2 Differential Geometry of Riemann Surfaces

2.1 The Concept of a Riemann Surface

Definition 2.1.1 A two-dimensional manifold is called a surface.

Definition 2.1.2 An atlas on a surface S with charts $z_\alpha : U_\alpha \to \mathbb{C}$ is called conformal if the transition maps

$$z_\beta \circ z_\alpha^{-1} : z_\alpha(U_\alpha \cap U_\beta) \longrightarrow z_\beta(U_\alpha \cap U_\beta)$$

are holomorphic. A chart is compatible with a given conformal atlas if adding it to the atlas again yields a conformal atlas. A conformal structure is obtained by adding all compatible charts to a conformal atlas. A Riemann surface is a surface together with a conformal structure.

Definition 2.1.3 A continuous map $h : S_1 \to S_2$ between Riemann surfaces is said to be holomorphic[1] if, in local coordinates $\{U_\alpha, z_\alpha\}$ on S_1 and $\{U'_\beta, z'_\beta\}$ on S_2, all the maps $z'_\beta \circ h \circ z_\alpha^{-1}$ are holomorphic wherever they are defined. A holomorphic map h with nowhere vanishing derivative $\frac{\partial h}{\partial z}$ is called conformal.

We shall usually identify $U_\alpha \subset S$ with $z_\alpha(U_\alpha)$. The subscript is usually unnecessary, and we shall then identify $p \in U$ with $z(p) \in \mathbb{C}$. This will not cause any difficulties, since we only study local objects and concepts which are invariant under conformal maps. For example, this holds for holomorphic functions and maps, for meromorphic functions, for harmonic and subharmonic[2] functions, and for differentiable or rectifiable curves.
(The conformal invariance of (sub)harmonicity follows from the formula

$$\frac{\partial^2}{\partial z \partial \overline{z}}(f \circ h) = \left(\frac{\partial^2}{\partial w \partial \overline{w}} f \right) (h(z)) \frac{\partial}{\partial z} h \, \frac{\partial}{\partial \overline{z}} \overline{h}$$

for smooth f and holomorphic h.). In particular, all the local theorems of function theory carry over to holomorphic functions on Riemann surfaces (Riemann's theorem on removable singularities of holomorphic functions, the local form of a holomorphic function, local power-series expansions etc.).

[1] We shall also use the word "analytic" with the same significance.

[2] A function f on a Riemann surface is called (sub) harmonic if in a local conformal coordinate z, $\frac{\partial^2}{\partial z \partial \overline{z}} f = (\geq) 0$.

Examples.

1) Here is a trivial example: $\mathbb{C}$ and open subsets of $\mathbb{C}$ are Riemann surfaces. (More generally, any open nonempty subset of a Riemann surface is itself a Riemann surface.)
2) Here is the **most important example** of a compact Riemann surface: The Riemann sphere. $S^2 \subset \mathbb{R}^3$. We choose U_1 and U_2 as in the discussion of the sphere in Sec 1.1, and set

$$z_1 = \frac{x_1 + ix_2}{1 - x_3} \text{ on } U_1, \quad z_2 = \frac{x_1 - ix_2}{1 + x_3} \text{ on } U_2.$$

We then have $z_2 = \frac{1}{z_1}$ on $U_1 \cap U_2$, so that the transition map is indeed holomorphic.
It is also instructive and useful for the sequel to consider this example in the following manner: If we consider z_1 on all of S^2, and not only on $S^2 \backslash \{(0,0,1)\}$, then z_1 maps S^2 onto $\mathbb{C} \cup \{\infty\}$, the extended complex plane (this map z_1 then is called stereographic projection), in a bijective manner. While

$$z_1(U_1) = \mathbb{C} =: V_1$$

we have

$$z_2(U_2) = (\mathbb{C}\backslash\{0\}) \cup \{\infty\} =: V_2.$$

In that manner, the extended complex plane $\mathbb{C} \cup \{\infty\}$ is equipped with the structure of a Riemann surface with coordinate charts

