## §0. Introduction

Our emphasis is on closed orientable surfaces, usually denoted by F, with the *genus* of F, the number of torus summands, denoted by g. An *automorphism* of a surface F we take to be a homeomorphism $h:F \to F$, usually orientation-preserving. The Nielsen-Thurston theory generalizes the well-known classification of toral automorphisms to the automorphisms of an arbitrary orientable surface. We begin by briefly recalling this classification.

Regard the torus as the quotient of the Euclidean plane $\mathbb{R}^2$ by the integer lattice $\mathbb{Z}^2$, endowed with a fixed orientation. Hence $\pi_1(T^2) \cong \mathbb{Z}\oplus\mathbb{Z}$. The homeomorphisms of $T^2$ correspond to the elements of the general linear group $GL_2(\mathbb{Z})$ as any element $\alpha$ in $GL_2(\mathbb{Z})$ maps $\mathbb{Z}^2$ to itself and so induces a continuous map $h_\alpha: T^2 \to T^2$. The homeomorphism $h_\alpha$ has inverse $h_{\alpha^{-1}}$ and $(h_\alpha)^*: \pi_1(T^2) \to \pi_1(T^2)$ has matrix $\alpha$. The map $h_\alpha$ preserves orientation if and only if $\det(\alpha) = 1$. i.e. $\alpha$ is an element of the special linear group $SL_2(\mathbb{Z})$.

If $\alpha$ in $SL_2(\mathbb{Z})$ is represented by the 2x2 matrix $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$, the characteristic polynomial of $\alpha$, $t^2-(p+s)t+(ps-rq)$, can be written as $t^2-(\mathrm{trace}(\alpha))t+1$. The eigenvalues $\lambda, \lambda^{-1}$ of $\alpha$ are either

1) complex, i.e. $\mathrm{trace}(\alpha) = 0, 1$, or $-1$

or 2) both $\pm 1$, i.e. $\mathrm{trace}(\alpha) = \pm 2$

or 3) distinct reals, i.e. $|\mathrm{trace}(\alpha)| > 2$.

Consider each of these cases in turn.

In case 1) an easy exercise in the Cayley-Hamilton theorem shows that $\alpha$ is of finite order and $(h_\alpha)^{12} = 1$. The map $h_\alpha$ is said to be *periodic*.

In case 2) $\alpha$ has an integral eigenvector which projects to an essential simple closed curve C under the quotient map $\mathbb{R}^2 \to T^2$. The curve C is invariant under the map $h_\alpha$, but possibly with reversed string orientation. In this case the map $h_\alpha$ is a power of a *Dehn twist* in C. More precisely, the curve C has a regular neighborhood A homeomorphic to an annulus; which we consider parameterized as $\{[r,\theta] \mid 1 \leq r \leq 2\}$. The *Dehn twist* in C is defined to be the homeomorphism given by the identity off A and by the map $[r,\theta] \to [r,\theta+2\pi r]$ on A, see Figure 0.1.

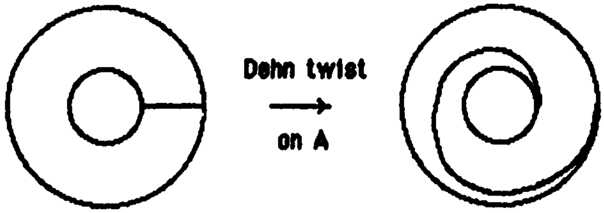


Figure 0.1

In this case the map $h_\alpha$ is said to be *reducible.*

For case 3) suppose that $|\lambda| > 1 > |\lambda^{-1}|$ and that x and x′ are the corresponding real eigenvectors. The map $h_\alpha$ has infinite order and leaves no simple closed curve invariant. Translating the vectors x,x′ yields vector fields $\mathfrak{F}, \mathfrak{F}'$ which are carried by $h_\alpha$ to vector fields $\lambda\mathfrak{F}$ and $\lambda^{-1}\mathfrak{F}'$ respectively. That is, $h_\alpha$ is a linear homeomorphism which stretches by a factor $\lambda$ in one direction and shrinks by the same factor in a complementary direction. When this occurs $h_\alpha$ is called *Anosov.*

## §1. The Hyperbolic Plane $\mathbb{H}^2$

This chapter contains all the necessary background material on hyperbolic plane geometry. Much of the work generalizes to higher dimensions. We use the Poincare disk model which identifies the hyperbolic plane with the interior of the unit disk $D^2$ in the Euclidean plane $\mathbb{R}^2$. The boundary of $D^2$ is the *circle at infinity* $S^1_\infty$; notice that $\mathbb{H}^2 \cap S^1_\infty$ is empty.

