Chapter 2
Hyperbolic Surfaces

For the purposes of this book, a *surface* is a connected, orientable two-dimensional smooth manifold, without boundary unless otherwise specified. Throughout the book we will restrict our attention to surfaces which are *topologically finite*, meaning that the surface is homeomorphic to a compact surface with finitely many points excised. An *end* is an equivalence class of neighborhoods which are contractible to one of these excised points. Topologically finite surfaces are classified up to diffeomorphism by the genus g and the number of ends n. The corresponding value of the Euler characteristic is $\chi = 2 - 2g - n$. An example is shown in Figure 2.1.

Definition 2.1. A *hyperbolic surface* is a smooth surface equipped with a complete Riemannian metric of constant Gaussian curvature -1.

For $\chi \geq 0$ there are only a few special cases of hyperbolic surfaces (the plane and cylinders), but any topological surface with $\chi < 0$ admits a family of hyperbolic metrics. After a brief introduction to plane hyperbolic geometry, the main point of this chapter will be a classification of hyperbolic surfaces. For the later analysis we are particularly interested in the structure of the ends.

2.1 The Hyperbolic Plane

Up to isometry, there is a unique simply connected hyperbolic surface, called the hyperbolic plane, for which there are several standard models. The model we will use most frequently is the upper half-plane,

$$\mathbb{H} := \{z = x + iy \in \mathbb{C} : \ y > 0\}, \qquad ds^2 = \frac{dx^2 + dy^2}{y^2}. \tag{2.1}$$

D. Borthwick, *Spectral Theory of Infinite-Area Hyperbolic Surfaces*, Progress in Mathematics 318, DOI 10.1007/978-3-319-33877-4_2

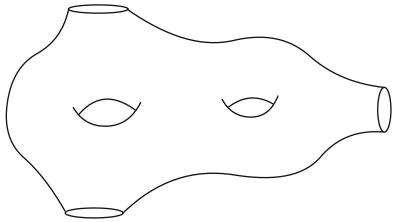

Fig. 2.1 A surface of genus two with three ends, for which $\chi = -5$.

The other standard alternative is the unit disk model (or Poincaré disk),

$$\mathbb{B} := \left\{z \in \mathbb{C} : \ |z| < 1\right\}, \qquad ds^2 = 4\,\frac{dx^2 + dy^2}{(1 - |z|^2)^2}. \tag{2.2}$$

Most calculations are simpler in $\mathbb{H}$, but $\mathbb{B}$ has the advantage that the boundary is treated uniformly.

In either model, the *Möbius transformations* provide a natural set of orientation-preserving maps. Given the matrix,

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{2.3}$$

the corresponding Möbius transformation is

$$z \mapsto Tz := \frac{az + b}{cz + d}.$$

Note that T is invertible as a map if and only if $\det T \neq 0$ as a matrix. And rescaling $T \to \lambda T$ does not change the action. Hence Möbius transformations are naturally identified with the matrix group,

$$\mathrm{PSL}(2, \mathbb{C}) := \mathrm{SL}(2, \mathbb{C})/\{\pm I\}.$$

A map $T \in \mathrm{PSL}(2, \mathbb{C})$ will preserve $\mathbb{H}$ if and only if its coefficients are real, so the group of Möbius automorphisms of $\mathbb{H}$ is $\mathrm{PSL}(2, \mathbb{R})$.

Proposition 2.2. *The group of orientation-preserving isometries of $\mathbb{H}$ is the group $PSL(2, \mathbb{R})$ of Möbius transformations preserving the upper half-plane.*

Proof. Because the hyperbolic metric is conformally related to the Euclidean metric, an isometry $\mathbb{H} \to \mathbb{H}$ preserves Euclidean angles in particular and so must be a conformal automorphism of the upper half-plane. The Schwarz lemma implies that the only such automorphisms are Möbius transformations. Thus isometries must be Möbius.

To see the converse, note that in complex coordinates the hyperbolic metric can be written

$$ds^2 = \frac{|dz|^2}{(\operatorname{Im} z)^2}.$$

Suppose that $T \in \mathrm{PSL}(2, \mathbb{R})$ is represented as in (2.3), with $\det T = 1$. We simply compute,

$$T'(z) = \frac{1}{(cz+d)^2}, \qquad \operatorname{Im}(Tz) = \frac{\operatorname{Im} z}{|cz+d|^2}, \tag{2.4}$$

where T' denotes the complex derivative. (In the notation we distinguish between the action $z \to Tz$ and the function $T'(z)$.) Using these to compute the pullback of the metric gives

$$T^*(ds^2) = \frac{|T'(z)\, dz|^2}{(\operatorname{Im} Tz)^2} = ds^2,$$

which shows that T is an isometry. □

Any Möbius transformation from the upper half-plane onto the unit disk, for example

$$z \mapsto \frac{z-i}{z+i}, \tag{2.5}$$

will give an isometry $\mathbb{H} \to \mathbb{B}$. From this we can immediately deduce that the (orientation-preserving) isometry group of $\mathbb{B}$ is the group of Möbius transformations preserving the unit disk. This is identified with the matrix group $\mathrm{PSU}(1, 1)$.

We will make frequent use of the topology of the unit sphere metric on the Riemann sphere $\mathbb{C} \cup \{\infty\}$. For $z \in \mathbb{C}$, $w \in \mathbb{C} \cup \{\infty\}$, the unit sphere distance function is given by

$$d_\infty(z, w) := \begin{cases} \dfrac{2|z-w|}{\sqrt{(1+|z|^2)(1+|w|^2)}} & w \in \mathbb{C}, \\ \dfrac{2}{\sqrt{1+|z|^2}} & w = \infty. \end{cases}$$

For example, we define the boundary of $\mathbb{H}$ with respect to this topology, as the one-point compactification of the real line,

$$\partial\mathbb{H} := \mathbb{R} \cup \{\infty\}.$$

For the $\mathbb{B}$ model the Riemann sphere topology is equivalent to the Euclidean topology, and we simply have $\partial\mathbb{B} := S^1$.

When considering Möbius transformations, it is convenient to define a *circle* in $\mathbb{C}$ in the generalized sense of a circle with respect to d_∞. Any Euclidean circle or straight line in $\mathbb{C}$ is a circle in this sense.

The large isometry group makes it easy to determine the geodesics of the hyperbolic plane, which turn out to be circles of a certain type.

Proposition 2.3. *The geodesics of $\mathbb{H}$ are precisely the arcs of circles intersecting $\partial\mathbb{H}$ orthogonally. Similarly the geodesics of $\mathbb{B}$ are circles intersecting $\partial\mathbb{B}$ orthogonally.*

Proof. First, we claim that the positive y-axis is a geodesic. Let $\eta : [t_1, t_2] \to \mathbb{H}$ be some curve connecting ia to ib, where $a < b$. The hyperbolic length of the curve is given by integrating ds along η, so if we write $\eta(t) = x(t) + iy(t)$, then

$$\begin{aligned}\ell(\eta) &= \int_{t_1}^{t_2} \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)}\, dt \\ &\geq \int_{t_1}^{t_2} \frac{|y'(t)|}{y(t)}\, dt \\ &\geq \int_{t_1}^{t_2} (\log y(t))'\, dt \\ &= \log(b/a).\end{aligned}$$

It's clear from this calculation that the minimum is achieved if and only if $y'(t) > 0$ and $x'(t) = 0$ (which implies $x(t) = 0$). Thus the y-axis is a path of shortest distance and hence geodesic.

Now suppose that $\gamma : \mathbb{R} \to \mathbb{H}$ is an arbitrary geodesic. By a Möbius transformation R we can send $\gamma(0)$ to i and rotate $\gamma'(0)$ to i also. By uniqueness of the geodesic with given starting position and velocity, this implies that $R \circ \gamma$ parametrizes the y-axis. The characterization of γ follows easily because $\mathrm{PSL}(2, \mathbb{R})$ preserves circles as well as angles and fixes $\partial\mathbb{H}$.

Conversely, any arc of a generalized circle intersecting $\partial\mathbb{H}$ orthogonally could be mapped to the y-axis by an isometry and is therefore geodesic. The same reasoning applies to $\mathbb{B}$. □

From Proposition 2.3 it follows that there is a unique geodesic arc connecting any two distinct points $z, w \in \mathbb{H} \cup \partial\mathbb{H}$. We will denote this segment by $[z, w]$. When $z, w \in \mathbb{H}$, the hyperbolic distance is defined by

$$d(z, w) := \ell([z, w]).$$

Proposition 2.4. *For $z, z' \in \mathbb{H}$ the hyperbolic distance is given by*

$$\cosh d(z,z') = 1 + \frac{|z-z'|^2}{2yy'} \tag{2.6}$$

Proof. A simple exercise using the formula for $T'(z)$ from (2.4) shows that

$$|Tz - Tw|^2 = \left|T'(z)T'(w)(z-w)^2\right|.$$

The second identity in (2.4) then makes it obvious that the right-hand side of (2.6) is invariant under isometries. Since the distance function is invariant by definition, it suffices to check the formula for two general points on the y-axis. The computation in the proof of Proposition 2.3 shows that $d(ia, ib) = \log(b/a)$, which verifies (2.6). □

Elements of $\mathrm{PSL}(2,\mathbb{R})$ are classified by their fixed points. The solutions of the equation $z = Tz$ are roots of the polynomial $cz^2 + (d-a)z - b$, whose discriminant is $(d-a)^2 + 4bc = (\operatorname{tr} T)^2 - 4$. The sign of the discriminant determines how the fixed points are situated within $\mathbb{H}$.

Definition 2.5. A transformation $T \in \mathrm{PSL}(2,\mathbb{R})$ is:

1. *elliptic* if $\operatorname{tr} T < 2$, implying one fixed point within $\mathbb{H}$ (with a matching point in the lower half-plane);
2. *parabolic* if $\operatorname{tr} T = 2$ (and $T \neq I$), with a single degenerate fixed point in $\partial\mathbb{H}$;
3. *hyperbolic* if $\operatorname{tr} T > 2$, yielding two distinct fixed points in $\partial\mathbb{H}$, one attracting and one repelling.

(The double usage of the term "hyperbolic" here is standard but potentially confusing; note that all three types of transformations could reasonably be called "hyperbolic isometries.") Figure 2.2 shows the fixed points and circles preserved by each type of isometry. Since traces are preserved under conjugation, the same classification by traces applies in $\mathrm{PSU}(1,1)$ as well.

Consider an elliptic transformation T, with fixed point $z_0 \in \mathbb{H}$. Let Q be a Möbius transformation mapping $\mathbb{H}$ onto $\mathbb{B}$ such that $Q(z_0) = 0$. Then QTQ^{-1} fixes the origin and so must be a rotation of the form $z \mapsto e^{i\theta}z$, by the Schwarz lemma. Hence the conjugacy class of an elliptic transformation is determined by the rotation angle.

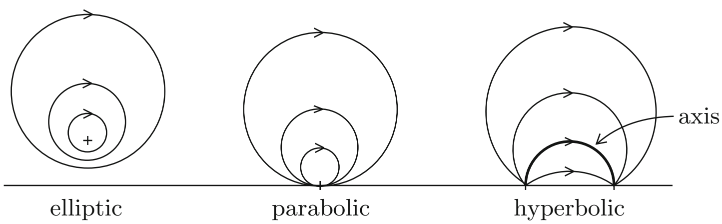

Fig. 2.2 Isometries of $\mathbb{H}$.

A parabolic transformation can be conjugated to a map whose fixed point is ∞. The only such maps are horizontal translations $z \mapsto z + b$ for $b \in \mathbb{R}$. A further conjugation by the dilation $R : z \mapsto |b|^{-1}z$ reduces this translation to $z \mapsto z \pm 1$. Thus, within $\mathrm{PSL}(2, \mathbb{R})$ there are two conjugacy classes of parabolic transformations, corresponding to left or right translations.

The standard form for a hyperbolic transformation is given by conjugating the repelling fixed point to 0 and the attracting fixed point to ∞. The resulting map must be a dilation $z \mapsto e^{\ell}z$ with $\ell > 0$. The conjugacy classes of hyperbolic elements are indexed by the positive number $\ell = \ell(T)$, called the *displacement length* of T.

There is a unique geodesic $\alpha(T)$ called the *axis* connecting the fixed points of a hyperbolic transformation T, as shown in Figure 2.2. By conjugating T to the standard dilation form as above, we see immediately that the displacement length $\ell(T)$ is the distance by which points on $\alpha(T)$ are translated. Since conjugation preserves traces, this implies

$$\mathrm{tr}(T) = 2\cosh(\ell(T)/2).$$

By applying (2.6) to give a simple expression for $\cosh d(z, e^{\ell}z)$, we can easily see that the displacement length satisfies

$$\ell(T) = \min_{z\in\mathbb{H}} d(z, Tz), \tag{2.7}$$

with the minimum achieved if and only if z lies on $\alpha(T)$.