$$\text{id} \ : V_1 \to \mathbb{C}$$

and

$$\begin{aligned} V_2 &\to \mathbb{C} \\ z &\mapsto \frac{1}{z}. \end{aligned}$$

Thus, we have two equivalent pictures or models of the Riemann sphere, namely the sphere $S^2 \subset \mathbb{R}^3$ on one hand and the extended complex plane $\mathbb{C} \cup \{\infty\}$ on the other hand. The stereographic projection $z_1 : S^2 \to \mathbb{C} \cup \{\infty\}$ is a conformal map. In Chapter 5, we shall see a third model or interpretation of the Riemann sphere, namely 1-dimensional complex projective space $\mathbb{P}^1$.
The model provided by $\mathbb{C} \cup \{\infty\}$ also offers the interpretation of a meromorphic function on an open subset of $\mathbb{C}$, or more generally, of a meromorphic function of a Riemann surface S, as a holomorphic map

$$g : S \to \mathbb{C} \cup \{\infty\}.$$

Namely, a function g on S is meromorphic precisely if every point $p \in S$ possesses a coordinate neighborhood U such that either g or $\frac{1}{g}$ is holomorphic on U. This, however, is the same as saying that we can find a neighborhood U of p for which g maps U holomorphically either to V_1 or V_2.

3) The torus, also introduced in Sec. 1.1, is a Riemann surface; the charts introduced there satisfy the conditions for a conformal atlas.
4) If S is a Riemann surface with conformal charts $\{U_\beta, z_\beta\}$, and $\pi : S' \to S$ a local homeomorphism, then there is a unique way of making S' a Riemann surface such that π becomes holomorphic. The charts $\{U'_\alpha, z'_\alpha\}$ for S' are constructed such that $\pi \mid U'_\alpha$ is bijective, and the $z_\beta \circ \pi \circ {z'}_\alpha^{-1}$ are holomorphic wherever they are defined. Thus $h \circ \pi$ will be holomorphic on S' if and only if h is holomorphic on S.
5) If $\pi : S' \to S$ is a (holomorphic) local homeomorphism of Riemann surfaces, then every covering transformation φ is conformal. Indeed, we can assume by 4) that $z'_\alpha = z_\alpha \circ \pi$. To say that φ is conformal means that $z'_\beta \circ \varphi \circ {z'}_\alpha^{-1}$ is conformal wherever it is defined. But $z'_\beta \circ \varphi \circ {z'}_\alpha^{-1} = z_\beta \circ \pi \circ \varphi \circ \pi^{-1} \circ z_\alpha^{-1} = z_\beta \circ z_\alpha^{-1}$, which is indeed conformal.

Exercises for § 2.1

1) Let S' be a Riemann surface, and $\pi : S' \to S$ a covering for which every covering transformation is conformal. Introduce on S the structure of a Riemann surface in such a way that π becomes holomorphic.
   Discuss a torus and H/Γ of exercise 3) in § 1.3 as examples.
2) Let S be a Riemann surface. Show that one may find a conformal atlas $\{U_\alpha, z_\alpha\}$ (compatible with the one defining the conformal structure of S) for which for every α, z_α maps U_α onto the unit disk $D := \{z \in \mathbb{C} : |z| < 1\}$. Thus, every U_α is conformally equivalent to D.

2.2 Some Simple Properties of Riemann Surfaces

Lemma 2.2.1 *On a compact Riemann surface S, every subharmonic function (hence also every harmonic or holomorphic function) is constant.*

Proof. Let $f : S \to \mathbb{R}$ be subharmonic. Since S is compact, f as a continuous function on S attains its maximum at some $p \in S$. Let $z : U \to \mathbb{C}$ be a local chart with $p \in U$. Then $f \circ z^{-1}$ is subharmonic on $z(U)$ and attains its maximum at an interior point, and is therefore constant by the maximum principle.
Thus the closed subset of S where f attains its maximum is also open, and hence is all of S. □

Lemma 2.2.2 *Let S be a simply-connected surface, and $F : S \to \mathbb{C}$ a continuous function, nowhere vanishing on S. Then* $\log F$ *can be defined on S, i.e. there exists a continuous function f on S with* $\mathrm{e}^f = F$.