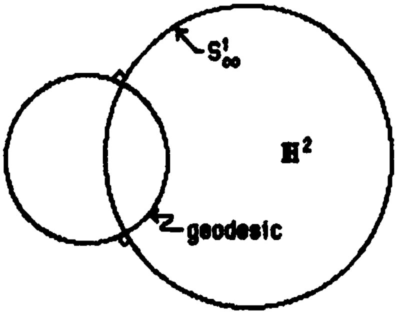


Figure 1.1

It is convenient to regard $\mathbb{R}^2$ as embedded in its one-point compactification $\mathbb{R}^2 \cup \{\infty\}$, so that Euclidean lines can be regarded as circles through $\infty$. A *geodesic* (or straight line) in $\mathbb{H}^2$ is $C \cap \mathbb{H}^2$, where $C$ is a circle in $\mathbb{R}^2 \cup \{\infty\}$ meeting $S^1_\infty$ orthogonally. There is a unique geodesic joining any two points in $\mathbb{H}^2$. The *angle* between two intersecting geodesics is the Euclidean angle between their defining circles.

Next, the operation of *reflection* in a geodesic of $\mathbb{H}^2$ will be defined in terms of Euclidean inversion. If C is a circle in $\mathbb{R}^2$ with center O and radius r, *inversion* in C carries a point $P \in \mathbb{R}^2-\{O\}$ to the unique point P' on the ray OP such that $OP \cdot OP' = r^2$, and interchanges O with $\infty$. If C is a line in $\mathbb{R}^2$, inversion in C is just Euclidean reflection in C. In either case, inversion in C defines an involution of $\mathbb{R}^2 \cup \{\infty\}$.

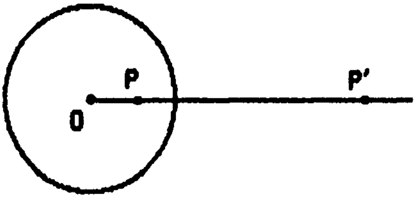


Figure 1.2

**Lemma 1.1:** Inversions preserve angles (but reverse orientation).

**Lemma 1.2:** Inversions carry circles in $\mathbb{R}^2 \cup \{\infty\}$ to circles in $\mathbb{R}^2 \cup \{\infty\}$.

The proofs are (well-known) exercises in Euclidean geometry.

If $C \cap \mathbb{H}^2$ is a geodesic in $\mathbb{H}^2$ then inversion in C induces an involution of $\mathbb{H}^2$, called *reflection* in $C \cap \mathbb{H}^2$. An *isometry* of $\mathbb{H}^2$ is defined to be a product of reflections. By 1.1, isometries preserve angles.

**Lemma 1.3:** The group of isometries acts transitively on $\mathbb{H}^2$, and the stabilizer of any point in $\mathbb{H}^2$ is isomorphic to O(2).

*Proof:* Let O be the center of D, A any point of $\mathbb{H}^2$. Then A can be carried to O by a single reflection in a geodesic $C \cap \mathbb{H}^2$, C having center on the ray OA. So any two points of $\mathbb{H}^2$ are "connected" by an isometry that is the product of at most two reflections.

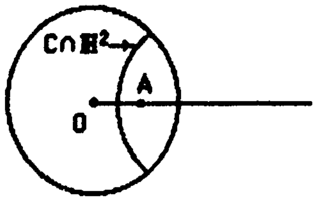


Figure 1.3

In light of this it follows that the stabilizers of any two points in $\mathbb{H}^2$ are isomorphic via conjugation. Thus we only need determine Stab(O). This group contains the reflections in the lines through O. A rotation about O can be expressed as the product of two reflections and these generate O(2).

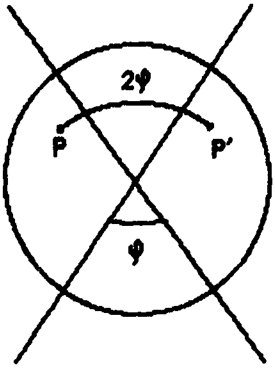


Figure 1.4

To see that O(2) is the entire stabilizer of 0, notice that isometries extend in a unique manner to $S^1_\infty$. So it is enough to show that any isometry extending to the identity on $S^1_\infty$ is actually the identity. But for any point P in $\mathbb{H}^2$ which is the intersection of geodesics $\gamma,\gamma'$, the "ends" of $\gamma,\gamma'$ are fixed, hence $\gamma,\gamma'$ are fixed, and hence P.

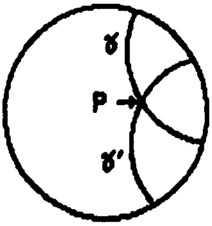


Figure 1.5

**Lemma 1.4:** Isometries leave $ds/(1-r^2)$ invariant where r denotes Euclidean distance from the center of D and ds=d(Euclidean distance).

*Proof:* We must show that if $P' = f(P)$ and $Q' = f(Q)$, where f is an isometry of $\mathbb{H}^2$, then

$$\frac{P'Q'}{1-r'^2} \approx \frac{PQ}{1-r^2}.$$

It follows from Lemma 1.3 that it is enough to check this when P is the center of the Poincare disk and f is a reflection in $C \cap \mathbb{H}^2$ (where C is a Euclidean circle with center O and radius k).