Other geometric features of $\mathbb{H}$ which will be important to us are the area form,

$$dg(z) = \frac{dx\,dy}{y^2}, \tag{2.8}$$

and the formula for the Laplacian. The (positive) Laplacian on a Riemannian manifold is defined globally by $\Delta = -\mathrm{div}\,\mathrm{grad}$. In local coordinates x^i, with the metric given by $ds^2 = g_{ij}dx^i dx^j$, this translates to

$$\Delta = -\frac{1}{\sqrt{\det g}}\partial_i\Big(g^{ij}\sqrt{\det g}\,\partial_j\Big),$$

where g^{ij} denotes the components of the inverse matrix to g_{ij}. For the hyperbolic metric on $\mathbb{H}$, the resulting operator is

$$\Delta = -y^2(\partial_x^2 + \partial_y^2).$$

In addition to the $\mathbb{H}$ and $\mathbb{B}$ models, we will make frequent use of *geodesic normal coordinates* for hyperbolic metrics. These are coordinates (r, t) for which the r-coordinate curves are unit-speed geodesics and the t-coordinate curves are orthogonal to them. This implies a metric of the form

$$ds^2 = dr^2 + \varphi^2 dt^2, \tag{2.9}$$

for some function $\varphi(r,t)$. In any such coordinate system, the Gaussian curvature is given by the simple formula

$$K = -\frac{\partial_r^2 \varphi}{\varphi}.$$

For a hyperbolic metric written in the form (2.9), φ must therefore satisfy

$$\partial_r^2 \varphi = \varphi. \tag{2.10}$$

Geodesic *polar* coordinates $(r,\theta) \in \mathbb{R}_+ \times S^1$ are defined so as to be asymptotic to Euclidean polar coordinates as $r \to 0$. This means that if we write the metric in the form $dr^2 + \varphi^2\, d\theta^2$, then $\varphi \sim r$ as $r \to 0$. In the hyperbolic case, (2.10) then implies $\varphi(r,\theta) = \sinh r$. Hence the geodesic polar form of the metric of a hyperbolic surface is

$$ds^2 = dr^2 + \sinh^2 r\, d\theta^2. \tag{2.11}$$

Two other obvious solutions of (2.10) will be important for us as well: $\varphi = \cosh r$ and $\varphi = e^{-r}$ are the model metrics for funnel and cusp ends, respectively.

We let $B(w;r)$ denote an open neighborhood with respect to the hyperbolic metric: for $w \in \mathbb{H}$ and $r > 0$,

$$B(w;r) := \{z : \ d(z,w) < r\}.$$

In geodesic polar coordinates $dg = \sinh r\, dr\, d\theta$, so that

$$\operatorname{area}(B(w;r)) = 2\pi \int_0^r \sinh r\, dr = 2\pi(\cosh r - 1). \tag{2.12}$$

2.2 Fuchsian Groups

Given the large isometry group of $\mathbb{H}$, a natural way to obtain a hyperbolic surface is as a quotient $\Gamma\backslash\mathbb{H}$, for some subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$. Points in the quotient correspond to orbits of Γ, and there is a natural projection

$$\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}$$

given by $\pi(z) = \Gamma z$. For the quotient to be well defined as a metric space, the action needs to be *properly discontinuous*, which means that the orbits are locally finite (any compact subset of $\mathbb{H}$ contains only finitely many orbit points).

Conveniently, we can characterize the groups which act properly discontinuously on $\mathbb{H}$ by their topology as subsets of $\mathrm{PSL}(2,\mathbb{R})$. On $\mathrm{PSL}(2,\mathbb{R})$ we use the standard matrix topology defined by the norm $\|T\| := (\operatorname{tr} T^*T)^{1/2}$.

Definition 2.6. A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$.

One easy way to obtain examples of Fuchsian groups is to choose an even number of Euclidean disks centered on the real axis, with mutually disjoint closures. Divide the disks up into pairs, and for each pair choose a hyperbolic transformation mapping the exterior of one disk to the interior of the other. These transformations generate a particular kind of Fuchsian group called a *Schottky group*. We will study this class in more detail in §15.1.

Proposition 2.7. *A subgroup $\Gamma \subset PSL(2,\mathbb{R})$ acts properly discontinuously on $\mathbb{H}$ if and only if it is Fuchsian.*

Proof. If a subgroup Γ is Fuchsian, then it is easy to see that any orbit Γz is discrete. For any compact $K \subset \mathbb{H}$, the set $\Gamma z \cap K$ is both discrete and compact and therefore finite. Hence Γ acts properly discontinuously.

On the other hand, assume that Γ acts properly discontinuously. We claim that there are points in $\mathbb{H}$ not fixed by any element of Γ except I. Indeed, if $Tw = w$, then for any $z \in \mathbb{H}$ we have

$$d(Tz,z) \leq d(Tz,Tw) + d(Tw,z) = 2d(z,w),$$

by the triangle inequality. Proper discontinuity therefore implies that only finitely many points in any neighborhood of z could be fixed by elements of $\Gamma - \{I\}$.

Hence we can pick a point w not fixed by any element of Γ except I. If Γ is not discrete, then there exists a sequence $\{T_k\} \subset \Gamma$ of distinct elements such that $T_k \to I$. By the choice of w, the sequence $\{T_k w\}$ contains only distinct points, and $T_k w \to w$ contradicts the proper discontinuity of the action. □

Our requirement that the quotient space to be smooth corresponds to the condition that Γ act without fixed points. Since only elliptic transformations fix points within $\mathbb{H}$, this is equivalent to the absence of elliptic elements in Γ. If Γ had elliptic elements, then the quotient would be an *orbifold*, with conical singularities corresponding to the elliptic fixed points. Orbifolds are not intractable from a spectral theory point of view, because one can always pass to a finite cover. We omit this case mainly to avoid excessive notational complexity later on.

Hopf's theorem on the classification of manifolds of constant sectional curvature implies, in the two-dimensional case, that all hyperbolic surfaces are associated with Fuchsian groups.

Theorem 2.8 (Hopf). *For any hyperbolic surface X there is a Fuchsian group Γ with no elliptic elements and a Γ-invariant Riemannian covering map $\pi : \mathbb{H} \to X$ realizing the isometry $X \cong \Gamma\backslash\mathbb{H}$.*

Proof (Sketch). For $p \in X$ the exponential map $\exp_p : T_pX \to X$ defines geodesic polar coordinates, in which the metric takes the form $ds^2 = dr^2 + \sinh^2 r\, d\theta^2$ by the assumption of Gaussian curvature -1. The lack of singularity in the metric for $r \in (0, \infty)$ implies that $\exp_p : T_pX \to X$ is an immersion. With the geodesic polar coordinates we can identify $T_pX \cong \mathbb{H}$, and $\exp_p$ induces a local isometry $\pi : \mathbb{H} \to X$. A local isometry of complete Riemannian manifolds is automatically a covering map. And since X is a smooth surface, the group of covering transformations must be Fuchsian with no elliptic elements. (The details of these arguments involve some differential geometry that will not be needed for the rest of this book; see, e.g., [155] or [223].) □

A *hyperbolic structure* on a surface is defined by an atlas of coordinate patches identified with open subsets of $\mathbb{H}$, with transition maps given by orientation-preserving isometries. Theorem 2.8 shows that any hyperbolic metric is induced by a hyperbolic structure. Of course, since isometries are Möbius transformations, the hyperbolic structure also induces a complex structure.

A *Riemann surface* is a one-dimensional complex manifold, so the fact that a hyperbolic structure on a surface induces a complex structure implies that hyperbolic surfaces are a subcategory of Riemann surfaces. One might expect complex structure to be a more general concept than hyperbolic structure, since analytic functions need not be Möbius. But the uniformization theorem for Riemann surfaces says that a smooth Riemann surface is either the Riemann sphere or a quotient of $\mathbb{C}$ or $\mathbb{H}$ by a discrete group of conformal automorphisms (see, e.g., [82]). The Riemann sphere and flat tori are the only compact examples of Riemann surfaces with $\chi \geq 0$. Every Riemann surface with $\chi < 0$ is a hyperbolic surface, so in some sense most of the Riemann surfaces are hyperbolic.

2.2.1 The Limit Set

A fundamental object associated with a Fuchsian group is the set of accumulation points of orbits of the group action.

Definition 2.9. For a Fuchsian group Γ, the *limit set* $\Lambda(\Gamma) \subset \partial\mathbb{H}$ is the set of limit points (in the Riemann sphere topology) of all orbits Γz for $z \in \mathbb{H}$. The complement of the limit set in $\partial\mathbb{H}$ is the set of *ordinary points*.

To analyze the limit set, we introduce some basic concepts that will help us understand the structure of the orbits. A *fundamental domain* $\mathcal{F} \subset \mathbb{H}$ for a Fuchsian group Γ is a closed region such that

$$\Gamma\mathcal{F} := \bigcup_{T\in\Gamma} T\mathcal{F} = \mathbb{H},$$

and for each $T \in \Gamma - \{I\}$, the interiors of $\mathcal{F}$ and $T\mathcal{F}$ do not intersect. A convenient construction of fundamental domain is given by the *Dirichlet domain* of a point $w \in \mathbb{H}$, defined by

$$\mathcal{D}_w := \big\{z \in \mathbb{H} : \ d(z, w) \le d(z, Tw) \text{ for all } T \in \Gamma\big\}. \tag{2.13}$$

Convexity in $\mathbb{H}$ is interpreted in terms of hyperbolic geodesics, i.e., a set $U \subset \mathbb{H}$ is *convex* if for any $z, w \in U$ the geodesic arc $[z, w]$ is a subset of U.

Lemma 2.10. *If w is not the fixed point of an elliptic element of Γ, then the Dirichlet domain $\mathcal{D}_w$ is a fundamental domain for Γ. Moreover, $\mathcal{D}_w$ is convex and bounded by a union of geodesics.*

Proof. Fix such a w with domain $\mathcal{D}_w$. Given $z_0 \in \mathbb{H}$, we can minimize $d(z, w)$ for $z \in \Gamma z_0$ by the discreteness of the orbit. This gives at least one $z \in \Gamma z_0 \cap \mathcal{D}_w$, implying that $z_0 \in \Gamma\mathcal{D}_w$. Hence $\Gamma\mathcal{D}_w = \mathbb{H}$.

Suppose now that both $z \in \mathcal{D}_w$ and $Rz \in \mathcal{D}_w$ for $R \in \Gamma - \{I\}$. Then $z \in \mathcal{D}_w$ implies

$$d(z, w) \le d(z, Rw),$$

and $Rz \in \mathcal{D}_w$ implies

$$d(Rz, w) \le d(Rz, Rw) = d(z, w).$$

Hence $d(z, w) = d(z, Rw)$, so z lies on the boundary of $\mathcal{D}_w$. This shows that the interiors of $\mathcal{D}_w$ and $R\mathcal{D}_w$ do not intersect, and thus $\mathcal{D}_w$ is a fundamental domain.

Note that $\mathcal{D}_w$ is an intersection of closed half-planes of the form,

$$H_w(T) := \{z \in \mathbb{H} : \ d(z, w) \le d(z, Tw)\},$$

for $T \in \Gamma$. Thus to prove the second statement, it suffices to check that such half-planes have geodesic boundary. By conjugation we can assume $w = i$ and $T : z \mapsto \lambda^2 z$. Then $z \in \partial H_w(T)$ is characterized by $d(z, i) = d(z, i\lambda^2)$. By (2.6) this can easily be reduced to $|z| = \lambda$, which defines a geodesic. □

The action of Γ on a fundamental domain gives a tessellation of $\mathbb{H}$. An example is shown in Figure 2.3. The corresponding quotient surface is a *funneled torus* of genus one with a single end.

Lemma 2.11. *The tessellation $\{T\mathcal{D}_w : \ T \in \Gamma\}$ is locally finite, meaning that any compact region of $\mathbb{H}$ meets only finitely many copies of $\mathcal{D}_w$, and contains only finitely many vertices and sides of any particular copy.*

Proof. Suppose that $\overline{B(w; r)}$ contained infinitely many points of the form $z_j = T_j(w_j)$ for $w_j \in \mathcal{D}_w$. Then

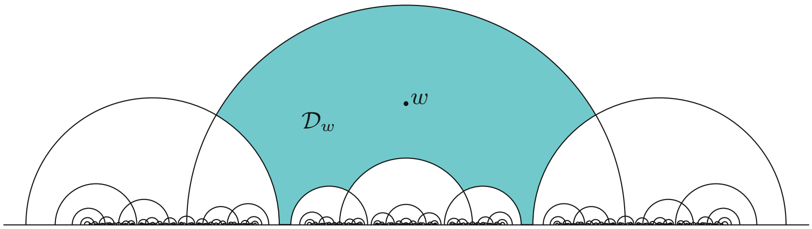

Fig. 2.3 A Dirichlet tessellation of $\mathbb{H}$.

$$\begin{aligned} d(w, T_j w) &\le d(w, z_j) + d(z_j, T_j w) \\ &= d(w, z_j) + d(w_j, w) \\ &\le 2r. \end{aligned}$$

Thus $\overline{B(w; 2r)}$ would contain infinitely many images of w, contradicting the properly discontinuous action of Γ. □

Lemma 2.12. *If $w \in \mathbb{H}$ is not an elliptic fixed point of Γ, then $\Lambda(\Gamma)$ is the set of limit points of the single orbit Γw. It follows immediately that $\Lambda(\Gamma)$ is closed and invariant under Γ.*

Proof. Let $\mathcal{D}_w$ be the Dirichlet domain centered at w. Suppose $q \in \Lambda(\Gamma)$. Then there is a $z \in \mathcal{D}_w$ and a sequence $\{T_j\} \subset \Gamma$ such that $T_j z \to q$. Applying the triangle inequality for the Riemann sphere metric d_∞ gives

$$d_\infty(T_j w, q) \le d_\infty(T_j w, T_j z) + d_\infty(T_j z, q).$$

We claim that the first term on the right converges to zero as $j \to \infty$. The second term does so by assumption, so this would imply $T_j w \to q$, establishing that q is a limit point of Γw.