Proof. Every $p_0 \in S$ has an open connected neighbourhood U with

$$\|F(p) - F(p_0)\| < \|F(p_0)\|$$

for $p \in U$, since $F(p_0) \neq 0$. Let $\{U_\alpha\}$ be the system consisting of these neighbourhoods, $(\log F)_\alpha$ a continuous branch of the logarithm of F in U_α, and $F_\alpha = \{(\log F)_\alpha + 2n\pi i,\, n \in \mathbb{Z}\}$. Then the assumptions of Lemma 1.4.1 are satisfied, hence there exists an f such that, for all α,

$$f_{|U_\alpha} = (\log F)_\alpha + n_\alpha \cdot 2\pi i, \qquad n_\alpha \in \mathbb{Z}.$$

Then f is continuous, and $\mathrm{e}^f = F$. □

Lemma 2.2.2 can also be proved as follows.
We consider the covering $\exp = e^z : \mathbb{C} \to \mathbb{C}\backslash\{0\}$. By Theorem 1.3.1, the continuous map $F : S \to \mathbb{C}\backslash\{0\}$ can be lifted to a continuous map $f : S \to \mathbb{C}$ with $\mathrm{e}^f = F$, since S is simply connected.

Lemma 2.2.3 *Let S be a simply connected Riemann surface, and $u : S \to \mathbb{R}$ a harmonic function. Then there exists a harmonic conjugate to u on the whole of S.*
(v is called a harmonic conjugate of u if $u + \mathrm{i}v$ is holomorphic.)

Proof. Let the U_α be conformally equivalent to the disc, and v_α a harmonic conjugate of u in U_α. Let $F_\alpha := \{v_\alpha + c, \quad c \in \mathbb{R}\}$. Then, by Lemma 1.4.1, there exists v such that, for all α,

$$v_{|U_\alpha} = v_\alpha + c_\alpha \qquad \text{for a } c_\alpha \in \mathbb{R}.$$

Such a v is harmonic, and conjugate to u. □

2.3 Metrics on Riemann Surfaces

We begin by introducing some general concepts:

Definition 2.3.1 A conformal Riemannian metric on a Riemann surface Σ is given in local coordinates by

$$\lambda^2(z)\ \mathrm{d}z\ \mathrm{d}\overline{z}, \qquad \lambda(z) > 0$$

(we assume λ is C^∞; this class of metrics is sufficient for our purposes). If $w \to z(w)$ is a transformation of local coordinates, then the metric should transform to

$$\lambda^2(z)\,\frac{\partial z}{\partial w}\frac{\partial \overline{z}}{\partial \overline{w}}\ \mathrm{d}w\ \mathrm{d}\overline{w}$$

(with $w = u + \mathrm{i}v$, $\frac{\partial}{\partial w} = \frac{1}{2}\left(\frac{\partial}{\partial u} - \mathrm{i}\frac{\partial}{\partial v}\right)$, $\frac{\partial}{\partial \overline{w}} = \frac{1}{2}\left(\frac{\partial}{\partial u} + \mathrm{i}\frac{\partial}{\partial v}\right)$.)

The length of a rectifiable curve γ on Σ is given by

$$\ell(\gamma) := \int_\gamma \lambda(z)\ |\mathrm{d}z|,$$

and the area of a measurable subset B of Σ by

$$\text{Area }(B) := \int_B \lambda^2(z)\, \frac{\mathrm{i}}{2}\ \mathrm{d}z \wedge \mathrm{d}\overline{z}$$

(the factor $\frac{\mathrm{i}}{2}$ arises because $\mathrm{d}z \wedge \mathrm{d}\overline{z} = (\mathrm{d}x + \mathrm{i}\mathrm{d}y) \wedge (\mathrm{d}x - \mathrm{i}\,\mathrm{d}y) = -2\mathrm{i}\,\mathrm{d}x \wedge \mathrm{d}y$). We shall usually write

$$\frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \quad \text{in place of} \quad \frac{\mathrm{i}}{2}\ \mathrm{d}z \wedge \mathrm{d}\overline{z}.$$

The distance between two points z_1, z_2 of Σ is defined as

$$d(z_1, z_2) := \inf\{\ell(\gamma) :\ \gamma : [0,1] \to \Sigma$$
$$\text{a (rectifiable) curve with}\, \gamma(0) = z_1,\ \gamma(1) = z_2\}.$$

The metric is said to be complete if every sequence $(t_n)_{n\in\mathbb{N}}$ in Σ which is Cauchy with respect to $d(\cdot\,,\cdot)$ (i.e. for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $\mathrm{d}(t_n, t_m) < \varepsilon$ for all $n, m \geq n_0$) has a limit in Σ. We leave it as an exercise to the reader to verify that the metric topology defined by the distance function $d(\cdot\,,\cdot)$ coincides with the original topology of Σ as a manifold.

