where ∂r_i projects to a train track τ. (Equivalently, one may think of $S - U_i r_i$ as made of very wide corridors, with the horizontal direction given approximately by γ).

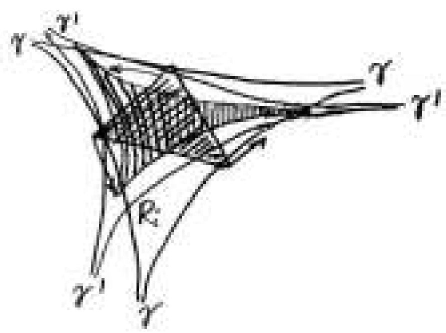


9.65

If we take shrinking sequences of regions $r_{i,j}$ in this manner, we obtain a sequence of train tracks τ_j which obviously have the property that τ_j carries τ_k when $j > k$. Let $\gamma' \in \mathcal{PL}_0(S) - \Delta_\gamma$ be any lamination not topologically equivalent to γ. From the density in γ of ends of leaves of γ, it follows that whenever leaves of γ and γ' cross, they cross at an angle. There is a lower bound to this angle. It also follows that $\gamma \cup \gamma'$ cuts S into pieces which are compact except for cusps of S.

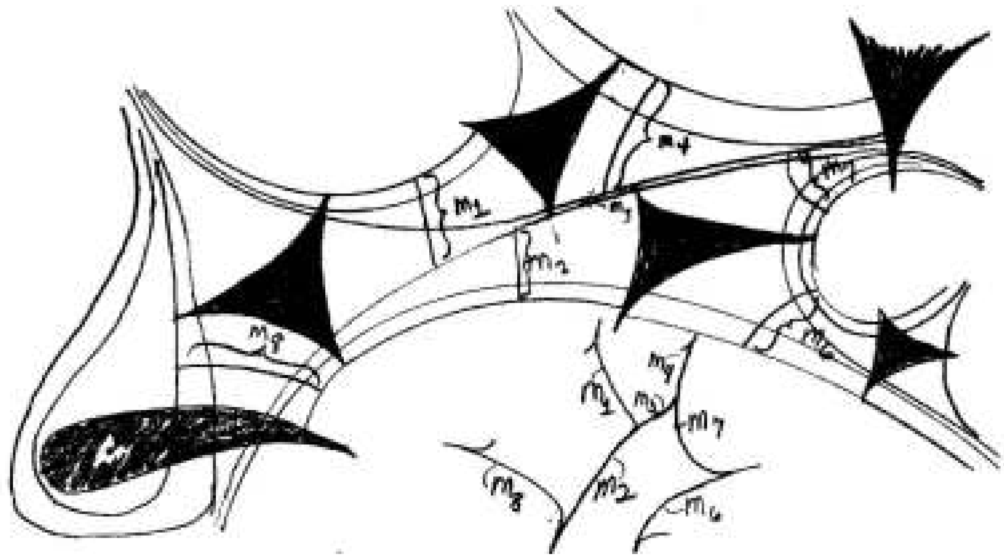


When R_i is an asymptotic triangle, for instance, it contains exactly one region of $S - \gamma - \gamma'$ which is a hexagon, and all other regions of $S - \gamma - \gamma'$ are rectangles. For sufficiently high j, the r_{ij} can be isotoped, without changing the leaves of γ which they touch, into the complement of γ'. It follows that γ' projects nicely to τ_j.



□

Stereographic coordinates give a method of computing and understanding intersection number. The transverse measure for γ projects to a "tangential" measure ν_γ on each of the train tracks τ_i: i.e., $\nu_\gamma(b)$ is the γ-transverse length of the sides of the rectangle projecting to b.



It is clear that for any $\alpha \in \mathcal{ML}_0$ which is determined by a measure μ_α on τ_i

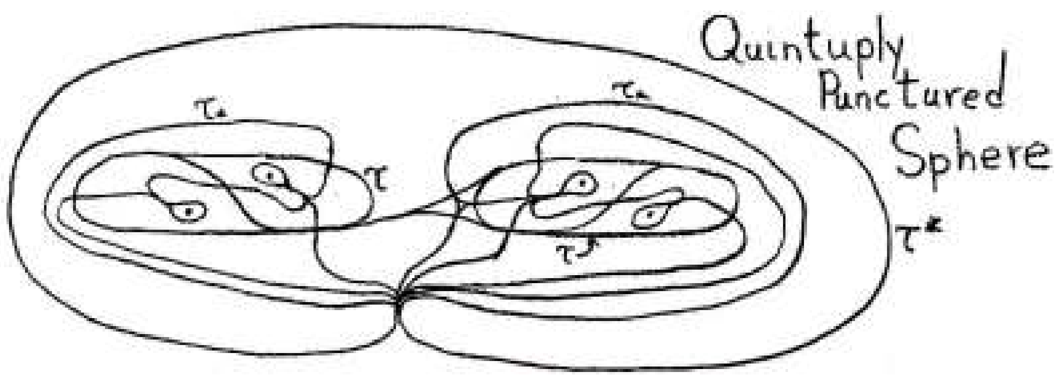
9.7.5. $$i(\alpha, \gamma) = \sum_b \mu_\alpha(b) \cdot \nu_\gamma(b).$$

Thus, in the coordinate system V_{τ_i} in $\mathcal{ML}_0$, intersection with γ is a linear function.

To make this observation more useful, we can reverse the process of finding a family of "transverse" train tracks τ_i depending on a lamination γ. Suppose we are given 9.67
an essentially complete train track τ, and a non-negative function (or "tangential" measure) ν on the branches of b, subject only to the triangle inequalities

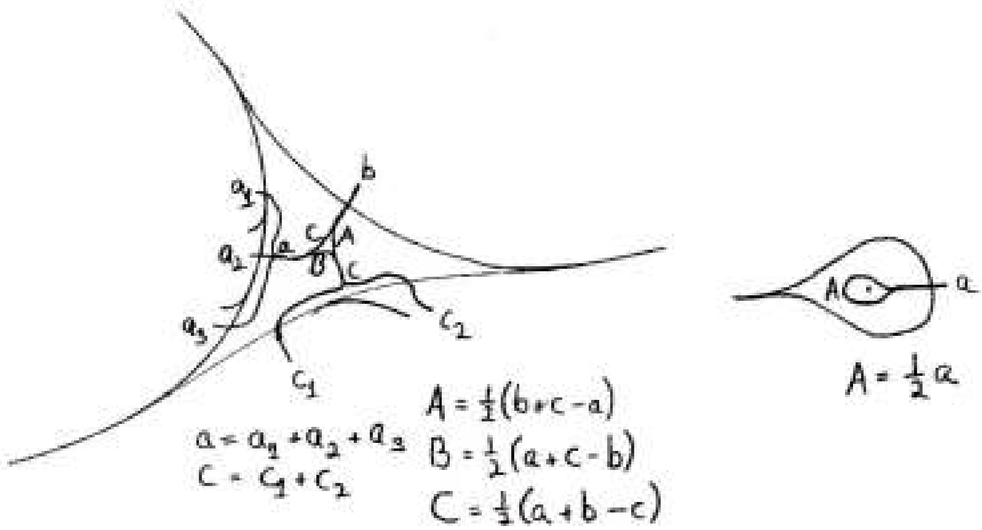
$$a + b - c \geq 0 \quad a + c - b \geq 0 \quad b + c - a \geq 0$$

whenever a, b and c are the total ν-lengths of the sides of any triangle in $S - \tau$. We shall construct a "train track" τ^* dual to τ, where we permit regions of $S - \tau^*$ to be bigons as well as ordinary types of admissible regions—let us call τ^* a *bigon track*.

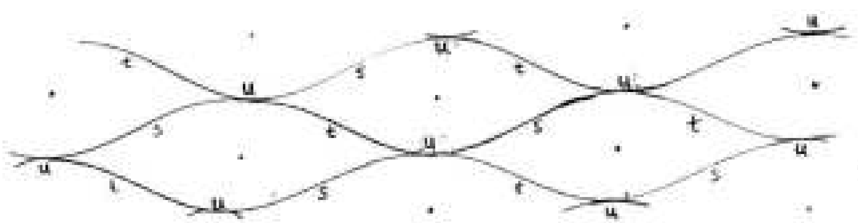


τ^* is constructed by shrinking each region R_i of $S-\tau$ and rotating to obtain a region $R_i^* \subset R_i$ whose points alternate with points of R_i. These points are joined using one more branch b^* crossing each branch b of τ; branches b_1^* and b_2^* are confluent at a vertex of R^* whenever b_1 and b_2 lie on the same side of R. Note that there is a bigon in $S-\tau^*$ for each switch in τ.

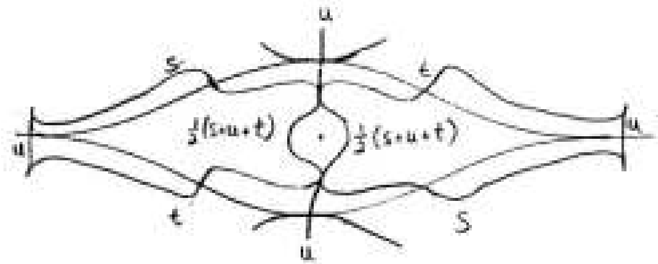
The tangential measure ν for τ determines a transverse measure defined on the branches of τ^* of the form b^*. This extends uniquely to a transverse for τ^* when S is not a punctured torus. 9.68



When S is the punctured torus, then τ must look like this, up to the homeomorphism (drawn on the abelian cover of $T-p$):



Note that each side of the punctured bigon is incident to each branch of τ. Therefore, the tangential measure ν has an extension to a transverse measure ν^* for τ^*, which is unique if we impose the condition that the two sides of R^* have equal transverse measure.

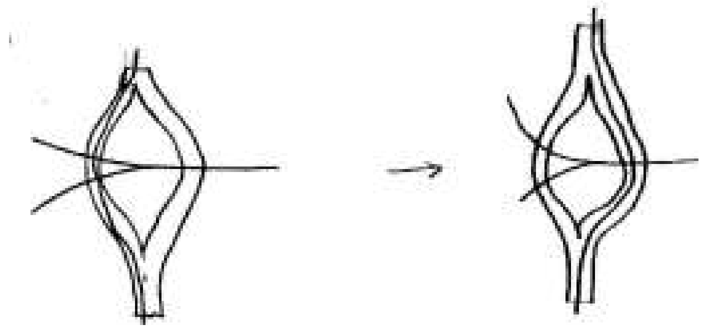


9.69

A transverse measure on a bigon track determines a measured geodesic lamination, by the reasoning of 8.9.4. When τ is an essentially complete train track, an open subset of $\mathcal{ML}_0$ is determined by a function μ on the branches of τ subject to a condition for each switch that

$$\sum_{b \in \mathcal{I}} \mu(b) = \sum_{b \in \mathcal{O}} \mu(b),$$

where $\mathcal{I}$ and $\mathcal{O}$ are the sets of "incoming" and "outgoing" branches. Dually, "tangential" measure ν on the branches of τ determines an element of $\mathcal{ML}_0$ (via ν^*), but two functions ν and ν' determine the same element if ν is obtained from ν' by a process of adding a constant to the incoming branches of a switch, and subtracting the same constant from the outgoing branches—or, in other words, if $\nu - \nu'$ annihilates all transverse measures for τ (using the obvious inner product $\nu \cdot \mu = \sum \nu(b)\mu(b)$). In fact, this operation on ν merely has the effect of switching "trains" from one side of a bigon to the other.



(Some care must be taken to obtain ν' from ν by a sequence of elementary "switching" operations without going through negative numbers. We leave this as an exercise to 9.70
the reader.)

Given an essentially complete train track τ, we now have two canonical coordinate systems V_τ and V_τ^* in $\mathcal{ML}_0$ or $\mathcal{PL}_0$. If $\gamma \in V_\tau$ and $\gamma^* \in V_\tau^*$ are defined by measures μ_γ and ν_{γ^*} on τ, then $i(\gamma, \gamma^*)$ is given by the inner product

$$i(\gamma, \gamma^*) = \sum_{b \in \tau} \mu_\gamma(b) \nu_{\gamma^*}(b).$$

To see this, consider the universal cover of S. By an Euler characteristic or area argument, no path on $\tilde{\tau}$ can intersect a path on $\tilde{\tau}^*$ more than once. This implies the formula when γ and γ' are simple geodesics, hence, by continuity, for all measured geodesic laminations.

PROPOSITION 9.7.4. *Formula* 9.7.3 *holds for all* $\gamma \in V_\tau$ *and* $\gamma^* \in V_\tau^*$. *Intersection number is a bilinear function on* $V_\tau \times V_\tau^*$ *(in* $\mathcal{ML}_0$*).* □

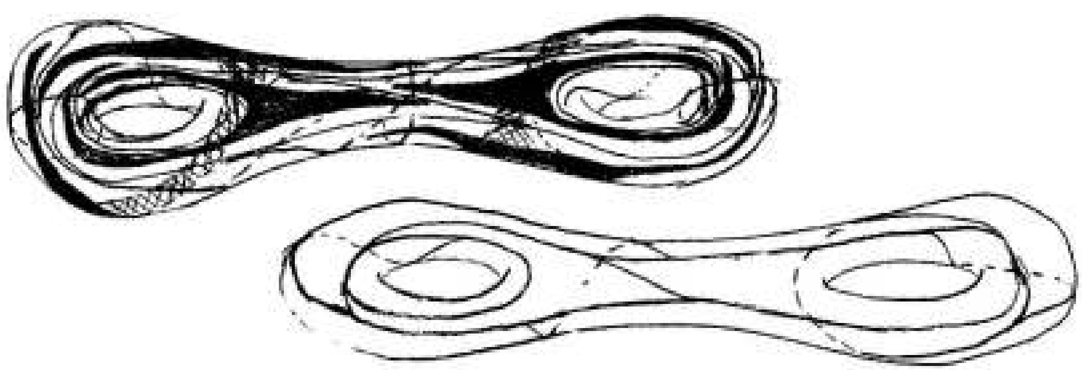
This can be interpreted as a more intrinsic justification for the linear structure on the coordinate systems V_τ—the linear structure can be reconstructed from the embedding of V_τ in the dual space of the vector space with basis $\gamma^* \in V_\tau^*$.

COROLLARY 9.7.5. *If* $\gamma, \gamma' \in \mathcal{ML}_0$ *are not topologically conjugate and if at least one of them is essentially complete, then there are neighborhoods* U *and* U' *of* γ *and* γ' *with linear structures in which intersection number is bilinear.*

9.71

PROOF. Apply 9.7.4 to one of the train tracks τ_i constructed in 9.7.2. □

REMARK. More generally, the only requirement for obtaining this local bilinearity near γ and γ' is that the complementary regions of $\gamma \cup \gamma'$ are "atomic" and that $S - \gamma$ have no closed non-peripheral curves. To find an appropriate τ, simply burrow out regions of r_i, "transverse" to γ with points going between strands of γ', so the regions r_i cut all leaves of γ into arcs. Then collapse to a train track carrying γ' and "transverse" to γ, as in 9.7.2.



What is the image of $\mathbb{R}^n$ of stereographic coordinates S_γ for $\mathcal{ML}_0(S)$? To understand this, consider a system of train tracks

$$\tau_1 \longrightarrow \tau_2 \longrightarrow \cdots \longrightarrow \tau_k \longrightarrow \cdots$$

defining S_γ. A "transverse" measure for τ_i pushes forward to a "transverse" measure for τ_j, for $j > i$. If we drop the restriction that the measure on τ_i is non-negative, still it often pushes forward to a positive measure on τ_j. The image of S_γ is the set of 9.72

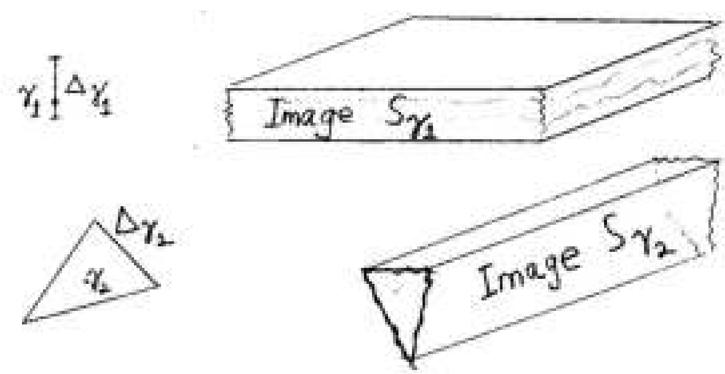
such arbitrary "transverse" measures on τ_1 which eventually become positive when pushed far enough forward.

For $\gamma' \in \Delta_\gamma$, let $\nu_{\gamma'}$ be a "tangential" measure on τ_1 defining γ'.

Proposition 9.7.6. *The image of S_γ is the set of all "transverse," not necessarily positive, measures μ on τ_1 such that for all $\gamma' \in \Delta_\gamma$, $\nu_{\gamma'} \cdot \mu > 0$.*

(Note that the functions $\nu_{\gamma'} \cdot \mu$ and $\nu_{\gamma''} \cdot \mu$ are distinct for $\gamma' \neq \gamma''$.)

In particular, note that if $\Delta_\gamma = \gamma$, the image of stereographic coordinates for $\mathcal{ML}_0$ is a half-space, or for $\mathcal{PL}_0$ the image is $\mathbb{R}^n$. If Δ_γ is a k-simplex, then the image of S_γ for $\mathcal{PL}_0$ is of the form $\operatorname{int}(\Delta^k) \times \mathbb{R}^{n-k}$. (This image is defined only up to projective equivalence, until a normalization is made.)



Proof. The condition that $\nu_{\gamma'} \cdot \mu > 0$ is clearly necessary: intersection number $i(\gamma', \gamma'')$ for $\gamma' \in \Delta_\gamma$, $\gamma'' \in S_\gamma$ is bilinear and given by the formula 9.73

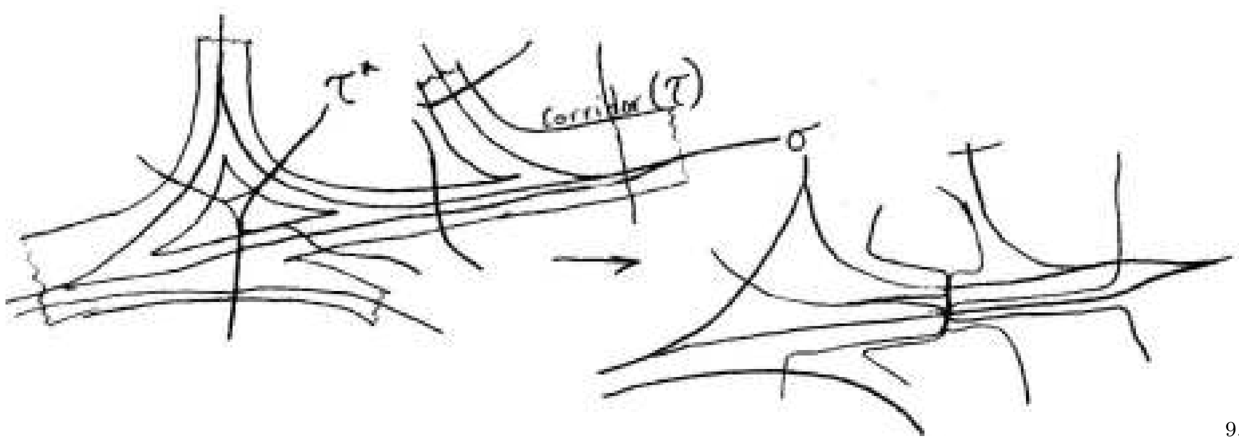
$$i(\gamma', \gamma'') = \nu_{\gamma'} \cdot \mu_{\gamma''}.$$

Consider any transverse measure μ on τ_1 such that μ is always non-positive when pushed forward to τ_i. Let b_i be a branch of τ_i such that the push-forward of μ is non-positive on b_i. This branch b_i, for high i, comes from a very long and thin rectangle ρ_i. There is a standard construction for a transverse measure coming from a limit of the average transverse counting measures of one of the sides of ρ_i. To make this more concrete, one can map ρ_i in a natural way to τ_j^* for $j \leq i$.

(In general, whenever an essentially complete train track τ carries a train track σ, then σ^* carries τ^*

$$\sigma \to \tau$$
$$\sigma^* \leftarrow \tau^*.$$

To see this, embed σ in a narrow corridor around τ, so that branches of τ^* do not pass through switches of σ. Now σ^* is obtained by squeezing all intersections of branches of τ^* with a single branch of σ to a single point, and then eliminating any bigons contained in a single region of $S - \sigma$.)



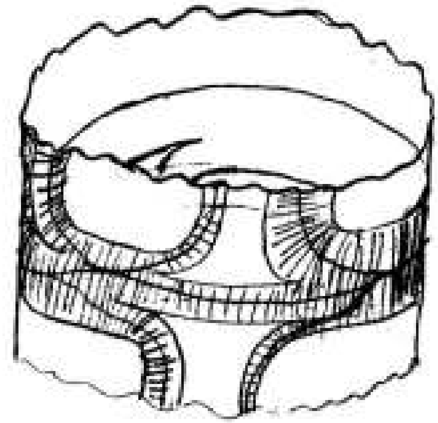
9.74

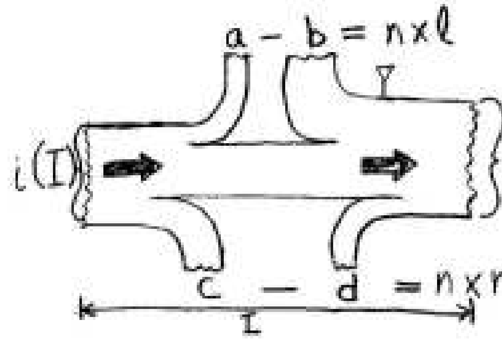
On τ_1^*, ρ_i is a finite but very long path. The average number of times ρ_i tranverses a branch of τ_1^* gives a function ν_i which almost satisfies the switch condition, but not quite. Passing to a limit point of $\{\nu_i\}$ one obtains a "transverse" measure ν for τ_1^*, whose lamination topologically equals γ, since it comes from a transverse measure on τ_i^*, for all i. Clearly $\nu \cdot \mu \leq 0$, since ν_i comes frm a function supported on a single branch b_i^* of τ_i^*, and $\mu(b_i) < 0$. $\square$

For $\gamma \in \mathcal{ML}_0$ let $Z_\gamma \subset \mathcal{ML}_0$ consist of γ' such that $i(\gamma, \gamma') = 0$. Let C_γ consist of laminations γ' not intersecting γ, i.e., such that support of γ' is disjoint from the support of γ. An arbitrary element of Z_γ is an element of C_γ, together with some measure on γ. The same symbols will be used to denote the images of these sets in $\mathcal{PL}_0(S)$.

Proposition 9.7.6. *The intersection of Z_γ with any of the canonical coordinate systems X containing γ is convex.* (In $\mathcal{ML}_0$ or $\mathcal{PL}_0$.)

Proof. It suffices to give the proof in $\mathcal{ML}_0$. First consider the case that γ is a simple closed curve and $X = V_\tau$, for some train track τ carrying γ. Pass to the cylindrical covering space C of S with fundamental group generated by γ. The path of γ on C is embedded in the train track $\tilde{\tau}$ covering τ. From a "transverse" measure m on $\tilde{\tau}$, construct corridors on C with a metric giving them the proper widths. 9.75





For any subinterval I of γ, let $\mathrm{nxr}(I)$ and $\mathrm{nxl}(I)$ be (respectively) the net right hand exiting and the net left hand exiting in the corresponding to I; in computing this, we weight entrances negatively. (We have chosen some orientation for γ). Let $i(I)$ be the initial width of I, and $f(I)$ be the final width.

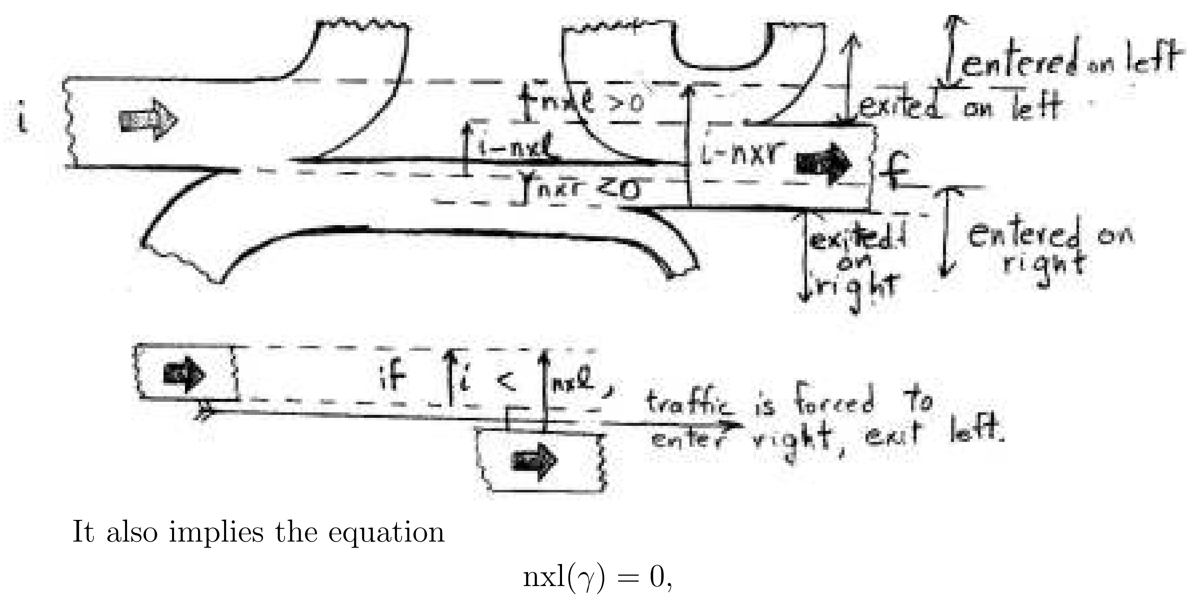
If the measure m comes from an element γ', then $\gamma' \in Z_\gamma$ if and only if there is no "traffic" entering the corridor of γ on one side and exiting on the other. This implies the inequalities

$$i(I) \geq \mathrm{nxl}(I)$$

and

$$i(I) \geq \mathrm{nxr}(I)$$

for all subintervals I.



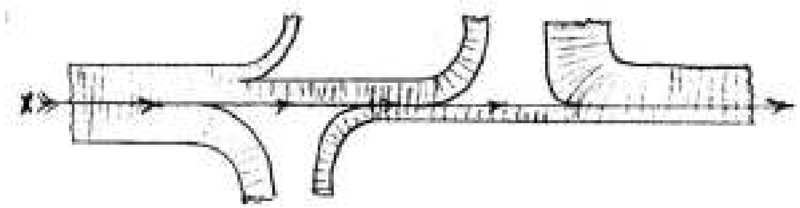
9.76

It also implies the equation

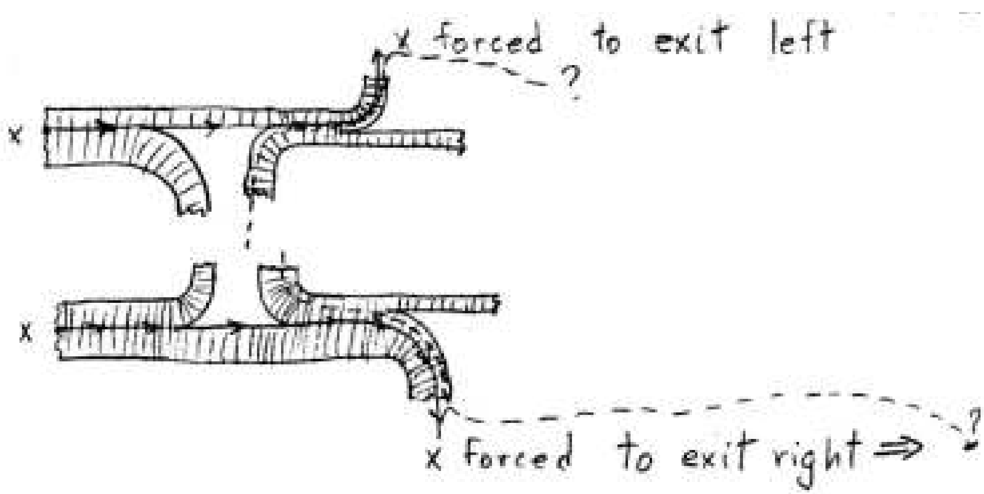
$$\mathrm{nxl}(\gamma) = 0,$$

so that any traffic travelling once around the corridor returns to its inital position. (Otherwise, this traffic would spiral around to the left or right, and be inexorably forced off on the side opposite to its entrance.)

Conversely, if these inequalities hold, then there is some trajectory going clear around the corridor and closing up. To see this, begin with any cross-section of the corridor. Let x be the supremum of points whose trajectories exit on the right. Follow the trajectory of x as far as possible around the corridor, always staying in the corridor whenever there is a choice.



The trajectory can never exit on the left—otherwise some trajectory slightly lower would be forced to enter on the right and exit on the left, or vice versa. Similarly, it can't exit on the right. Therefore it continues around until it closes up.



9.77

Thus when γ is a simple closed curve, $Z_\gamma \cap V_\tau$ is defined by linear inequalities, so it is convex.

Consider now the case $X = V_\tau$ and γ is connected but not a simple geodesic. Then γ is associated with some subsurface $M_\gamma \subset S$ with geodesic boundary defined to be the minimal convex surface containing γ. The set C_γ is the set of laminations not intersecting int (M_γ). It is convex in V_τ, since

$$C_\gamma = \bigcap \{ Z_\alpha | \alpha \text{ is a simple closed curve } \subset \text{ int } (M_\gamma) \}.$$

A general element γ' of Z_γ is a measure on $\gamma \cup \gamma''$, so Z_γ consists of convex combinations of Δ_γ and C_γ: hence, it is convex.