To prove the claim, suppose that $d_\infty(T_j w, T_j z)$ doesn't converge to zero. Because $\mathbb{H} \cup \partial\mathbb{H}$ is compact in the topology of d_∞, by passing to a subsequence we can assume that $T_j w \to p \in \partial\mathbb{H}$ and $T_j z \to p' \in \partial\mathbb{H}$, where $p \neq p'$ by assumption. Then the geodesic arcs $[T_j w, T_j z]$ accumulate on $[p, p']$, contradicting Lemma 2.11. □

2.2.2 *Classification of Fuchsian Groups*

The standard classification of Fuchsian groups is based on the following characterization of the limit set.

Theorem 2.13 (Poincaré, Fricke-Klein). *The possibilities for the limit set of a Fuchsian group Γ are:*

1. *$\Lambda(\Gamma)$ contains 0, 1, or 2 points.*
2. *$\Lambda(\Gamma)$ is a perfect nowhere dense subset of $\partial\mathbb{H}$.*
3. *$\Lambda(\Gamma) = \partial\mathbb{H}$.*

Proof. Assume that $\Lambda(\Gamma)$ contains more than two points. Our first claim is that then Γ must contain non-elliptic elements. If Γ was purely elliptic, then a straightforward exercise shows that all elements of Γ have the same fixed point. (The product of elliptic transformations with different fixed points is hyperbolic; see, e.g., [142, Thm. 2.4.1]). A group for which all elements fix a single point must be finite cyclic by discreteness, so $\Lambda(\Gamma)$ would be empty in this case.

Now suppose Γ contains a parabolic element. By conjugation we can assume this parabolic element is $T : z \mapsto z + 1$ and therefore $\infty \in \Lambda(\Gamma)$. If every element of Γ fixed ∞ then this would imply that Γ was parabolic cyclic by discreteness, which would also mean that $\Lambda(\Gamma) = \{\infty\}$. Hence, under the assumption that $\Lambda(\Gamma)$ has more than two points, Γ must contain in addition to T some transformation that does not fix infinity, say

$$S = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

with $c \neq 0$. Since

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

a simple computation shows that

$$\operatorname{tr}(T^k S) = a + kc + d.$$

For k sufficiently large, this implies that $|\operatorname{tr} T^k S| > 2$, so that $T^k S$ must be hyperbolic. Hence Γ contains hyperbolic elements.

At this stage, under the assumption that $\Lambda(\Gamma)$ contains at least three points, we have shown that Γ contains at least one hyperbolic element. Our next claim is that $\Lambda(\Gamma)$ is perfect (every point is a limit point). An arbitrary point in $\Lambda(\Gamma)$ can be moved to 0 by conjugation of the group. So it suffices to assume $0 \in \Lambda(\Gamma)$ and prove that this is a limit point of $\Lambda(\Gamma)$. This is easy if 0 is a hyperbolic fixed point. In this case Γ contains a dilation $T : z \mapsto e^{-\lambda} z$ for some $\lambda > 0$. Since $\Lambda(\Gamma)$ contains at least three points, there must exist some $q \in \Lambda(\Gamma)$, not equal to 0 or ∞, Then $T^k q \to 0$, showing that 0 is a limit point.

Suppose that $0 \in \Lambda(\Gamma)$ is not a hyperbolic fixed point. We know that Γ contains some hyperbolic element T with fixed points p_1, p_2. Choose some point w on the axis $\alpha(T)$. By Lemma 2.12, since $0 \in \Lambda(\Gamma)$ we can assume that $R_j w \to 0$ for some sequence $\{R_j\} \subset \Gamma$. Given $\varepsilon > 0$, we can insist that $|R_j w| < \varepsilon$ for all j, by passing to a subsequence if needed. Then because $R_j w$ lies on the half-circle $R_j \alpha(T)$, at least one of the endpoints $R_j p_1$ or $R_j p_2$ must lie in the interval $(-\varepsilon, \varepsilon)$. Since these endpoints are the fixed points of $R_jTR_j^{-1} \in \Gamma$, this shows that 0 is a limit point of hyperbolic fixed points. In particular, 0 is a limit point of $\Lambda(\Gamma)$. This completes the argument that $\Lambda(\Gamma)$ is perfect if it contains at least three points.

It remains to show that $\Lambda(\Gamma)$ is either $\partial\mathbb{H}$ or nowhere dense. Assume that $\Lambda(\Gamma) \neq \partial\mathbb{H}$. Then we have at least one ordinary point $a \in \partial\mathbb{H} - \Lambda(\Gamma)$. Given $q \in \Lambda(\Gamma)$ and $\varepsilon > 0$, we need to show that there is an ordinary point within ε of q (assuming $q \neq \infty$ without loss of generality). By the arguments above we can find a hyperbolic fixed point p within $\varepsilon/2$ of q. Let $T \in \Gamma$ have p as an attractive hyperbolic fixed point. Then $T^k a$ converges to p as $k \to \infty$. Choosing k so that $|T^k a - p| < \varepsilon/2$ then implies $|T^k a - q| < \varepsilon$. Note that $T^k a$ is ordinary since $\Lambda(\Gamma)$ is Γ-invariant. This shows that there is an ordinary point within every neighborhood of any point of $\Lambda(\Gamma)$. Therefore, if $\Lambda(\Gamma) \neq \partial\mathbb{H}$ and it contains at least three points, $\Lambda(\Gamma)$ is nowhere dense in $\partial\mathbb{H}$. □

With Theorem 2.13 in mind, we introduce some further terminology:

Definition 2.14. A Fuchsian group Γ is said to be:

1. *elementary* if $\Lambda(\Gamma)$ is finite;
2. *of the first kind* if $\Lambda(\Gamma) = \partial\mathbb{H}$;
3. *of the second kind* if $\Lambda(\Gamma)$ is perfect and nowhere dense.

An alternate definition for elementary group is the condition that Γ has a finite orbit in $\mathbb{H} \cup \partial\mathbb{H}$. This sounds more general than the definition above, but turns out to be equivalent. Cyclic Fuchsian groups are obviously elementary, with $\Lambda(\Gamma)$ consisting of 0, 1, and 2 points in the elliptic, parabolic, and hyperbolic cases, respectively. The only other elementary possibility is a group conjugate to the group generated by $z \mapsto \lambda z$ and $z \mapsto -1/z$, for which $\Lambda(\Gamma)$ has 2 points also (see e.g. [142, Thm. 2.4.3]). Since we assume smoothness, the elementary hyperbolic surfaces consist only of $\mathbb{H}$ and its quotients by hyperbolic or parabolic cyclic groups.

If the quotient $\Gamma\backslash\mathbb{H}$ has finite area, then Γ is called *cofinite*. Fuchsian groups of the first kind are precisely the cofinite groups (see e.g. [142, §4.5]). A cofinite Fuchsian group is called *cocompact* if the quotient $\Gamma\backslash\mathbb{H}$ is compact. For most of this book we are concerned with surfaces of infinite area, so our attention will be focused on Fuchsian groups of the second kind.

2.3 Geometrically Finite Groups

We turn next to the question of what conditions are imposed on the group Γ by the assumption of topological finiteness of the quotient $\Gamma\backslash\mathbb{H}$. The answer can be given in terms of a nice geometric condition.

Definition 2.15. A Fuchsian group (or corresponding hyperbolic surface) is said to be *geometrically finite* if there exists a fundamental domain which is a finite-sided convex polygon.

There is also an algebraic finiteness condition—we say Γ is *finitely generated* if there exists a finite list of transformations that generate the group.

Theorem 2.16 (Geometric Finiteness). *For a Fuchsian group Γ the following are equivalent:*

1. *$\Gamma\backslash\mathbb{H}$ is topologically finite (i.e., finite Euler characteristic).*
2. *Γ is finitely generated.*
3. *Γ is geometrically finite.*

A related result which we won't prove here is Siegel's theorem, which says that all cofinite Fuchsian groups are geometrically finite (see e.g. [142, Thm. 4.1.1]).

For the proof of Theorem 2.16 we need to establish some connections between the structure of the group and the geometry of the Dirichlet domain. In Lemma 2.10, we saw that the boundary of $\mathcal{D}_w$ is a union of geodesics. Since $\mathcal{D}_w$ is convex, each geodesic meets $\mathcal{D}_w$ either in a point or in a geodesic segment. The segments in the boundary are called *sides* and must take the form

$$\begin{aligned} \sigma_w(R) &:= \left\{z \in \partial\mathcal{D}_w : \ d(z,w) = d(z,Rw)\right\} \\ &= \mathcal{D}_w \cap R\mathcal{D}_w \end{aligned} \tag{2.14}$$

for some $R \in \Gamma$. By Lemma 2.11, the vertices of $\mathcal{D}_w$ are isolated.

Two sides of a Dirichlet domain $\mathcal{D}_w$ are called *congruent* if they are related by an element of Γ. Notice that if a side is given by $\sigma_w(R) \neq \emptyset$, then

$$R^{-1}\sigma_w(R) = (R^{-1}\mathcal{D}_w) \cap \mathcal{D}_w = \sigma_w(R^{-1}).$$

Since this is nonempty as well, $\sigma_w(R^{-1})$ must also be a side of $\mathcal{D}_w$, congruent to the original. This is illustrated in Figure 2.4. It follows that the sides of $\mathcal{D}_w$ come in congruent pairs of the form $\sigma_w(R), \sigma_w(R^{-1})$.

Lemma 2.17. *The side-pairing congruences of a Dirichlet domain generate the group Γ.*

Proof. Suppose $\mathcal{D}_w$ is a Dirichlet domain. Let $\Gamma_s \subset \Gamma$ be the subgroup generated by side-pairing congruences. Clearly, if we define

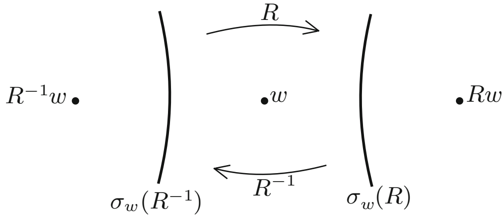

Fig. 2.4 Side-pairing congruences.

$$A := \bigcup_{T \in \Gamma_s} T\mathcal{D}_w, \qquad B := \bigcup_{T \in \Gamma - \Gamma_s} T\mathcal{D}_w,$$

then $A \cup B = \mathbb{H}$. Furthermore, A and B are closed, since $\mathcal{D}_w$ is closed and any compact region contains only finitely many copies of $\mathcal{D}_w$ by Lemma 2.11.

Thus if we can show that A and B are disjoint, the connectedness of $\mathbb{H}$ would imply that $B = \emptyset$ (since A is clearly not empty). To prove disjointness, suppose A intersects B. This could happen only at a side or vertex of the Dirichlet tessellation. To rule out a shared side, suppose $T\mathcal{D}_w$ is adjacent to $R\mathcal{D}_w$ and $T\mathcal{D}_w \subset A$, with $T \in \Gamma_s$. This implies $R^{-1}T\mathcal{D}_w$ is adjacent to $\mathcal{D}_w$, hence $R^{-1}T$ is a generator of Γ_s, hence $R \in \Gamma_s$. Thus $R\mathcal{D}_w \in A$, which shows that A cannot meet B along a side.

Ruling out a shared vertex is similar. Suppose that $T\mathcal{D}_w$ shares a vertex with $R\mathcal{D}_w$ and $T \in \Gamma_s$. There can be only finitely many sides of the Dirichlet tessellation sharing the same vertex. Therefore $T\mathcal{D}_w$ is connected to $R\mathcal{D}_w$ by a chain of side-sharing faces of the tessellation. We saw above that faces in A only share sides with other faces in A, so we find $R \in \Gamma_s$. Hence A does not intersect B at a vertex, and this finishes the proof that $B = \emptyset$. □

Proof of Theorem 2.16. Lemma 2.17 shows in particular that 3 ⇒ 2. The implication 1 ⇒ 2 holds because the fundamental group of a finitely punctured compact surface is finitely generated and $\pi_1(X) \cong \Gamma$. And 3 ⇒ 1 is also clear, because a surface assembled out of a finite-sided polygon by gluing the sides together in pairs must have finite Euler characteristic.

The hard part to prove is 2 ⇒ 3. We will follow the approaches from Beardon [20] and Katok [142]. Assume that Γ is finitely generated, and choose a Dirichlet domain $\mathcal{D}_w$. By Lemma 2.17 we know that the side pairing transformations of $\mathcal{D}_w$ generate Γ. By assumption, Γ can be generated by finitely many of the side pairing transformations, say $T_1, \ldots, T_k$.

The strategy is to choose a disk $B(w; r)$ that includes arcs of positive length of the $2k$ sides of $\mathcal{D}_w$ paired by the T_j's. By local finiteness of the sides and vertices (Lemma 2.11), we can choose r so that the boundary circle $\partial B(w; r)$ does not intersect vertices of $\mathcal{D}_w$ and is not tangent to any side. Our goal will be to show that $\mathcal{D}_w - B(w; r)$ is the union of finitely many connected components, each of which meets only finitely many sides of $\mathcal{D}_w$. Thus $\mathcal{D}_w$ has only finitely many sides outside of $B(w; r)$. Since only finitely many sides of $\mathcal{D}_w$ meet the interior of $B(w; r)$, by Lemma 2.11, this will imply that the total number of sides of $\mathcal{D}_w$ is finite.