Definition 2.3.2 A potential for the metric $\lambda^2(z)\mathrm{d}z\mathrm{d}\overline{z}$ is a function $F(z)$ such that

$$4\,\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}F(z) = \lambda^2(z).$$

The following lemma is immediate:

Lemma 2.3.1 *Arc lengths, areas and potentials do not depend on the local coordinates.* □

A metric is most simply described by means of a potential. Since a potential is invariant under coordinate transformations (and hence also under isometries, cf. Def. 2.3.5 and Lemma 2.3.2 below), it provides the easiest method of studying the transformation behaviour of the metric.

Definition 2.3.3 The Laplace-Beltrami operator with respect to the metric $\lambda^2(z)\,\mathrm{d}z\mathrm{d}\overline{z}$ is defined by

$$\begin{aligned} \Delta &:= \frac{4}{\lambda^2}\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}} \\ &= \frac{1}{\lambda^2}\left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}\right), \quad z = x + \mathrm{i}y. \end{aligned}$$

Definition 2.3.4 The curvature of the metric $\lambda^2(z)\,\mathrm{d}z\mathrm{d}\overline{z}$ is defined by

$$K = -\Delta \log \lambda.$$

Remark. With $z = x + \mathrm{i}y$, we have

$$\lambda^2(z)\ \mathrm{d}z\ \mathrm{d}\overline{z} = \lambda^2 \left(\mathrm{d}x^2 + \mathrm{d}y^2\right).$$

Thus the metric differs from the Euclidian metric only by the conformal factor λ^2. In particular, the angles with respect to $\lambda^2 \mathrm{d}z\mathrm{d}\overline{z}$ are the same as those with respect to the Euclidian metric.

Definition 2.3.5 A bijective map $h : \Sigma_1 \to \Sigma_2$ between Riemann surfaces, with metrics $\lambda^2\,\mathrm{d}z\mathrm{d}\overline{z}$ and $\varrho^2\,\mathrm{d}w\mathrm{d}\overline{w}$ respectively, is called an isometry if it preserves angles and arc-lengths.

Remark. We have assumed here that angles are oriented angles. Thus, anti-conformal maps cannot be isometries in our sense. Usually, the concept of an isometry permits orientation-reversing maps as well, for instance reflections. Thus, what we have called an isometry should be more precisely called an orientation-preserving isometry.

Lemma 2.3.2 *With the notation of Def. 2.3.5, $h = w(z)$ is an isometry if and only if it is conformal and*

$$\varrho^2(w(z))\,\frac{\partial w}{\partial z}\frac{\partial \overline{w}}{\partial \overline{z}} = \lambda^2(z)$$

(in local coordinates). If F_1 and F_2 are the respective potentials, then $F_1(z) = F_2(w(z))$ for an isometry. The Laplace-Beltrami operator and the curvature K are invariant under isometries.

Remark. An isometry has thus the same effect as a change of coordinates.

Proof. Conformality is equivalent to the preservation of angles, and the transformation formula of the lemma is equivalent to the preservation of arc-length. Finally,

$$\begin{aligned}\frac{4}{\varrho^2}\frac{\partial}{\partial w}\frac{\partial}{\partial \overline{w}}\log \varrho^2 &= 4\lambda^2 \frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}\log\left(\lambda^2 \frac{\partial z}{\partial w}\frac{\partial \overline{z}}{\partial \overline{w}}\right)\\ &= \frac{4}{\lambda^2}\left(\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}\right)\log \lambda^2,\end{aligned}$$

since the conformality of f implies that

$$\frac{\partial}{\partial \overline{z}}\frac{\partial z}{\partial w} = 0 = \frac{\partial}{\partial z}\frac{\partial \overline{z}}{\partial \overline{w}}.$$

This is equivalent to the invariance of the curvature. □

The trivial example is of course that of the Euclidian metric

$$\mathrm{d}z\ \mathrm{d}\overline{z}\ \left(= \mathrm{d}x^2 + \mathrm{d}y^2\right)$$

on $\mathbb{C}$. This has $K \equiv 0$.