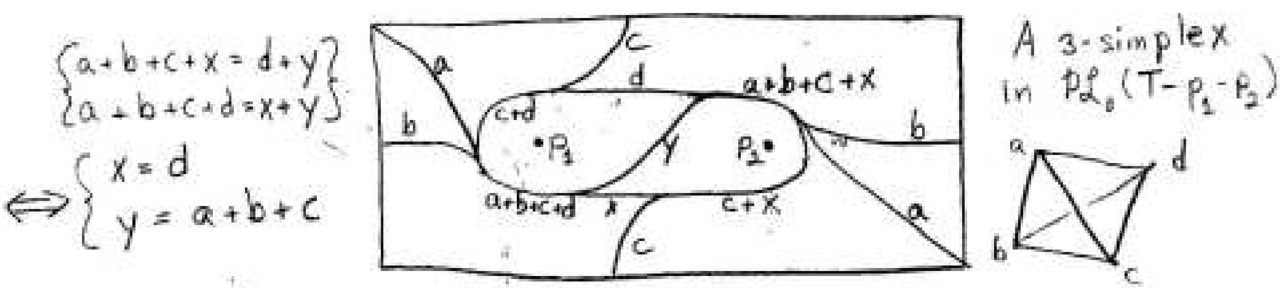
If γ is not connected, then Z_γ is convex since it is the intersection of $\{Z_{\gamma_i}\}$, where the γ_i are the components of γ.

The case X is a stereographic coordinate system follows immediately. When $X = V_\tau^*$, consider any essentially complete $\gamma \in V_\tau$. From 9.7.5 it follows that V_τ^* is linearly embedded in S_γ. (Or more directly, construct a train track (without bigons) carrying τ^*; or, apply the preceding proof to bigon track τ^*.) $\square$

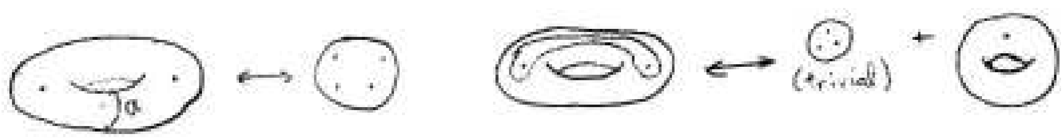
Remark. Note that when γ is a union of simple closed curves, C_γ in $\mathcal{PL}_0(S)$ is homeomorphic to $\mathcal{PL}_0(S - \gamma)$, regarded as a complete surface with finite area—i.e., 9.78
C_γ is a sphere. When γ has no component which is a simple closed curve, C_γ is convex. Topologically, it is the join of $\mathcal{PL}_0(S - \bigcup S_\gamma)$ with the simplex of measures on the boundary components of the S_{γ_i}, where the S_{γ_i} are subsurfaces associated with the components γ_i of γ.

Now we are in a position to form an image of the set of unrealizable laminations for $\rho\pi_1 S$. Let $U_+ \subset \mathcal{PL}_0$ be the union of laminations containing a component of χ_+ and define U_- similarly, so that γ is unrealizable if and only if $\gamma \in U_+ \cup U_-$. U_+ is a union of finitely many convex pieces, and it is contained in a subcomplex of $\mathcal{PL}_0$ of codimension at least one. It may be disjoint from U_-, or it may intersect U_- in an interesting way.

Example. Let S be the twice punctured torus. From a random essentially complete train track,



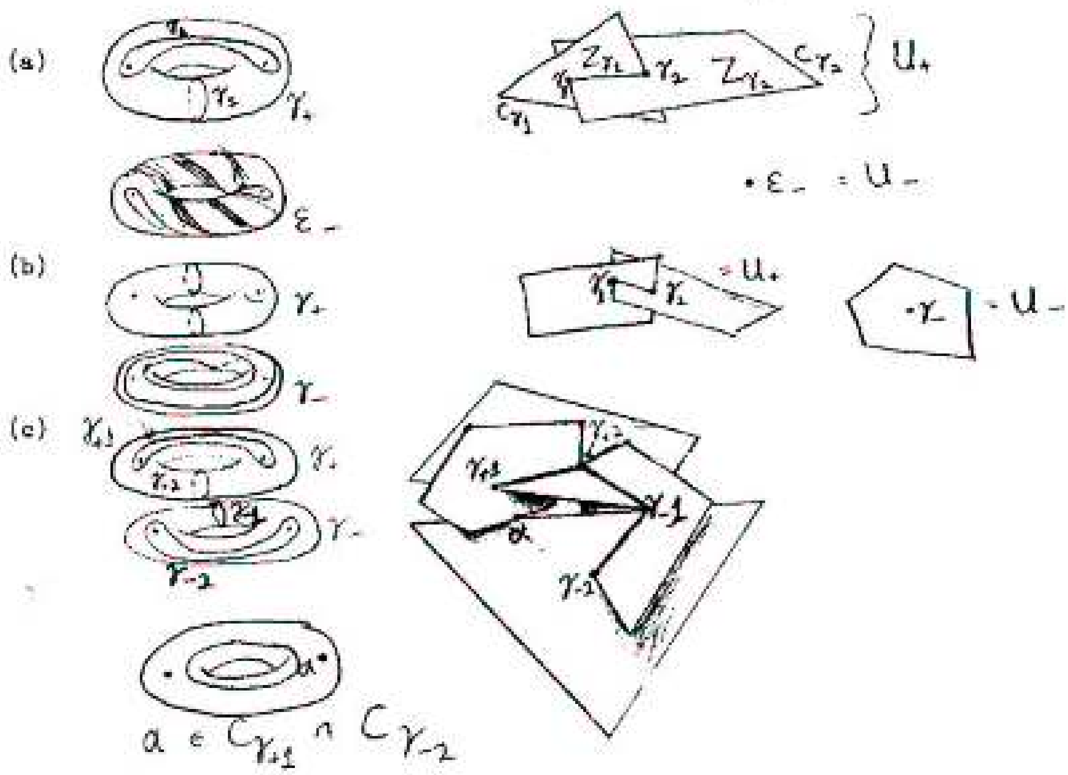
we compute that $\mathcal{ML}_0$ has dimension 4, so $\mathcal{PL}_0$ is homeomorphic to S^3. For any simple closed curve α on S, C_α is $\mathcal{PL}_0(S - \alpha)$,



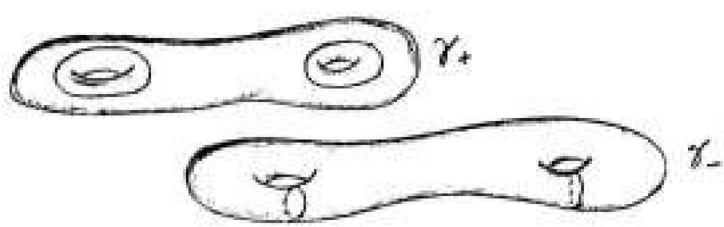
9.79

where $S - \alpha$ is either a punctured torus union a (trivial) thrice punctured sphere, or a 4-times punctured sphere. In either case, C_α is a circle, so Z_α is a disk.

Here are some sketches of what U_+ and $U-$ can look like.

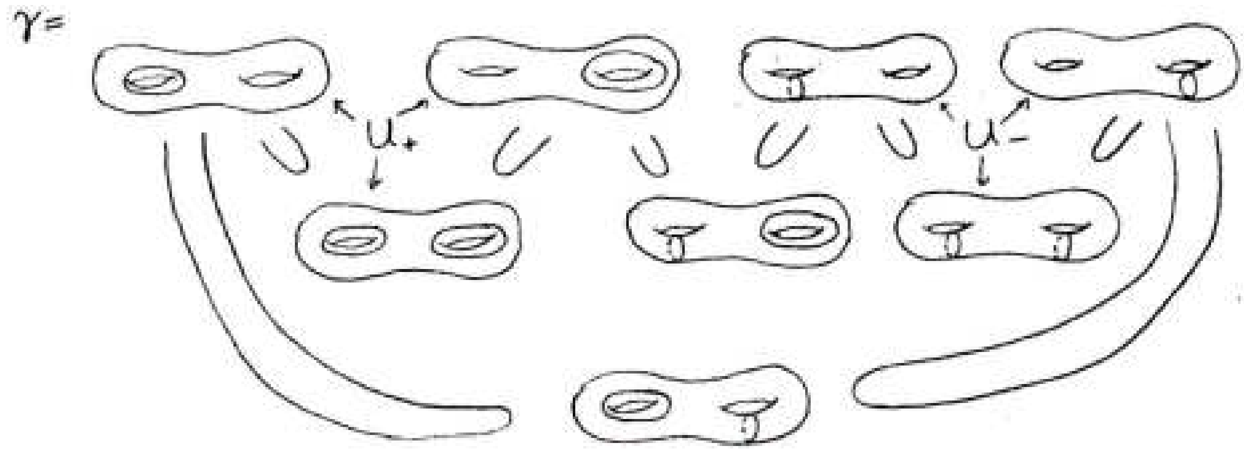


Here is another example, where S is a surface of genus 2, and $U_+(S) \cup U_-(S)$ has the homotopy type of a circle (although its closure is contractible):

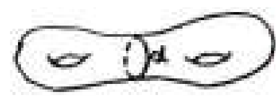


9.80

In fact, $U_+ \cup U_-$ is made up of convex sets $Z_\gamma - C_\gamma$, with relations of inclusion as diagrammed:



The closures all contain the element α; hence the closure of the union is starlike:



9.9-1

9.9. Ergodicity of the geodesic flow

We will prove a theorem of Sullivan (1979): There is no §9.8

THEOREM 9.9.1. *Let M^n be a complete hyperbolic manifold (of not necessarily finite volume). Then these four conditions are equivalent:*

(a) *The series*

$$\sum_{\gamma \in \pi_1 M^n} \exp\big(-(n-1)\, d(x_0, \gamma x_0)\big)$$

diverges. (Here, $x_0 \in H^n$ is an arbitrary point, γx_0 is the image of x_0 under a covering transformation, and $d(\ ,\ )$ is hyperbolic distance).

(b) *The geodesic flow is not dissipative. (A flow ϕ_t on a measure space (X, μ) is dissipative if there exists a measurable set $A \subset X$ and a $T > 0$ such that $\mu(A \cap \phi_t(A)) = 0$ for $t > T$, and $X = \cup_{t\phi_t(A)}$.)*

(c) *The geodesic flow on $T_1(M)$ is recurrent. (A flow ϕ_t on a measure space (X, μ) is recurrent when for every measure set $A \subset X$ of positive measure and every $T > 0$ there is a $t \geq T$ such that $\mu(A \cap \phi_t(A)) > 0$.)*

(d) *The geodesic flow on $T_1(M)$ is ergodic.*

Note that in the case M has finite volume, recurrence of the geodesic flow is immediate (from the Poincaré recurrence lemma). The ergodicity of the geodesic flow in this case was proved by Eberhard Hopf, in ??. The idea of (c) $\to$ (d) goes back to Hopf, and has been developed more generally in the theory of Anosov flows ??. 9.9-2

COROLLARY 9.9.2. *If the geodesic flow is not ergodic, there is a non-constant bounded superharmonic function on M.*

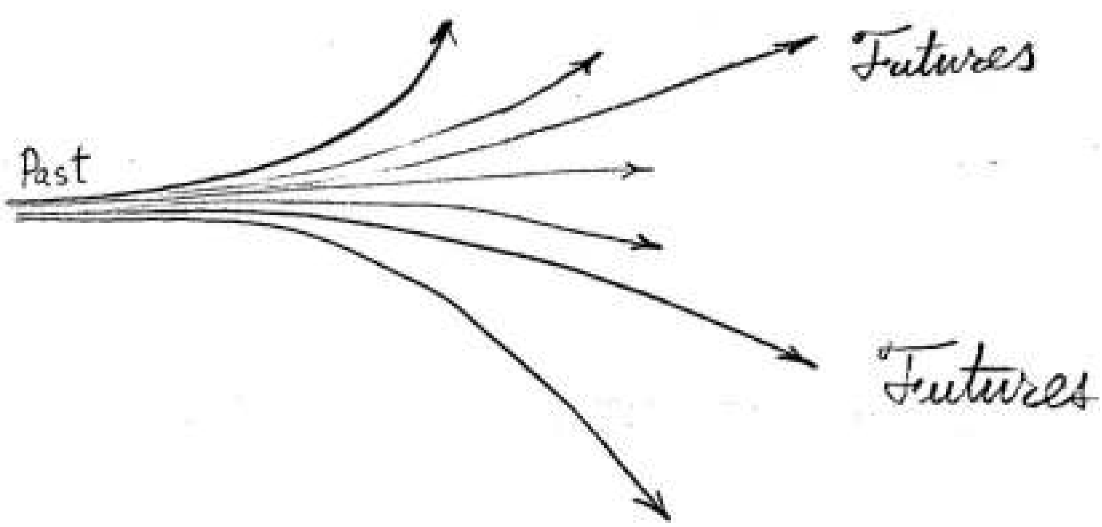
PROOF OF 9.9.2. Consider the Green's function $g(x) = \int_{d(x,x_0)}^{\infty} \sin h^{1-n} t \, dt$ for hyperbolic space. (This is a harmonic function which blows up at x_0.) By (a), the series $\sum_{\gamma \in \pi_1 M} g \circ \gamma$ converges to a function, invariant by γ, which projects to a Green's function G for M. The function $f = \arctan G$ (where $\arctan \infty = \pi/2$) is a bounded superharmonic function, since arctan is convex. □

REMARK. The convergence of the series (a) is actually equivalent to the existence of a Green's function on M, and also equivalent to the existence of a bounded superharmonic function. See (Ahlfors, Sario) for the case $n = 2$, and [ ] for the general case.

COROLLARY 9.9.3. *If Γ is a geometrically tame Kleinian group, the geodesic flow on $T_1(H^n/\Gamma)$ is ergodic if and only if $L_\Gamma = S^2$.*

PROOF OF 9.9.3. From 9.9.2 and 8.12.3. □

PROOF OF 9.9.1. Sullivan's proof of 9.9.1 makes use of the theory of Brownian motion on M^n. This approach is conceptually simple, but takes a certain amount of technical background (or faith). Our proof will be phrased directly in terms of geodesics, but a basic underlying idea is that a geodesic behaves like a random path: its future is "nearly" independent of its past. 9.9-2a



9.9-3

(d) → (c). This is a general fact. If a flow ϕ_t is not recurrent, there is some set A of positive measure such that only for t in some bounded interval is $\mu(A \cap \phi_t(A)) > 0$.

Then for any subset $B \subset A$ of small enough measure, $\cup_t \phi_t(B)$ is an invariant subset which is proper, since its intersection with A is proper.

(c) $\to$ (b). Immediate.

(b) $\to$ (a). Let B be any ball in H^n, and consider its orbit ΓB where $\Gamma = \pi_1 M$. For the series of (a) to diverge means precisely that the total apparent area of ΓG as seen from a point $x_0 \in H^n$, (measured with multiplicity) is infinite.

In general, the underlying space of a flow is decomposed into two measurable parts, $X = D \cup R$, where ϕ_t is dissipative on D (the union of all subsets of X which eventually do not return) and recurrent on R. The reader may check this elementary fact. If the recurrent part of the geodesic flow is non-empty, there is some ball B in M^n such that a set of positive measure of tangent vectors to points of B give rise to geodesics that intersect B infinitely often. This clearly implies that the series of (a) diverges.

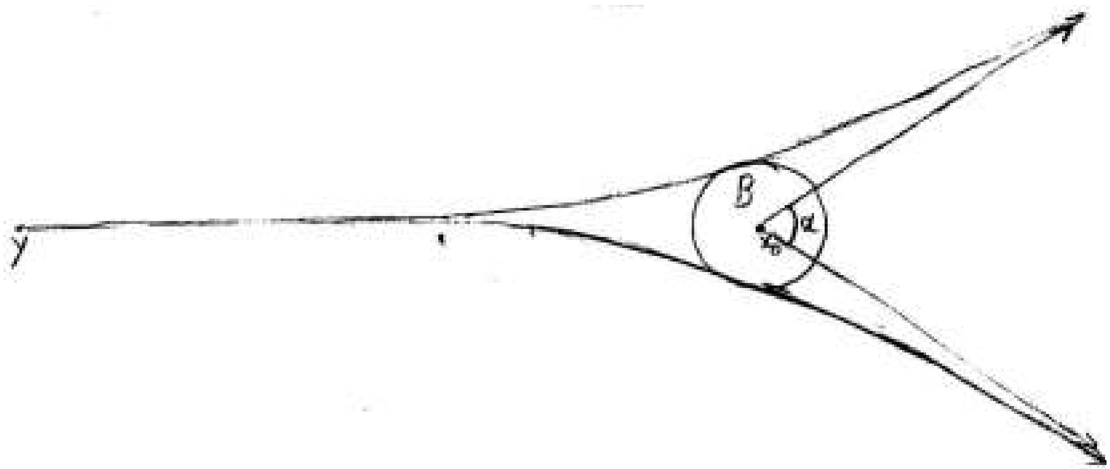
The idea of the reverse implication (a) $\to$ (b) is this: if the geodesic flow is dissipative there are points x_0 such that a positive proportion of the visual sphere is not covered infinitely often by images of some ball. Then for *each* "group" of geodesics that return to B, a definite proportion must eventually escape ΓB, because future and past are nearly independent. The series of (a) can be regrouped as a geometric progression, so it converges. We now make this more precise.

Recall that the term "visual sphere" at x_0 is a synonym to the "set of rays" emanating from x_0. It has a metric and a measure obtained from its identification with the unit sphere in the tangent space at x_0. 9.9-4

Let $x_0 \in M^n$ be any point and $B \subset M^n$ any ball. If a positive proportion of the rays emanating from x_0 pass infinitely often through B, then for a slightly larger ball B', a definite proportion of the rays emanating from *any* point $x \in M^n$ spend an infinite amount of time in B', since the rays through x are parallel to rays through x_0. Consequently, a subset of $T_1(B')$ of positive measure consists of vectors whose geodesics spend an infinite total time in $T_1(B')$; by the Poincaré recurrence lemma, the set of such vectors is a recurrent set for the geodesic flow. (b) holds so (a) $\to$ (b) is valid in this case. To prove (a) $\to$ (b), it remains to consider the case that almost every ray from x_0 eventually escapes B; we will prove that (a) fails, i.e., the series of (a) converges.

Replace B by a slightly smaller ball. Now almost every ray from almost every point $x \in M$ eventually escapes the ball. Equivalently, we have a ball $B \subset H^n$ such that for every point $x \in H^n$, almost no geodesic through x intersects ΓB, or even $\Gamma(N_\epsilon(B))$, more than a finite number of times.

Let x_0 be the center of B and let α be the infimum, for $y \in H^n$, of the diameter of the set of rays from x_0 which are parallel to rays from y which intersect B. This infimum is positive, and very rapidly approached as y moves away from x_0.



9.9-5

Let R be large enough so that for every ball of diameter greater than α in the visual sphere at x_0, at most (say) half of the rays in this ball intersect $\Gamma N_\in(B)$ at a distance greater than R from x_0. R should also be reasonably large in absolute terms and in comparison to the diameter of B.

Let x_0 be the center of B. Choose a subset $\Gamma' \subset \Gamma$ of elements such that: (i) for every $\gamma \in \Gamma$ there is a $\gamma' \in \Gamma'$ with $d(\gamma' x_0, \gamma x_0) < R$. (ii) For any γ_1 and γ_2 in Γ', $d(\gamma_1 x_0, \gamma_2 x_0) \geq R$.

Any subset of Γ maximal with respect to (ii) satisfies (i).

We will show that $\sum_{\gamma' \in \Gamma'} \exp(-(n-1)\, d(x_0, \gamma' x_0))$ converges. Since for any γ' there are a bounded number of elements $\gamma \in \Gamma$ so that $d(\gamma x_0, \gamma' x_0) < R$, this will imply that the series of (a) converges.

Let $<$ be the partial ordering on the elements of Γ' generated by the relation $\gamma_1 < \gamma_2$ when $\gamma_2 B$ eclipses $\gamma_1 B$ (partially or totally) as viewed from x_0; extend $<$ to be transitive.

Let us denote the image of γB in the visual sphere of x_0 by B_γ. Note that when $\gamma' < \gamma$, the ratio $\operatorname{diam}(B_{\gamma'})/\operatorname{diam}(B_\gamma)$ is fairly small, less than 1/10, say. Therefore $\cup_{\gamma' < \gamma} B_{\gamma'}$ is contained in a ball concentric with B_γ of radius 10/9 that of B_γ.

Choose a maximal independent subset $\Delta_1 \subset \Gamma'$ (this means there is no relation $\delta_1 < \delta_2$ for any $\delta_1, \delta_2 \in \Delta_1$). Do this by successively adjoining any γ whose B_γ has largest size among elements not less than any previously chosen member. Note that area $(\cup_{\delta \in \Delta} B_\delta)/\operatorname{area}(\cup_{\gamma \in \Gamma'} B_\gamma)$ is greater than some definite (a priori) constant: $(9/10)^{n-1}$ in our example. Inductively define $\Gamma'_0 = \Gamma'$, $\gamma'_{i+1} = \Gamma'_i - \Delta_{i+1}$ and define $\Delta_{i+1} \subset \Gamma_i$ similarly to Δ_1. Then $\Gamma' = \cup_{i=1}^{\infty} \Delta_i$. 9.9-6

For any $\gamma \in \Gamma'$, we can compare the set B_γ of rays through x_0 which intersect $\gamma(B)$ to the set C_γ of parallel rays through γX_0.

Any ray of B_γ which re-enters $\Gamma'(B)$ after passing through $\gamma'(B)$, is within ϵ of the parallel ray of C_γ by that time. At most half of the rays of C_γ ever enter $N_\epsilon(\Gamma' B)$.

The distortion between the visual measure of B_γ and that of C_γ is modest, so we can conclude that the set of reentering rays, $B_\gamma \cap \bigcup_{\gamma'<\gamma} B_{\gamma'}$, has measure less than 2/3 the measure of B_γ.

We conclude that, for each i,

$$\begin{aligned}
&\operatorname{area}\big(\bigcup_{\gamma\in\Gamma'_{i+1}} B_\gamma\big) - \operatorname{area}\big(\bigcup_{\gamma\in\Gamma'_i} B_\gamma\big)\\
&\geq 1/3 \operatorname{area}\big(\bigcup_{\delta\in\Delta_{i+1}} B_\delta\big)\\
&\geq 1/3\cdot(9/10)^{n-1}\operatorname{area}\big(\bigcup_{\gamma\in\Gamma'_i} B_\gamma\big).
\end{aligned}$$

The sequence $\{\operatorname{area}(\bigcup_{\gamma\in\Gamma'_i} B_\gamma)\}$ decreases geometrically. This sequence dominates the terms of the series $\sum_i \operatorname{area}\cup_{\delta\in\Delta_i} B_\delta = \sum_{\gamma\in\Gamma'} \operatorname{area}(B_\gamma)$, so the latter converges, which completes the proof of (a) $\to$ (b). 9.9-7

(b) $\to$ (c). Suppose $R \subset T_1(M^n)$ is any recurrent set of positive measure for the geodesic flow ϕ_t. Let B be a ball such that $R \cap T_1(B)$ has positive measure. Almost every forward geodesic of a vector in R spends an infinite amount of time in B. Let $A \subset T_1(B)$ consist of all vectors whose forward geodesics spend an infinite time in B and let ψ_t, $t \geq 0$, be the measurable flow on A induced from ϕ_t which takes a point leaving A immediately back to its next return to A.

Since ψ_t is measure preserving, almost every point of A is in the image of ψ_t for all t and an inverse flow ψ_{-t} is defined on almost all of A, so the definition of A is unchanged under reversal of time. Every geodesic parallel in either direction to a geodesic in A is also in A; it follows that $A = T_1(B)$. By the Poincaré recurrence lemma, ψ_t is recurrent, hence ϕ_t is also recurrent.

(c) $\to$ (d). It is convenient to prove this in the equivalent form, that if the action of Γ on $S^{n-1}_\infty \times S^{n-1}_\infty$ is recurrent, it is ergodic. "Recurrent" in this context means that for any set $A \subset S^{n-1}\times S^{n-1}$ of positive measure, there are an infinite number of elements $\gamma\in\Gamma$ such that $\mu(\gamma A\cap A) > 0$. Let $I \subset S^{n-1}\times S^{n-1}$ be any measurable set invariant by Γ. Let $-B_1$ and $B_2 \subset S^{n-1}$ be small balls. Let us consider what I must look like near a general point $x = (x_1, x_2) \in B_1 \times B_2$. If γ is a "large" element of Γ such that γx is near x, then the preimage of γ of a product of small ϵ-ball around γx_1 and γx_2 is one of two types: it is a thin neighborhood of one of the factors, $(x_1 \times B_2)$ or $(B_1 \times x_2)$. (γ must be a translation in one direction or the other along an axis from approximately x_1 to approximately x_2.) Since Γ is recurrent, almost every point $x \in B_1 \times B_2$ is the preimage of elements γ of both types, of an infinite number of 9.9-8 points where I has density 0 or 1. Define

$$f(x_1) = \int_{B_2} \chi_I(x_1, x_2)\, dx_2,$$

where χ_I is the characteristic function of I, for $x_1 \in B_1$ (using a probability measure on B_2). By the above, for almost every x_1 there are arbitrarily small intervals around x_1 such that the average of f in that interval is either 0 or 1. Therefore f is a characteristic function, so $I \cap B_1 \times B_2$ is of the form $S \times B_2$ (up to a set of measure zero) for some set $S \subset B_1$.

Similarly, I is of the form $B_1 \times R$, so I is either $\emptyset \times \emptyset$ or $B_1 \times B_2$ (up to a set of measure zero). $\square$

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

NOTE

Since a new academic year is beginning, I am departing from the intended order in writing these notes. For the present, the end of chapter 9 and chapters 10, 11 and 12, which depend heavily on chapters 8 and 9, are to be omitted. The tentative plan for the omitted parts is to cover the following topics:

The end of chapter 9—a more general discussion of algebraic convergence.

Chapter 10—Geometric convergence: an analysis of the possibilities for geometric limits.

Chapter 11. The Riemann mapping theorem; parametrizing quasi-conformal deformations. Extending quasi-conformal deformations of S^2_∞ to quasi-isometric deformations of H^3. Examples; conditions for the existence of limiting Kleinian groups.

Chapter 12. Boundaries for Teichmüller space, classification of diffeomorphisms of surfaces, algorithms involving the mapping class group of a surface.

CHAPTER 11

Deforming Kleinian manifolds by homeomorphisms of the sphere at infinity

A pseudo-isometry between hyperbolic three-manifolds gives rise to a quasi-conformal map between the spheres at infinity in their universal covering spaces. This is a key point in Mostow's proof of his rigidity theorem (Chapter 5). In this chapter, we shall reverse this connection, and show that a k-quasi-conformal map of S^2_∞ to itself gives rise to a k-quasi-isometry of hyperbolic space to itself. A self-map $f : X \to X$ of a metric space is a *k-quasi-isometry* if

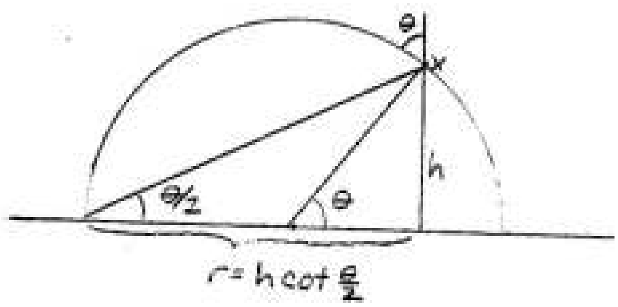
$$\frac{1}{k} d(fx, fy) \leq d(x, y) \leq k d(fx, fy)$$

for all x and y. By use of a version of the Riemann mapping theorem, the space of quasi-conformal maps of S^2 can be parametrized by the non-conformal part of their derivatives. In this way we obtain a remarkable global parametrization of quasi-isometric deformations of Kleinian manifolds by the Teichmüller spaces of their boundaries. 11.2

11.1. Extensions of vector fields

In §§8.4 and 8.12, we made use of the harmonic extensions of measurable functions on S_∞ to study the limit set of a Kleinian group. More generally, any tensor field on S^2_∞ extends, by a visual average, over H^3. To do this, first identify S^2_∞ with the unit sphere in $T_x(H^3)$, where x is a given point in H^3. If $y \in S^2_\infty$, this gives an identification $i : T_y(S^2_\infty) \to T_x(H^3)$. There is a reverse map $p : T_x(H^3) \to T_y(S^2_\infty)$ coming from orthogonal projection to the image of i. We can use i_* and p^* to take care of covariant tensor fields, like vector fields, and contravariant tensor fields, like differential forms and quadratic forms, as well as tensor fields of mixed type. The visual average of any tensor field T on S^2_∞ is thus a tensor field $\operatorname{av} T$, of the same type, on H^3. In general, $\operatorname{av} T$ needs to be modified by a constant to give it the right boundary behavior.

We need some formulas in order to make computations in the upper half-space model. Let x be a point in upper half-space, at Euclidean height h above the bounding plane $\mathbb{C}$. A geodesic through x at angle θ from the vertical hits $\mathbb{C}$ at a distance $r = h\cot(\theta/2)$ from the foot z_0 of the perpendicular from x to $\mathbb{C}$.

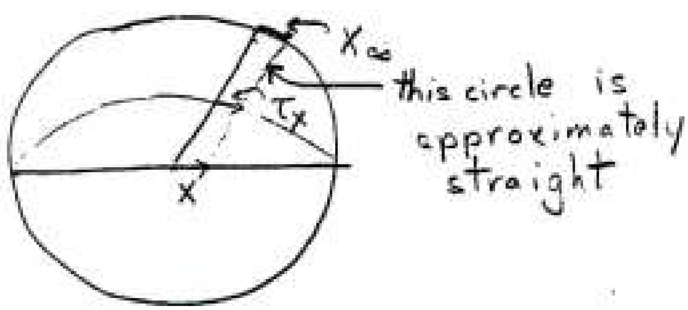


Thus, $dr = -(h/2)\csc^2(\theta/2)\,d\theta = -\frac{1}{2}(h + r^2/h)\,d\theta$. Since the map from the visual sphere at x to S^2_∞ is conformal, it follows that 11.3

$$dV_x = 4\Big(h + \frac{r^2}{h}\Big)^{-2} d\mu,$$

where μ is Lebesgue measure on $\mathbb{C}$ and V_x is visual measure at x.

Any tensor T at the point x pushes out to a tensor field T_∞ on $S^2_\infty = \hat{C}$ by the maps i^* and p_*. When X is a vector, then X_∞ is a holomorphic vector field, with derivative field, with derivative $\pm\|X\|$ at its zeros. To see this, let τ_X be the vector field representing the infinitesimal isometry of translation in the direction X. The claim is that $X_\infty = \tau_X|S_\infty$. This may be seen geometrically when X is at the center in the Poincaré disk model.