By properties of inversion, $\dfrac{P'Q'}{PQ} \approx \dfrac{OP'}{OP}$ when Q is near P.

From Figure 1.6 $\dfrac{OP'}{OP} = \dfrac{k^2}{(OP)^2} = 1-r'^2$.

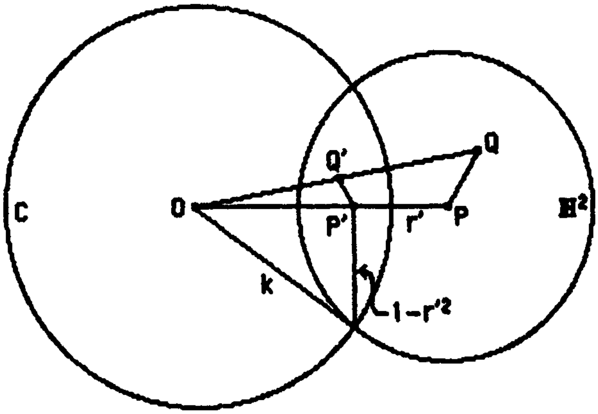


Figure 1.6

Hence $\dfrac{P'Q'}{1-r'^2} \approx \dfrac{PQ}{1-0^2}$.

**Exercise:** Show that geodesics with respect to this metric really are geodesics.

**Remark:** It suffices to prove that lines through the center of the Poincare disk minimize hyperbolic arc length.

**Definition:** The *hyperbolic metric* on $\mathbb{H}^2$ is given by $2ds/(1-r^2)$.

**Convention:** Greek letters denote hyperbolic distance, Roman letters denote Euclidean distance.

**Examples:**

1. Hyperbolic distance

$$OP = \rho = \int_0^r \frac{2dx}{1-x^2} = 2\tanh^{-1} r$$

Therefore, $r = \tanh \rho/2$.

2. The circle centered at O of hyperbolic radius $\rho$ has circumference

$$2\pi \frac{2r}{1-r^2} = \frac{4\pi r}{1-r^2} = 2\pi \tanh \rho/2 \cosh^2 \rho/2 = 2\pi \sinh\rho.$$

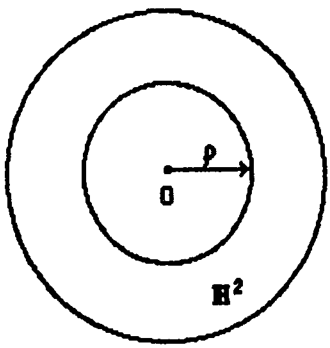


Figure 1.7

3. The circle of the previous example has area

$$\int_0^r \frac{4\pi x}{(1-x^2)^2}\,2dx = 4\pi\left(\frac{1}{1-r^2}-1\right) = 2\left(\frac{2r^2}{1-r^2}\right)$$

In terms of hyperbolic distance, this is

$$\int_0^\rho 2\pi \sinh\tau\, d\tau = 2\pi(\cosh\rho - 1).$$

**Theorem 1.5:** (Gauss–Bonnet) A geodesic triangle in $\mathbb{H}^2$ with angles $\alpha, \beta, \gamma$ has area $\pi-(\alpha+\beta+\gamma)$.

*Proof:* Without loss of generality the $\alpha$ vertex is at the center of the Poincare disk, and as we can subdivide our triangles into right triangles we may assume $\gamma = \pi/2$. First note

$$k\sin\beta\, d\theta = r\,d\alpha \Rightarrow \frac{d\theta}{d\alpha} = \frac{r}{k\sin\beta} = \frac{2r}{1/r - r} = \frac{2r^2}{1-r^2}.$$

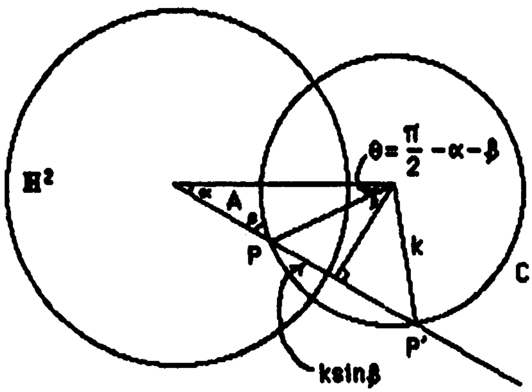


Figure 1.8

Further note: $\frac{dA}{d\alpha} = \frac{2r^2}{1-r^2}$ .

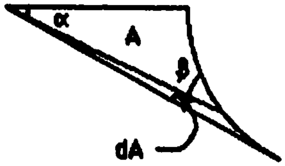


Figure 1.9

Thus A = θ + constant = $\pi/2 - \alpha - \beta$ + constant. When $\alpha=0$ then $\beta=0$, so the constant is zero.

**Corollary 1.5.1:** An n-gon with angles $\alpha_1, \ldots, \alpha_n$ has area