First we show that $\Gamma B(w; r)$ is connected. Clearly $B(w; r)$ overlaps $T_j B(w; r)$ for $j = 1, \ldots, k$, since T_j pairs sides of $\mathcal{D}_w$ and $B(w; r)$ was chosen to include arcs of such sides. Then we can argue $T_j B(w; r)$ overlaps $T_j T_i B(w; r)$, by translation, and so on. Since the T_j's generate Γ, by continuing this process we see that $\Gamma B(w; r)$ is connected.

Let $\eta_1, \ldots, \eta_m$ be the arcs of $\partial B(w; r) \cap \mathcal{D}_w$. If z is an endpoint of η_j then it lies in some side of $\mathcal{D}_w$. Therefore there is a side-pairing $R \in \Gamma$ such that $Rz \in \mathcal{D}_w$ also. By definition of the Dirichlet domain, $z \in \mathcal{D}_w$ implies

$$r = d(z, w) \leq d(z, R^{-1}w) = d(Rz, w).$$

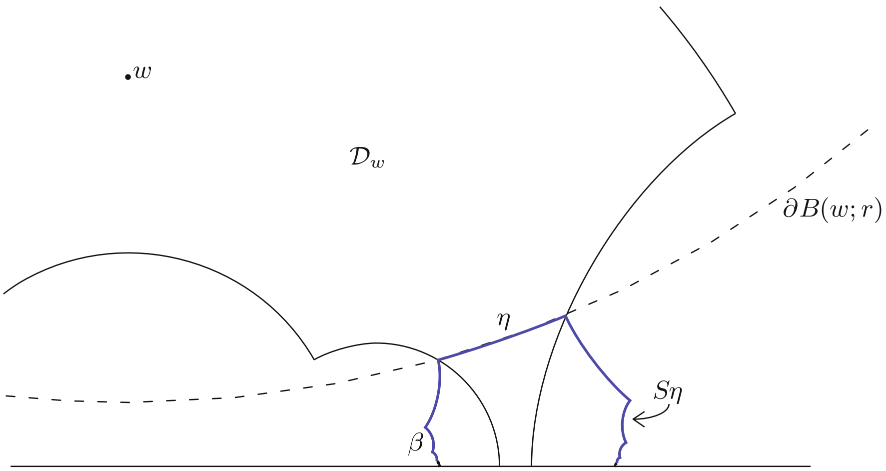

Fig. 2.5 Construction of β.

On the other hand $Rz \in \mathcal{D}_w$ implies

$$d(Rz, w) \leq d(Rz, Rw) = r.$$

This shows that $d(Rz, w) = r$, so that $Rz \in \partial B(w; r)$ also. Therefore Rz must be an endpoint of some side η_i (possibly the other endpoint of the same η_j).

It suffices to focus on a single arc $\eta = \eta_j$. Given an endpoint of η we can find another endpoint of some η_i congruent to it, and translate that η_i by some element of Γ to add an arc to our original η. Iterating this process in both directions results in a uniquely defined continuous curve β which is a union of arcs each congruent to some η_i. Since there are only finitely many η_i's, β must eventually include two arcs congruent to each other, hence there is some nontrivial $S \in \Gamma$ which preserves β. This setup is illustrated in Figure 2.5.

Let E be the component of $\mathcal{D}_w - B(w; r)$ meeting η. Our goal is to show that E meets $\mathcal{D}_w$ in only finitely many sides. For this purpose it suffices to show that E does not meet $\Lambda(\Gamma)$. Observe that the curve β divides $\mathbb{H}$ into two components, one of which contains E and the other w. Since $\Gamma B(w; r)$ is connected, β separates all of $\Gamma B(w; r)$ from E. The limit points of Γ are all limit points of $\Gamma w \subset \Gamma B(w; r)$. So the only limit points we need to worry about being close to E are the endpoints of β.

Suppose first that S is hyperbolic. Then β must run between its two fixed points. By the definition of a Dirichlet domain, $\mathcal{D}_w$ is contained in the closed half-planes $\{z : \ d(z, w) \leq d(z, Sw)\}$ and $\{z : \ d(z, w) \leq d(z, S^{-1}w)\}$. In the notation (2.14), these half-planes are bounded by arcs $\sigma_w(S)$ and $\sigma_w(S^{-1})$. These two arcs don't intersect (obvious if one conjugates S to a dilation). Since S maps $\sigma_w(S^{-1})$ to $\sigma_w(S)$, neither contains a fixed point of S. Therefore E is separated from the limit points of Γ (in the d_∞ metric). If E met infinitely many sides of $\mathcal{D}_w$ then there would have to be a limit point of Γ on its boundary. Since this doesn't happen, E meets only finitely many sides of $\mathcal{D}_w$.

Now consider the case when S is parabolic. If p denotes the fixed point of S, then the curve β is a closed loop from p to itself. If w lies inside β, then $\Gamma B(w;r)$ does also, and this would imply that p was the only point in $\Lambda(\Gamma)$. Then Γ would be parabolic cyclic and obviously geometrically finite. So assume that w lies on the outside of β, in which case E must be contained inside the loop. If E doesn't meet the boundary $\partial\mathbb{H}$ then it is separated from the limit points of Γ and then we argue as above that E meets only finitely many sides of $\mathcal{D}_w$.

So let us suppose that E lies inside β and meets $\partial\mathbb{H}$ at p. We want to control the shape of $\mathcal{D}_w$ nearby. As in the hyperbolic case above, $\mathcal{D}_w$ lies between the arcs $\sigma_w(S)$ and $\sigma_w(S^{-1})$. These arcs do not intersect in $\mathbb{H}$ but are tangent to each other at p. This implies that a small neighborhood of p meets exactly two sides of $\mathcal{D}_w$ Since otherwise E is bounded away from $\Lambda(\Gamma)$, E meets only finitely many sides of $\mathcal{D}_w$.

Finally, suppose that S is elliptic, in which case β is a closed loop. If E is contained in the interior, then it is obviously bounded away from the limit set. But if w lies in the interior, then $\Lambda(\Gamma) = \emptyset$ and the group is cyclic. □

2.4 Classification of Hyperbolic Ends

Geometric finiteness imposes strong restrictions on the ends of a hyperbolic surface. We will show that the only possibilities, beyond the hyperbolic plane itself, are the ends of cylinders, i.e., quotients of $\mathbb{H}$ by hyperbolic and parabolic cyclic groups. We start by examining these model cases.

A hyperbolic transformation $T \in \mathrm{PSL}(2,\mathbb{R})$ generates a cyclic hyperbolic group $\langle T\rangle$. The quotient $C_\ell := \langle T\rangle\backslash\mathbb{H}$ is a *hyperbolic cylinder* of diameter $\ell = \ell(T)$. By conjugation, we can identify the generator T with the map $z \mapsto e^\ell z$, and we define Γ_ℓ to be the corresponding cyclic group. A natural fundamental domain for Γ_ℓ would be the region $\mathcal{F}_\ell := \{1 \le |z| \le e^\ell\}$. The y-axis is the lift of the only simple closed geodesic on C_ℓ, whose length is ℓ.

Definition 2.18. A *funnel* is a closed half of a hyperbolic cylinder, with boundary given by the central geodesic.

Let T be a hyperbolic transformation of $\mathbb{H}$ with displacement length ℓ. If H denotes one of the open half-planes of $\mathbb{H}$ bounded by the axis $\alpha(T)$, then

$$F_\ell := \langle T\rangle\backslash\overline{H},$$

is a funnel of diameter $\ell(T)$. This is illustrated in Figure 2.6, which also shows the Riemannian embedding of a portion the funnel into $\mathbb{R}^3$. From the hyperbolic area form dg given in (2.8), it is clear that

$$\operatorname{area}(F_\ell) = \infty.$$

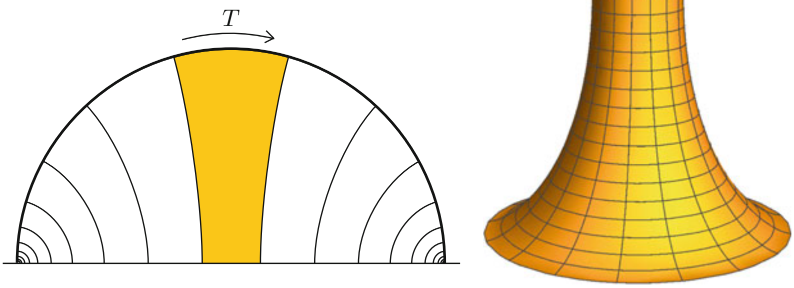

Fig. 2.6 Hyperbolic funnel.

The quotient of $\mathbb{H}$ by a parabolic cyclic group $\langle T\rangle$ will be called a *parabolic cylinder*. We can always conjugate $\langle T\rangle$ to the group Γ_∞ generated by $z \mapsto z+1$, so the parabolic cylinder is unique up to isometry. A natural fundamental domain for Γ_∞ is $\mathcal{F}_\infty := \{0 \le \operatorname{Re} z \le 1\} \subset \mathbb{H}$. A circle lying in $\mathbb{H}$ and tangent to $\partial\mathbb{H}$ is called a *horocycle*. The curves stabilized by a parabolic transformation, as shown in Figure 2.2, are horocycles tangent at the fixed point.

Definition 2.19. A *cusp* is the small end of a parabolic cylinder, with boundary the unique closed horocycle of length 1.

There is no canonical choice of boundary for a cusp, but it is convenient to standardize the definition by fixing the boundary length. To get a cusp from Γ_∞ as defined above, we take the quotient

$$C := \Gamma_\infty \backslash \big\{0 \le \operatorname{Re} z \le 1, \operatorname{Im} z \ge 1\big\}.$$

To obtain the cusp corresponding to a general parabolic generator T with fixed point p, we would first find the unique horocycle σ tangent to $\partial\mathbb{H}$ at p, such that $\langle T\rangle\backslash\sigma$ has length one. If O denotes the interior of σ then $\langle T\rangle\backslash\overline{O}$ is the cusp associated with T. The cusp can be fully embedded into Euclidean $\mathbb{R}^3$, as illustrated in Figure 2.7, where it forms a portion of the classical pseudosphere. Using the fundamental domain for Γ_∞ we compute that

$$\operatorname{area}(\text{cusp}) = \int_1^\infty \int_0^1 \frac{dx\,dy}{y^2} = 1.$$

The large end of the parabolic cylinder furnishes yet another type of hyperbolic cylindrical end. This end in some sense just a special case of the funnel. This type will not play much of a role in our discussion, because it does not occur in any other hyperbolic surface.

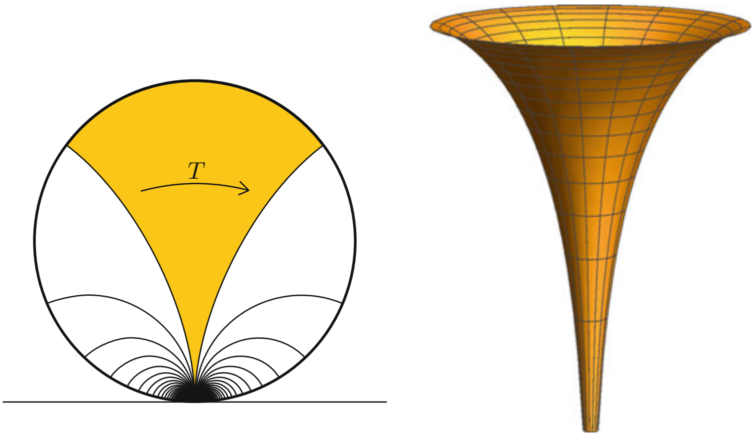

Fig. 2.7 Cusp.

2.4.1 Nielsen Regions

Let us focus on the case of a hyperbolic surface $X = \Gamma\backslash\mathbb{H}$ where Γ is non-elementary. By Theorem 2.13, $\Lambda(\Gamma)$ is either perfect and nowhere dense or equal to $\partial\mathbb{H}$. In the former case, $\partial\mathbb{H} - \Lambda(\Gamma)$ is a countable union of open intervals I_j. For each j suppose that γ_j is the geodesic whose endpoints are the endpoints of I_j, and H_j is the open half-plane bounded by γ_j and I_j. If $\Lambda(\Gamma) = \partial\mathbb{H}$ then we take the convention that $\{H_j\} = \emptyset$.

Definition 2.20. The *Nielsen region* of a Fuchsian group Γ is the set

$$\tilde{N} := \mathbb{H} - \left(\cup H_j\right). \tag{2.15}$$

The quotient

$$N := \Gamma\backslash\tilde{N}$$

is called the *convex core* of X.

Figure 2.8 shows a sample construction of the Nielsen region, pictured in the unit disk model for the sake of clarity; the dotted lines mark the boundary of the fundamental domain. The Nielsen region is also commonly (and equivalently) defined as the convex hull of the limit set $\Lambda(\Gamma)$ in $\mathbb{H}$, meaning the union of geodesic arcs $[p,q]$ for all $p,q \in \Lambda(\Gamma)$. The term "convex core" refers to the fact that N is the smallest closed, nonempty convex subset of X. If Γ is of the first kind, i.e., $\Lambda(\Gamma) = \partial\mathbb{H}$, then $\tilde{N} := \mathbb{H}$ and $N = X$.