Alternatively if X is a vertical unit vector in the upper half-space, then we can compute that

$$X_\infty = -\sin\theta\,\frac{\partial}{\partial\theta} = \frac{h}{2}\;\frac{\sin\theta}{\sin^2\theta/2}\;\frac{\partial}{\partial r} = r\frac{\partial}{\partial r} = (z - z_0)\,\frac{\partial}{\partial z},$$

where z_0 is the foot of the perpendicular from x to $\mathbb{C}$. This clearly agrees with the corresponding infinitesimal isometry. (As a "physical" vector field, $\partial/\partial z$ is the same as the unit horizontal vector field, $\partial/\partial x$, on $\mathbb{C}$. The reason for this notation is that the differential operators $\partial/\partial x$ and $\partial/\partial z$ have the same action on holomorphic functions: they are directional derivatives in the appropriate direction. Even though the complex notation may at first seem obscure, it is useful because it makes it meaningful to multiply vectors by complex numbers.) 11.4

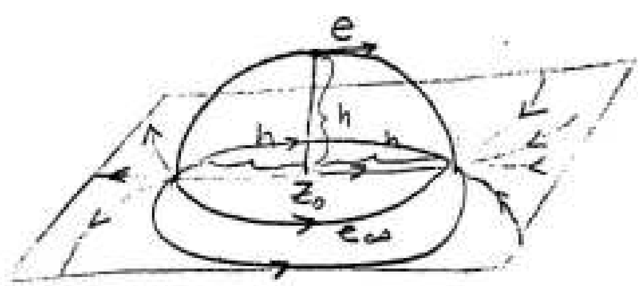
When g is the standard inner product on $T_x(H^3)$, then

$$g_\infty(Y_1, Y_2) = 4\Big(h + \frac{r^2}{h}\Big)^{-2} Y_1 \cdot Y_2,$$

where $Y_1 \cdot Y_2$ is the inner product of two vectors on $\mathbb{C}$.

Let us now compute $\mathrm{av}(\partial/\partial z)$. By symmetry considerations, it is clear that $\mathrm{av}(\partial/\partial z)$ is a horizontal vector field, parallel to $\partial/\partial z$. Let e be the vector of unit hyperbolic length, parallel to $\partial/\partial z$ at a point x in upper half-space. Then

$$e_\infty = -\frac{1}{2h}(z - z_0 - h)(z - z_0 + h)\frac{\partial}{\partial z}.$$



We have

$$\mathrm{av}\,\frac{\partial}{\partial z} = \frac{1}{4\pi}\int_{S^2} i_x\Big(\frac{\partial}{\partial z}\Big)\, dV_x,$$

so

$$\begin{aligned}\mathrm{av}\,\frac{\partial}{\partial z}\cdot e &= \frac{1}{4\pi}\int_{\mathbb{C}} g_\infty\Big(\frac{\partial}{\partial z}, e_\infty\Big)\, dV_x \\ &= \frac{1}{4\pi}\int_{\mathbb{C}} \mathrm{Re}\Big(-\frac{1}{2h}(z - z_0)^2 - h^2\Big)\, 16\Big(h + \frac{r^2}{h}\Big)^{-4} d\mu.\end{aligned}$$

Clearly, by symmetry, the term involving $\mathrm{Re}(z - z_0)^2$ integrates to zero, so we have 11.5

$$\begin{aligned}\mathrm{av}\,\frac{\partial}{\partial z}\cdot e &= \frac{1}{4\pi}\int_0^\infty \int_0^{2\pi} r\, d\theta \cdot 8h\Big(h + \frac{r^2}{h}\Big)^{-4} dr \\ &= -\frac{2h^2}{3}\Big(h + \frac{r^2}{h}\Big)^{-3}\Big|_0^\infty = \Big(\frac{2}{3}\Big)\frac{1}{h}.\end{aligned}$$

Note that the hyperbolic norm of $\mathrm{av}(\partial/\partial z)$ goes to ∞ as $h \to 0$, while the Euclidean norm is the constant $\frac{2}{3}$.

We now introduce the fudge factor by defining the extension of a vector field X on S^2_∞ to be

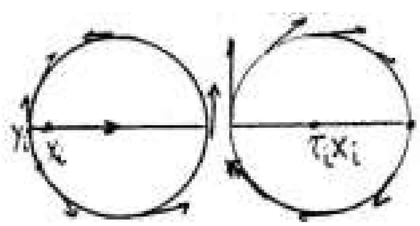
$$\operatorname{ex}(X) = \begin{cases} \frac{3}{2}\operatorname{av}(X) & \text{in } H^3, \\ X & \text{on } S^2_\infty. \end{cases}$$

PROPOSITION 11.1.1. *If X is continuous or Lipschitz, then so is* $\operatorname{ex}(X)$. *If X is holomorphic, then* $\operatorname{ex}(X)$ *is an infinitesimal isometry.*

PROOF. When X is an infinitesimal translation of $\mathbb{C}$, then $\operatorname{ex}(X)$ is the same infinitesimal translation of upper half-space. Thus every "parabolic" vector field (with a zero of order 2) on S^2_∞ extends to the correct infinitesimal isometry. A general holomorphic vector field on S^2_∞ is of the form $(az^2 + bz + c)(\partial/\partial z)$ on $\mathbb{C}$. Since such a vector field can be expressed as a linear combination of the parabolic vector fields $\partial/\partial z$, $z^2\,\partial/\partial z$ and $(z-1)^2\,\partial/\partial z$, it follows that every holomorphic vector field extends to the correct infinitesimal isometry.

Suppose X is continuous, and consider any sequence $\{x_i\}$ of points in H^3 converging to a point at infinity. Bring x_i back to the origin O by the translation τ_i along 11.6
the line $\overline{Ox_i}$. If x_i is close to S^2_∞, τ_i spreads a small neighborhood of the endpoint y_i of the geodesic from O to x_i over almost all the sphere. $\tau_{i*}X$ is large on most of the sphere, except near the antipodal point to y_i, so it is close to a parabolic vector field P_i, in the sense that for any ϵ, and sufficiently high i,

$$\|\tau_{i*}X - P_i\| \le \epsilon \cdot \lambda_i,$$



where λ_i is the norm of the derivative of τ_i at y_i. Here P_i is the parabolic vector field agreeing with $\tau_{i*}X$ at y_i, and 0 at the antipodal point of y_i. It follows that

$$\operatorname{ex} X(x_i) - X(y_i) \to 0,$$

so X is continuous along ∂B^3. Continuity in the interior is self-evident (if you see the evidence).

Suppose now that X is a vector field on $S^2_\infty \subset \mathbb{R}^3$ which has a global Lipschitz constant

$$k = \sup_{y,y' \in S^2} \frac{\|X_y - X_{y'}\|}{\|y - y'\|}.$$

Then the translates $\tau_{i*}X$ satisfy

$$\|\tau_{i*}X - P_i\| \le B,$$

where B is some constant independent of i. This may be seen by considering stereographic projection from the antipodal point of y_i. The part of the image of $X - \tau_{i*}^{-1}P_i$ in the unit disk is Lipschitz and vanishes at the origin. When τ_{i*} is applied, the resulting vector field on $\mathbb{C}$ satisfies a linear growth condition (with a uniform growth constant). This shows that, on S^2_∞, $\|\tau_{i*}X - P_i\|$ is uniformly bounded in all but a 11.7
neighborhood of the antipodal point of Y, where boundedness is obvious. Then

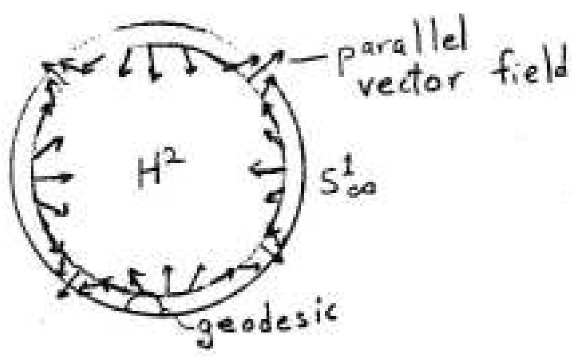
$$\|\operatorname{ex} X(x_i) - \operatorname{ex} \tau_{i*}P_i(x_i)\| \le B \cdot \mu_i,$$

where μ_i is the norm of the derivative of τ_i^{-1} at the origin in B^3, or $1/\lambda_i$ up to a bounded factor.

Since μ_i is on the order of the (Euclidean) distance of x_i from y_i, it follows that $\operatorname{ex} X$ is Lipschitz along S^2_∞.

To see that $\operatorname{ex} X$ has a global Lipschitz constant in B^3, consider $x \in B^3$, and let τ be a translation as before taking x to O, and P a parabolic vector field approximating $\tau_* X$. The vector fields $\tau_* X - P$ obtained in this way are uniformly bounded, so it is clear that the vector fields $\operatorname{ex}(\tau_* X - P)$ have a uniform Lipschitz constant at the origin in B^3. By comparison with the upper half-space model, where τ_* can be taken to be a similarity, we obtain a uniform bound on the local Lipschitz constant for $\operatorname{ex}(X - \tau_*^{-1}P)$ at an arbitrary point x. Since the vector fields $\tau_*^{-1}P$ are uniformly Lipschitz, it follows that X is globally Lipschitz. □

Note that the stereographic image in $\mathbb{C}$ of a uniformly Lipschitz vector field on S^2_∞ is not necessarily uniformly Lipschitz—consider $z^2\,\partial/\partial z$, for example. This is explained by the large deviation of the covariant derivatives on S^2_∞ and on $\mathbb{C}$ near the point at infinity. Similarly, a uniformly Lipschitz vector field on B^3 is not generally uniformly Lipschitz on H^3. In fact, because of the curvature of H^3, a uniformly Lipschitz vector field on H^3 must be bounded; such vector fields correspond precisely to those Lipschitz vector fields on B^3 which vanish on ∂B^3. 11.8



A hyperbolic parallel vector field along a curve near S^1_∞ appears to turn rapidly.

The significance of the Lipschitz condition stems from the elementary fact that Lipschitz vector fields are uniquely integrable. Thus, any isotopy φ_t of the boundary of a Kleinian manifold $O_\Gamma = (B^3 - L_\Gamma)/\Gamma$ whose time derivative $\dot{\varphi}_t$ is Lipschitz as a vector field on $I \times \partial O_\Gamma$ extends canonically to an isotopy $\operatorname{ex} \varphi_t$ on O_Γ. One may see this most simply by observing that the proof that $\operatorname{ex} X$ is Lipschitz works locally.

A *k-quasi-isometric* vector field is a vector field whose flow, φ_t, distorts distances at a rate of at most k. In other words, for all x, y and t, φ_t must satisfy

$$e^{-kt} d(x, y) \leq d(\varphi_t x, \varphi_t y) \leq e^{kt} d(x, y).$$

A k-Lipschitz vector field on a Riemannian manifold is k-quasi-isometric. In fact, a Lipschitz vector field X on B^3 which is tangent to ∂B^3 is quasi-isometry as a vector field on $H^3 = \operatorname{int} B^3$. This is clear in a neighborhood of the origin in B^3. To see this for an arbitrary point x, approximate X near x by a parabolic vector field, as in the proof of 11.1.1, and translate x to the origin.

In particular, if φ_t is an isometry of ∂O_Γ with Lipschitz time derivative, then $\operatorname{ex} \varphi_t$ has a quasi-isometric time derivative, and φ_1 is a quasi-isometry.

Our next step is to study the derivatives of $\operatorname{ex} X$, so we can understand how a 11.9
more general isotopy such as $\operatorname{ex} \varphi_t$ distorts the hyperbolic metric. From the definition of $\operatorname{ex} X$, it is clear that ex is natural, or in other words,

$$\operatorname{ex}(T_* X) = T_*\big(\operatorname{ex}(X)\big)$$

where T is an isometry of H^3 (extended to S^2_∞ where appropriate).

If X is differentiable, we can take the derivative at $T = \mathrm{id}$, yielding

$$\operatorname{ex}[Y, X] = [Y, \operatorname{ex} X]$$

for any infinitesimal isometry Y. If Y is a pure translation and X is any point on the axis of Y, then $\nabla_X Y_x = 0$. (Here, ∇ is the hyperbolic covariant derivative, so $\nabla_Z W$ is the directional derivative of a vector field W in the direction of the vector field Z.) Using the formula

$$[Y, X] = \nabla_Y X - \nabla_X Y,$$

we obtain:

Proposition 11.1.2. *The direction derivative of $\operatorname{ex} X$ in the direction Y_x, at a point $x \in H^3$, is*

$$\nabla_{Y_x} \operatorname{ex} X = \operatorname{ex}[Y, X],$$

where Y is any infinitesimal translation with axis through x and value Y_x at x. □

The covariant derivative ∇X_x, which is a linear transformation of the tangent space $T_x(H^3)$ to itself, can be expressed as the sum of its symmetric and antisymmetric parts, 11.10

$$\nabla X = \nabla^s X + \nabla^a X,$$

where

$$\nabla^s_Y X \cdot Y' = \tfrac{1}{2}(\nabla_Y X \cdot Y' + \nabla_{Y'} X \cdot Y)$$

and

$$\nabla^a_Y X \cdot Y' = \tfrac{1}{2}(\nabla_Y X \cdot Y' - \nabla_{Y'} X \cdot Y).$$

The anti-symmetric part $\nabla^a X$ describes the infinitesimal rotational effect of the flow generated by X. It can be described by a vector field curl X pointing along the axis of the infinitesimal rotation, satisfying the equation

$$\nabla^a_Y X = \tfrac{1}{2}\operatorname{curl} X \times Y$$

where $\times$ is the cross-product. If e_0, e_1, e_2 forms a positively oriented orthonormal frame at X, the formula is

$$\operatorname{curl} X = \sum_{i \in \mathbb{Z}/3} (\nabla_{e_i} X \cdot e_{i+1} - \nabla_{e_{i+1}} X \cdot e_i)\, e_{i+2}.$$

Consider now the contribution to $\operatorname{ex} X$ from the part of X on an infinitesimal area on S^2_∞, centered at y. This part of $\operatorname{ex} X$ has constant length on each horosphere about y (since the first derivative of a parabolic transformation fixing y is the identity), and it scales as e^{-3t}, where t is a parameter measuring distance between horospheres and increasing away from y. (Linear measurements scale as e^{-t}. Hence, there is a factor of e^{-2t} describing the scaling of the apparent area from a point in H^3, and a factor of $-e^t$ representing the scaling of the lengths of vectors.) Choose positively oriented coordinates t, x_1, x_2, so that $ds^2 = dt^2 + e^{2t}(dx_1^2 + dx_2^2)$, and this infinitesimal contribution to $\operatorname{ex} X$ is in the $\partial/\partial x_1$ direction. Let e_0, e_1 and e_2 be unit vectors in the three coordinate directions. The horospheres $t =$ constant are parallel surfaces, of constant normal curvature 1 (like the unit sphere in $\mathbb{R}^3$), so you can see that 11.11

$$\nabla_{e_0} e_0 = \nabla_{e_0} e_1 = \nabla_{e_0} e_2 = 0$$

$$\nabla_{e_1} e_0 = +e_1, \nabla_{e_1} e_1 = -e_0, \nabla_{e_1} e_2 = 0$$

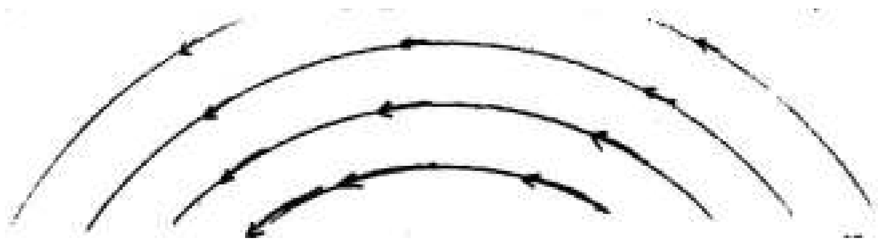
and

$$\nabla_{e_2} e_0 = e_2, \nabla_{e_2} e_2 = -e_0, \nabla_{e_2} e_1 = 0.$$

(This information is also easy to compute by using the Cartan structure equations.) The infinitesimal contribution to $\operatorname{ex} X$ is proportional to $Z = e^{-3t} e_1$, so

$$\begin{aligned}\operatorname{curl} Z &= (\nabla_{e_0} Z \cdot e_1 - \nabla_{e_1} Z \cdot e_0)\, e_2 \\ &= -2e^{-3t} e_2.\end{aligned}$$

(The curl is in the opposite sense from the curving of the flow lines because the effect of the flow speeding up on inner horospheres is stronger.)



This is proportional to the contribution of iX to $\operatorname{ex} iX$ from the same infinitesimal region, so we have

PROPOSITION 11.1.3.

$$\operatorname{Curl}(\operatorname{ex} X) = 2\operatorname{ex}(iX),$$

and consequently

$$\operatorname{Curl}^2(\operatorname{ex} X) = -4\operatorname{ex} X$$

and

$$\operatorname{Div}(\operatorname{ex} X) = 0.$$

PROOF. The first statement follows by integration of the infinitesimal contributions to $\operatorname{curl}\operatorname{ex} X$. The second statement

$$\operatorname{curl}\operatorname{curl}\operatorname{ex} X = 2\operatorname{curl}\operatorname{ex} iX = 4\operatorname{ex} i^2X = -4\operatorname{ex} X,$$

is immediate. The third statement follows from the identity $\operatorname{div}\operatorname{curl} Y = 0$, or by considering the infinitesimal contributions to $\operatorname{ex} X$. ☐

The differential equation $\operatorname{curl}^2\operatorname{ex} X + \operatorname{ex} X = 0$ is the counterpart to the statement that $\operatorname{ex} f = \operatorname{av} f$ is harmonic, when f is a function. The symmetric part $\nabla^s X$ of the covariant derivative measures the infinitesimal strain, or distortion of the metric, of the flow generated by X. That is, if Y and Y' are vector fields invariant by the flow of X, so that $[X, Y] = [X, Y'] = 0$, then $\nabla_Y X = \nabla_X Y$ and $\nabla_{Y'} X = \nabla_X Y'$, so the derivative of the dot product of Y and Y' in the direction X, by the Leibniz rule is

$$\begin{aligned} X(Y \cdot Y') &= \nabla_X Y \cdot Y' + Y \cdot \nabla_X Y' \\ &= \nabla_Y X \cdot Y' + \nabla_{Y'} X \cdot Y \\ &= 2(\nabla^s_Y X \cdot Y'). \end{aligned}$$

The symmetric part of ∇ can be further decomposed into its effect on volume and a part with trace 0,

$$\nabla^s X = \tfrac{1}{3}\operatorname{Trace}(\nabla^s X) \cdot I + \nabla^{s_0} X.$$

Here, I represents the identity transformation (which has trace 3 in dimension 3). Note that $\operatorname{trace} \nabla^s X = \operatorname{trace} \nabla X = \operatorname{divergence} X = \Sigma \nabla_{e_i} X \cdot e_i$ where $\{e_i\}$ is an orthonormal basis, so for a vector field of the form $\operatorname{ex} X$, $\nabla^s \operatorname{ex} X = \nabla^{s_0} \operatorname{ex} X$. 11.13

Now let us consider the analagous decomposition of the covariant derivative ∇X of a vector field on the Riemann sphere (or any surface). There is a decomposition

$$\nabla X = \nabla^a X + \tfrac{1}{2}(\operatorname{trace} \nabla X) I + \nabla^{s_0} X.$$

Define linear maps ∂ and $\overline{\partial}$ of the tangent space to itself by the formulas

$$\partial X(Y) = \tfrac{1}{2}\{\nabla_Y X - i \nabla_{iY} X\}$$

and

$$\overline{\partial} X(Y) = \tfrac{1}{2}\{\nabla_Y X + i \nabla_{iY} X\}$$

for any vector field Y. (On a general surface, i is interpreted as a $90°$ counter-clockwise rotation of the tangent space of the surface.)

Proposition 11.1.4.

$$\begin{aligned} \partial X &= \tfrac{1}{2}(\operatorname{trace} \nabla X) I + \nabla^a X \\ &= \tfrac{1}{2}\{(\operatorname{div} X) I + (\operatorname{curl} X) i I\} \end{aligned}$$

and

$$\overline{\partial} X = \nabla^{s_0} X.$$

∂X is invariant under conformal changes of metric.

Remark (Notational remark). Any vector field on $\mathbb{C}$ be written $X = f(z)\partial/\partial z$, in local coordinates. The derivative of f can be written $df = f_x dx + f_y dy$. This can be re-expressed in terms of $dz = dx + idy$ and $d\bar{z} = dx - idy$ as

$$df = f_z dz + f_{\bar{z}} d\bar{z}$$

where

$$f_z = \tfrac{1}{2}(f_x - i f_y)$$

and 11.14

$$f_{\bar{z}} = \tfrac{1}{2}(f_x + i f_y).$$

Then $\partial f = f_z dz$ and $\bar{\partial} f = f_{\bar{z}} d\bar{z}$ are the complex linear and complex conjugate linear parts of the real linear map df. Similarly, $\partial X = f_z \, dz \, \partial/\partial z$ and $\bar{\partial} X = f_{\bar{z}} \, d\bar{z} \, \partial/\partial z$ are the complex linear and conjugate linear parts of the map $dX = \nabla X$.

Proof. If $L : \mathbb{C} \to \mathbb{C}$ is any real linear map, then

$$L = \tfrac{1}{2}(L - i \circ L \circ i) + \tfrac{1}{2}(L + i \circ L \circ i)$$

is clearly the decomposition into its complex linear and conjugate linear parts. A complex linear map, in matrix form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

is an expansion followed by a rotation, while a conjugate linear map in matrix form

$$\begin{bmatrix} a & b \\ b & -a \end{bmatrix},$$

is a symmetric map with trace 0.

To see that $\bar{\partial}X$ is invariant under conformal changes of metric, note that $\nabla_X iY = i\nabla_X Y$ and write $\bar{\partial}X$ without using the metric as

$$\begin{aligned} \bar{\partial}X(Y) &= \tfrac{1}{2}\{\nabla_Y X + i\nabla_{iY} X\} \\ &= \tfrac{1}{2}\{\nabla_Y X - \nabla_X Y + i\nabla_{iY} X - i\nabla_X iY\} \\ &= \tfrac{1}{2}\{[Y,X] + i[iY,X]\}. \qquad \square \end{aligned}$$

We can now derive a nice formula for $\nabla^s \operatorname{ex} X$:

Proposition 11.1.5. *For any vector $Y \in T_x(H^3)$ and any C^1 vector field X on S^2_∞,* 11.15

$$\nabla^s_Y \operatorname{ex} X = 3/4\pi \int_{S^2_\infty} i_*\big(\bar{\partial}X(Y_\infty)\big)\, dV_x.$$

Proof. Clearly both sides are symmetric linear maps applied to Y, so it suffices to show that the equation gives the right value for $\nabla_Y \operatorname{ex} X \cdot Y$. From 11.1.2, we have

$$\begin{aligned} \nabla_Y \operatorname{ex} X \cdot Y &= \operatorname{ex}[Y_\infty, X] \cdot Y \\ &= 3/8\pi \int_{S^2} [Y_\infty, X] \cdot Y_\infty \, dV_x \end{aligned}$$

and also, at the point x (where $\operatorname{ex} iY_\infty = 0$),

$$\begin{aligned} 0 &= [\operatorname{ex} iY_\infty, X] \cdot \operatorname{ex} iY_\infty \\ &= 3/8\pi \int_{S^2} [iY_\infty, X] \cdot iY_\infty \, dV_x \\ &= 3/8\pi \int_{S^2} -i\,[iY_\infty, X] \cdot Y_\infty \, dV_x. \end{aligned}$$

Therefore

$$\begin{aligned}\nabla^s_Y \operatorname{ex} X \cdot Y &= \nabla_Y \operatorname{ex} X \cdot Y \\ &= 3/8\pi \int_{S^2} [Y_\infty, X] \cdot Y_\infty + i\,[iY_\infty, X] \cdot Y_\infty \, dV_x \\ &= 3/4\pi \left(\int \bar\partial X (Y_\infty) \, dV_x \right) Y.\end{aligned}$$

$\square$

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TEX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

CHAPTER 13

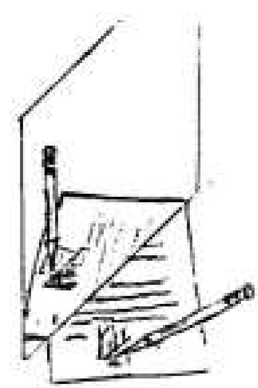
Orbifolds

As we have had occasion to see, it is often more effective to study the quotient manifold of a group acting freely and properly discontinuously on a space rather than to limit one's image to the group action alone. It is time now to enlarge our vocabulary, so that we can work with the quotient spaces of groups acting properly discontinuously but not necessarily freely. In the first place, such quotient spaces will yield a technical device useful for showing the existence of hyperbolic structures on many three-manifolds. In the second place, they are often simpler than three-manifolds tend to be, and hence they often give easy, graphic examples of phenomena involving three-manifolds. Finally, they are beautiful and interesting in their own right.

13.1. Some examples of quotient spaces.

We begin our discussion with a few examples of quotient spaces of groups acting properly discontinuously on manifolds in order to get a taste of their geometric flavor.

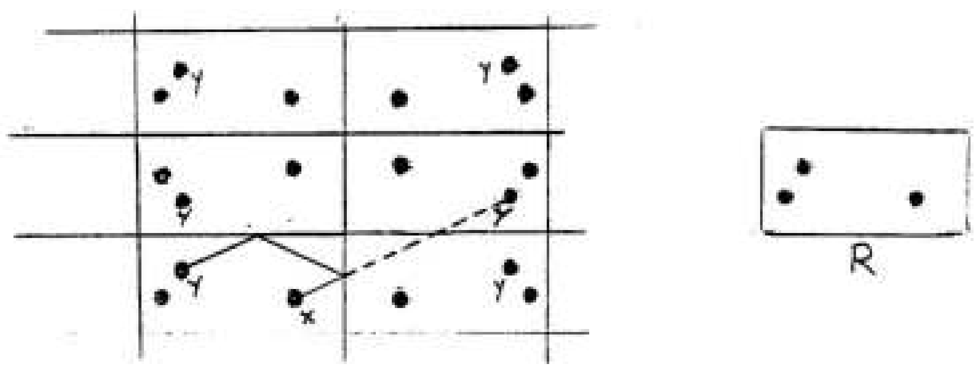
Example 13.1.1 (A single mirror). Consider the action of $\mathbb{Z}_2$ on $\mathbb{R}^3$ by reflection in the $y-z$ plane. The quotient space is the half-space $x \geq 0$. Physically, one may imagine a mirror placed on the $y-z$ wall of the half-space $x \geq 0$. The scene as viewed by a person in this half-space is like all of $\mathbb{R}^3$, with scenery invariant by the $\mathbb{Z}_2$ symmetry.



13.2

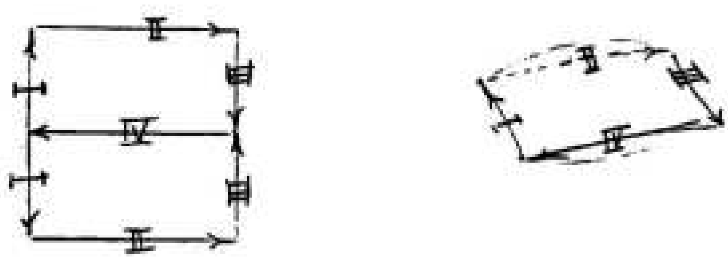
Example 13.1.2 (A barber shop). Consider the group G generated by reflections in the planes $x = 0$ and $x = 1$ in $\mathbb{R}^3$. G is the infinite dihedral group $D_\infty = \mathbb{Z}_2 * \mathbb{Z}_2$. The quotient space is the slab $0 \leq x \leq 1$. Physically, this is related to two mirrors on parallel walls, as commonly seen in a barber shop.

Example 13.1.3 (A billiard table). Let G be the group of isometries of the Euclidean plane generated by reflection in the four sides of a rectangle R. G is isomorphic to $D_\infty \times D_\infty$, 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 $\mathbb{R}^2$, invariant by G. (Each side of the billiard table should be one ball diameter larger than the corresponding side of R so that the *centers* of the balls can take any position in R. A ball may intersect its images in $\mathbb{R}^2$.)



Ignoring spin, in order to make ball x hit ball y it suffices to aim it at any of the images of y by G. (Unless some ball is in the way.)

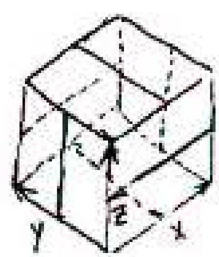
Example 13.1.4 (A rectangular pillow). Let H be the subgroup of index 2 which preserves orientation in the group G of the preceding example. A fundamental do- 13.3
main for H consists of two adjacent rectangles. The quotient space is obtained by identifying the edges of the two rectangles by reflection in the common edge.



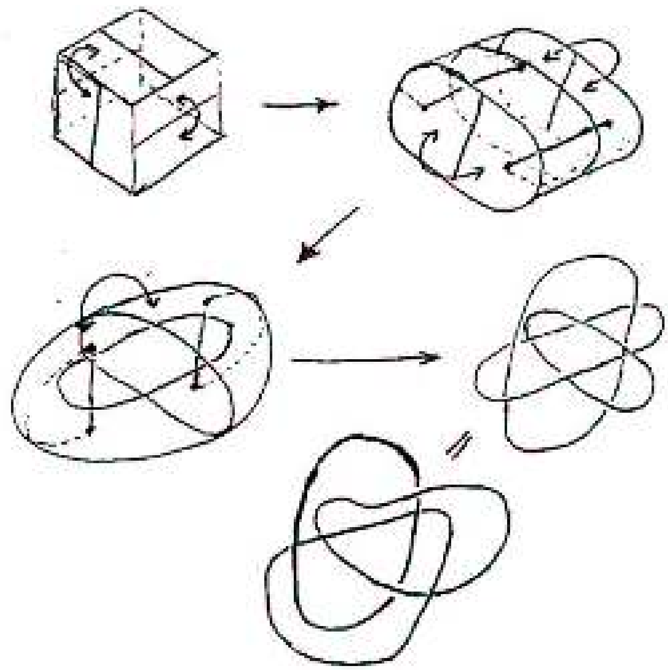
Topologically, this quotient space is a sphere, with four distinguished points or singular points, which come from points in $\mathbb{R}^2$ with non-trivial isotropy ($\mathbb{Z}_2$). The sphere inherits a Riemannian metric of 0 curvature in the complement of these 4 points, and it has curvature $K_{p_i} = \pi$ concentrated at each of the four points p_i. In other words, a neighborhood of each point p_i is a cone, with cone angle $\pi = 2\pi - K_{p_i}$.

EXERCISE. On any tetrahedron in $\mathbb{R}^3$ all of whose four sides are congruent, every geodesic is simple. This may be tested with a cardboard model and string or with strips of paper. Explain.

EXAMPLE 13.1.5 (An orientation-preserving crystallographic group). Here is one more three-dimensional example to illustrate the geometry of quotient spaces. Con- 13.4
sider the 3 families of lines in $\mathbb{R}^3$ of the form $(t, n, m+\frac{1}{2})$, $(m+\frac{1}{2}, t, n)$ and $(n, m+\frac{1}{2}, t)$ where n and m are integers and t is a real parameter. They intersect a cube in the unit lattice as depicted.



Let G be the group generated by 180° rotations about these lines. It is not hard to see that a fundamental domain is a unit cube. We may construct the quotient space by making all identifications coming from non-trivial elements of G acting on the faces of the cube. This means that each face must be folded shut, like a book. In doing this, we will keep track of the images of the axes, which form the singular locus.



13.5

As you can see by studying the picture, the quotient space is S^3 with singular locus consisting of three circles in the form of Borromean rings. S^3 inherits a Euclidean structure (or metric of zero curvature) in the complement of these rings, with a cone-type singularity with cone angle π along the rings.

In these examples, it was not hard to construct the quotient space from the group action. In order to go in the opposite direction, we need to know not only the quotient space, but also the singular locus and appropriate data concerning the local behavior of the group action above the singular locus.

13.2. Basic definitions.

An *orbifold** O is a space locally modelled on $\mathbb{R}^n$ modulo finite group actions. Here is the formal definition: O consists of a Hausdorff space X_O, with some additional structure. X_O is to have a covering by a collection of open sets $\{U_i\}$ closed under finite intersections. To each U_i is associated a finite group Γ_i, an action of Γ_i on an open subset $\tilde{U}_i$ of $\mathbb{R}^n$ and a homeomorphism $\varphi_i : U_i \approx \tilde{U}_i/\Gamma_i$. Whenever $U_i \subset U_j$,

*This terminology should not be blamed on me. It was obtained by a democratic process in my course of 1976-77. An orbifold is something with many folds; unfortunately, the word "manifold" already has a different definition. I tried "foldamani," which was quickly displaced by the suggestion of "manifolded." After two months of patiently saying "no, not a manifold, a manifol*dead*," we held a vote, and "orbifold" won.

there is to be an injective homomorphism

$$f_{ij} : \Gamma_i \hookrightarrow \Gamma_j$$

and an embedding

$$\tilde{\varphi}_{ij} : \tilde{U}_i \hookrightarrow \tilde{U}_j$$

equivariant with respect to f_{ij} (i.e., for $\gamma \in \Gamma_i$, $\tilde{\varphi}_{ij}(\gamma x) = f_{ij}(\gamma)\tilde{\varphi}_{ij}(x)$) such that the 13.6
diagram below commutes.†

$$\begin{tikzcd}
\tilde{U}_i \arrow[r, "\tilde{\varphi}_{ij}"] \arrow[d] & \tilde{U}_j \arrow[d] \\
\tilde{U}_i/\Gamma_i \arrow[r, "\varphi_{ij} = \tilde{\varphi}_{ij}/\Gamma_i"] & \tilde{U}_j/\Gamma_i \arrow[d, "f_{ij}"] \\
 & \tilde{U}_j/\Gamma_j \\
U_i \arrow[uu, "\varphi_i"] \arrow[r, phantom, "\subset"] & U_j \arrow[u, "\varphi_j"']
\end{tikzcd}$$

We regard $\tilde{\varphi}_{ij}$ 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 $\tilde{\varphi}_{ik} = \tilde{\varphi}_{jk} \circ \tilde{\varphi}_{ij}$ when $U_i \subset U_j \subset U_k$, but there should exist an element $\gamma \in \Gamma_k$ such that $\gamma\tilde{\varphi}_{ik} = \tilde{\varphi}_{jk} \circ \tilde{\varphi}_{ij}$ and $\gamma \cdot f_{ik}(g) \cdot \gamma^{-1} = f_{jk} \circ f_{ij}(g)$.

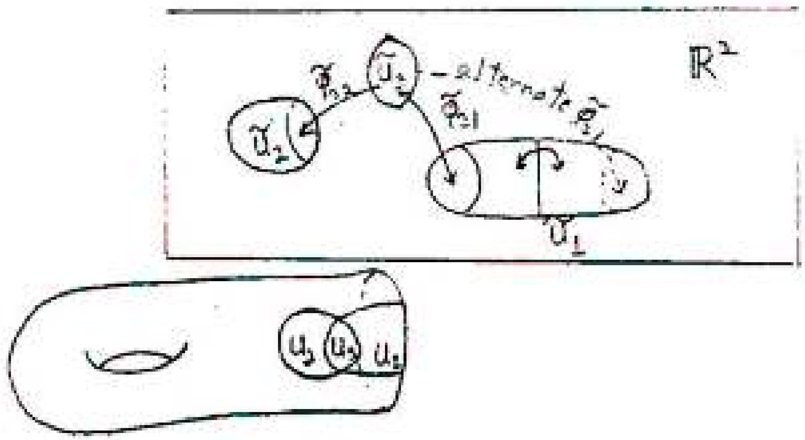
Of course, the covering $\{U_i\}$ is not an intrinsic part of the structure of an orbifold: two coverings give rise to the same orbifold structure if they can be combined consistently to give a larger cover still satisfying the definitions.

A $\mathcal{G}$-orbifold, where $\mathcal{G}$ is a pseudogroup, means that all maps and group actions respect $\mathcal{G}$. (See chapter 3).

Example 13.2.1. A closed manifold is an orbifold, where each group Γ_i is the trivial group, so that $\tilde{U} = U$.

Example 13.2.2. 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 modelled on $\mathbb{R}^n/\mathbb{Z}_2$, where $\mathbb{Z}_2$ acts by reflection in a hyperplane. 13.7

†The commutative diagrams in Chapter 13 were made using Paul Taylor's diagrams.sty package (available at ftp://ftp.dcs.qmw.ac.uk/pub/tex/contrib/pt/diagrams/). —SL



Proposition 13.2.1. *If M is a manifold and Γ is a group acting properly discontinuously on M, then M/Γ has the structure of an orbifold.*

Proof. For any point $x \in M/\Gamma$, choose $\tilde{x} \in M$ projecting to x. Let I_x be the isotropy group of $\tilde{x}$ (I_x depends of course on the particular choice $\tilde{x}$.) There is a neighborhood $\tilde{U}_x$ of $\tilde{x}$ invariant by I_x and disjoint from its translates by elements of Γ not in I_x. The projection of $U_x = \tilde{U}_x/I_x$ is a homeomorphism. To obtain a suitable cover of M/Γ, augment some cover $\{U_x\}$ by adjoining finite intersections. Whenever $U_{x_1} \cap \ldots \cap U_{x_k} \neq \emptyset$, this means some set of translates $\gamma_1 \tilde{U}_{x_1} \cap \ldots \cap \gamma_k \tilde{U}_{k_k}$ has a corresponding non-empty intersection. This intersection may be taken to be

$$\widetilde{\overbrace{U_{x_1} \cap \cdots \cap U_{x_k}}},$$

with associated group $\gamma_1 I_{x_1} \gamma_1^{-1} \cap \cdots \cap \gamma_k I_{x_k} \gamma_k^{-1}$ acting on it. □

The orbifold mM arises in this way, for instance: it is obtained as the quotient space of the $\mathbb{Z}_2$ action on the double dM of M which interchanges the two halves. 13.8

Henceforth, we shall use the terminology M/Γ to mean M/Γ as an orbifold.

Note that each point x in an orbifold O is associated with a group Γ_x, well-defined up to isomorphism: in a local coordinate system $U = \tilde{U}/\Gamma$, Γ_x is the isotropy group of any point in $\tilde{U}$ corresponding to x. (Alternatively Γ_x may be defined as the smallest group corresponding to some coordinate system containing x.) The set $\Sigma_O = \{x | \Gamma_x \neq \{1\}\}$ is the *singular locus* of O. *We shall say that O is a manifold when* $\Sigma_O = \emptyset$. *Warning.* It happens much more commonly that the underlying space X_O is a topological manifold, especially in dimensions 2 and 3. Do not confuse properties of O with properties of X_O.

The singular locus is a closed set, since its intersection with any coordinate patch is closed. Also, it is nowhere dense. This is a consequence of the fact that a non-trivial homeomorphism of a manifold which fixes an open set cannot have finite order. (See Newman, 1931. In the differentiable case, this is an easy exercise.)

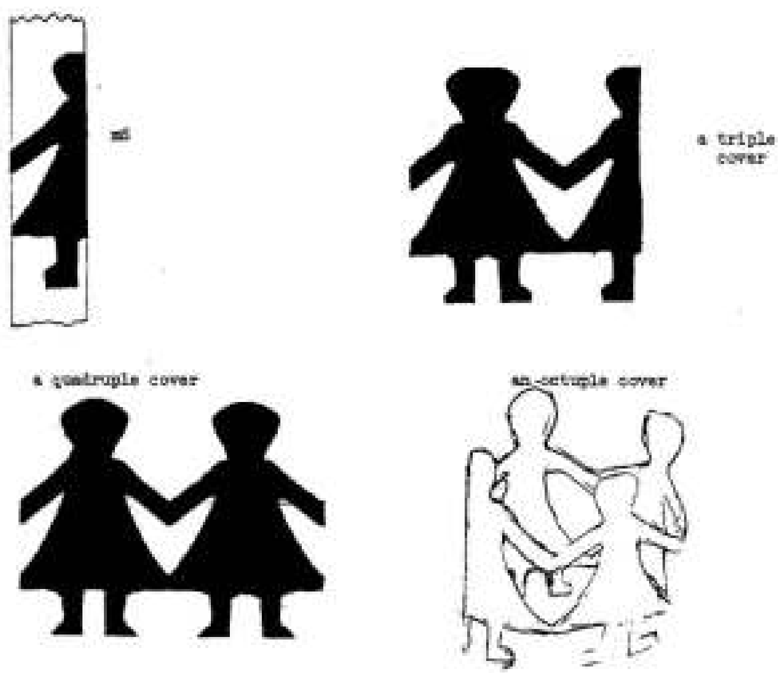
When M in the proposition is simply connected, then M plays the role of universal covering space and Γ plays the role of the fundamental group of the orbifold M/Γ, (even though the underlying space of M/Γ may well be simply connected, as in the examples of §13.1). To justify this, we first define the notion of a covering orbifold.

Definition 13.2.2. A *covering orbifold* of an orbifold O is an orbifold $\tilde{O}$, with a projection $p : X \to X_O$ between the underlying spaces, such that each point $x \in X_O$ has a neighborhood $U = \tilde{U}/\Gamma$ (where $\tilde{U}$ is an open subset of $\mathbb{R}^n$) for which each 13.9
component v_i of $p^{-1}(U)$ is isomorphic to $\tilde{U}/\Gamma_i$, where $\Gamma_i \subset \Gamma$ is some subgroup. The isomorphism must respect the projections.

Note that the underlying space $X_{\tilde{O}}$ is not generally a covering space of X_O.

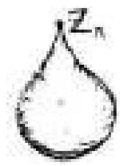
As a basic example, when Γ is a group acting properly discontinuously on a manifold M, then M is a covering orbifold of M/Γ. In fact, for any subgroup $\Gamma' \subset \Gamma$, M/Γ' is a covering orbifold of M/Γ. Thus, the rectangular pillow (13.1.4) is a two-fold covering space of the billiard table (13.1.3).

Here is another explicit example to illustrate the notion of covering orbifold. Let S be the infinite strip $0 \leq x \leq 1$ in $\mathbb{R}^2$; consider the orbifold mS. Some covering spaces of S are depicted below.



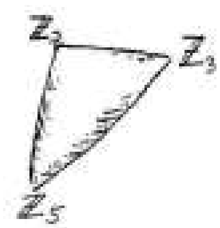
Definition 13.2.3. An orbifold is *good* if it has some covering orbifold which is a manifold. Otherwise it is *bad*.

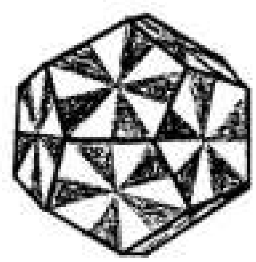
The teardrop is an example of a bad orbifold. The underlying space for a teardrop is S^2. Σ_O consists of a single point, whose neighborhood is modelled on $\mathbb{R}^2/\mathbb{Z}_n$, where $\mathbb{Z}_n$ acts by rotations.



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 coverings.

Similarly, you may verify that an orbifold O with underlying space $X_O = S^2$ having only two singular points associated with groups $\mathbb{Z}_n$ and $\mathbb{Z}_n$ is bad, unless $n = m$. The orbifolds with three or more singular points on S^2, as we shall see, are always good. For instance, the orbifold below is S^2 modulo the orientation-preserving symmetries of a dodecahedron.





13.11

Proposition 13.2.4. *An orbifold O has a universal cover $\tilde{O}$. In other words, if $* \in X_O - \Sigma_O$ is a base point for O,*

$$\tilde{O} \xrightarrow{p} O$$

is a connected covering orbifold with base point $\tilde{*}$ which projects to $*$, such that for any other covering orbifold

$$\tilde{O}' \xrightarrow{p'} O$$

with base point $\tilde{*}'$, $p'(\tilde{*}') = *$, there is a lifting $q : \tilde{O} \to \tilde{O}'$ of p to a covering map of $\tilde{O}'$.

$$\begin{tikzcd} \tilde{O} \arrow[dd, "p"'] \arrow[dr, "q"] & \\ & \tilde{O}' \arrow[dl, "p'"] \\ O & \end{tikzcd}$$

The universal covering orbifold $\tilde{O}$, in some contexts, is often called the universal branched cover. There is a simple way to prove 13.2.4 in the case Σ_O has codimension 2 or more. In that case, any covering space of O is determined by the induced covering space of $X_O - \Sigma_O$ as its metric completion. Whether a covering Y space of $X_O - \Sigma_O$ comes from a covering space of O is a local question, which is expressed algebraically by saying that $\pi_1(Y)$ maps to a group containing a certain obvious normal subgroup of $\pi_1(X - \Sigma_O)$.

When O is a good orbifold, then it is covered by a simply connected manifold, M. It can be shown directly that M is the universal covering orbifold by proving that every covering orbifold is isomorphic to M/Γ', for some $\Gamma' \subset \Gamma$, where Γ is the group of deck transformations of M over O.

Proof of 13.2.4. One proof of the existence of a universal cover for a space X goes as follows.

Consider pointed, connected covering spaces

$$\tilde{X}_i \xrightarrow{p_i} X.$$

For any pair of such covering spaces, the component of the base point in the fiber product of the two is a covering space of both. 13.12

$$\begin{tikzcd}
 & \tilde{X}_3 \arrow[dl] \arrow[dr] & \\
\tilde{X}_1 \arrow[dr] & & \tilde{X}_2 \arrow[dl] \\
 & X &
\end{tikzcd}$$

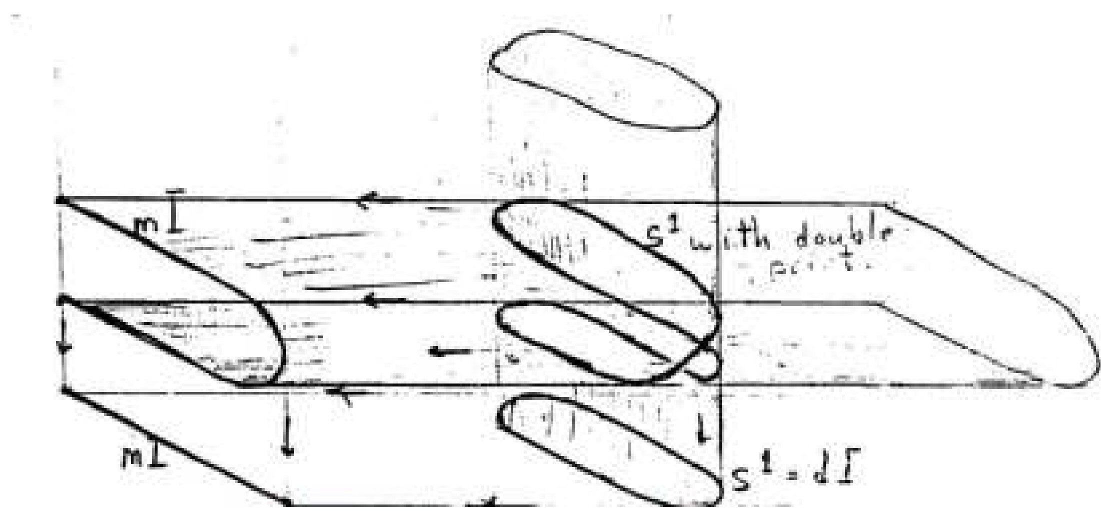
(Recall that the fiber product of two maps $f_i : X_i \to X$ is the space $X_1 \times_X X_2 = \{(x_1, x_2) \in X_1 \times X_2 : f_1(x_1) = f_2(x_2)\}$.)

If X is locally simply connected, or more generally, if it has the property that every $x \in X$ has a neighborhood U such that every covering of X induces a trivial covering of U (that is, each component of $p^{-1}(U)$ is homeomorphic to U), then one can take the inverse limit over some set of pointed, connected covering spaces of X which represents all isomorphism classes to obtain a universal cover for X.

We can follow this same outline with orbifolds, but we need to refine the notion of fiber product. The difficulty is best illustrated by example. Two covering maps

$$S^1 = dI \xrightarrow{f_2} m_i \quad \text{and} \quad m_i \to m_i$$

are sketched below, along with the fiber product of the underlying maps of spaces.



(This picture is sketched in $\mathbb{R}^3 = \mathbb{R}^2 \times_{\mathbb{R}_1} \mathbb{R}^2$.) The fiber product of spaces is a circle but with a double point. In the definition of fiber product of orbifolds, we must eliminate such double points, which always lie above Σ_O.

To do this, we work in local coordinates. Let $U \approx \tilde{U}/\Gamma$ be a coordinate system. 13.13
We may suppose that U is small enough so in every covering of O, $p^{-1}(U)$ consists

of components of the form $\tilde{U}/\Gamma'$, $\Gamma' \subset \Gamma$. Let

$$O_i \xrightarrow{p_i} O$$

be covering orbifolds ($i = 1, 2$), and consider components of $p_i^{-1}(U)$, which for notational convenience we identify with $\tilde{U}/\Gamma_1$ and $\tilde{U}/\Gamma_2$. Formally, we can write $\tilde{U}/\Gamma_1 = \{\Gamma_1 y \,|\, y \in \tilde{U}\}$. [It would be more consistent to use the notation $\Gamma_1 \backslash \tilde{U}$ instead of $\tilde{U}/\Gamma_1$]. For each pair of elements γ_1 and $\gamma_2 \in \Gamma$, we obtain a map

$$f_{\gamma_1,\gamma_2} : \tilde{U} \to \tilde{U}/\Gamma_1 \times \tilde{U}/\Gamma_2,$$

by the formula

$$f_{\gamma_1,\gamma_2} y = (\Gamma_1 \gamma_1 y, \Gamma_2 \gamma_2 y).$$

In fact, f_{γ_1,γ_2} factors through

$$\tilde{U}/\gamma_1^{-1}\Gamma_1\gamma_1 \cap \gamma_2^{-1}\Gamma_2\gamma_2.$$

Of course, f_{γ_1,γ_2} depends only on the cosets $\Gamma_2\gamma_1$ and $\Gamma_2\gamma_2$. Furthermore, for any $\gamma \in \Gamma$, the maps f_{γ_1,γ_2} and $f_{\gamma_1\gamma,\gamma_2\gamma}$ differ only by a group element acting on $\tilde{U}$; in particular, their images are identical so only the product $\gamma_1\gamma_2^{-1}$ really matters. Thus, the "real" invariant of f_{γ_1,γ_2} is the double coset

$$\Gamma_1\gamma_1\gamma_2^{-1}\Gamma_2 \in \Gamma_1 \backslash \Gamma / \Gamma_2.$$

(Similarly, in the fiber product of coverings X_1 and X_2 of a space X, the components are parametrized by the double cosets $\pi_1 X_1 \backslash \pi_1 X / \pi_1 X_2$.) The fiber product of $\tilde{U}/\Gamma_1$ and $\tilde{U}/\Gamma_2$ over $\tilde{U}/\Gamma$, is defined now to be the disjoint union, over elements γ representing double cosets $\Gamma_1\backslash\Gamma/\Gamma_2$ of the orbifolds $\tilde{U}/\Gamma_1 \cap \gamma^{-1}\Gamma_2\gamma$. We have shown above how this canonically covers $\tilde{U}/\Gamma_1$ and $\tilde{U}/\Gamma_2$, via the map $f_{1,\gamma}$. This definition agrees with the usual definition of fiber product in the complement of Σ_O. These locally defined patches easily fit together to give a fiber product orbifold $O_1 \times_O O_2$. As in the case of spaces, a universal covering orbifold $\tilde{O}$ is obtained by taking the inverse limit over some suitable set representing all isomorphism classes of orbifolds. $\square$

The universal cover $\tilde{O}$ of an orbifold O is automatically a regular cover: for any preimage of $\tilde{x}$ of the base point $*$ there is a deck transformation taking $\tilde{*}$ to $\tilde{x}$. 13.14

Definition 13.2.5. The *fundamental group* $\pi_1(O)$ of an orbifold O is the group of deck transformations of the universal cover $\tilde{O}$.

The fundamental groups of orbifolds can be computed in much the same ways as fundamental groups of manifolds. Later we shall interpret $\pi_1(O)$ in terms of loops on O.

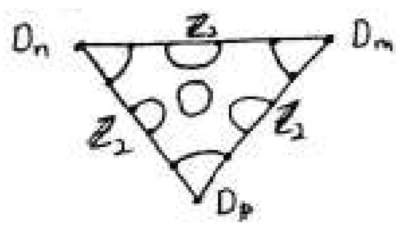
Here are two more definitions which are completely parallel to definitions for manifolds.

Definition 13.2.6. *An orbifold with boundary* means a space locally modelled on $\mathbb{R}^n$ modulo finite groups and $\mathbb{R}^n_+$ modulo finite groups.

When X_O is a topological manifold, be careful not to confuse ∂X_O with ∂O or $X_{\partial O}$.

Definition 13.2.7. A *suborbifold* O_1 of an orbifold O_2 means a subspace $X_{O_1} \subset X_{O_2}$ locally modelled on $\mathbb{R}^d \subset \mathbb{R}^n$ modulo finite groups.

Thus, a triangle orbifold has seven distinct "closed" one-dimensional suborbifolds, up to isotopy: one S^1 and six mI's.



Note that each of the seven is the boundary of a suborbifold with boundary (defined in the obvious way) with universal cover D^2. 13.15

13.3. Two-dimensional orbifolds.

To avoid technicalities, we shall work with differentiable orbifolds from now on.

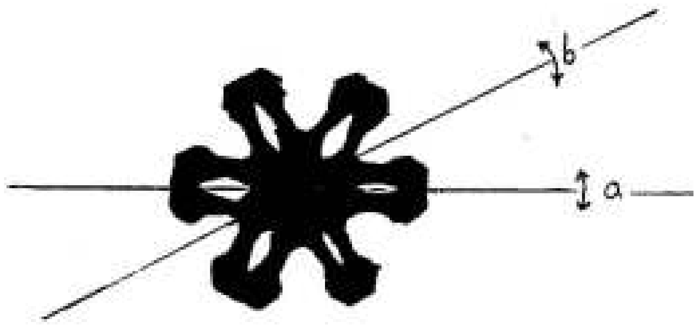
The nature of the singular locus of a differentiable orbifold may be understood as follows. Let $U = \tilde{U}/\Gamma$ be any local coordinate system. There is a Riemannian metric on $\tilde{U}$ invariant by Γ: such a metric may be obtained from any metric on $\tilde{U}$ by averaging under Γ. For any point $\tilde{x} \in \tilde{U}$ consider the exponential map, which gives a diffeomorphism from the ϵ ball in the tangent space at $\tilde{x}$ to a small neighborhood of $\tilde{x}$. Since the exponential map commutes with the action of the isotropy group of $\tilde{x}$, it gives rise to an isomorphism between a neighborhood of the image of $\tilde{x}$ in O, and a neighborhood of the origin in the orbifold $\mathbb{R}^n/\Gamma$, where Γ is a finite subgroup of the orthgonal group O_n.

Proposition 13.3.1. *The singular locus of a two-dimensional orbifold has these types of local models:*

(i) *The mirror: $\mathbb{R}^2/\mathbb{Z}_2$, where $\mathbb{Z}_2$ acts by reflection in the y-axis.*
(ii) *Elliptic points of order n: $\mathbb{R}^2/\mathbb{Z}_n$, with $\mathbb{Z}_n$ acting by rotations.*
(iii) *Corner reflectors of order n: $\mathbb{R}^2/D_n$, with D_n is the dihedral group of order $2n$, with presentation*

$$\langle a, b : a^2 = b^2 = (ab)^n = 1\rangle.$$

The generators a and b correspond to reflections in lines meeting at angle π/n. 13.16



PROOF. These are the only three types of finite subgroups of O_2. $\square$

It follows that the underlying space of a two-dimensional orbifold is always a topological surface, possibly with boundary. It is easy to enumerate all two-dimensional orbifolds, by enumerating surfaces, together with combinatorial information which determines the orbifold structure. From a topological point of view, however, it is not completely trivial to determine which of these orbifolds are good and which are bad.

We shall classify two-dimensional orbifolds from a geometric point of view. When G is a group of real analytic diffeomorphisms of a real analytic manifold X, then the elementary properties of (G, X)-orbifolds are similar to the case of manifolds (see §3.5). In particular a developing map

$$D : \tilde{O} \to X$$

can be defined for a (G, X)-orbifold O. Since we do not yet have a notion of paths in O, this requires a little explanation. Let $\{U_i\}$ be a covering of O by a collection of open sets, closed under intersections, modelled on $\tilde{U}_i/\Gamma_i$, with $\tilde{U}_i \subset X$, such that the 13.17
inclusion maps $U_i \subset U_j$ come from isometries $\tilde{\varphi}_{ij} : \tilde{U}_i \to \tilde{U}_j$. Choose a "base" chart $\tilde{U}_0$. When $U_0 \supset U_{i_1} \subset U_{i_2} \supset \cdots \subset U_{i_{2n}}$ is a chain of open sets (a simplicial path in the one-skeleton of the nerve of $\{U_i\}$), then for each choice of isometries of the form

$$\tilde{U}_0 \xleftarrow{\gamma_0 \tilde{\varphi}_{i_1,0}} \tilde{U}_{i_1} \xrightarrow{\gamma_2' \tilde{\varphi}_{i_1,i_2}} \tilde{U}_{i_2} \longleftarrow \cdots \longrightarrow \tilde{U}_{i_{2n}}$$

one obtains an isometry of $\tilde{U}_{i_{2n}}$ in X, obtained by composing the transition functions (which are globally defined on X). A covering space $\tilde{O}$ of O is defined by the covering $\{(\varphi, \varphi(\tilde{U}_i))\} \subset G \times X$, where φ is any isometry of $\tilde{U}_i$ obtained by the above construction.

These are glued together by the obvious "inclusion" maps, $(\varphi, \varphi\tilde{U}_i) \hookrightarrow (\psi, \psi\tilde{U}_j)$ whenever $\psi^{-1} \circ \varphi$ is of the form $\gamma_j \circ \tilde{\varphi}_{ij}$ for some $\gamma_j \in \Gamma_j$.

The reader desiring a picture may construct a "foliation" of the space $\{(x, y, g) \mid x \in X,\ y \in X_O,\ g$ is the germ of a G-map between neighborhoods of x and $y\}$. Any leaf of this foliation gives a developing map.

Proposition 13.3.2. *When G is an analytic group of diffeomorphisms of a manifold X, then every (G, X)-orbiifold is good. A developing map*

$$D : \tilde{O} \to X$$

and a holonomy homomorphism

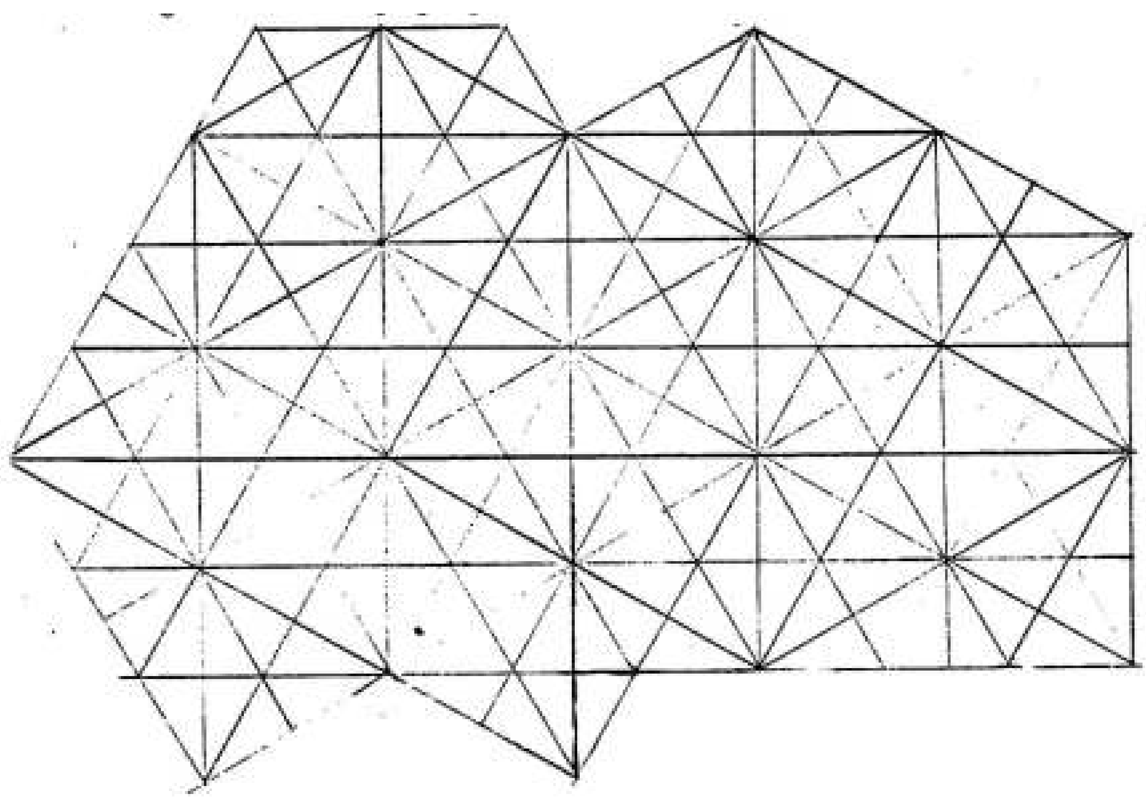
$$H : \pi_1(O) \to G$$

are defined.