In Theorem 2.23 we will show that $X - N$ is a finite collection of funnels. But before we get to that, let us develop a way to isolate the cusps also. The basic idea

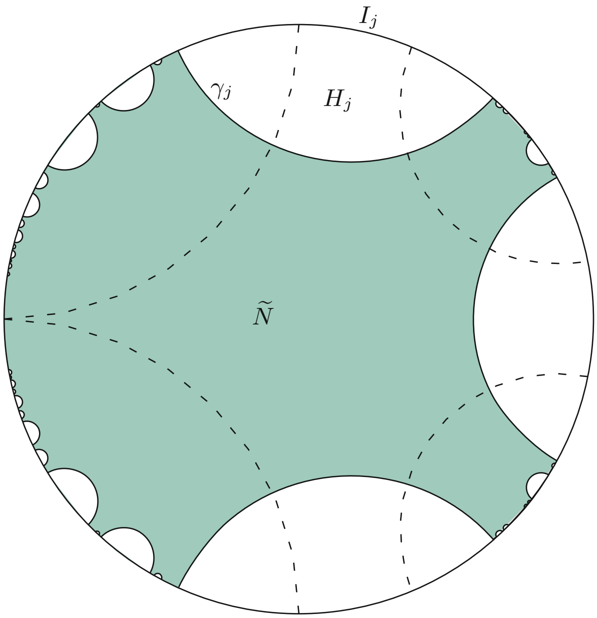

Fig. 2.8 Nielsen region.

is that each parabolic fixed point in $\Lambda(\Gamma)$ should have a cusp fundamental region attached to it. Given a parabolic fixed point $p \in \partial\mathbb{H}$, let Γ_p be the parabolic cyclic subgroup of Γ fixing p. We take σ_p to be the unique horocycle tangent to $\partial\mathbb{H}$ at p such that $\Gamma_p\backslash\sigma_p$ has length 1. Then let O_p be the open region bounded by σ_p, so that $\Gamma_p\backslash\overline{O}_p$ is a cusp fitting our convention of boundary length 1.

Lemma 2.21. *For σ_p and O_p defined as above, the following statements hold:*

1. *If two points of O_p are related by $T \in \Gamma$, then $T \in \Gamma_p$.*
2. *The horocycles σ_p for different parabolic fixed points do not intersect.*
3. *The horocycles σ_p do not intersect the half-planes H_j defining the Nielsen region, so that each $\overline{O}_p \subset N$.*

Proof. Let T be the map $z \mapsto z + 1$. By conjugation, we assume that $p = \infty$ and $\Gamma_p = \langle T\rangle$, so that $O_p = \{\operatorname{Im} z > 1\}$. Let $S \in \Gamma$ be given by

$$S = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

with $ad - bc = 1$. Assume that S does not fix ∞, which implies $c \neq 0$. We claim that in fact $|c| \geq 1$. This will prove the first assertion, because then the inequality

$$\operatorname{Im}(Sz) = \frac{\operatorname{Im} z}{|cz+d|^2} \le \frac{1}{c^2 \operatorname{Im} z}$$

shows that any point inside O_p is mapped to $\{\operatorname{Im} z < 1\}$ by S.

For the proof that $|c| \ge 1$ we follow Kra [145, Lemma II.2.4]. Suppose that $|c| < 1$. Define a recursive sequence of $S_n \in \Gamma$ by setting $S_0 := S$ and

$$S_{n+1} := S_n T S_n^{-1}.$$

If the matrix elements of S_n are denoted a_n, b_n, c_n, d_n, then the recursive condition becomes

$$\begin{pmatrix} a_{n+1} & b_{n+1} \\ c_{n+1} & d_{n+1} \end{pmatrix} = \begin{pmatrix} 1 - a_n c_n & a_n^2 \\ -c_n^2 & 1 + a_n c_n \end{pmatrix}. \tag{2.16}$$

In particular, the assumption $|c| < 1$ implies that

$$c_n = -c^{2^n} \to 0,$$

as $n \to \infty$. Also, from the equation $a_{n+1} = 1 - a_n c_n$, it is easy to prove inductively that a_n is bounded for $|c| < 1$, and hence that $a_n \to 1$. Then by (2.16) we conclude immediately that $b_n \to 1$ and $d_n \to 1$ also. This shows that $S_n T^{-1} \to I$ within Γ, contradicting the discreteness of Γ. We conclude that $|c| \ge 1$, and this proves the first claim.

Still assuming that $p = \infty$, with T and O_p as above, let $q \in \mathbb{R}$ be some other parabolic fixed point of Γ. Then σ_q is a Euclidean circle tangent to $\mathbb{R}$, and the first part of the proof shows that no two points of σ_q could be related by $T : z \mapsto z + 1$. This means in particular that the Euclidean diameter of σ_q must be strictly less than 1, and thus σ_q is too short to intersect O_p.

The proof of the third claim is similar to the second. Suppose H_j is one of the half-planes in question. If ∞ is assumed to be a parabolic fixed point, then H_j cannot include ∞ and so must be a Euclidean half-disk centered on $\mathbb{R}$. The full collection $\cup H_i$ is invariant under Γ by the invariance of $\Lambda(\Gamma)$. The map T clearly does not fix H_j, hence no two points of H_j are related by T. This implies that the Euclidean diameter of H_j is less than 1. Thus the half-planes H_j are contained in the region $\{\operatorname{Im} z \le \frac{1}{2}\}$, whereas $O_p = \{\operatorname{Im} z > 1\}$. □

Using Lemma 2.21, we can now modify the definition of the Nielsen region so as to isolate the cusps as well as the funnels.

Definition 2.22. The *truncated Nielsen region* is

$$\tilde{K} := \tilde{N} - \cup O_p,$$

with the union taken over all parabolic fixed points p of Γ. When Γ is geometrically finite, the corresponding quotient region,

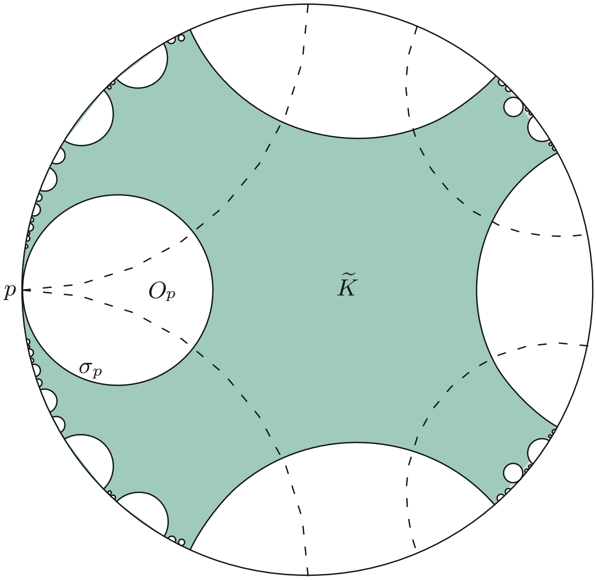

Fig. 2.9 Truncated Nielsen region.

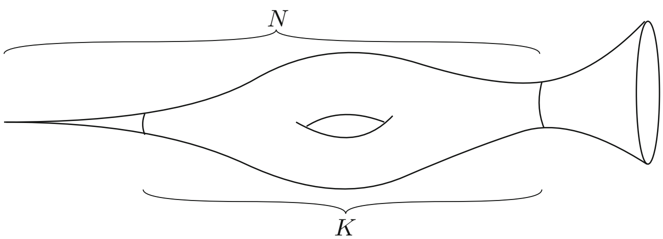

Fig. 2.10 Convex core *N* and compact core *K*.

$$K := \Gamma\backslash\tilde{K},$$

is called the *compact core* of X.

Figure 2.9 shows an example of the truncated Nielsen region, for the same Fuchsian group whose Nielsen region was pictured in Figure 2.8. The distinction between the convex core and compact core is illustrated in Figure 2.10.

Use of the term "compact core" is justified by the following, which is the main result of this section:

Theorem 2.23 (Classification of Hyperbolic Ends). *Let $X = \Gamma\backslash\mathbb{H}$ be a non-elementary geometrically finite hyperbolic surface. Then the region K defined above is compact, and $X - K$ is a finite disjoint union of cusps and funnels.*

Proof. Suppose Γ is a geometrically finite, non-elementary Fuchsian group. Let $\mathcal{D}_w$ be a Dirichlet fundamental domain for Γ, which can intersect $\partial\mathbb{H}$ only in a finite number of intervals or isolated points. We saw in the proof of Theorem 2.16 that $\mathcal{D}_w$ could meet $\Lambda(\Gamma)$ only at parabolic fixed points. At such a point p, two sides of $\mathcal{D}_w$ must meet tangentially. So if O_p is the corresponding horocyclic region O_p from Lemma 2.21, then $\mathcal{D}_w - O_p$ is bounded away from p in the d_∞ metric. If $\mathcal{D}_w$ meets $\partial\mathbb{H}$ in an arc η (possibly just a point) consisting of ordinary points, then η must be included in one of the half-planes H_j used to define $\tilde{N}$ in (2.15). Since the boundary of each H_j meets $\partial\mathbb{H}$ in $\Lambda(\Gamma)$, $\mathcal{D}_w - H_j$ is bounded away from η with respect to d_∞. These arguments show that $\mathcal{D}_w \cap \tilde{K}$ is bounded away from $\partial\mathbb{H}$ in the d_∞ metric, and therefore compact. Hence K is compact also.

We have shown also that the components of $\mathcal{D}_w - \tilde{K}$ are either be contained in either half-planes H_j or in horocyclic regions O_p. What remains to be seen is that the former give rise to finitely many funnels, and the latter to finitely many cusps.

First the funnel case. Let $\tau_1, \ldots, \tau_k$ denote the geodesic segments of the form $\partial H_j \cap \mathcal{D}_w$. Any point in ∂H_j is congruent to a point in $\mathcal{D}_w$, and these points must lie in some τ_i. In other words, ∂H_j is covered by segments each of which is congruent to one of the τ_i's. Since the τ_i's have finite length, ∂H_j must in fact contain multiple segments congruent to some particular τ_i. Therefore there are hyperbolic elements of Γ which relate points of ∂H_j, and because the collection $\cup H_j$ is invariant under Γ, such transformations must then preserve ∂H_j. The subgroup $\Gamma_j \subset \Gamma$ which preserves ∂H_j is thus nontrivial, and by discreteness it must be cyclic. Then we have $\Gamma\backslash H_j = \Gamma_j\backslash H_j$, which is by definition a funnel. Because the set of τ_k's was finite to begin with, we conclude that $X - N$ is a finite disjoint union of funnels.

For any parabolic fixed point p, $\Gamma\backslash O_p$ is a cusp bounded by a horocycle of length 1 by Lemma 2.21. Since $\mathcal{D}_w$ meets $\Lambda(\Gamma)$ at only finitely many points, $N - K$ is a finite disjoint union of cusps. □

The compactness of K is equivalent to geometric finiteness; see, e.g., [84, §15.1]. We will further subdivide the compact core K into a "pants" decomposition in Theorem 2.38.

If the convex core is compact, i.e. $N = K$, then Γ is said to be *convex cocompact*. This is equivalent to the quotient being geometrically finite with no cusps.

2.5 Length Spectrum and Selberg's Zeta Function

The Euler characteristic, genus, and numbers of funnels and cusps are the most basic invariants of a hyperbolic surface. An additional set of natural geometric invariants is provided by the lengths of closed geodesics.

For compact surfaces of negative curvature, each closed geodesic is uniquely associated with a free homotopy class of closed curves, as the representative of minimum length within the class. For a non-compact surface we must be a little careful about this; the horocycle bounding a cusp has no geodesic within its homotopy class. To account for this exception, we say a curve is *cuspidal* if it is freely homotopic to the horocyclic boundary of a cusp, and non-cuspidal otherwise.