If G is a group of isometries acting transitively on X, then if O is closed or metrically complete, it is complete (i.e., D is a covering map). In particular, if X is simply connected, then $\tilde{O} = X$ and $\pi_1(O)$ is a discrete subgroup of G. 13.18

Proof. See §3.5. □

Here is an example. $\triangle_{2,3,6}$ has a Euclidean structure, as a 30°, 60°, 90° triangle. The developing map looks like this:



Here is a definition that will aid us in the geometric classification of two-dimensional orbifolds.

DEFINITION 13.3.3. When an orbifold O has a cell-division of X_O such that each open cell is in the same stratum of the singular locus (i.e., the group associated to the interior points of a cell is constant), then the *Euler number* $\chi(O)$ is defined by 13.19
the formula

$$\chi(O) = \sum_{c_i} (-1)^{\dim(c_i)} \frac{1}{|\Gamma(c_i)|},$$

where c_i ranges over cells and $|\Gamma(c_i)|$ is the order of the group $\Gamma(c_i)$ associated to each cell. The Euler number is not always an integer.

The definition is concocted for the following reason. Define the *number of sheets* of a cover to be the number of preimages of a non-singular point.

PROPOSITION 13.3.4. *If $\tilde{O} \to O$ is a covering map with k sheets, then*

$$\chi(\tilde{O}) = k\chi(O).$$

PROOF. It is easily verified that the number of sheets of a cover can be computed by the ratio

$$\#\text{ sheets } = \sum_{\tilde{x} \ni p(\tilde{x}) = x} (\,|\Gamma_x| \quad / \,|\Gamma_{\tilde{x}}|\,),$$

where x is *any* point. The formula [???] for the Euler number of a cover follows immediately. □

As an example, a triangle orbifold Δ_{n_1,n_2,n_3} has Euler number $\frac{1}{2}\big(\sum(1/n_i) - 1\big)$: here $+1$ comes from the 2-cell, three $-\frac{1}{2}$'s from the edges, and $1/(2n_i)$ from each vertex. 13.20

Thus, $\Delta_{2,3,5}$ has Euler number $+1/60$. Its universal cover is S^2, with deck transformations the group of symmetries of the dodecahedron. This group has order $120 = 2/(1/60)$. On the other hand, $\chi(\Delta_{2,3,6}) = 0$ and $\chi(\Delta_{2,3,7}) = -1/84$. These orbifolds cannot be covered by S^2.

The general formula for the Euler number of an orbifold O with k corner reflectors of orders $n_1, \dots, n_k$ and l elliptic points of orders $m_1, \dots, m_l$ is

13.3.4. $$\chi(O) = \chi(X_O) - \tfrac{1}{2}\sum(1 - 1/n_i) - \sum(1 - 1/m_i).$$

Note in particular that $\chi(O) \le \chi(X_O)$, with equality if and only if O is the surface χ_O or if $O = m\chi_O$.

If O is equipped with a metric coming from invariant Riemannian metrics on the local models $\tilde{U}$, then one may easily derive the Gauss-Bonnet theorem,

13.3.5. $$\int_O K\,dA = 2\pi\chi(O).$$

One way to prove this is by excising small neighborhoods of the singular locus, and applying the usual Gauss-Bonnet theorem for manifolds with boundary. For O to have an elliptic, parabolic or hyperbolic structure, $\chi(O)$ must be respectively positive, zero or negative. If O is elliptic or hyperbolic, then area $(O) = 2\pi|\chi(O)|$.

THEOREM 13.3.6. *A closed two-dimensional orbifold has an elliptic, parabolic or hyperbolic structure if and only if it is good. An orbifold O has a hyperbolic structure if and only if $\chi(O) < 0$, and a parabolic structure if and only if $\chi(O) = 0$. An orbifold* 13.21
is elliptic or bad if and only if $\chi(O) > 0$.

All bad, elliptic and parabolic orbifolds are tabulated below, where

$$(n_1, \dots, n_k; m_1, \dots, m_l)$$

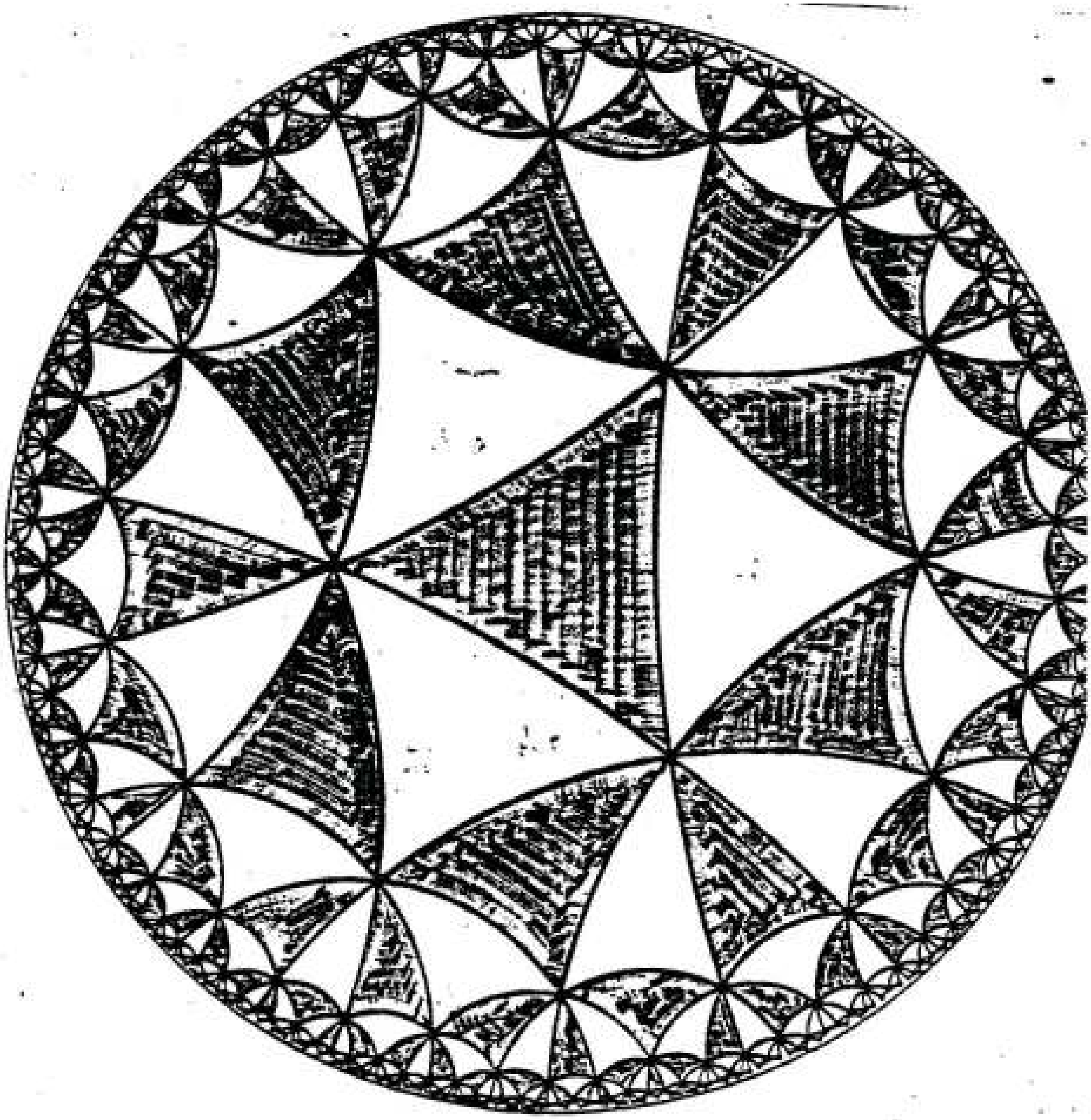
denotes an orbifold with elliptic points of orders $n_1, \dots, n_k$ (in ascending order) and corner reflectors of orders $m_1, \dots, m_l$ (in ascending order). Orbifolds not listed are hyperbolic.

- Bad orbifolds:
  - $X_O = S^2$: (n), (n_1, n_2) with $n_1 < n_2$.
  - $X_O = D^2$: $(\,;n)$, $(\,;n_1, n_2)$ with $n_1 < n_2$.
- Elliptic orbifolds:
  - $X_O = S^2$: $(\;)$, (n, n), $(2, 2, n)$, $(2, 3, 3)$, $(2, 3, 4)$, $(2, 3, 5)$.
  - $X_O = D^2$: $(\,;)$, $(\,;n, n)$, $(\,;2, 2, n)$, $(\,;2, 3, 3)$, $(\,;2, 3, 4)$, $(\,;2, 3, 5)$, $(n;)$, $(2; m)$, $(3; 2)$.
  - $X_O = \mathbb{P}^2$: $(\;)$, (n).
- Parabolic orbifolds:
  - $X_O = S^2$: $(2, 3, 6)$, $(2, 4, 4)$, $(3, 3, 3)$, $(2, 2, 2, 2)$.
  - $X_O = D^2$: $(\,;2, 3, 6)$, $(\,;2, 4, 4)$, $(\,;3, 3, 3)$, $(\,;2, 2, 2, 2)$, $(2; 2, 2)$, $(3; 3)$, $(4; 2)$, $(2; 2;)$.
  - $X_O = \mathbb{P}^2$: $(2, 2)$.
  - $X_O = T^2$: $(\;)$
  - $X_O =$ Klein bottle: $(\;)$
  - $X_O =$ annulus: $(\,;)$
  - $X_O =$ Möbius band: $(\,;)$

13.21.a

The universal covering space of $D^2_{(;4,4,4)}$ and $S^2_{(4,4,4)} \cdot \pi_1(D^2_{(;4,4,4)})$ is generated by reflections in the faces of one of the triangles. The full group of symmetries of this tiling of H^2 is $\pi_1(D^2_{(;2,3,8)})$.

This picture was drawn with a computer by Peter Oppenheimer.



Proof. It is routine to list all orbifolds with non-negative Euler number, as in the table. We have already indicated an easy, direct argument to show the orbifolds listed as bad are bad; here is another. First, by passing to covers, we only need consider the case that the underlying space is S^2, and that if there are two elliptic 13.22

points their orders are relatively prime. These orbifolds have Riemannian metrics of curvature bounded above zero,

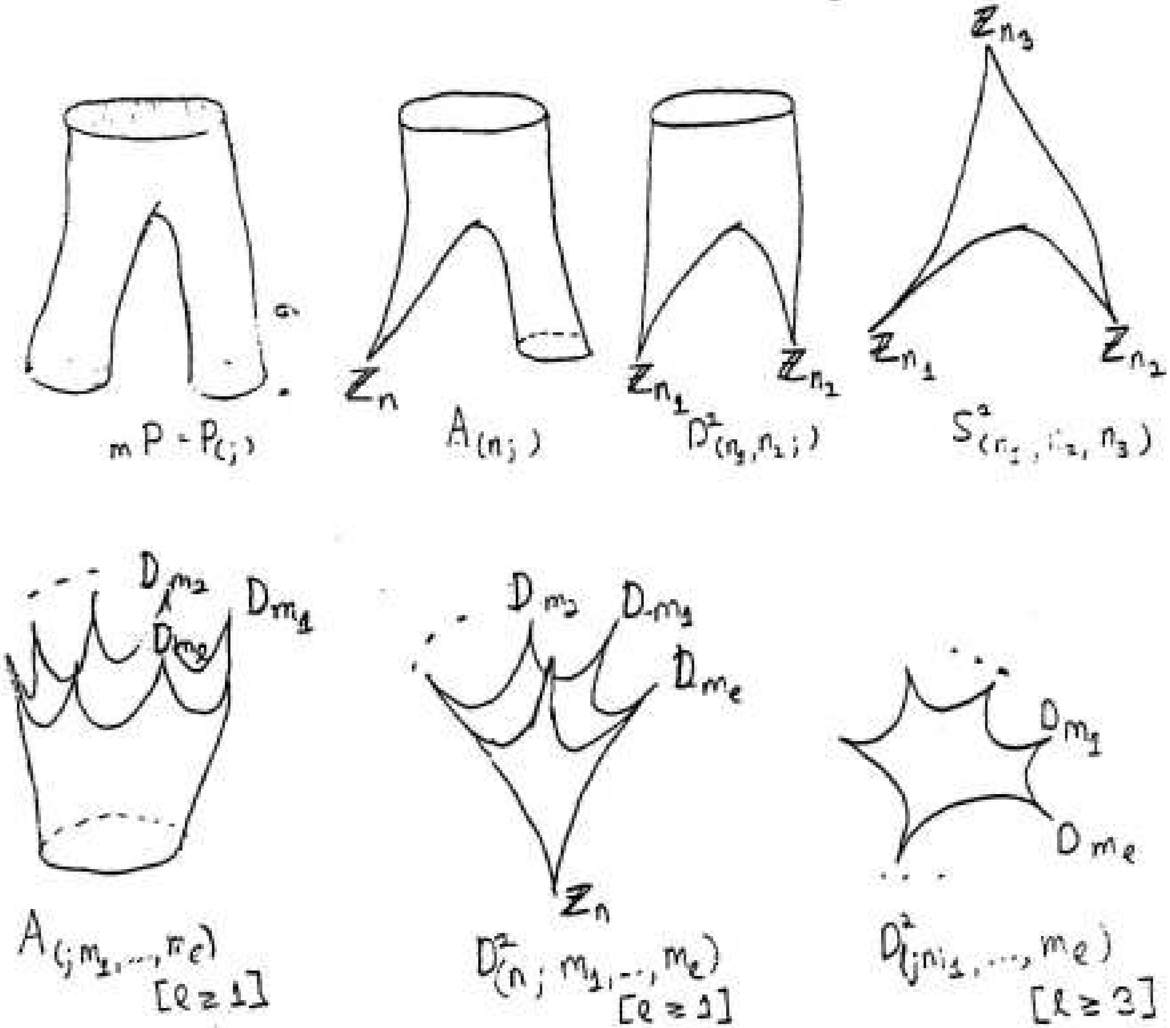
which implies (by elementary Riemannian geometry) that any surface covering them must be compact. But the Euler number is either $1 + 1/n$ or $1/n_1 + 1/n_2$, which is a rational number with numerator > 2.

Since no connected surface has an Euler number greater than 2, these orbifolds must be bad.

Question. What is the best pinching constant for Riemannian metrics on these orbifolds?

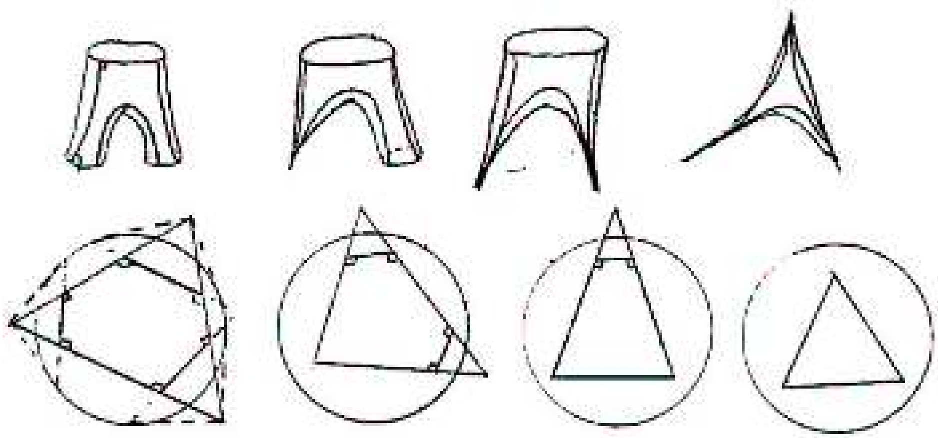
All the orbifolds listed as elliptic and parabolic may be readily identified as the quotient of S^2 or E^2 modulo a discrete group. The 17 parabolic orbifolds correspond to the 17 "wallpaper groups." The reader should unfold these orbifolds for himself, 13.23
to appreciate their beauty. Another pleasant 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 (§5.3). Given an orbifold O with $\chi(O) < 0$, we may repeatedly cut it along simple closed curves and then "mirror" these curves (to remain in the class of closed orbifolds) until we are left with pieces of the form below. (If the underlying surface is unoriented, then make the first cut so the result is oriented.)

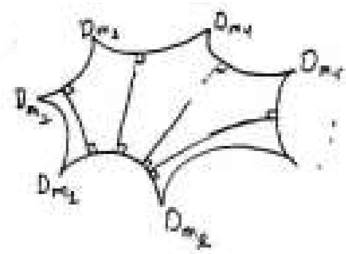


13.24

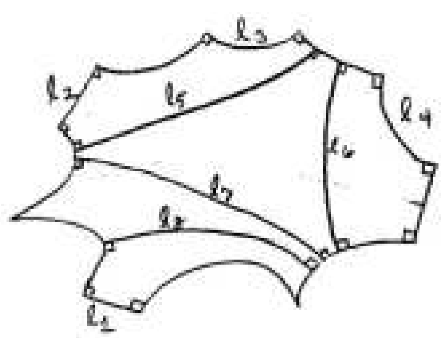
The orbifolds mP, $A_{(n;)}$ and $D_{(n_1,n_2;)}$ (except the degenerate case $A_{(2,2;)}$) and $S^2_{(n_1,n_2,n_3)}$ have hyperbolic structures paremetrized by the lengths of their boundary components. The proof is precisely analogous to the classification of shapes of pants in §5.3; one decomposes these orbifolds into two congruent "generalized triangles" (see §2.6).



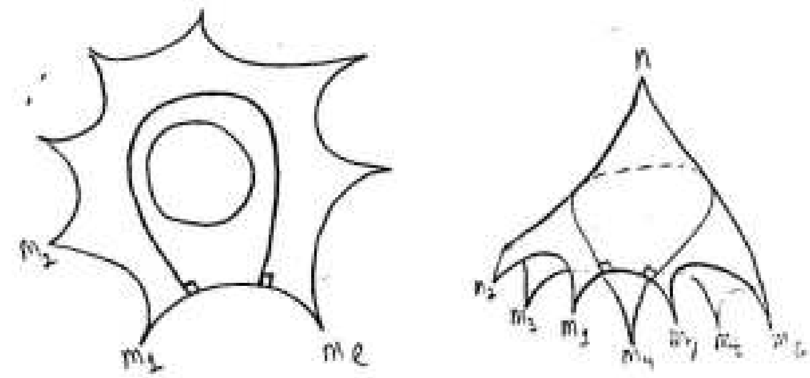
The orbifold $D^2_{(;m_1,\dots,m_l)}$ also can be decomposed into "generalized triangles,"



for instance in the pattern above. One immediately sees that the orbifold has hyperbolic structures (provided $\chi < 0$) parametrized by the lengths of the cuts; that is, $(\mathbb{R}_+)^{l-3}$. Special care must be taken when, say, $m_1 = m_2 = 2$. Then one of 13.25
the cuts must be omitted, and an edge length becomes a parameter. In general any disjoint set of edges with ends on order 2 corner reflectors can be taken as positive real parameters, with extra parameters coming from cuts not meeting these edges:

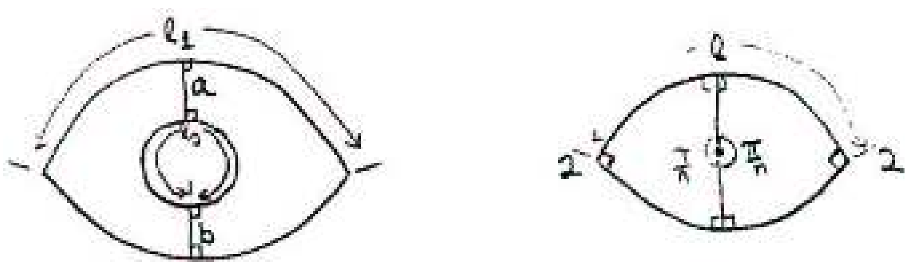


The annulus with more than one corner reflector on one boundary component should be dissected, as below, into $D_{(\,;n_1,\dots,n_k)}$ and an annulus with two order two corner reflectors. $D^2_{(n;\,m_1,\dots,m_l)}$ is analogous.

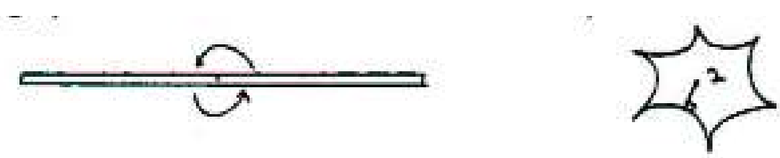


13.26

Hyperbolic structures on an annulus with two order two corner reflectors on one boundary component are parametrized by the length of the other boundary component, and the length of one of the edges:

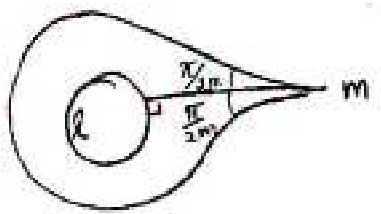


(The two all right pentagons agree on a and b, so they are congruent; thus they are determined by their edges of length $l_1/2$ and $l_2/2$). Similarly, $D^2_{(n;\,2,2)}$ is determined by one edge length, provided $n > 2$. $D^2_{(2;\,2,2)}$ is not hyperbolic. However, it has a degenerate hyperbolic structure as an infinitely thin rectangle, modulo a rotation of order 2—or, an interval.



This is consistent with the way in which it arises in considering hyperbolic structures, in the dissection of $D^2_{(2;m_1,\ldots,m_l)}$. One can cut such an orbifold along the perpendicular arc from the elliptic point to an edge, to obtain $D^2_{(\,;\,2,2,m_1,\ldots,m_l)}$. In the case of an annulus with only one corner reflector, 13.27

note first that it is symmetric, since it can be dissected into an isosceles "triangle." Now, from a second dissection, we see hyperbolic structures are paremetrized by the length of the boundary component without the reflector.



By the same argument, $D^2_{(n;m)}$ has a unique hyperbolic structure.

All these pieces can easily be reassembled to give a hyperbolic structure on O. $\square$

From the proof of 13.3.6 we derive

COROLLARY 13.3.7. *The Teichmüller space $\mathcal{T}(O)$ of an orbifold O with $\chi(O) < 0$ is homeomorphic to Euclidean space of dimension $-3\chi(X_O) + 2k + l$, where k is the number of elliptic points and l is the number of corner reflectors.*

PROOF. O can be dissected into primitive pieces, as above, by cutting along disjoint closed geodesics and arcs perpendicular to ∂X_O: i.e., one-dimensional hyperbolic 13.28
suborbifolds. The lengths of the arcs, and lengths and twist parameters for simple closed curves form a set of parameters showing that $\mathcal{T}(O)$ is homeomorphic to Euclidean space of some dimension. The formula for the dimension is verified directly for the primitive pieces, and so for disjoint unions of primitive pieces. When two circles are glued together, neither the formula nor the dimension of the Teichmüller space changes—two length parameters are replaced by one length parameter and one first parameter. When two arcs are glued together, one length parameter is lost, and the formula for the dimension decreases by one. $\square$

13.4. Fibrations.

There is a very natural way to define the tangent space $T(O)$ of an orbifold O. When the universal cover $\tilde{O}$ is a manifold, then the covering transformations act on $T(\tilde{O})$ by their derivatives. $T(O)$ is then $T(\tilde{O})/\pi_1(O)$. In the general case, O is made up of pieces covered by manifolds, and the tangent space of O is pieced together from the tangent space of the pieces. Similarly, any natural fibration over manifolds gives rise to something over an orbifold.

DEFINITION 13.4.1. A *fibration*, E, with generic fiber F, over an orbifold O is an orbifold with a projection

$$p : X_E \to X_O$$

between the underlying spaces, such that each point $x \in O$ has a neighborhood $U = \tilde{U}/\Gamma$ (with $\tilde{U} \subset \mathbb{R}^n$) 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. 13.29

$$\begin{tikzcd} \tilde{U} \times F \arrow[r] \arrow[d] & p^{-1}(U) \arrow[d] \\ \tilde{U} \arrow[r] & U \end{tikzcd}$$

With this definition, natural fibrations over manifolds give rise to natural fibrations over orbifolds.

The *tangent sphere bundle* $TS(M)$ is the fibration over M with fiber the sphere of rays through O in $T(M)$. When M is Riemannian, this is identified with the unit tangent bundle $T_1(M)$.

PROPOSITION 13.4.2. *Let O be a two-orbifold. If O is elliptic, then $T_1(O)$ is an elliptic three-orbifold. If O is Euclidean, then $T_1(O)$ is Euclidean. If O is bad, then $TS(O)$ admits an elliptic structure.*

PROOF. The unit tangent bundle $T_1(S^2)$ can be identified with the grup SO_3 by picking a "base" tangent vector V_O and parametrizing an element $g \in SO_3$ by the image vector $Dg(V_O)$. SO_3 is homeomorphic to $\mathbb{P}^3$, and its universal covering group os S^3. This correspondence can be seen by regarding S^3 as the multiplicative group of unit quaternions, which acts as isometries on the subspace of purely imaginary quaternions (spanned by i, j and k) by conjugation. The only elements acting trivially are ± 1. The action of SO_3 on $T_1(S^2) = SO_3$ corresponds to left translation so that for an orientable $O = S^2/\Gamma$, $T_1(O) = T_1(S^2/\Gamma) = \Gamma \backslash SO_3 = \tilde{\Gamma} \backslash S^3$ is clearly elliptic. Here $\tilde{\Gamma}$ is the preimage of Γ in S^3. (Whatever Γ stands for, $\tilde{\Gamma}$ is generally called "the binary Γ"—e.g., the binary dodecahedral group, etc.)

When O is not oriented, then we use the model $T_1(S^2) = O_3/\mathbb{Z}_2$, where $\mathbb{Z}_2$ is generated by the reflection r through the geodesic determined by V_O. Again, the 13.29
action of O_3 on $T_1(S^2)$ comes from left multiplication on $O_3/\mathbb{Z}_2$. An element gr, with $g \in SO_3$, thus takes $g'V_O$ to $grg'rV_O$. But $rg'r = sg's$, where $s \in SO_3$ is $180°$ rotation of the geodesic through V_O, so the corresponding transformations of S^3,

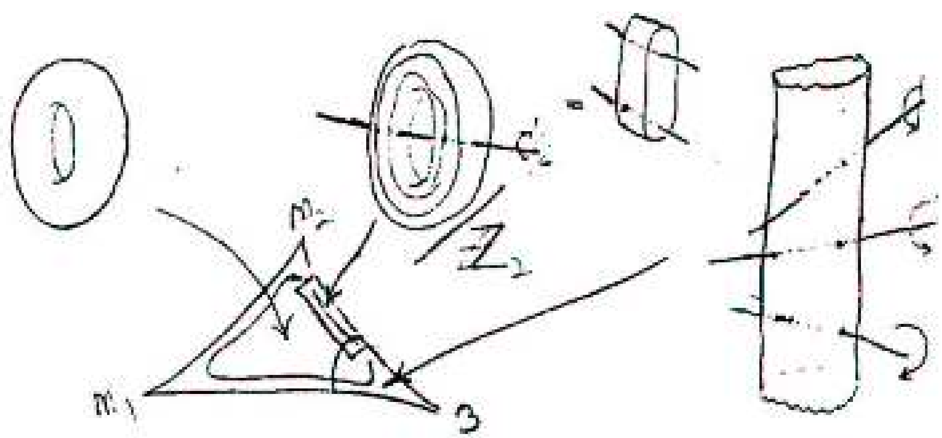
$\tilde{g} \mapsto (\tilde{g}\tilde{s})\, \tilde{g}'\, (\tilde{s})$, are compositions of left and right multiplication, hence isometries.

For the case of a Euclidean orbifold O, note that T_1E^2 has a natural product structure as $E^2 \times S^1$. From this, a natural Euclidean structure is obtained on T_1E^2, hence on $T_1(O)$.

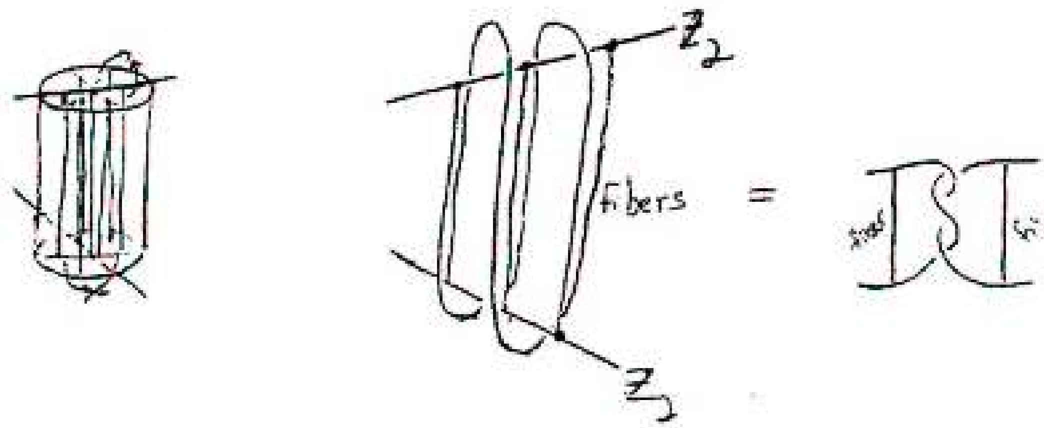
The bad orbifolds are covered by orbifolds $S^2_{(n)}$ or $S^2_{(n_1,n_2)}$. Then $TS(H)$, where H is either hemisphere, is a solid torus, so the entire unit tangent space is a lens space—hence it is elliptic. $TS(D^2_{(\,;n)})$, or $TSD^2_{(\,;n_1,n_2)}$, is obtained as the quotient by a $\mathbb{Z}_2$ action on these lens spaces. $\square$

As an example, $T_1(S^2_{(2,3,5)})$ is the Poincaré dodecahedral space. This follows immediately from one definition of the Poincaré dodecahedral space as S^3 modulo the binary dodecahedral group. In general, observe that $TS(O^2)$ is always a manifold if O^2 is oriented; otherwise it has elliptic axes of order 2, lying above mirrors and consisting of vectors tangent to the mirrors. In more classical terminology, the Poincaré dodecahedral space is a Seifert fiber space over S^2 with three singular fibers, of type $(2,1)$, $(3,1)$ and $(5,1)$.

When O has the combinatorial type of a polygon, it turns out that $X_{TS(O)}$ is S^3, with singular locus a certain knot or two-component link. There is an a priori reason to suspect that $X_{TS(O)}$ be S^3, since $\pi_1 O$ is generated by reflections. These reflections 13.31
have fixed points when they act on $TS(\tilde{O})$, so $\pi_1(X_{TS(O)})$ is the surjective image of $\pi_1 TS(\tilde{O})$. The image is trivial, since a reflection folds the fibers above its axis in half. Every easily producible simply connected closed three-manifold seems to be S^3. We can draw the picture of $TS(O)$ by piecing.

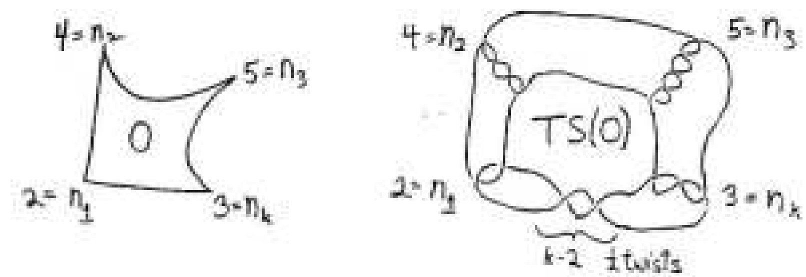


Over the non-singular part of O, we have a solid torus. Over an edge, we have $mI \times I$, with fibers folded into mI; nearby figures go once around these mI's. Above a corner reflector of order n, the fiber is folded into mI. The fibers above the nearby edges weave up and down n times, and nearby circles wind around $2n$ times.



13.32

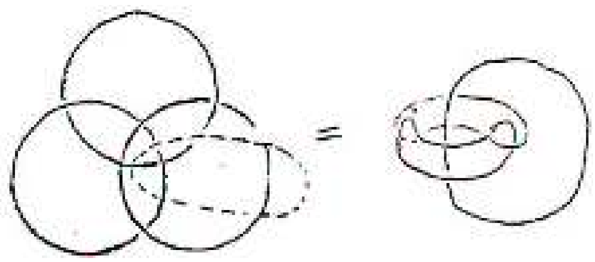
When the pieces are assembled, we obtain this knot or link:



When O is a Riemannian orbifold, this gives $T_1(O)$ a canonical flow, the geodesic flow. For the Euclidean orbifolds with X_O a polygon, this flow is physically realized (up to friction and spin) by the motion of a billiard ball. The flow is tangent to the singular locus. Thus, the phase space for the familiar rectangular billiard table is S^3:

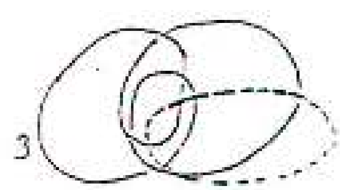
There are two invariant annuli, with boundary the singular locus, corresponding to trajectories orthogonal to a side. The other trajectories group into invariant tori.

Note the two-fold symmetry in the tangent space of a billiard table, which in the picture is 180° rotation about the axis perpendicular to the paper. The quotient orbifold is the same as example 13.1.5. 13.33

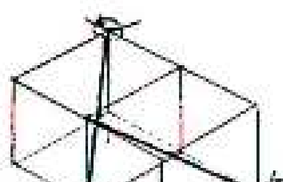


You can obtain many other examples via symmetries and covering spaces. For instance, the Borromean rings above have a three-fold axis of symmetry, with quotient orbifold:

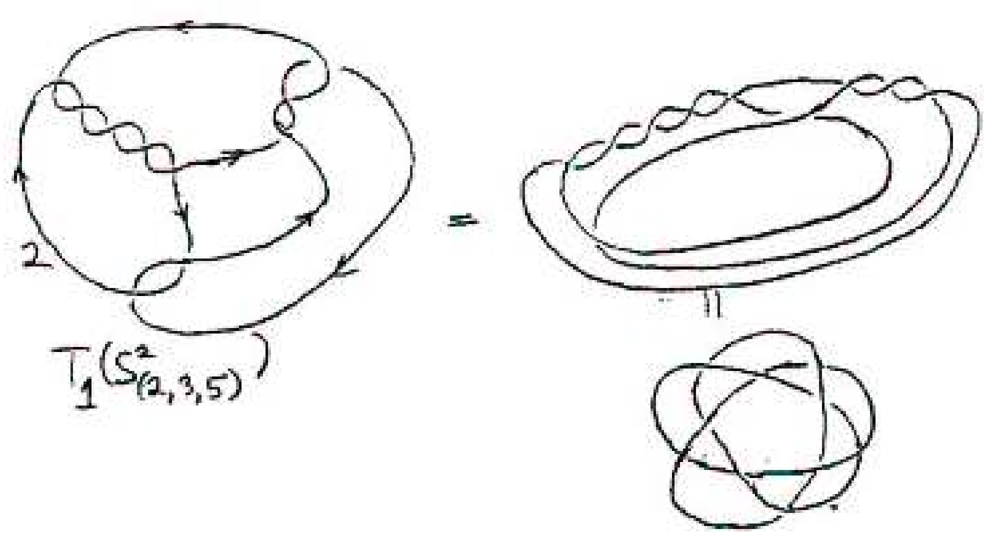
We can pass to a two-fold cover, unwrapping around the $\mathbb{Z}_3$ elliptic axis, to obtain the figure-eight knot as a $\mathbb{Z}_3$ elliptic axis.



This is a Euclidean orbifold, whose fundamental group is generated by order 3 rotations in main diagonals of two adjacent cubes (regarded as fundamental domains for example 13.1.5).



When O is elliptic, then all geodesics are closed, and the geodesic flow comes from a circle action. It follows that $T_1(O)$ is a fibration in a second way, by projecting to the quotient space by the geodesic flow! For instance, the singular locus of $T_1(D^2_{(2,3,5)})$ is a torus knot of type $(3,5)$: 13.34



Therefore, it also fibers over $S^2_{(2,3,5)}$. In general, an oriented three-orbifold which fibers over a two-orbifold, with general fiber a circle, is determined by three kinds of information:

(a) The base orbifold.
(b) For each elliptic point or corner reflector of order n, an integer $0 \leq k < n$ which specifies the local structure. Above an elliptic point, the $\mathbb{Z}_n$ action on $\tilde{U} \times S^1$ is generated by a $1/n$ rotation of the disk U and a k/n rotation of the fiber S^1. Above a corner reflector, the D_n action on $\tilde{U} \times S^1$ (with S^1 taken as the unit circle in $\mathbb{R}^2$) is generated by reflections of $\tilde{U}$ in lines making an angle of π/n and reflections of S^1 in lines making an angle of $k\pi/n$.
(c) A rational-valued Euler number for the fibration. This is defined as the obstruction to a rational section—i.e., a multiple-valued section, with rational weights for the sheets summing to one. (This is necessary, since there is not usually even a local section near an elliptic point or corner reflector).

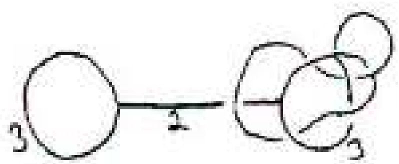
The Euler number for $TS(O)$ equals $\chi(O)$. It can be shown that a fibration of non- 13.35
zero Euler number over an elliptic or bad orbifold is elliptic, and a fibration of zero Euler number over a Euclidean orbifold is Euclidean.

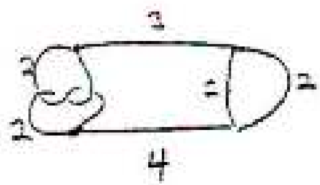
13.5. Tetrahedral orbifolds.

The next project is 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_O is the three-disk is interesting—the fundamental group of such an orbifold (if it is good) is called a *reflection group*. It turns out that the case when O has the combinatorial type of a tetrahedron is quite different from the general case. Geometrically, the case of a tetrahedron is subtle, although there is a simple way to classify such orbifolds with the aid of linear algebra.

The explanation for this distinction seems to come from the fact that orbifolds of the type of a simplex are non-Haken. First, we define this terminology.

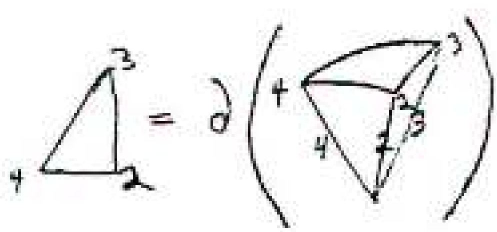
A closed three-orbifold is *irreducible* if it has no bad two-suborbifolds and if every two-suborbifold with an elliptic structure bounds a three-suborbifold with an elliptic structure. Here, an elliptic orbifold with boundary is meant to have totally geodesic boundary—in other words, it must be D^3/Γ, for some $\Gamma \subset O_3$. (For a non-oriented three-manifold,this definition entails being irreducible *and* $\mathbb{P}^2$-*irreducible*, in the usual terminology.) Observe that any three-dimensional orbifold with a bad suborbifold must itself be bad—it is conjectured that this is a necessary and sufficient condition for badness. 13.36

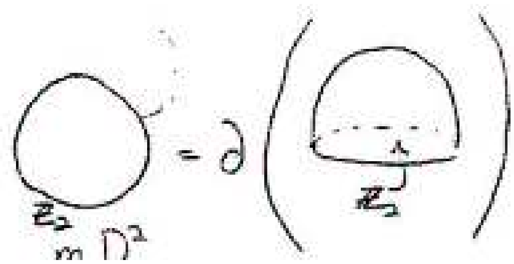




Frequently in three dimensions it is easy to see that certain orbifolds are good but hard to prove much more about them. For instance, the orbifolds with singular locus a knot or link in S^3 are always good: they always have finite abelian covers by manifolds.

Each elliptic two-orbifold is the boundary of exactly one elliptic three-orbifold, which may be visualized as the cone on it.





An *incompressible suborbifold* of a three-orbifold O, when X_O is oriented, is a two-suborbifold $O' \subset O$ with $\chi(O') \leq 0$ such that every one-suborbifold $O'' \subset O'$ which bounds an elliptic suborbifold of $O - O'$ bounds an elliptic suborbifold of O'. O is *Haken* if it is irreducible and contains an incompressible suborbifold.

Proposition 13.5.1. *Suppose $X_O = D^3$, $\Sigma_O = \partial D^3$. Then O is irreducible if and only if:*

(a) *The one-dimensional singular locus* Σ_O^1 *cannot be disconnected by the removal of zero, one, or two edges, and* 13.37

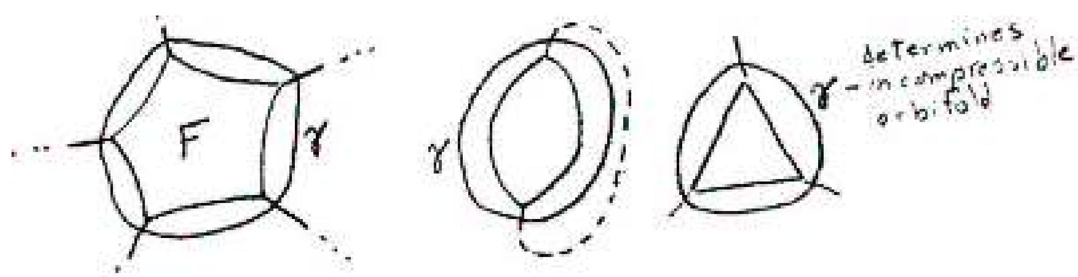
(b) *if the removal of* γ_1, γ_2 *and* γ_3 *disconnects* Σ_O^1, *then either they are incident to a commen vertex or the orders* n_1, n_2 *and* n_3 *satisfy*

$$1/n_1 + 1/n_2 + 1/n_3 \leq 1.$$

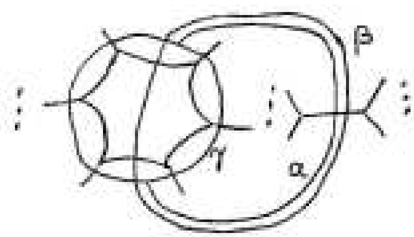
PROOF. For any bad or elliptic suborbifold $O' \subset O$, $X_{O'}$ must be a disk meeting Σ_O^1 in $1, 2$ or 3 points. $X_{O'}$ separates X_O into two three-disks; one of these gives an elliptic three-orbifold with boundary O' if and only if it contains no one-dimensional parts of Σ_O other than the edges meeting $\partial X_{O'}$. For any set E of edges disconnecting Σ_O^1 there is a simple closed curve on ∂X_O meeting only edges in E, meeting such an edge at most once, and separating $\Sigma_O^1 - E$. Such a curve is the boundary of a disk in X_O, which determines a suborbifold. Any closed elliptic orbifold S^n/Γ of dimension $n \geq 2$ can be *suspended* to give an elliptic orbifold S^{n+1}/Γ of dimension $n+1$, via the canonical inclusion $O_{n+1} \subset O_{n+2}$. □

PROPOSITION 13.5.2. *An orbifold* O *with* $X_O = D^3$ *and* $\Sigma_O = \partial D^3$ *is Haken if and only if it is irreducible, it is not the suspension of an elliptic two-orbifold and it does not have the type of a tetrahedron.*

PROOF. First, suppose that O satisfies the conditions. Let F be any *face* of O, that is a component of Σ_O minus its one dimensional part. The closure $\bar{F}$ is a disk or sphere, for otherwise O would not be irreducible. If F is the entire sphere, then O is the suspension of $D^2_{(;)}$. Otherwise, consider a curve γ going around just outside 13.38
F, and meeting only edges of Σ_O^1 incident to $\bar{F}$.



If γ meets no edges, then $\Sigma_O^1 = \partial F$ (since O is irreducible) and O is the suspension of $D^2_{(;n,n)}$. The next case is that γ meets two edges of order n; then they must really be the same edge, and O is the suspension of an elliptic orbifold $D^2_{(;n,n_1,n_2)}$. If γ meets three edges, then γ determines a “triangle” suborbifold $D^2_{(;n_1,n_2,n_3)}$ of O. O' cannot be elliptic, for then the three edges would meet at a point and O would have the type of a tetrahedron. Since $D^2_{(;n_1,n_2,n_3)}$ has no non-trivial one-suborbifolds, it is automatically incompressible, so O is Haken. If γ meets four or more edges, then the two-suborbifold it determines is either incompressible or compressible. But if it is compressible, then an automatically incompressible triangle suborbifold of O can be constructed.



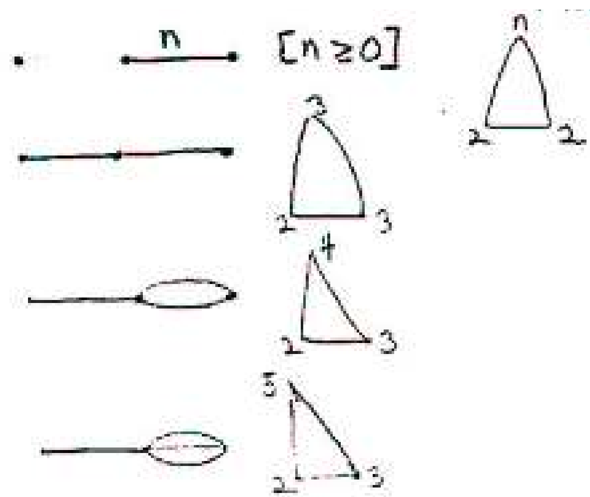
If α determines a "compression," then β determines a triangle orbifold.

The converse assertion, that suspensions of elliptic orbifolds and tetrahedral orbifolds are not Haken, is fairly simple to demonstrate. In general, for a curve γ on 13.39
∂X_O to determine an incompressible suborbifold, it can never enter the same face twice, and it can enter two faces which touch only along their common edge. Such a curve is evidently impossible in the cases being considered. $\square$

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 labelled 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 $X_O = D^n$ and $\Sigma_O \subset \partial D^n$, by using dotted lines to denote the faces which are not incident. However, for a complicated polyhedron—even the dodecahedron—this becomes quite unwieldy.

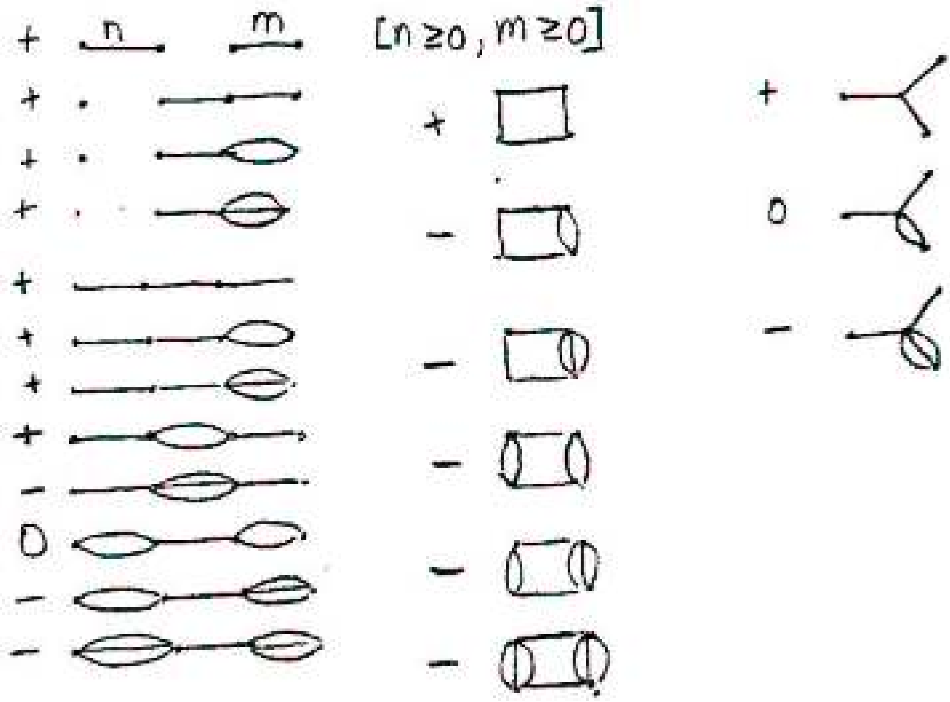
The condition for a graph with $n+1$ vertices to determine an orbifold (of the type of an n-simplex) is that each complete subgraph on n vertices is the Coxeter diagram for an elliptic $(n-1)$-orbifold.

Here are the Coxeter diagrams for the elliptic triangle orbifolds:



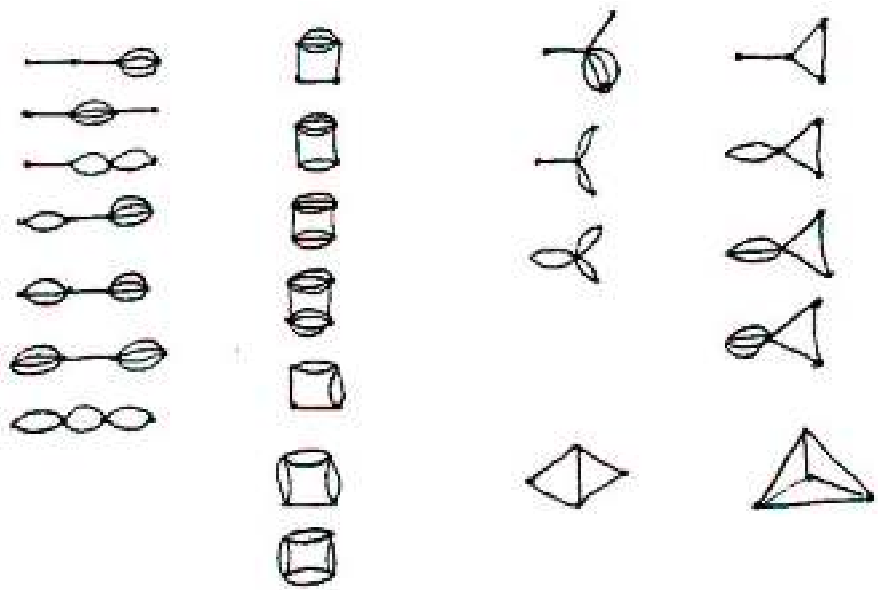
13.40

THEOREM 13.5.3. *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 below:*



This statement may be slightly generalized to include non-compact orbifolds of the combinatorial type of a simplex with some vertices deleted.

THEOREM 13.5.4. *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 struc-* 13.41
ture of finite volume. The three-dimensional examples are listed below:



PROOF OF 13.5.3 AND 13.5.4. 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 $V_O, \dots, V_n$ be unit vectors orthogonal to these subspaces in the direction away from σ^n. Clearly, the V_i are linearly indpendent. 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 contained in the intersection of a subspace of $E^{n,1}$ with the sphere of imaginary radius $X_1^2 + \cdots + X_n^2 - X_{n+1}^2 = -1$ (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 $V_0, \dots, V_n$ 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 $V_O, \dots, V_n$ be outward unit 13.42
vectors in directions orthogonal to the faces on σ^n. Once again, $V_i \cdot V_j = -\cos \alpha_{ij}$.

Given a collection $\{\alpha_{ij}\}$ of angles, we now try to construct a simplex. For the matrix M of presumed inner products, with l's down the diagonal and $-\cos \alpha_{ij}$'s 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 $V_O, \dots, V_n$ having the specified inner product matrix. The intersection of the half-spaces $X \cdot V_i \leq O$ 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 hyperbolic case, the cone determines a simplex in $\mathbb{RP}^n$, but the simplex may not be contained in $H^n \subset \mathbb{RP}^n$. To determine the positions of the vertices, observe that each vertex v_i determines a one-dimensional subspace, whose orthogonal subspace is spanned by $V_O, \dots, \hat{V}_i, \dots, V_n$. 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 subspace is positive definite, degenerate, or of type $(n-1, 1)$. Thus, the angles $\{\alpha_{ij}\}$ are the angles of an ordinary hyperbolic simplex if and only if M has type $(n, 1)$, and for each i the submatrix obtained by deleting the ith 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}_∞) if and only if 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 if and only if M is positive semidefinite of rank n. 13.43

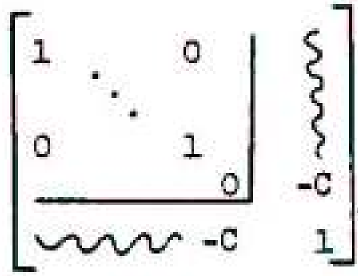
When the angles $\{\alpha_{ij}\}$ are derived from the Coxeter diagram of an orbifold, then each submatrix of M obtained by deleting the i-th row and the i-th column corresponds to an elliptic orbifold of dimension $n - 1$, hence it is positive definite. The full matrix must be either positive definite, of type $(n, 1)$ or positive semidefinite with rank n. It is routine to list the examples in any dimension. The sign of the determinant of M is a practical invariant of the type. We have thus proven theorem 13.5.

In the Euclidean case, it is not hard to see that the subspace of vectors of zero length with respect to M is spanned by $(a_0, \dots, a_n)$, where a_i is the $(n-1)$-dimensional area of the i-th face of σ.

To establish 13.5.4, 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

$$\begin{bmatrix} 1 & & & & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & 1 & \\ 0 & & & & 0 \end{bmatrix}$$

using $(a_0, \dots, \hat{a}_i, \dots, a_n)$ as the last basis vector. When the basis vector V_i is put back, the quadratic form determined by M becomes



where $-C = -\sum_{j \ni j \neq i} a_i \cos \alpha_{ij}$ 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. □

13.44

13.6. Andreev's theorem and generalizations.

There is a remarkably clean statement, due to Andreev, describing hyperbolic reflection groups whose fundamental domains are not tetrahedra.

THEOREM 13.6.1 (Andreev, 1967). (a) *Let O be a Haken orbifold with*

$$X_O = D^3, \quad \Sigma_0 = \partial D^3.$$

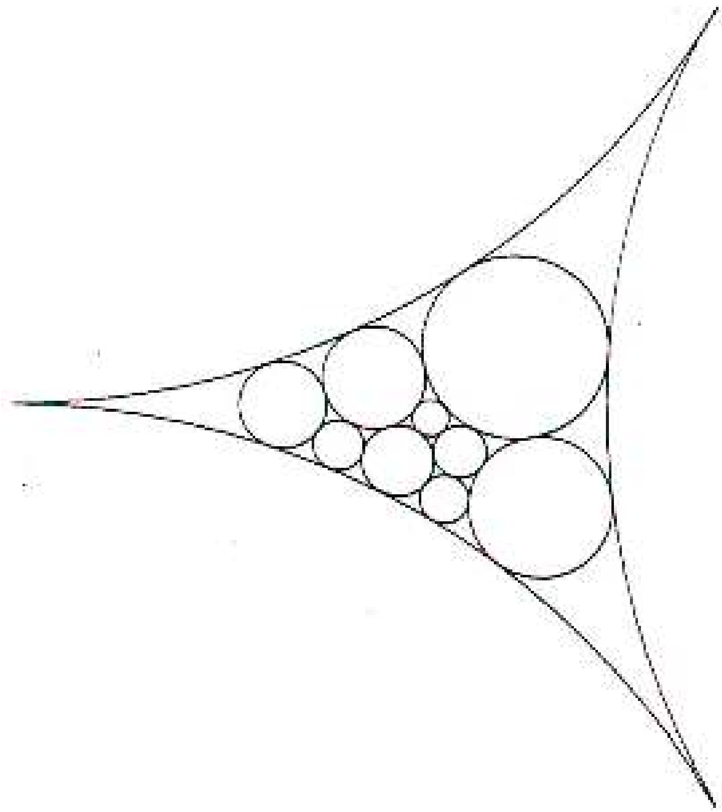
Then O has a hyperbolic structure if and only if O has no incompressible Euclidean suborbifolds.

(b) *If O is a Haken orbifold with $X_O = D^3$—(finitely many points) and $\Sigma_O = \partial X_O$, and if a neighborhood of each deleted point is the product of a Euclidean orbifold with an open interval, (but O itself is not such a product) then O has a complete hyperbolic structure with finite volume if and only if each incompressible Euclidean suborbifold can be isotoped into one of the product neighborhoods.*

The proof of 13.6.1 will be given in §??.

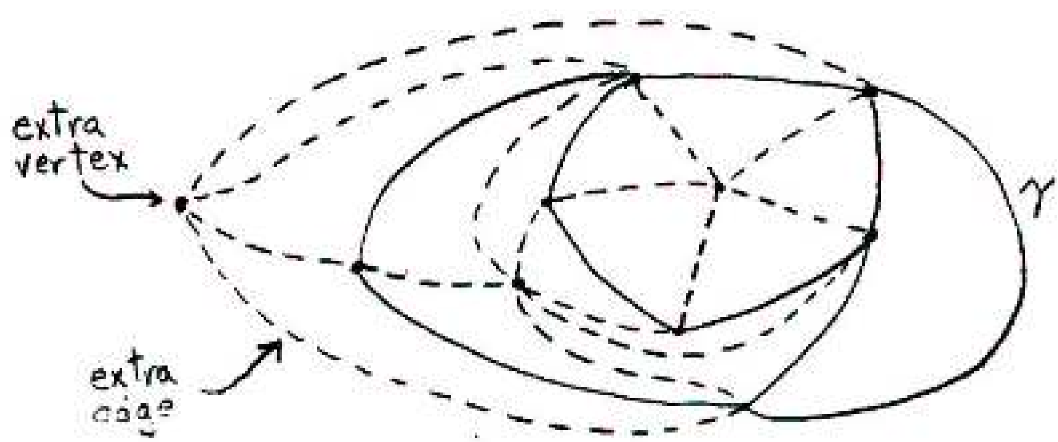
COROLLARY 13.6.2. *Let γ be any graph in $\mathbb{R}^2$, such that each edge has distinct ends and no two vertices are joined by more than one edge. Then there is a packing of circles in $\mathbb{R}^2$ whose nerve is isotopic to γ. If γ is the one-skeleton of a triangulation of S^2, then this circle packing is unique up to Moebius transformation.*

A *packing* of circles means a collection of circles with disjoint interiors. The nerve of a packing is then a graph, whose vertices correspond to circles, and whose edges correspond to pairs of circles which intersect. This graph has a canonical embedding in the plane, by mapping the vertices to the centers of the circles and the edges to straight line segments which will pass through points of tangency of circles. 13.44a

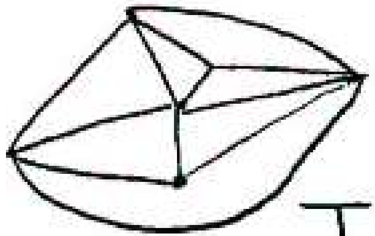


13.45

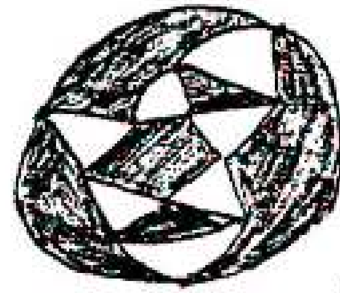
Proof of 13.6.2. We transfer the problem to S^2 by stereographic projection. Add an extra vertex in each non-triangular region of $S^2 - \gamma$, and edges connecting it to neighboring vertices, so that γ becomes the one-skeleton of a triangulation T of S^2.



Let P be the polyhedron obtained by cutting off neighborhoods of the vertices of T, down to the middle of each edge of T.