Proposition 2.24. *Let η be a homotopically nontrivial curve on a hyperbolic surface X. If η is non-cuspidal, then there is a unique closed geodesic γ which is the shortest closed curve freely homotopic to η.*

Proof. Let $\tilde{\eta}$ be a maximal continuous curve in $\mathbb{H}$ obtained by joining successive lifts of η. There is some $T \in \Gamma$ that preserves $\tilde{\eta}$ and corresponds to moving through one period of η. This T must be hyperbolic, or else η would be cuspidal. The axis $\alpha(T)$ descends to a closed geodesic $\gamma = \langle T\rangle\backslash\alpha(T)$, in the free homotopy class of η. Since $\alpha(T)$ is the unique geodesic in $\mathbb{H}$ fixed by T, there is no other geodesic in the free homotopy class of η. It follows from (2.7) that γ is the shortest curve in this class. □

The proof of Proposition 2.24 reveals an association between closed geodesics and axes of hyperbolic elements of Γ which will turn out to be of great importance to us. We can express this more precisely in the following:

Proposition 2.25. *There is a one-to-one correspondence between the closed, oriented geodesics of a hyperbolic surface $X = \Gamma\backslash\mathbb{H}$ and the conjugacy classes of hyperbolic elements of Γ. The length of the geodesic corresponding to the conjugacy class $[T]$ is the displacement length $\ell(T)$.*

Proof. Suppose $T \in \Gamma$ is hyperbolic. The axis $\alpha(T)$ of T is preserved by T and so projects to a closed geodesic under $\pi : \mathbb{H} \to X$. The length of $\pi(\alpha(T))$ is equal to the displacement length $\ell(T)$. Note that the axis has a natural orientation because T maps points away from one fixed point toward the other. The projected geodesic inherits the orientation. The axis of any other element of the conjugacy class of T will also project to $\pi(\alpha(T))$, since

$$\alpha(RTR^{-1}) = R(\alpha(T)). \tag{2.17}$$

For the converse statement, suppose that γ is a closed oriented geodesic in X, with $\gamma(t) = \gamma(t+\ell)$ for some ℓ. We can construct a complete oriented geodesic arc $\tilde{\gamma}$ in $\mathbb{H}$ by successive lifts of γ. Associated with $\tilde{\gamma}$ is a unique hyperbolic $T \in \mathrm{PSL}(2,\mathbb{R})$ with axis and displacement length given by $\tilde{\gamma}$ and ℓ, respectively. To see that T must be an element of Γ, we observe that since $\tilde{\gamma}(0)$ and $\tilde{\gamma}(\ell)$ project to the same point of X, we must have $\tilde{\gamma}(\ell) = R\tilde{\gamma}(0)$ for some $R \in \Gamma$. In this case $R^{-1}T$ fixes $\tilde{\gamma}(0)$, implying $T = R$ since Γ acts freely. Hence $T \in \Gamma$. To complete the argument, note that (2.17) shows that any other lift of γ must be the axis of a hyperbolic transformation conjugate to T in Γ. □

Given a closed geodesic, we can generate a family of iterates which traverse the same path multiple times. We define a *primitive* closed geodesic to be the root element of such a family, a closed geodesic which is not an iterate of a shorter closed geodesic. Similarly, an element of Γ is called primitive if it is not the power of some other element.

It is trivial to see that an oriented closed geodesic γ is uniquely represented as the iterate of a primitive oriented closed geodesic. Starting from $\gamma(0)$, we simply follow the curve and find the first value of $t > 0$ such that $\gamma(t) = \gamma(0)$ and $\gamma'(t) = \gamma'(0)$. The corresponding result for group elements is:

Lemma 2.26. *Given a Fuchsian group Γ, each element $S \in \Gamma$ can be written uniquely as a power T^k where $T \in \Gamma$ is primitive and $k \geq 1$. The centralizer Z_S of S in Γ is the cyclic group $\langle T \rangle$.*

Proof. Suppose S is hyperbolic. By conjugation, we can assume that it has the standard form $S : z \mapsto e^{\ell} z$ with $\ell > 0$. The commutation relation,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e^{\ell/2} & 0 \\ 0 & e^{-\ell/2} \end{pmatrix} = \begin{pmatrix} e^{\ell/2} & 0 \\ 0 & e^{-\ell/2} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

implies $b = c = 0$, so any element of Z_S is a dilation also. The signed displacement length, given by $\log|R(i)|$ for $R \in Z_S$, therefore defines a homomorphism $Z_S \to \mathbb{R}$. The discreteness of Γ implies that the image must be a lattice $\ell_0\mathbb{Z}$, for some minimum displacement length $\ell_0 > 0$. The unique choice for T is then $z \mapsto e^{\ell_0} z$, and we let $k = \ell/\ell_0$.

The proof for parabolic S is very similar. □

2.5.1 Length Spectrum

The full set of lengths of closed geodesics includes integer multiples of each length corresponding to iterates of the curve. For our purposes it is convenient to restrict our attention to the primitive elements.

Definition 2.27. The (primitive) *length spectrum* of a hyperbolic surface X is the set

$$\mathcal{L}_X = \left\{ \ell(\gamma) : \ \gamma \text{ is a primitive oriented closed geodesic on } X \right\},$$

with lengths repeated according to multiplicity.

Note that the values of ℓ in $\mathcal{L}_X$ come in pairs corresponding to the two possible orientations of each closed geodesic. This might seem redundant, but it proves convenient because of the association with conjugacy classes.

The corresponding *length counting function* is given by

$$\pi_X(t) = \#\{\ell \in \mathcal{L}_X : \ \ell \leq t\}. \tag{2.18}$$

In Chapter 14 we will develop a precise asymptotic formula for $\pi_X(t)$, but for the moment we need only a basic bound.

Proposition 2.28. *For X a geometrically finite hyperbolic surface,*

$$\pi_X(t) = O(e^t).$$

Proof. Let K be the compact core of X as introduced in Theorem 2.23. Closed geodesics on X are contained in the convex core N, though not necessarily in K. It is clear, however, that any closed geodesic must at least pass through K, since a cusp cannot contain a closed geodesic completely. Given a realization $X \cong \Gamma\backslash\mathbb{H}$, the lift of K to $\mathbb{H}$ is the truncated Nielsen region $\tilde{K}$.

For some $w \in \tilde{K}$, let Z be the compact region $\tilde{K} \cap \mathcal{D}_w$. If γ is a primitive closed geodesic on X, then because γ passes through K it can be covered by a geodesic $\tilde{\gamma}$ in $\mathbb{H}$ that passes through some point $q \in Z$. This curve $\tilde{\gamma}$ is the axis of some primitive $T \in \Gamma$. If the length of γ is $\ell = \ell(T)$ and we set a equal to the diameter of Z, then by the triangle inequality,

$$\begin{aligned} d(w, Tw) &\le d(w, q) + d(q, Tq) + d(Tq, Tw) \\ &\le \ell + 2a. \end{aligned} \tag{2.19}$$

This means that for each primitive closed geodesic of length ℓ there is a transformation T which maps w to a point of distance less than $\ell + 2a$ away. We can therefore bound $\pi_X(t)$ by the number of images of w lying within distance $t+2a$ of the point w. Alternatively, we could use the number of images of Z lying within distance $t + 3a$ of the point w, which gives a bound

$$\pi_X(t) \le \frac{\operatorname{area}(B(w; t + 3a))}{\operatorname{area} Z}.$$

The result follows, because $\operatorname{area}(B(w; r)) = O(e^r)$ by (2.12). □

2.5.2 Zeta Function

For a Fuchsian group Γ, the function given by summing $e^{-sd(z,Tw)}$ over $T \in \Gamma$ is called the (absolute) *Poincaré series* for Γ. This sum converges for all $\operatorname{Re} s$ above some threshold value, which we single out in the following:

Definition 2.29. The *exponent of convergence* of a Fuchsian group Γ is

$$\delta := \inf\left\{ s \ge 0 : \sum_{T \in \Gamma} e^{-sd(z,Tw)} < \infty \right\}, \tag{2.20}$$

for some $z, w \in \mathbb{H}$.

To see that the definition does not depend on the choice of z, w, we use the triangle inequality to show that

$$e^{-sd(z,w)}e^{-sd(w,Tw)} \leq e^{-sd(z,Tw)} \leq e^{sd(z,w)}e^{-sd(w,Tw)}. \tag{2.21}$$

It is easy to check that $\delta = 0$ when Γ is elementary. For Γ geometrically finite, a slight modification of the argument from Proposition 2.28 shows that

$$\#\{T \in \Gamma : d(z, Tw) \leq t\} = O(e^t). \tag{2.22}$$

This implies in particular that $\delta \leq 1$. It turns out that $\delta = 1$ precisely when $X = \Gamma\backslash\mathbb{H}$ has finite area (Γ is of the first kind).

The interesting case is when X is non-elementary but infinite area (Γ is of the second kind). Under this assumption, Beardon [18, 19] established that $0 < \delta < 1$, with $\delta > \frac{1}{2}$ if Γ has parabolic elements (i.e., if X has cusps). Patterson [208] and Sullivan [260] proved that δ is the Hausdorff dimension of the limit set when Γ is geometrically finite. We will explore their theory, along with interesting applications to spectral theory, in Chapter 14.

By setting $z = w$ in (2.20) and using the bound (2.19), we see that

$$\sum_{\ell \in \mathcal{L}_X} e^{-s\ell} < \infty, \quad \text{for } \operatorname{Re} s > \delta.$$

This gives the range of convergence for the following:

Definition 2.30. For a hyperbolic surface X, the *Selberg zeta function* is defined for $\operatorname{Re} s > \delta$ by the product

$$Z_X(s) := \prod_{\ell \in \mathcal{L}_X} \prod_{k=0}^{\infty} \left(1 - e^{-(s+k)\ell}\right). \tag{2.23}$$

The product expression for $Z_X(s)$ is analogous to the Euler product form of the Riemann zeta function, with the role of the prime numbers being played by the primitive length spectrum. Like the Riemann zeta function, the Selberg zeta function admits an analytic continuation to a meromorphic function of $s \in \mathbb{C}$. This can be derived from the Selberg trace formula if X has finite area. For the full geometrically finite case, meromorphic continuation was proven by Guillopé [113]. We will essentially follow the same route to give the proof in Proposition 10.13. A simpler proof by dynamical methods is available if X has no cusps (Γ is convex cocompact); see Chapter 15.

2.6 Hyperbolic Trigonometry

One of the most satisfying results in differential geometry is the Gauss-Bonnet theorem relating the shape of a polygonal region to the curvature of its interior. We present here only the hyperbolic version, for which we can take advantage of the characterization of geodesics given in Proposition 2.3.

The first step in our proof is to compute the area of a triangle in $\mathbb{H}$. Note that since the geodesic arc $[p, q]$ is uniquely defined even for p or q in $\partial\mathbb{H}$, it makes sense to allow "degenerate" triangles with some vertices in $\partial\mathbb{H}$. Two geodesics meeting at a point of $\partial\mathbb{H}$ must be tangent there, so the interior angle at a vertex is zero if and only if it lies in $\partial\mathbb{H}$.

Lemma 2.31 (Triangle Area). *Let T be a triangle in $\mathbb{H}$ with interior angles $\alpha, \beta, \gamma \geq 0$. Then*

$$\text{area}(T) = \pi - (\alpha + \beta + \gamma).$$

Proof. This simple proof is taken from Katok [142]. First consider a triangle with at least one point on $\partial\mathbb{H}$. We can apply a Möbius transformation to map this point to ∞ while sending the opposite side to an arc of the unit circle. Let α, β be the interior angles at the other two vertices. From the diagram in Figure 2.11 we can see that the two sides meeting at infinity are the vertical lines $\operatorname{Re} z = -\cos\alpha$ and $\operatorname{Re} z = \cos\beta$. To compute the area we simply integrate

$$\begin{aligned}\text{area}(T) &= \int_{-\cos\alpha}^{\cos\beta} \int_{\sqrt{1-x^2}}^{\infty} \frac{dy\, dx}{y^2} \\ &= \int_{-\cos\alpha}^{\cos\beta} \frac{dx}{\sqrt{1-x^2}} \\ &= \pi - \alpha - \beta.\end{aligned}$$

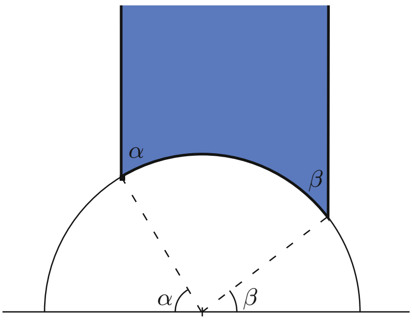

Fig. 2.11 Triangle with vertex at ∞.

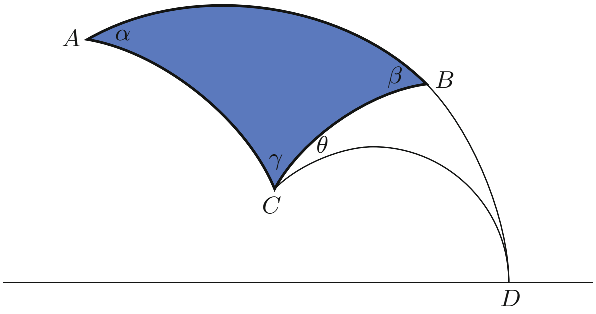

Fig. 2.12 Triangle extension.

If a triangle has all three vertices A, B, C in $\mathbb{H}$, we simply draw an auxiliary triangle by extending the segment $[A, B]$ until it meets $\mathbb{H}$ at a new vertex D, as shown in Figure 2.12. Applying the above computation to the triangles ACD and BCD gives

$$\begin{aligned}\operatorname{area}(ABC) &= \operatorname{area}(ACD) - \operatorname{area}(BCD)\\ &= (\pi - \alpha - (\gamma + \theta)) - (\pi - \theta - (\pi - \beta))\\ &= \pi - \alpha - \beta - \gamma.\end{aligned}$$

□

The computation of Lemma 2.31 already gives a local version of the Gauss-Bonnet theorem. For the global version we allow a polygonal region with possibly nontrivial topology.

Theorem 2.32 (Gauss-Bonnet). *Suppose Z is a region of finite area in some geometrically finite hyperbolic surface X, with boundary (if any) consisting of n geodesic arcs meeting at interior angles $\alpha_1, \ldots, \alpha_n$. Then*

$$\operatorname{area}(Z) = -2\pi\chi(Z) + \sum_{j=1}^{n} (\pi - \alpha_j).$$

In particular, the area of the convex core N of a geometrically finite non-elementary hyperbolic surface X is given by

$$\operatorname{area}(N) = -2\pi\chi(X).$$

Proof. First assume Z is compact. If $\mathcal{F}_X$ is a Dirichlet fundamental region for X then Z is represented inside $\mathcal{F}_X$ by a polygonal region $\tilde{Z}$. By subdividing this polygonal region with geodesic arcs, we can produce a triangulation of Z with all edges are geodesic. Let V, E, F be the number of vertices, edges, and faces of the triangulation.

By definition, $\chi(Z) = V - E + F$. Note that there are n exterior edges and vertices in the triangulation. Since each interior edge bounds two faces, while an exterior edge bounds a single face, we have

$$3F = 2E - n. \tag{2.24}$$

Applying Lemma 2.31 to the F triangles and summing gives

$$\operatorname{area}(Z) = \pi F - \sum_{i=1}^{3F} \theta_i,$$

where the θ_i are the interior angles of the triangles. For each of the $V - n$ interior vertices the sum of the θ_i's contributes 2π, and of course α_j is the sum of the θ_i's at the exterior vertex j. Hence

$$\operatorname{area}(Z) = \pi F - 2\pi(V - n) - \sum_{j=1}^{n} \alpha_j.$$

By (2.24), $F - 2V + n = -2\chi(Z)$, so this completes the proof for Z compact.

For a non-compact region Z there are two cases to consider. The first is when Z has a vertex at the cusp point, meaning that two sides of Z extend tangentially out the cusp. This case requires no change from the above argument; we simply allow degenerate triangles in the triangulation and assign interior angle zero to any cusp vertices.

The second possibility is that some complete ends of cusps are contained within Z. Suppose that Z encompasses the ends of k cusps. For each of these ends we introduce a geodesic loop (with one new vertex) to cut off the end of the cusp, as shown in Figure 2.13. Let $Z_1, \dots, Z_k$ be the regions cut off in this way, so that $Y = Z - \cup Z_j$ is a compact region with $n + k$ vertices. Because the regions Y and Z have the same diffeomorphism type, $\chi(Y) = \chi(Z)$. If the interior angles of Y at the added vertices are denoted $\beta_1, \dots, \beta_k$, then by the formula above for the compact case we have

$$\operatorname{area}(Y) = -2\pi\chi(Z) + \sum_{j=1}^{n}(\pi - \alpha_j) + \sum_{j=1}^{k}(\pi - \beta_j).$$

On the other hand, by Lemma 2.31 we can see that

$$\operatorname{area}(Z_j) = \beta_j - \pi,$$

so that the extra terms cancel out in the formula for area(Z). (Note that a fully enclosed cusp end is not counted as a vertex of Z.) □

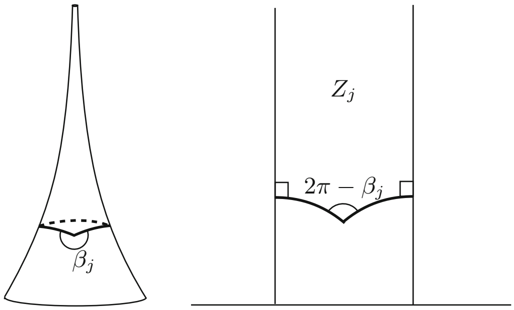

Fig. 2.13 Geodesic loop around a cusp end.

We next consider the geometry of hyperbolic hexagons. As with triangles, it makes sense to include cases where geodesics meet on the boundary and the interior angle is zero. We can also allow sides to have length zero, so that the two adjacent sides meet at the boundary. If multiple sides have length zero then they must be nonadjacent.

Lemma 2.33 (Right-Angled Hexagons). *Given any $a, b, c \geq 0$, there is a unique right-angled hexagon in $\mathbb{H}$ (up to isometry) such that a, b, c are the lengths of three nonadjacent sides.*

Proof. Start with an arbitrary geodesic γ_1, on which we mark off a segment of length a. From the endpoints draw geodesic arcs γ_2 and γ_6 which are perpendicular to the original segment. (If $a = 0$ we let γ_1 be a boundary point and take for γ_2 and γ_6 any two geodesic arcs meeting at γ_1.) Inside the region enclosed by these three curves, let η_2 be the locus of points whose distance to γ_2 is b, and η_6 the locus of points at distance c from γ_6, as shown in Figure 2.14. (In either the $\mathbb{H}$ or $\mathbb{B}$ model, a curve lying at a fixed distance from a geodesic is a circle meeting the boundary $\partial\mathbb{H}$ or $\partial\mathbb{B}$ at the endpoints of the geodesic.)

We claim that there is a unique geodesic arc tangent to both η_2 and η_6, which we label γ_4. To complete the construction, we fill in γ_3 as the arc of shortest distance between γ_2 and γ_4. This meets γ_4 at its intersection point with η_2. Similarly γ_5 is the shortest arc between γ_4 and γ_6. By the construction of γ_4, the segments γ_3 and γ_5 have lengths b and c, respectively. The hexagon obtained by this procedure is uniquely determined by the starting segment (and the choice of γ_2 and γ_6 if $a = 0$). □

To prove other basic formulas of hyperbolic trigonometry, it is helpful to introduce yet another model for hyperbolic space, the Minkowski or *hyperboloid model*. We follow the treatment in Buser [51, §2.1] for this discussion. Three-dimensional Minkowski space is $\mathbb{R}^3$ equipped with the Lorentzian metric $h = dx_1^2 + dx_2^2 - dx_3^2$. (This type of metric is used in general relativity.) To obtain a model for the hyperbolic plane we restrict our attention to the hyperboloid,

$$H := \left\{x \in \mathbb{R} : \ x_1^2 + x_2^2 - x_3^2 = -1\right\}.$$

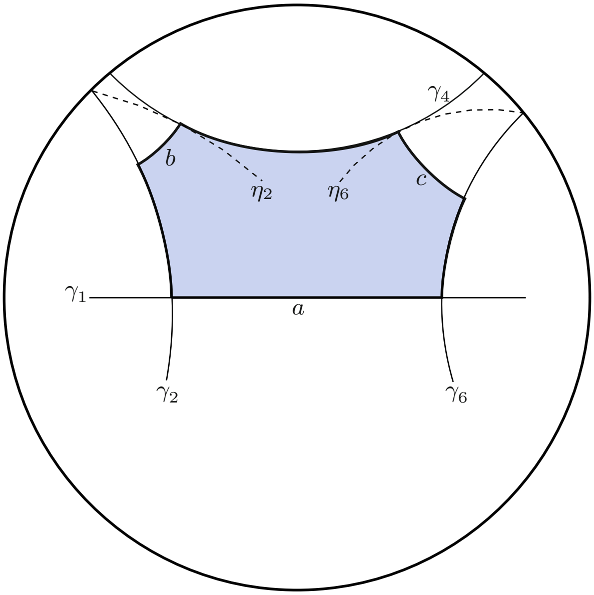

Fig. 2.14 Right-angled hexagon.

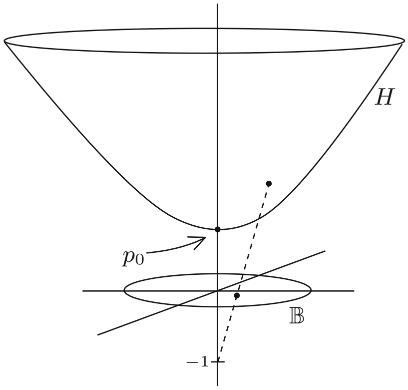

Fig. 2.15 Isometry from $\mathbb{B}$ to (H, g).

The restriction of h to H gives a positive definite metric g, and (H, g) is isometric to $\mathbb{B}$ by stereographic projection from $(0, 0, -1)$, as shown in Figure 2.15.

Isometries of (H, g) are generated by the linear transformations of $\mathbb{R}^3$ preserving h, so the orientation-preserving isometry group is identified with $\mathrm{SO}(2, 1)$. In particular, we can generate all isometries using

$$L_\theta := \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_r := \begin{pmatrix} \cosh r & 0 & \sinh r \\ 0 & 1 & 0 \\ \sinh r & 0 & \cosh r \end{pmatrix}, \tag{2.25}$$

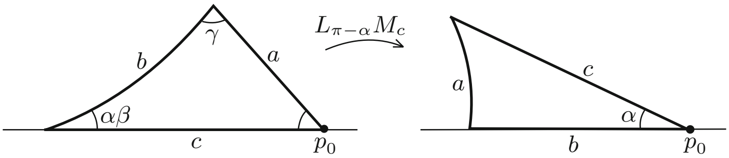

Fig. 2.16 Proving the sine rule.

for $\theta, r \in \mathbb{R}$. Fixing an origin $p_0 = (0,0,1)$, it is easy to check that the map $(r,\theta) \mapsto L_\theta M_r p_0$ defines a coordinate system on H in which g takes the geodesic polar form (2.11).

Lemma 2.34 (Sine Rule). *For a triangle ABC with geodesic sides, let α, β, γ denote the interior angles at the vertices, and a, b, c the respective lengths of the opposite sides. Then*

$$\frac{\sin\alpha}{\sinh a} = \frac{\sin\beta}{\sinh b} = \frac{\sin\gamma}{\sinh c}.$$

Proof. Regarding the triangle as a subset of H, we may assume vertex B is located at p_0 and that A is the point $M_{-c}p_0$. We first apply M_c to move vertex A to p_0, and then $L_{\pi-\alpha}$ to rotate so that C is located at $M_{-b}p_0$, as shown in Figure 2.16. Then we apply $L_{\pi-\gamma}M_b$ to shift C to p_0, followed by $L_{\pi-\beta}M_a$ to move B to p_0. Since this returns the triangle to its original position, we conclude that

$$L_{\pi-\beta}M_a L_{\pi-\gamma}M_b L_{\pi-\alpha}M_c = I. \tag{2.26}$$

Taking the equivalent statement

$$M_a L_{\pi-\gamma} M_b = L_{\beta-\pi} M_{-c} L_{\alpha-\pi},$$

and evaluating two particular matrix elements on either side gives

$$\begin{pmatrix} * & * & * \\ * & * & \sin\gamma\sinh b \\ * & -\sin\gamma\sinh a & * \end{pmatrix} = \begin{pmatrix} * & * & * \\ * & * & \sin\beta\sinh c \\ * & -\sin\alpha\sinh a & * \end{pmatrix}.$$

This proves our identity. ☐

We could have introduced rotations and translations in $\mathbb{B}$ or $\mathbb{H}$ and obtained a corresponding identity of the form (2.26), but it is much more difficult to read the sine and cosine rules from the matrix elements in those models. It is possible, though not easy, to prove the sine rule by more direct computation; see, e.g., [142].

Lemma 2.35 (Pentagon Rule). *For a right-angled pentagon with geodesic sides, with the lengths of consecutive sides labeled a, b, c, d, e,*

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

Proof. We apply the same strategy as in Lemma 2.34 to obtain the identity,

$$L_{\frac{\pi}{2}} M_a L_{\frac{\pi}{2}} M_b L_{\frac{\pi}{2}} M_c L_{\frac{\pi}{2}} M_d L_{\frac{\pi}{2}} M_e = I.$$

This implies the relation

$$L_{\frac{\pi}{2}} M_a L_{\frac{\pi}{2}} M_b = M_{-e} L_{-\frac{\pi}{2}} M_{-d} L_{-\frac{\pi}{2}} M_{-c} L_{-\frac{\pi}{2}},$$

and the claimed formula follows by comparing matrix entries on both sides. □

2.7 Fenchel-Nielsen Coordinates

Theorem 2.23 shows that a geometrically finite hyperbolic surface consists of a compact core with a finite number of funnels and cusps attached. In this section our goal is to develop some understanding of the space of hyperbolic metrics that could be put on the compact core in order to create such a surface.

2.7.1 Pants Decomposition

We will start by breaking the compact core of the surface up into components which are easy to parametrize.

Definition 2.36. A *pair of pants* is a hyperbolic surface diffeomorphic to a sphere with 3 punctures, with either geodesic boundary or cusp ends.

The Euler characteristic of a pair of pants is -1, so the Gauss-Bonnet theorem (Theorem 2.32) shows that the hyperbolic area is 2π. We can characterize each end with a boundary length ℓ, which is either the length of the closed geodesic or zero if the end is a cusp.

Lemma 2.37. *For each triple $\ell_1, \ell_2, \ell_3 \geq 0$, there is a unique pair of pants Y with these boundary lengths.*

Proof. Start with two identical right-angled hexagons with boundary lengths $\ell_1/2$, $\ell_2/2$, $\ell_3/2$, whose existence is guaranteed by Lemma 2.33. Because of the right angles, the hexagons can be glued together along seams given by the three edges whose lengths were not specified, to form a pair of pants with the appropriate boundary lengths, as shown in Figure 2.17.

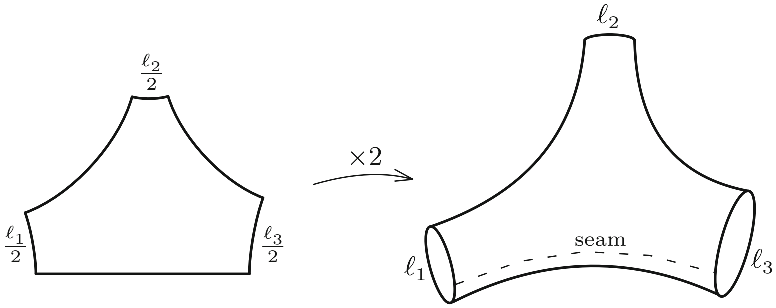

Fig. 2.17 Constructing a pair of pants.