T

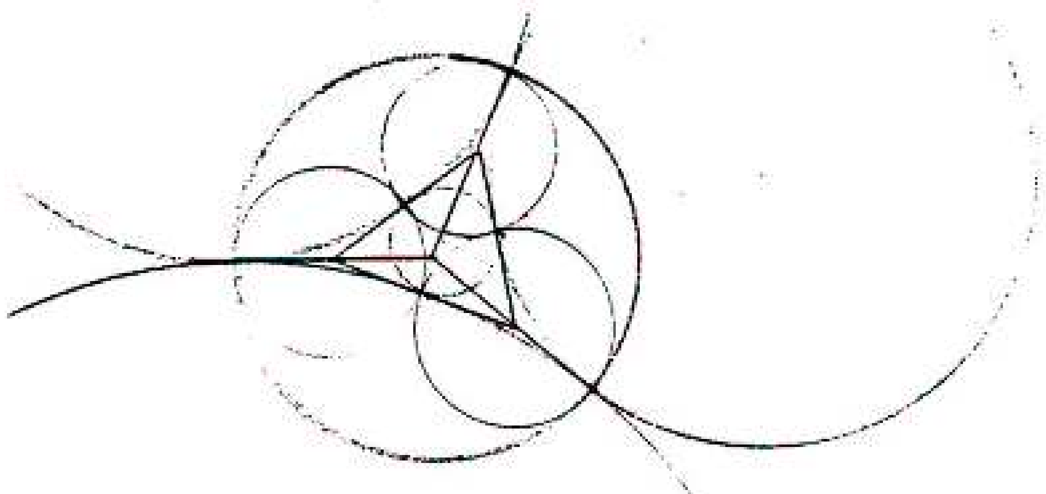


P

Let O be the orbifold with underlying space

$$X_O = D^3\text{-vertices of } P, \quad \text{and} \quad \Sigma^1_O = \text{ edges of } P,$$

each modelled on $\mathbb{R}^3/D_2$. For any incompressible Euclidean suborbifold O', ∂X_O must be a curve which circumnavigates a vertex. Thus, O satisfies the hypotheses of 13.6.1(b), and O has a hyperbolic structure. This means that P is realized as an ideal polyhedron in H^3, with all dihedral angles equal to 90°. The planes of the new faces of P (faces of P but not 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 13.46
by 13.6.2. The uniqueness statement is a consequence of Mostow's theorem, since the polyhedron P may be reconstructed from the packing of circles on S^2_∞. To make the reconstruction, observe that any three pairwise tangent circles have a unique common orthogonal circle. The set of planes determined by the packing of circles 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 O. □



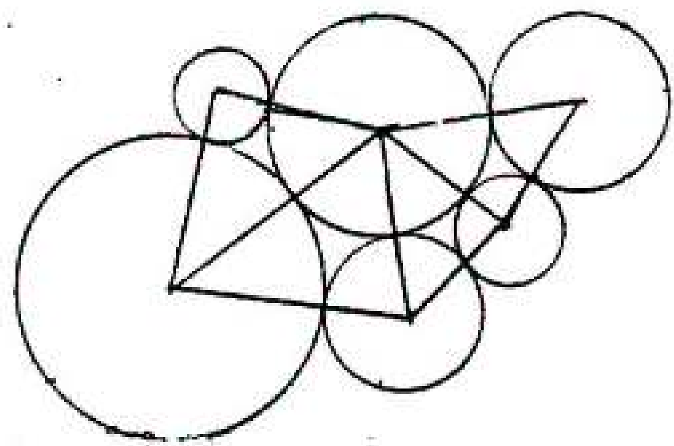
Remark. Andreev also gave a proof of uniqueness of a hyperbolic polyhedron with assigned concave angles, so the reference to Mostow's theorem is not essential.

Corollary 13.6.3. *Let T be any triangulation of S^2. Then there is a convex polyhedron in $\mathbb{R}^3$, 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 $\mathbb{R}^3 \subset \mathbb{P}^3$ which preserves the unit sphere.*

Proof of 13.6.3. Construct the ideal polyhedron P, as in the proof of 13.6.2. Embed H^3 in $\mathbb{P}^3$, as the projective model. The old faces of P (coming from faces of T) form a polyhedron in $\mathbb{P}^3$, combinatorially equivalent to T. Adjust by a projective transformation if necessary so that this polyhedron is in $\mathbb{R}^3$. (To do this, transform P so that the origin is in its interior.) □

Remarks. Note that the dual cell-division T^* to T is also a convex polyhedron in $\mathbb{R}^3$, with one-skeleton of T^* circumscribed about the unit sphere. The intersection $T \cap T^* = P$.

These three polyhedra may be projected to $\mathbb{R}^2 \subset \mathbb{P}^3$, by stereogrpahic projection, from the north pole of $S^2 \subset \mathbb{P}^3$. Stereographic projection is conformal on the tangent space of S^2, so 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 $\mathbb{R}^2$ of the nerve γ of the circle packing.



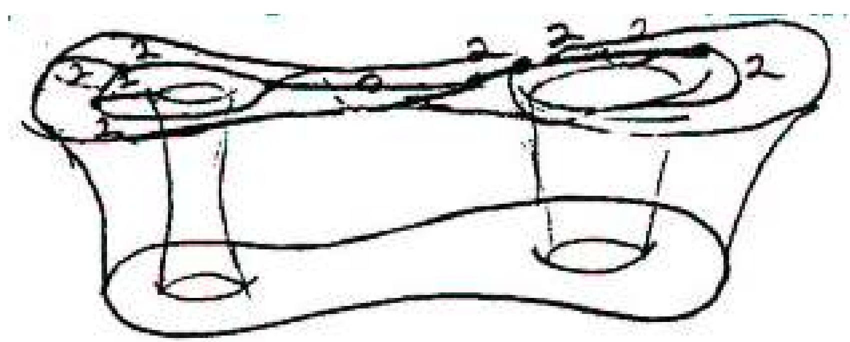
The existence of other geometric patterns of circles in $\mathbb{R}^2$ may also be deduced from Andreev's theorem. For instance, it gives necessary and sufficient condition for the existence of a family of circles meeting only orthogonally in a certain pattern, or 13.48
meeting at 60° angles.

One might also ask about the existence of packing circles on surfaces of constant curvature other than S^2. The answers are corollaries of the following theorems:

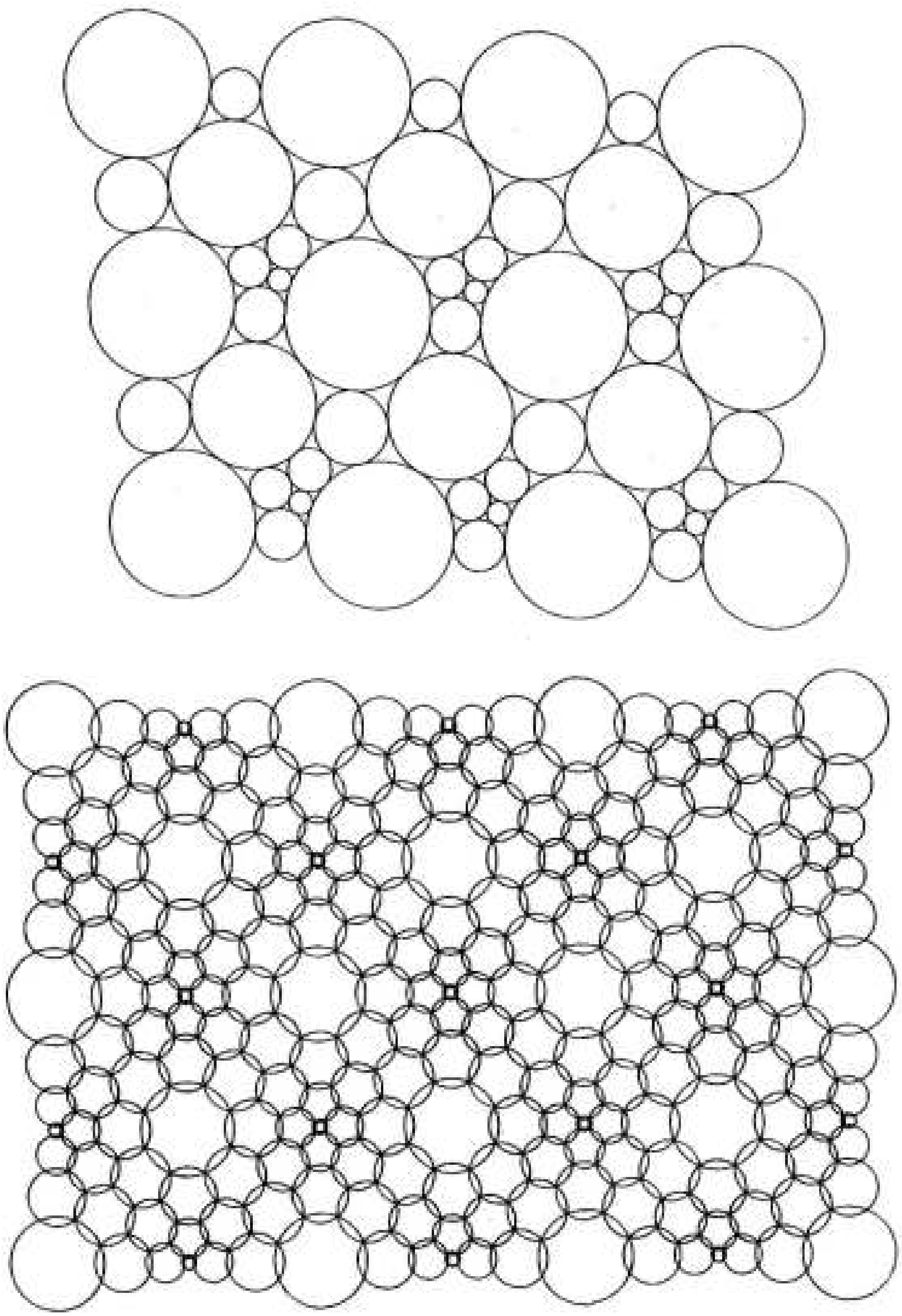
THEOREM 13.6.4. *Let O be an orbifold such that $X_O \approx T^2 \times [0, \infty)$, (with some vertices on $T^2 \times O$ having Euclidean links possibly deleted) and $\Sigma_O = \partial X_O$. Then O admits a complete hyperbolic structure of finite volume if and only if it is irreducible, and every incompressible complete, proper Euclidean suborbifold is homotopic to one of the ends.*

(Note that $mS^1 \times [0, \infty)$ is a complete Euclidean orbifold, so the hypothesis implies that every non-trivial simple closed curve on ∂X_O intersects Σ_O^1.)

THEOREM 13.6.5. *Let M^2 be a closed surface, with $\chi(M^2) < 0$. An orbifold O such that $X_O = M^2 \times [0, 1]$ (with some vertices on $M^2 \times 0$ having Euclidean links possibly deleted), $\Sigma_O = \partial X_O$ and $\Sigma_O^1 \subset M^2 \times O$. Then O has a hyperbolic structure if and only if it is irreducible, and every incompressible Euclidean suborbifold is homotopic to one of the ends.*



By considering $\pi_1 O$, O as in 13.6.4, as a Kleinian group in upper half space with $T^2 \times \infty$ at ∞, 13.6.4 may be translated into a statement about the existence of doubly periodic families of circles in the plane, or 13.48.a



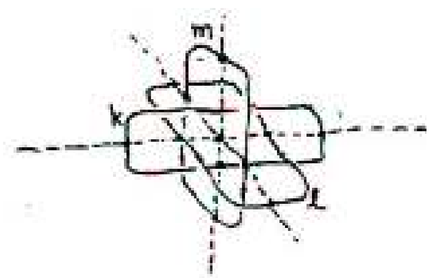
13.48.b

13.49

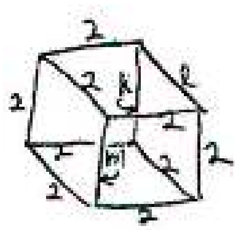
families of circles on flat toruses. Similarly, 13.6.5 is equivalent to a statement about families of circles in hyperbolic structures for M^2; in fact, since $M^2 \times 1$ has no one-dimensional singularities, it must be totally geodesic in any hyperbolic structure, so $\pi_1 M^2$ acts as a Fuchsian group. The face planes of $M^2 \times O$ give rise to a family of circles in the northern hemisphere of S^2_∞, invariant by this Fuchsian group, so each face corresponds to a circle in the hyperbolic structure for M^2.

Theorems 13.6.1, 13.6.4 and 13.6.5 will be proved in the next section, by studying patterns of circles on surfaces.

In example 13.1.5 we saw 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 Andreev's theorem, we may find all hyperbolic orbifolds which have 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. (Compare p. 13.4.)

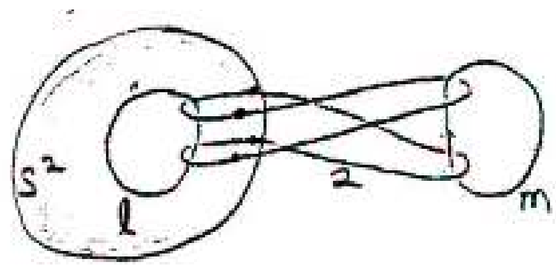


Thus, an orbifold O having the rings as elliptic axes of orders k, l and m is an eight-fold covering space of another orbifold, which has the combinatorial type of a cube.

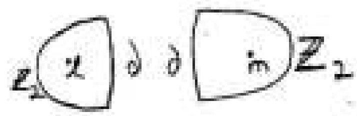


13.50

By Andreev'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 orders l and m and intersecting the elliptic axis of order 2 in four points. This forms an incompressible Euclidean suborbifold of O, which breaks O into



two halves, each fibering over two-orbifolds with boundary, but in incompatible ways (unless l or m is 2).



Base spaces of the fibrations

When $k = l = m = 4$, the fundamental domain, as in example 13.1.5, for $\pi_1 O$ acting on 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, to denote a parabolic cusp. When $l = m = \infty$, for instance, then O has a k-fold cover which is the complement of the untwisted $2k$-link chain D_{2k} of 6.8.7.

13.51

13.7. Constructing patterns of circles.

We will formulate a precise statement about patterns of circles on surfaces of non-positive Euler characteristic which gives theorems 13.6.4 and 13.6.5 as immediate consequences.

THEOREM 13.7.1. *Let S be a closed surface with $\chi(S) \leq 0$. Let τ be a cell-division of S into cells which are images of immersions of triangles and quadrangles which lift to embeddings in $\tilde{S}$. Let $\Theta : \mathcal{E} \to [0, \pi/2]$ (where $\mathcal{E}$ denotes the set of edges of τ) be any function satisfying the conditions below:*

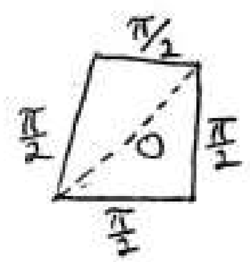
(i) *$\Theta(e) = \pi/2$ if e is an edge of a quadrilateral of τ.*

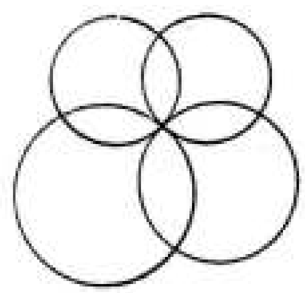
(ii) *If e_1, e_2, e_3 $[e_i \in \mathcal{E}]$ form a null-homotopic closed loop, and if $\sum_{i=1}^{3} \Theta(e_i) \geq \pi$, then these three edges form the boundary of a triangle of τ.*

(iii) *If e_1, e_2, e_3, e_4 form a null-homotopic closed loop and if $\sum_{i=1}^{4} \Theta(e_i) = 2\pi (\Leftrightarrow \Theta(e_i) = \pi/2)$, then the e_i form the boundary of a quadrilateral or of the union of two adjacent triangles.*

Then there is a metric of constant curvature on S, uniquely determined up to a scalar multiple, a uniquely determined geometric cell-division of S isotopic to τ so that the edges are geodesics, and a unique family of circles, one circle C_v for each vertex v of τ, so that C_{v_1} and C_{v_2} intersect at a positive angle if and only if v_1 and v_2 lie on a common edge. The angles in which C_{v_1} and C_{v_2} meet are determined by the common edges: there is an intersection point of C_{v_1} and C_{v_2} in a two-cell σ if and only if v_1 and v_2 are vertices of σ. If σ is a quadrangle and v_1 and v_2 are diagonally opposite, then C_{v_1} is tangent to C_{v_2}; otherwise, they meet at an angle of $\Theta(e)$, where 13.52
e is the edge joining them in σ.

Proof. First, observe that quadrangles can be eliminated by subdivision into two triangles by a new edge e with $\Theta(e) = 0$.

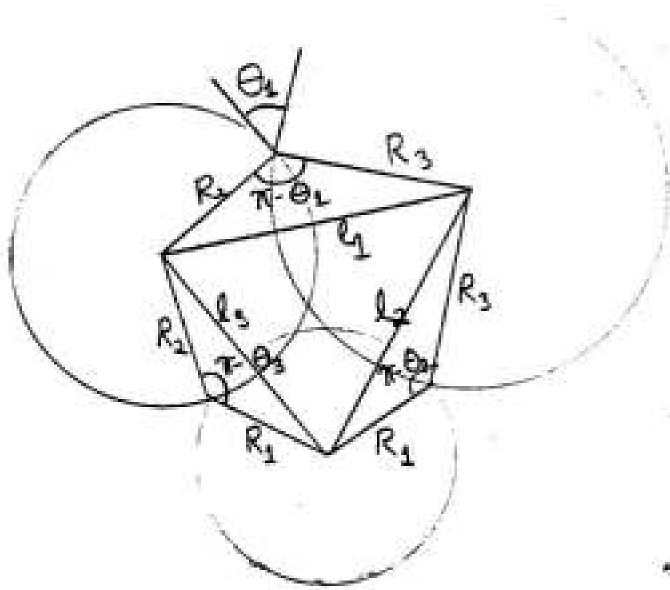




There is an extraneous tangency of circles here—in fact, all extraneous tangencies come from this situation. Henceforth, we assume τ has no quadrangles. The idea is to solve for the *radii* of the circles C_{v_1}. Given an arbitrary set of radii, we shall construct a Riemannian metric on S with cone 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 analogous to the idea that one can make a conformal change of any given Riemannian metric on a surface until it has constant curvature. Observe 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.

Lemma 13.7.2. *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 3 circles in both hyperbolic and Euclidean geometry, unique up to isometry, having radii R_i and meeting in angles θ_i.*

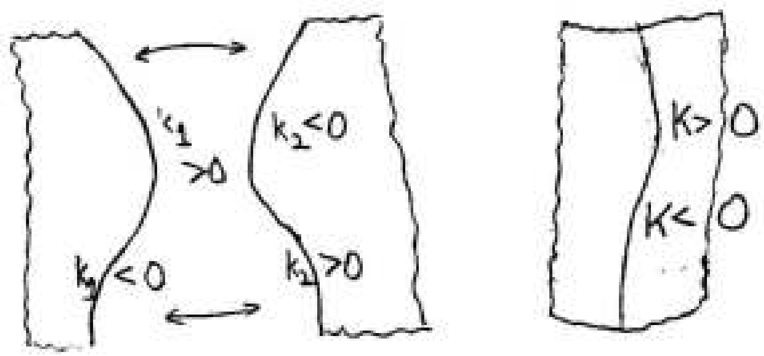
13.53



Proof of lemma. 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. $\square$

Proof of 13.7.1, continued. Let $\mathcal{V}$ denote the set of vertices of τ. For every element $R \in \mathbb{R}_+^{\mathcal{V}}$ (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 intersetcion as in 13.7.2. The triangles are taken in H^2 or 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 $K da$ becomes a measure κ on such 13.54
a surface. Tailors are of necessity familiar with curvature as a measure. Thus, a seam has curvature $(k_1 - k_2) \cdot \mu$, where μ 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 splitting 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.

Formally, we have a map

$$F : \mathbb{R}_+^{\mathcal{V}} \to \mathbb{R}^{\mathcal{V}}.$$

Given an element $R \in \mathbb{R}_+^{\mathcal{V}}$, 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 13.55
words, $F(R)(v) = \kappa_R(v)$. Our problem is to show that O is in the image of F, for then we will have a non-singular metric with the desired pattern of circles built in.

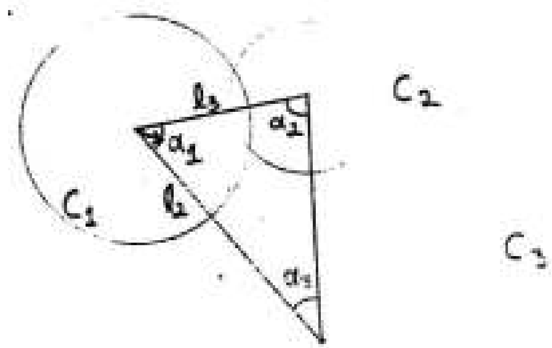
When $\chi(S) = 0$, then the shapes of the Euclidean triangles do not change when we multiply R by a constant, so $F(R)$ also does not change. Thus we may as well normalize so that $\sum_{v \in \mathcal{V}} R(v) = 1$. Let $\Delta \subset R_+^{\mathcal{V}}$ be this locus—Δ is the interior of the standard $|\mathcal{V}| - 1$ simplex. Observe, by the Guass-Bonnet theorem, that

$$\sum_{v \in \mathcal{V}} \kappa_R(v) = 0.$$

Let $Z \subset \mathbb{R}^{\mathcal{V}}$ 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 $\Delta \subset \mathbb{R}_+^{\mathcal{V}}$ 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). As one can easily believe, Δ intersects each ray through O 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 part of 13.7.1:

LEMMA 13.7.3. *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.*



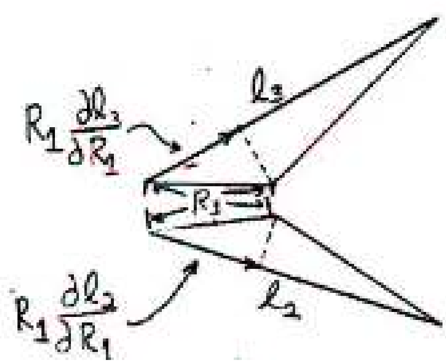
13.56

Thus we have

$$\frac{\partial \alpha_1}{\partial R_1} < 0 \quad , \quad \frac{\partial \alpha_2}{\partial R_1} > 0 \quad , \quad \frac{\partial \alpha_3}{\partial R_1} > 0,$$

where the α_i are the angles of the triangle of centers.

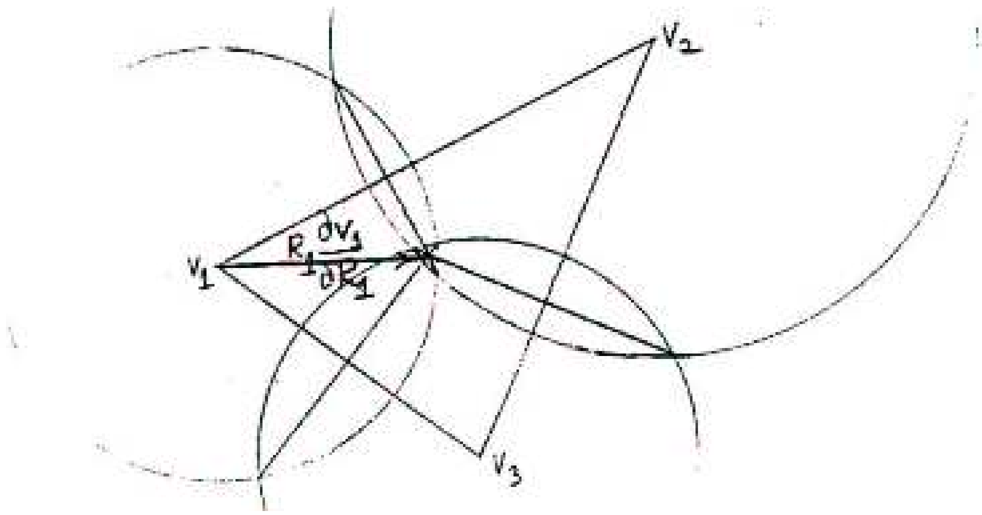
Proof of 13.7.3. Consider first the Euclidean case. Let l_1, l_2 and l_3 denote the lengths of the sides of the triangle of centers. The partial derivatives $\partial l_2/\partial R_1$ and $\partial l_3/\partial R_1$ can be computed geometrically.



If v_1 denotes the center of C_1, then $\partial v_1/\partial R_1$ is determined as the vector whose orthogonal projections sides 2 and 3 are $\partial l_2/\partial R_1$ and $\partial l_3/\partial R_1$. Thus,

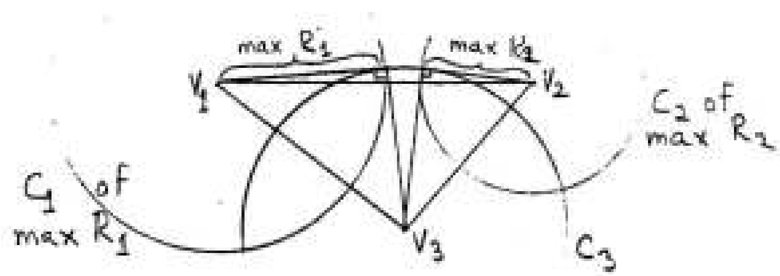
$$R_1 \, \frac{\partial v_1}{\partial R_1}$$

is the vector from v_1 to the intersection of the lines joining the pairs of intersection points of two circles.



13.57

When all angles of intersection of circles are acute, no circle meets the opposite side of the triangle of centers:



C_3 meets $\overline{v_1 v_2} \implies C_1$ and C_2 don't meet.

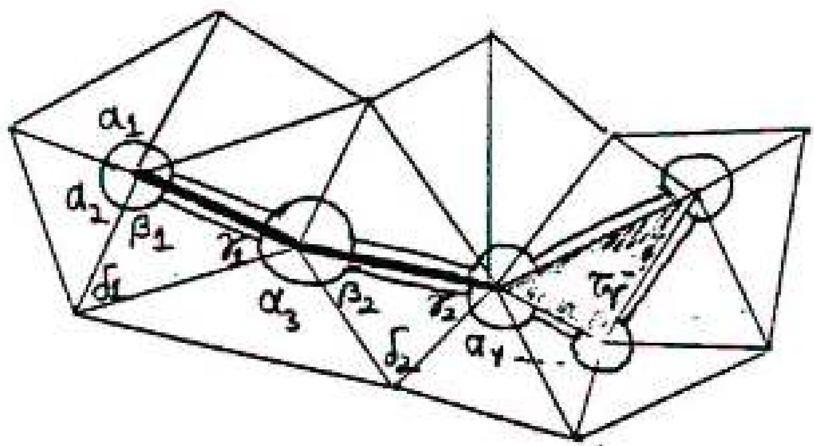
It follows that $\partial v_1/\partial R_1$ points to the interior of $\Delta v_1 v_2 v_3$.

The hyperbolic proof is similar, except that some of it takes place in the tangent space to H^2 at v_1. $\square$

Continuation of proof of 13.7.1. From lemma 13.7.3 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 $\mathbb{R}_+^{\mathcal{V}}$. The area near 0 is near 0, and near ∞ is near $\pi \times$ (# triangles $+ 2$# quadrangles); thus the ray intersects $\Delta = F^{-1}(Z)$ in a unique point.

It is now easy to prove that F is an embedding of Δ in Z. In fact, consider any two distinct points R and $R' \in \Delta$. Let $\mathcal{V}^- \subset \mathcal{V}$ be the set of v where $R'(v) < R(v)$. Clearly $\mathcal{V}^-$ is a proper subset. Let $\tau_{\mathcal{V}^-}$ be the subcomplex of τ spanned by $\mathcal{V}^-$. ($\tau_{\mathcal{V}^-}$ consists of all cells whose vertices are contained in $\mathcal{V}^-$). Let $S_{\mathcal{V}^-}$ be a small neighborhood of $\tau_{\mathcal{V}^-}$. We compare the geodesic curvature of $\partial S_{\mathcal{V}^-}$ in the two metrics. To do this, we

may arrange $\partial S_{\mathcal{V}^-}$ to be orthogonal to each edge it meets. Each arc of intersection
of $\partial S_{\mathcal{V}^-}$ with a triangle having one vertex in $\mathcal{V}^-$ contributes approximately α_i to the 13.58
total curvature, while each arc of intersection with a triangle having two vertices in $\mathcal{V}^-$ contributes approximately $\beta_i + \gamma_i - \pi$.



In view of 13.7.3, 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; $\pi - \delta_1 - (\beta_1 + \beta_2)$, 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 13.7.3, β_1 and γ_1 both increase. Hence, the geodesic curvature of $\partial S_{\mathcal{V}^-}$ increases.

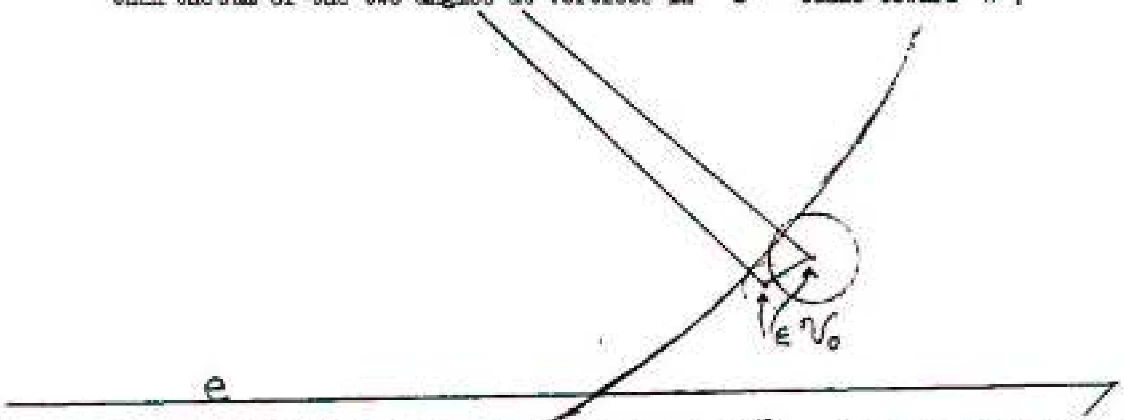
From the Gauss-Bonnet formula,

$$\sum_{v \in \mathcal{V}^-} \kappa(v) = \int_{\partial S_{\mathcal{V}^-}} d_g \, ds - \int_{S_{\mathcal{V}^-}} K \, dA + 2\pi\chi(S_{\mathcal{V}'})$$

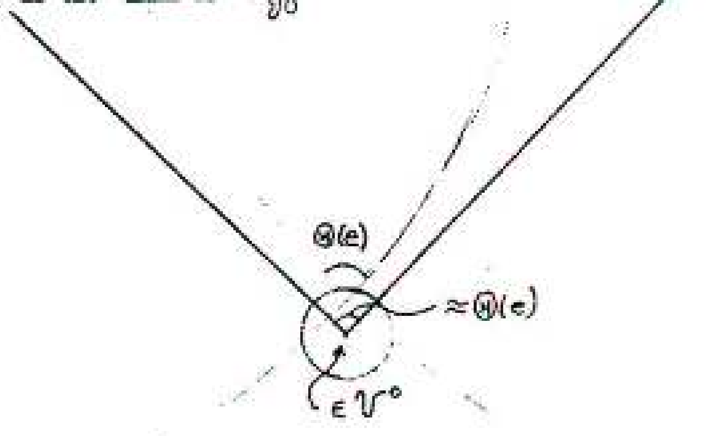
it follows that the total curvature at vertices in $\mathcal{V}^-$ must decrease in the R' metric. (Note that the area of $S_{\mathcal{V}^-}$ decreases, so if $k = -1$, the second term on the right
decreases.) In particular, $F(R) \neq F(R')$, which shows that F is an embedding of Δ. 13.59

The proof that O 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 $S_{\mathcal{V}^O}$, where $\mathcal{V}^O$ consists of the vertices v such that $R(v)$ is tending toward O. When a triangle of τ has two vertices in $\mathcal{V}^O$ and the third not in $\mathcal{V}^O$, then the sum of the two angles at vertices in $\mathcal{V}^O$ tends toward π.