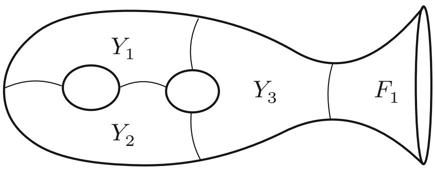

Fig. 2.18 Pants decomposition.

To prove uniqueness we observe that taking the shortest paths between the three boundary geodesics of a pair of pants gives three unique seams that split the pair of pants into two right-angled hexagons. (For a cusp, the "boundary geodesic" degenerates to a point at infinity and the hexagon has a side of length zero.) Since three nonadjacent side lengths (the seams) of the hexagons already match, the two hexagons are identical by Lemma 2.33. Therefore the lengths of the non-seam sides must be given by half the corresponding boundary lengths of the pair of pants. Lemma 2.33 thus shows that the hexagons are uniquely determined. □

Theorem 2.38 (Pants Decomposition). *The convex core of a geometrically finite, non-elementary hyperbolic surface X can be decomposed into a finite union of pairs of pants Y_j, $j = 1 \cdots, m$, where $m = -\chi(X)$, so that*

$$X = Y_1 \cup \cdots \cup Y_m \cup F_1 \cup \ldots F_{n_f}.$$

Proof. Figure 2.18 illustrates the claimed decomposition. From Theorem 2.23 we recall that the convex core N is X with funnels removed, and the compact core K is N minus the cusps. Since $\text{area}(N) = -2\pi\chi(X)$ and each pair of pants has area 2π, it is clear from the outset that at most m pairs of pants could be used. By induction, it suffices to show that we can cut a single pair of pants from N. (We are following the argument of Buser [51, Thm. 4.4.5] here.)

The boundary of K consists of finitely many closed geodesics or horocycles. For simplicity, we'll assume that there is at least one boundary geodesic, say γ. (The argument starting from a bounding horocycle is quite similar.) The neighborhood of points within distance a of γ,

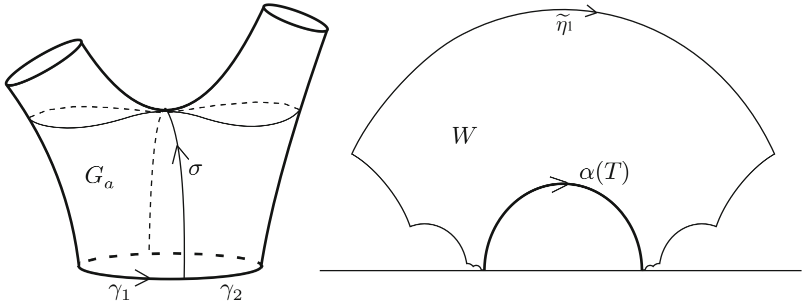

Fig. 2.19 Case 1 for pants decomposition.

$$G_a := \{z \in K : \ d(z, \gamma) \leq a\},$$

is isometric for small a to a half-collar $[0, a] \times S^1$, $ds^2 = dr^2 + \ell^2 \cosh^2 r\, d\theta^2$. (If we had started with a boundary horocycle, we'd get $ds^2 = dr^2 + e^{2r}\, d\theta^2$ instead.) As a increases, G_a must stop being isometric to a half-collar at some point. Otherwise the limit of G_a as $a \to \infty$ would be a funnel (or the big end of a parabolic cylinder). There are only two ways for the isometry to break down; for some value of a, either G_a meets itself (case 1) or G_a bumps into some other boundary curve of K (case 2).

Case 1: The perpendicular geodesic segments from γ to the first self-intersection point of G_a connect to form a geodesic arc σ dividing γ into two parts γ_1, γ_2. This set up is shown on the left in Figure 2.19. From the division of γ we form two simple closed curves, $\eta_1 = \gamma_1\sigma$ and $\eta_2 = \gamma_2\sigma^{-1}$. We can focus on η_1, as the argument is the same for either. Let $\tilde{\eta}_1$ denote a lift of η_1 to $\mathbb{H}$, as shown on the right in Figure 2.19. This lift is a union of segments meeting at right angles, which project down to γ_1 and σ in alternation. Since $\tilde{\eta}_1$ covers a closed curve on X, it is preserved by some maximal cyclic subgroup of Γ. Let $T \in \Gamma$ be the generator of this subgroup.

Assume first that T is parabolic, in which case $\tilde{\eta}_1$ would be a loop meeting $\partial\mathbb{H}$ at the fixed point p of T. For $R \in \Gamma - \langle T \rangle$, $R\tilde{\eta}_1$ cannot intersect $\tilde{\eta}_1$ because η_1 is simple. Also, p could not be fixed by R. (Lemma 2.21 shows that hyperbolic and parabolic fixed points cannot coincide, and if p were a parabolic fixed point of R this would contradict $R\tilde{\eta}_1 \cap \tilde{\eta}_1 = \emptyset$.) If W denotes the region enclosed by $\tilde{\eta}_1$, then since $R\tilde{\eta}_1 \cap \tilde{\eta}_1 = \emptyset$ and R does not fix p, we have $RW \cap W = \emptyset$. This means that $\Gamma\backslash W = \langle T \rangle\backslash W$, which therefore contains a single cusp.

On the other hand, suppose T is hyperbolic, with axis $\alpha(T)$. This is the case actually shown in Figure 2.19. Reasoning as above, for $R \in \Gamma - \langle T \rangle$ we have $R\tilde{\eta}_1 \cap \tilde{\eta}_1 = \emptyset$ and R cannot fix the endpoints of $\alpha(T)$. Together these imply that if W is the region bounded by $\tilde{\eta}_1$ and $\alpha(T)$, then $RW \cap W = \emptyset$. Hence $\Gamma\backslash W = \langle T \rangle\backslash W$ is an annulus bounded at one end by η_1 and at the other by the simple closed geodesic $\Gamma\backslash\alpha(T)$.

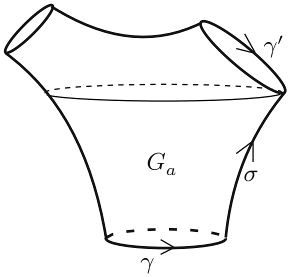

Fig. 2.20 Case 2 for pants decomposition.

With these two possibilities accounted for, and a similar argument applied to η_2, we have shown that γ is one of the boundary curves of a pair of pants contained within N.

Case 2: Let γ' be the boundary curve that G_a has bumped into, as shown in Figure 2.20. By linking γ and γ' through the geodesic arc σ connecting them and then expanding slightly, we can produce a simple closed curve $\eta := \sigma^{-1}\eta'\sigma\gamma$ which encompasses both boundary curves. We can argue exactly as in case 1 that η bounds either a cusp or an annulus with a closed geodesic at the other end. Thus γ and γ' are two of the boundary curves of a pair of pants contained within N. □

2.7.2 Moduli and Teichmüller Space

Let X be a non-elementary geometrically finite hyperbolic surface of genus g with n ends. The ends are already "marked" in the sense that they are divided into funnels and cusps. We also fix a pants decomposition as in Theorem 2.38. The combination of the pants decomposition with the labeling of the ends will be called a *marking* of X.

The number of pairs of pants in the marking is $2g - 2 + n$. Let n_{i} denote the number of "interior" bounding geodesics in the pants decomposition, meaning the boundaries between two pant legs. The decomposition contains a total of $2n_{\mathrm{i}} + n$ pant legs, and each pair of pants has 3 legs, which means that

$$n_{\mathrm{i}} = 3g - 3 + n.$$

Let us label the interior bounding geodesics of the pants decomposition by $\gamma_1, \ldots, \gamma_{n_{\mathrm{i}}}$. To each of these curves we can associated a two-parameter family of deformations of the surface. First, we can change the length ℓ_j of γ_j by changing the appropriate boundary lengths of the pairs of pants on both sides. Second, we can introduce a rotation in the gluing map between the two legs. To make the twist angle well defined, we can assume that the marking includes an orientation for each γ_j. A twist by angle $\theta_j \in \mathbb{R}$ can then be defined as a translation of the right side of the boundary geodesic by arclength $\theta_j\ell_j/2\pi$ relative to the left side. With this convention, a twist of $\pm 2\pi$ gives a surface isometric to the original.

There are an n_{f} additional geodesics bounding pairs of pants where funnels attach. The lengths of these geodesics, which we label $\ell_{n_{\mathrm{i}}+1}, \ldots, \ell_{r_{\mathrm{i}}+n_{\mathrm{f}}}$, could also be changed to deform the surface. However, since funnels are rotationally symmetric, there are no twist parameters for the funnels. The cusp ends correspond to pant legs of length zero, with no associated deformation parameters.

The parameters $\{\ell_j, \theta_j\}$ describing deformations of the hyperbolic structure on X are called *Fenchel-Nielsen coordinates*.

Proposition 2.39. *Let X be a geometrically hyperbolic surface with a marking as described above. Any complete hyperbolic surface diffeomorphic to X with ends of the same type can be realized by some combination of the Fenchel-Nielsen coordinates.*

Proof. Suppose Y is a hyperbolic surface diffeomorphic to X with ends of the same type. Since each non-cuspidal free homotopy class of Y contains a unique geodesic by Proposition 2.24, the marking of X can be transferred to Y. Now consider the deformation of X given by taking the boundary lengths ℓ_j to match those of Y. Since the boundary lengths determine the pairs of pants and funnels uniquely, it is clear that the angle parameters can then be chosen to create a deformation isometric to Y. □

The Fenchel-Nielsen coordinates are related to the *moduli space* $\mathcal{M}_X$, the set of isometry classes of complete hyperbolic metrics on X. In our convention the original surface X carries a hyperbolic metric, but $\mathcal{M}_X$ depends only on the topology and labeling of the ends. The moduli space is given a C^∞ topology, meaning that a sequence of isometry classes converges if and only if there exists representative metrics whose coordinate components and their derivatives converge uniformly on compact sets.

Let us define the full the Fenchel-Nielsen parameter space associated with a marking of X as

$$\mathcal{T}_X := \left\{(\ell_1, \ldots, \ell_{n_{\mathrm{i}}+n_{\mathrm{f}}}; \theta_1, \ldots \theta_{n_{\mathrm{i}}}) \in \mathbb{R}_+^{n_{\mathrm{i}}+n_{\mathrm{f}}} \times \mathbb{R}^d\right\} \tag{2.27}$$

(ignoring the periodicity of the twist angles). Proposition 2.39 gives a covering map

$$\mathcal{T}_X \to \mathcal{M}_X.$$

It is straightforward to see that the Euclidean topology on $\mathcal{T}_X$ is compatible with that of $\mathcal{M}_X$; see Buser [51, §3.2–3] for the details.

The space $\mathcal{T}_X$ is called *Teichmüller space*, and the standard definition is as the space of complex structures on X modulo pullback by diffeomorphisms isotopic to the identity. The Fenchel-Nielsen construction gives a set of coordinates for Teichmüller space by exploiting the connection between complex and hyperbolic structures on a surface. (See Buser [51, Ch. 6] for a proof of the identification of (2.27) with the definition in terms of complex structures.)

Notes

For the basic topology of differentiable manifolds assumed in this chapter (topology of surfaces, Euler characteristic, covering spaces, fundamental group, etc.), see e.g. Massey [168] or Munkres [188]. The differential geometry needed (metrics, Gaussian curvature, geodesics, etc.) can be found in an introductory book on surfaces, such as do Carmo [67] or Pressley [231]. Anderson [6] covers the basic geometry of the hyperbolic plane.

Our main sources for the theory of hyperbolic surfaces and Fuchsian groups were Beardon [20], Buser [51], Fenchel-Nielsen [84], and Katok [142]. Ratcliffe [233] gives a highly detailed introduction, with extensive historical notes. Milnor summarizes the history of hyperbolic geometry in [183].

Higher dimensional hyperbolic manifolds are obtained as quotients of $\mathbb{H}^n$ by discrete subgroups of isometries. Isometries of $\mathbb{H}^3$ can be realized by extending the action of Möbius transformations on the Riemann sphere to its interior. Thus in three dimensions the oriented isometry group is $\mathrm{PSL}(2,\mathbb{C})$. A discrete subgroup of $\mathrm{PSL}(2,\mathbb{C})$ is called a *Kleinian group*; see Maskit [167] for the basic theory. Limit sets of Kleinian groups are fascinating objects; see Mumford-Series-Wright [187].

Geometric finiteness is a more complicated issue in higher dimensions; see Bowditch [40] for an account of the various possible definitions.

There are many other approaches to Teichmüller theory. See, for example, Jost [138], Lehto [156], Seppälä-Sorvali [247], or Tromba [269].

http://www.springer.com/978-3-319-33875-0

Spectral Theory of Infinite-Area Hyperbolic Surfaces
Borthwick, D.
2016, XIII, 463 p. 64 illus., 37 illus. in color., Hardcover
ISBN: 978-3-319-33875-0
A product of Birkhäuser Basel