The proof that O is in the image of F is based on the same principle as the proof of uniqueness. We can extract information about the limiting behaviour of F as R approaches $\partial\Delta$ by studying the total curvature of the subsurface $S_{\mathcal{V}^0}$, where $\mathcal{V}^0$ consists of the vertices v such that $R(v)$ is tending toward 0. When a triangle of τ has two vertices in $\mathcal{V}^0$ and the third not in $\mathcal{V}^0$, then the sum of the two angles at vertices in $\mathcal{V}^0$ tends toward π.



When a triangle of τ has only one vertex in $\mathcal{V}^0$, then the angle at that vertex tends toward the value $\pi - \Theta(e)$, where e is the opposite edge. Thus, the total curvature of $\partial S_{\mathcal{V}^0}$ tends toward the value $\sum_{e \in L(\tau_{\mathcal{V}^0})} (\pi - \Theta(e))$, where $L(\tau_{\mathcal{V}^0})$ is the "link of $\tau_{\mathcal{V}^0}$."



When a triangle of τ has only one vertex in $\mathcal{V}^O$, then the angle at that vertex tends toward the value $\pi - \Theta(e)$, where e is the opposite edge. Thus, the total curvature of $\partial S_{\mathcal{V}^O}$ tends toward the value

$$\sum_{e \in L(\tau_{\mathcal{V}^O})} \big(\pi - \Theta(e)\big),$$

where $L(\tau_{\mathcal{V}^O})$ is the "link of $\tau_{\mathcal{V}^O}$."

The Gauss-Bonnet formula gives 13.60

$$\operatorname{Lim} \sum_{v \in \mathcal{V}^O} \kappa(v) = -\sum_{e \in L(\tau_{\mathcal{V}^O})} \big(\pi - \Theta(e)\big) + 2\pi\chi(S_{\mathcal{V}^O}) < 0.$$

(Note that area $(S_{\mathcal{V}^O}) \to 0$.) To see that the right hand side is always negative, it suffices to consider the case that $\tau_{\mathcal{V}^O}$ is connected. Unless $\tau_{\mathcal{V}^O}$ has Euler characteristic one, both terms are non-positive, and the sum is negative. If $L(\tau_{\mathcal{V}^O})$ has length 5 or more, then

$$\sum_{e \in L(\tau_{\mathcal{V}^O})} \pi - \Theta(e) > e\pi,$$

so the sum is negative. The cases when $L(\tau_{\mathcal{V}^O})$ has length 3 or 4 are dealt with in hypotheses (ii) and (iii) of theorem 13.7.1.

When $\mathcal{V}'$ is any proper subset of $\mathcal{V}^O$ and $R \in \Delta$ is an arbitrary point, we also have an inequality

$$\sum_{v \in \mathcal{V}'} \kappa_R(v) > -\sum_{e \in L(\tau_{\mathcal{V}'})} \big(\pi - \Theta(e)\big) + 2\pi\chi(S_{\mathcal{V}'}).$$

This may be deduced quickly by comparing the R metric with a metric R' in which $R'(\mathcal{V}')$ is near 0. In other words, the image $F(\Delta)$ is contained in the interior of the polyhedron $P \subset Z$ defined by the above inequalities. Since $F(\Delta)$ is an open set whose boundary is ∂P, $F(\Delta) =$ interior (P). Since $O \in \operatorname{int}(P)$, this completes the proof of 13.7.1, and also that of 13.6.4, and 13.6.5. □

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 proces converges reasonably well. 13.61

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 $\operatorname{isom}(H^2) \subset \operatorname{isom}(H^3)$ and $\operatorname{isom}(E^2) \subset \operatorname{isom}(H^3)$, there is associated with such a surface S a hyperbolic three-manifold M_S, homeomorphic to $S \times \mathbb{R}$, with cone type singularities along (the singularities of S) $\times \mathbb{R}$. Each circle on S determines a totally geodesic submanifold (a "plane") in M_S. These, together with the totally geodesic surface isotopic to S when S is hyperbolic, cut out a submanifold of M_S with finite volume—it is an orbifold as in 13.6.4 or 13.6.5 but with singularities along arcs or half-lines running from the top to the bottom.

Corollary 13.7.4. *Theorems* 13.6.4 *and* 13.6.5 *hold when S is a Euclidean or hyperbolic orbifold, instead of a surface.* (*The orbifold O is to have only singularities as in* 13.6.4 *or* 13.6.5*, plus* (singularities of S) $\times I$ *or* (singularities of S) $\times [0, \infty)$.)

Proof. Solve for pattern of circles on S in a metric of constant curvature on S—the underyling surface of S will have a Riemannian metric with cone type singularities of curvature $2\pi(1/n-1)$ at elliptic points of S, and angles at corner reflectors of S.

An alternative proof is to find a surface $\tilde{S}$ which is a finite covering space of the orbifold S, and find a hyperbolic structure for the corresponding covering space $\tilde{O}$ of O. The existence of a hyperbolic structure for O follows from the uniqueness of the hyperbolic structure on $\tilde{O}$ thence the invariance by deck transformations of $\tilde{O}$ over O. □ 13.62

13.8. A geometric compactification for the Teichmüller spaces of polygonal orbifolds

We will construct hyperbolic structures for a much greater variety of orbifolds by studying the quasi-isometric deformation spaces of orbifolds with boundary whose underlying space is the three-disk. In order to do this, we need a description of the limiting behavior of conformal structure on its boundary. We shall focus on the case when the boundary is a disjoint union of polygonal orbifolds. For this, the greatest clarity is attained by finding the right compactifications for these Teichmüller spaces.

When M is an orbifold, $M_{[\epsilon,\infty)}$ is defined to consist of points x in M such that the ball of radius $\epsilon/2$ about x has a finite fundamental group. Equivalently, no loop through x of length $< \epsilon$ has infinite order in $\pi_1(M)$. $M_{(0,\epsilon]}$ is defined similarly. It does *not*, in general, contain a neighborhood of the singular locus. With this definition, it follows (as in §5) that each component of $M_{(0,\epsilon]}$ is covered by a horoball or a uniform neighborhood of an axis, and its fundamental group contains $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$ with finite index.

In §5 we defined the geometric topology on sequences of hyperbolic three-manifolds of finite volume. For our present purpose, we want to modify this definition slightly. First, define a *hyperbolic structure with nodes* on a two-dimensional orbifold O to be a complete hyperbolic structure with finite volume on the complement of some one-dimensional suborbifold, whose components are the *nodes*. This includes the case when there are no nodes. A topology is defined on the set of hyperbolic structures with nodes, up to diffeomorphisms isotopic to the identity on a given surface, by saying that M_1 and M_2 have distance $\leq \epsilon$ if there is a diffeomorphism of O [isotopic to the identity] whose restriction to $M_{1[\epsilon',\infty)}$ is a (e^ϵ)-quasi-isometry to $M_{2[\epsilon',\infty)}$. Here, ϵ' is some fixed, small number. 13.63

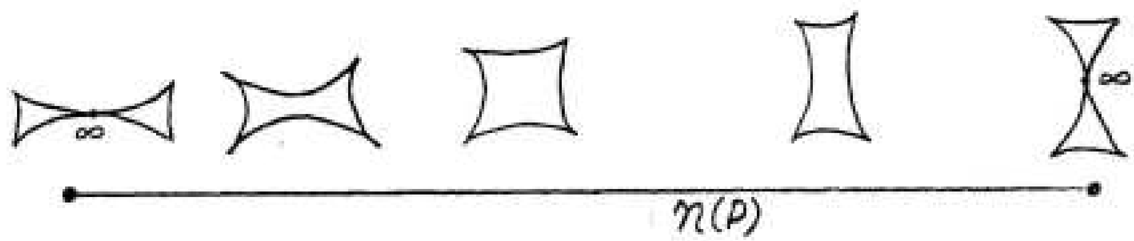
Remark. The related topology on hyperbolic structures with nodes *up to diffeomorphism* on a given surface is always compact. (Compare Jørgensen's theorem, 5.12, and Mumford's theorem, 8.8.3.) This gives a beautiful compactification for the modular space $\mathcal{T}(M)/\operatorname{Diff}(M)$, which has been studied by Bers, Earle and Marden and Abikoff. What we shall do works because a polygonal orbifold has a finite modular group.

For any two-dimensional orbifold O with $\chi(O) < 0$, let $\mathcal{N}(O)$ be the space of all hyperbolic structures with nodes (up to isotopy) on O.

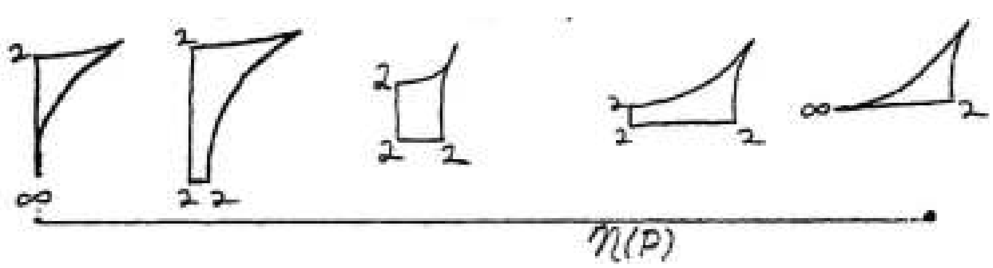
Theorem 13.8.1. *When P is an n-gonal orbifold, $\mathcal{N}(P)$ is homeomorphic to the (closed) disk, D^{n-3}, with interior $\mathcal{T}(P)$. It has a natural cell-structure with open cells parametrized by the set of nodes (up to isotopy).*

Here are the three simplest examples.

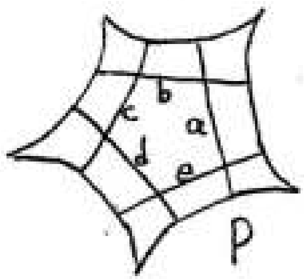
If P is a quadrilateral, then $\mathcal{T}(P)$ is $\mathbb{R}$. There are two possible nodes. $\mathcal{N}(P)$ looks like this:

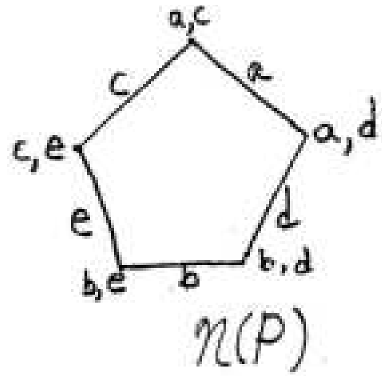


If there are two adjacent order 2 corner reflectors, the qualitative picture must be modified appropriately. For instance,



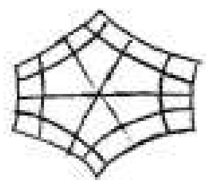
When P is a pentagon, $\mathcal{T}(P)$ is $\mathbb{R}^2$. There are five possible nodes, and the cell-structure is diagrammed below: 13.64



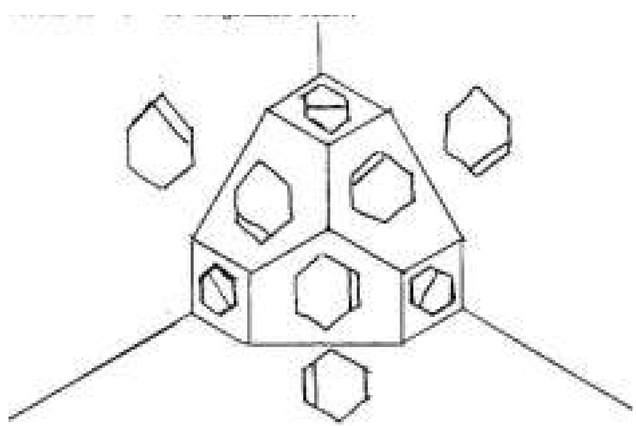


When there is only one node, the pentagon is pinched into a quadrilateral and a triangle, so there is still one degree of freedom.

When P is a hexagon, there are 9 possible nodes.



Each single node pinches the hexagon into a pentagon and a triangle, or into two quadrilaterals, so its associated 2-cell is a pentagon or a square. The cell division of ∂D^3 is diagrammed below:



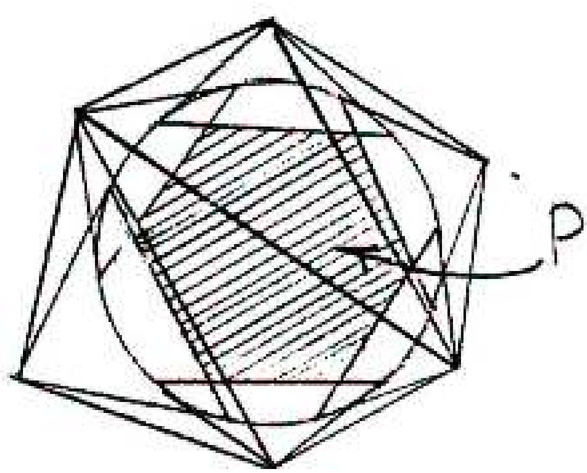
(The zero and one-dimensional cells are parametrized by the union of the nodes of 13.65
the incident 2-cells.)

PROOF OF 13.8.1. It is easy to see that $\mathcal{N}(P)$ is compact by familiar arguments, as in 5.12 and 8.8.3, for instance. In fact, choose ϵ sufficiently small so that $P_{(0,\epsilon]}$ is always a disjoint union of regular neighborhoods of short arcs. Given a sequence $\{P_i\}$, we can pass to a subsequence so that the core one-orbifolds of the components of $P_{i(0,\epsilon]}$ are constant. Extend this system of arcs to a maximal system of disjoint geodesic arcs $\{\alpha_1, \dots, \alpha_k\}$. The lengths of all such arcs remain bounded in $\{P_i\}$ (this follows from area considerations), so there is a subsequence so that all lengths converge—possibly to zero. But any set of $\{l(\alpha_i) | l(\alpha_i) \geq 0\}$ defines a hyperbolic structure with nodes, so our sequence converges in $\mathcal{N}(P)$.

Furthermore, we have described a covering of $\mathcal{N}(P)$ by neighborhoods diffeomorphic to quadrants, so it has the structure of a manifold with corners. Change of coordinates is obviously differentiable. Each stratum consists of hyperbolic structures with a prescribed set of nodes, so it is diffeomorphic to Euclidean space (this also follows directly from the nature of our local coordinate systems.)

Theorem 13.8.1 follows from this information. Here is a little overproof. An explicit homeomorphism to a disk can be constructed by observing that $\mathcal{PL}(P)$[‡] has a natural triangulation, which is dual to the cell structure of $\partial\mathcal{N}(P)$. This arises from the fact that any simple geodesic on P must be orthogonal to the mirrors, so a geodesic lamination on P is finite. The simplices in $\mathcal{PL}(P)$ are measures on a maximal family of geodesic one-orbifolds. 13.66

A projective structure for $\mathcal{PL}(P)$—that is, a piecewise projective[§] homeomorphism to a sphere—can be obtained as follows (compare Corollary 9.7.4). The set of geodesic laminations on P is in one-to-one correspondence with the set of cell divisions of P which have no added vertices. Geometrically, in fact, a geometric lamination extends in the projective (Klein) model to give a subdivision of the dual polygon.



Take the model P now to be a regular polygon in $\mathbb{R}^2 \subset \mathbb{R}^3$. Let V be the vertex set. For any function $f : V \to \mathbb{R}$, let C_f be the convex hull of the set of points

[‡]For definition, and other information, see p. 8.58

[§]See remark 9.5.9.

obtained by moving each vertex v of P to a height $f(v)$ (positive or negative along the perpendicular to $\mathbb{R}^2$ through v). The "top" of C_f gives a subdivision of P. The nature of this subdivision is unchanged if a function which extends to an affine function from $\mathbb{R}^2$ to $\mathbb{R}$ is added to f. Thus, we have a map $\mathbb{R}^V/\mathbb{R}^3 \to \mathcal{GL}(P)$. To lift the map to measured laminations, take the directional derivative at O of the bending measure for the top of the convex hull, in the direction f. The global description of this map is that a function f is associated to the measure which assigns to each edge e of the bending locus the change in slope of the intersection of the faces adjacent to e with a plane perpendicular to e.

It is geometrically clear that we thus obtain a piecewise linear homeomorphism, 13.67

$$e : \mathcal{ML}(P) \approx \mathbb{R}^{V-3} - 0.$$

The set of measures which assigns a maximal value of 1 to an edge gives a realization of $\mathcal{PL}(P)$ as a convex polyhedral sphere Q in $\mathbb{R}^{V-3}$. The dual polyhedron Q^*—which is, by definition, the set of vectors $X \in \mathbb{R}^{V-3}$ such that $\sup_{y \in Q} X \cdot Y = 1$—is the boundary of a convex disk, combinatorially equal to $\mathcal{N}(P)$. This seems explicit enough for now. $\square$

13.9. A geometric compactification for the deformation spaces of certain Kleinian groups.

Let O be an orbifold with underlying space $X_O = D^3$, $\Sigma_O \subset \partial D^3$, and $\partial \Sigma_O$ a union of polygons.

We will use the terminology *Kleinian structure* on O to mean a diffeomorphism of O to a Kleinian manifold $B^3 - L_\Gamma/\Gamma$, where Γ is a Kleinian group.

In order to describe the ways in which Kleinian structures on O can degenerate, we will also define the notion of a *Kleinian structure with nodes* on O. The nodes are meant to represent the limiting behavior as some one-dimensional suborbifold S becomes shorter and shorter, finally becoming parabolic. We shall see that this happens only when S is isotopic in one or more ways to ∂O; the geometry depends on the set of suborbifolds on ∂O isotopic to S which are being pinched in the conformal geometry of ∂O. To take care of the various possibilities, nodes are to be of one of these three types:

(a) An incompressible one-suborbifold of ∂O.
(b) An incompressible two-dimensional suborbifold of O, with Euler characteristic zero and non-empty boundary. In general, it would be one of these five: 13.68

but for the orbifolds we are considering only the last two can occur.

(c) An orbifold T modelled on $P_{2k} \times \mathbb{R}$, $k > 2$ where P_{2k} is a polygon with $2k$ sides. The sides of P_{2k} are to alternate being on ∂O and in the interior of O. (Cases a and b could be subsumed under this case by thickening them and regarding them as the cases $k = 1$ and $k = 2$.)

A Kleinian structure with nodes is now defined to be a Kleinian structure in the complement of a union of nodes of the above types, neighborhoods of the nodes in being horoball neighborhoods of cusps in the Kleinian structures. Of course, if O minus the nodes is not connected, each component is the quotient of a separate Kleinian group (so our definition was not general enough for this case).

Let $\mathcal{N}(O)$ denote the set of all Kleinian structure with nodes on O, up to homeomorphisms isotopic to the identity. As for surfaces, we define a topology on $\mathcal{N}(O)$, by saying that two structures K_1 and K_2 have distance $\leq \epsilon$ if there is a homeomorphism between them which is an $e^\epsilon -$ quasi-isometry on $K_{1[\epsilon,\infty)}$ intersected with the convex hull of K_1.

Theorem 13.9.1. *Let O be as above with O irreducible and ∂O incompressible. If O has one non-elementary Kleinian structure, then $\mathcal{N}(O)$ is compact. The conformal structure on ∂O is continuous, and it gives a homeomorphism to a disk,*

$$\mathcal{N}(O) \approx \mathcal{N}(\partial O).$$

Note: The necessary and sufficiently conditions for existence of a Kleinian structure will be given in [???] or they can be deduced from Andreev's theorem 13.6.1. 13.69
We will use 13.6.1 to prove existence.

Proof. We will study the convex hulls of the Kleinian structures with nodes on O. (When the Kleinian structure is disconnected, this is the union of convex hulls of the pieces.)

Lemma 13.9.2. *There is a uniform upper bound for the volume of the convex hull, H, of a Kleinian structure with nodes on O.*

Proof of 13.9.2. The bending lamination for ∂O has a bounded number of components. Therefore, H is (geometrically) a polyhedron with a bounded number of faces, each with a bounded number of sides. Hence the area of the boundary of the polyhedron is bounded. Its volume is also bounded, in view of the isoperimetric inequality,

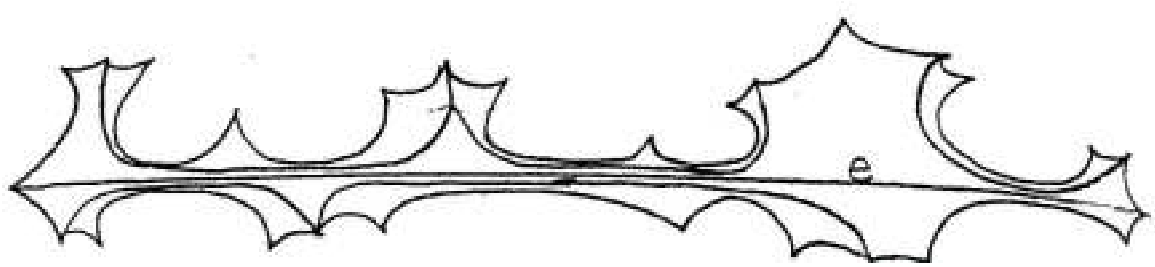
$$\text{volume }(S) \leq 1/2 \text{ area}(\partial S)$$

for a set $S \subset H^3$. (cf. §5.11). $\square$

Theorem 13.9.1 can now be derived by an adaptation of the proof of Jørgensen's theorem (5.12) to the present situation. It can also be proved by a direct analysis of the shape of H. We will carry through this latter course to make this proof more concrete and self-contained.

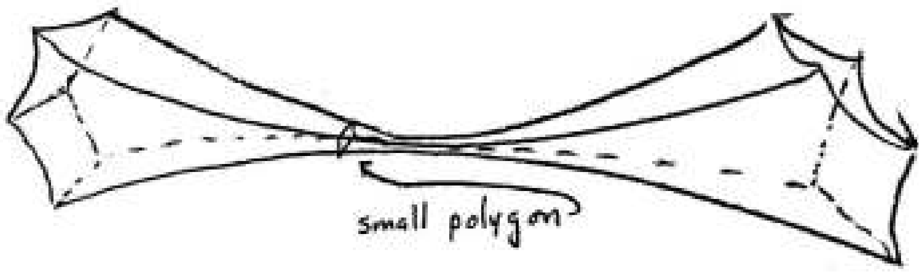
The first observation is that H can degenerate only when some edges of H become very long. When a face of H has vertices at infinity, "length" is measured here as the distance between canonical neighborhoods of the vertices. In fact, if the edges of H remain bounded in length, the faces remain bounded in shape by (§13.8, for 13.70
instance; the components of ∂H can be treated as single faces for this analysis). If we view X_H as a convex polyhedron in H^3 then as long as a sequence $\{H_i\}$ has all faces remaining bounded in shape, there is a subsequence such that the polyhedra $\{X_{H_i}\}$ converge, in the sense that the maps of each face into H^3 converge. One possibility is that the limiting map of X_H has a two-dimensional image: this happens in the case of a sequence of quasi-Fuchsian groups converging to a Fuchsian group, and we do not regard the limit as degenerate. The significant point is that two silvered faces of H (faces of H not on ∂H) which are not incident (along an edge or at a cusp) cannot come close together unless their diameter goes to infinity, because any points of close approach are deep inside $H_{(0,\epsilon]}$.

We can obtain a good picture of the degeneration which occurs as an edge becomes very long by the following analysis. We will consider only edges which are not in the interior of ∂H. Since the area of each face of H is bounded, any edge e of H which is very long must be close and nearly parallel, for most of its length all but a bounded part, of its length, on both sides, to other edges of its adjacent faces.



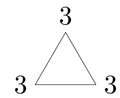
Similarly, these nearly parallel edges must be close and nearly parallel to still more edges on the far side from e. How long does this continue? Remember that H has an angle at each edge. In fact, if we ignore edges in the interior of ∂H, no angle exceeds $90°$. Special note should be made here of the angles between ∂H and mirrors 13.71

of H: the condition for convexity of H is that ∂H, together with its reflected image, is convex, so these angles also are $\leq 90°$. (If they are strictly less, then that edge of ∂H is part of the bending locus, and consequently it must have ends on order 2 corner reflectors.) Since H is geometrically a convex polyhedron, the only way that it can be bent so much along such closely spaced lines is that it be very thin. In other words, along most of the length of e, the planes perpendicular to $e \subset X_H \subset H^3$ intersect XH in a small polygon, which represents a suborbifold. It has $2, 3$ or 4 intersections with edges of XH not interior to ∂H.

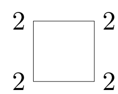


By area-angle considerations, this small suborbifold must have non-negative Euler characteristic. We investigate the cases separately.

(a) $\chi = 0, \quad \partial = \emptyset$

(i) 

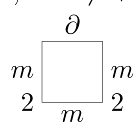
This is automatically incompressible, and since it is closed, it must be homotopic to a cusp. But this is supposed to be avoided by keeping our investigations away from the vertices of faces of P.

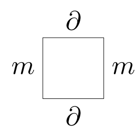
(ii) 

Either it is incompressible, and avoided as in (i), or compressible, so it is homotopic to some edge of H.

But since it is small, it must be very close to that edge. This contradicts the way it was chosen—or, in any case, it can account for only a small part 13.72
of the length of e.

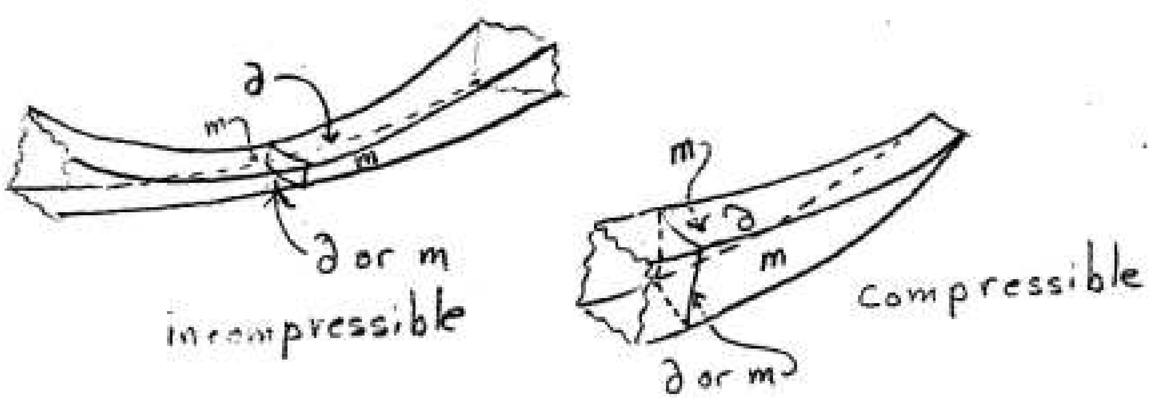
(b) $\chi = 0, \quad \partial \neq \emptyset$:

(i) 

(ii) 

where m denotes a mirror.

These can occur either as small ∂-incompressible suborbifolds (representing incipient two-dimensional nodes) or as small ∂-compressible orbifolds, representing the boundary of a neighborhood of an incipient one-dimensional node.



(c) $\chi > 0$. This can occur, since O is irreducible and ∂O incompressible.

We now can see that H is decomposed into reasonably wide convex pieces, joined together along long thin spikes whose cross-sections are two-dimensional orbifolds with boundary. There also may be some long thin spikes representing neighborhoods of short one-suborbifolds (arcs) of ∂O.

$H_{(0,\epsilon]}$ contains all the long spikes. It may also intersect certain regions between spikes, where two silvered faces of H come close together. If so, then $H_{(0,\epsilon]}$ contains the entire region, bounded by spikes (since each edge of the two nearby faces comes to a spike within a bounded distance, as we have seen). 13.73

The fundamental group of that part of H must be elementary: in other words, all faces represent reflections in planes perpendicular to or containing a single axis.

It should by now be clear that $\mathcal{N}(O)$ is compact. By [???], Kleinian structures with nodes of a certain type on O are parametrized, if they exist, by conformal structures with nodes of the appropriate type on ∂O. Given a Kleinian structure with nodes, K, and a nearby element K' in $\mathcal{N}(O)$, theer is a map with very small dilation from all but a small neighborhood of the nodes in ∂K to $\partial K'$, covering all but a long thin neck; this implies that $\partial K'$ is near ∂K in $\mathcal{N}(\partial O)$. Therefore, the map from $\mathcal{N}(O)$ to $\mathcal{N}(\partial O)$ is continuous. Since $\mathcal{N}(O)$ is compact, the image is all of $\mathcal{N}(\partial O)$. Since the map is one-to-one, it is a homeomorphism. $\square$

To be continued. . . .

William P. Thurston

The Geometry and Topology of Three-Manifolds

Electronic version 1.1 - March 2002

http://www.msri.org/publications/books/gt3m/

This is an electronic edition of the 1980 notes distributed by Princeton University. The text was typed in TeX by Sheila Newbery, who also scanned the figures. Typos have been corrected (and probably others introduced), but otherwise no attempt has been made to update the contents. Genevieve Walsh compiled the index.

Numbers on the right margin correspond to the original edition's page numbers.

Thurston's *Three-Dimensional Geometry and Topology*, Vol. 1 (Princeton University Press, 1997) is a considerable expansion of the first few chapters of these notes. Later chapters have not yet appeared in book form.

Please send corrections to Silvio Levy at levy@msri.org.

Index