We also have the following simple

Lemma 2.3.3 *Every compact Riemann surface Σ admits a conformal Riemannian metric.*

Proof. For every $z \in \Sigma$, there exists a conformal chart on some neighbourhood U_z

$$f_z : U_z \to \mathbb{C}.$$

We find some small open disk $D_z \subset f_z(U_z)$, and we consider the restricted chart

$$\varphi_z = f_{z|f_z^{-1}(D_z)} : V_z\ (:= U_z \cap f_z^{-1}(D_z)) \to \mathbb{C}.$$

Since Σ is compact, it can be covered by finitely many such neighbourhoods $V_{z_i},\ i = 1, \ldots, m$. For each i, we choose a smooth function $\eta_i : \mathbb{C} \to \mathbb{R}$ with

$$\eta_i > 0 \text{ on } D_{z_i}, \quad \eta_i = 0 \text{ on } \mathbb{C}\backslash D_{z_i}.$$

On D_{z_i}, we then use the conformal metric

$$\eta_i(w)\ \mathrm{d}w\ \mathrm{d}\overline{w}.$$

This then induces a conformal metric on $V_{z_i} = \varphi_{z_i}^{-1}(D_{z_i})$. The sum of all these local metrics over $i = 1, \ldots, n$ then is positive on all of Σ, and hence yields a conformal metric on Σ. □

We now want to consider the hyperbolic metric. For this purpose, we make some preparatory remarks.
Let

$$D := \{z \in \mathbb{C} :\ |z| < 1\} \text{ be the open unit disc}$$

and

$$H := \{z = x + \mathrm{i}y \in \mathbb{C} :\ y > 0\} = \text{ the upper half plane in } \mathbb{C}.$$

For $z_0 \in D$, $(z - z_0)/(1 - \overline{z}_0 z)$ defines a conformal self-map of D carrying z_0 to 0.
Similarly, for any $z_0 \in H$,

$$z \mapsto \frac{z - z_0}{z - \overline{z}_0}$$

is a conformal map of H onto D, mapping z_0 to 0. It follows in particular that H and D are conformally equivalent.

H and D are Poincaré's models of non-euclidean, or hyperbolic[3], geometry, of which we now give a brief exposition.
We shall need the following

Definition 2.3.6 A Möbius transformation is a map $\mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ of the form

$$z \mapsto \frac{az+b}{cz+d} \text{ with } a,b,c,d \in \mathbb{C}, \ ad-bc \neq 0.$$

We first recall the Schwarz lemma (see e.g. [A1]).

Theorem 2.3.1 *Let $f : D \to D$ be holomorphic, with $f(0) = 0$. Then*

$$|f(z)| \leq |z| \text{ and } |f'(0)| \leq 1.$$

If $|f(z)| = |z|$ for one $z \neq 0$, or if $|f'(0)| = 1$, then $f(z) = e^{i\alpha} z$ for an $\alpha \in [0, 2\pi)$.

An invariant form of this theorem is the theorem of Schwarz-Pick:

Theorem 2.3.2 *Let $f : D \to D$ be holomorphic. Then, for all $z_1, z_2 \in D$,*

$$\left| \frac{f(z_1) - f(z_2)}{1 - \overline{f(z_1)} f(z_2)} \right| \leq \frac{|z_1 - z_2|}{|1 - \bar{z}_1 z_2|}, \tag{2.3.1}$$

and, for all $z \in D$

$$\frac{|f'(z)|}{1 - |f(z)|^2} \leq \frac{1}{1 - |z|^2}. \tag{2.3.2}$$

Equality in (2.3.1) for some two distinct z_1, z_2 or in (2.3.2) for one z implies that f is a Möbius transformation (in which case both (2.3.1) and (2.3.2) are identities). (More precisely, f is the restriction to D of a Möbius transformation that maps D to itself.)

Proof. We reduce the assertions of the theorem to those of Theorem 2.3.1 by means of Möbius transformations, namely, with $w = f(z)$, and $w_1 = f(z_1)$, let

$$v := \omega^{-1}(z) := \frac{z_1 - z}{1 - \bar{z}_1 z}, \quad \xi(w) := \frac{w_1 - w}{1 - \overline{w}_1 w}.$$

Then $\xi \circ f \circ \omega$ satisfies the assumptions of Theorem 2.3.1. Hence

$$|\xi \circ f \circ \omega(v)| \leq |v|,$$

which is equivalent to (2.3.1). Further we can rewrite (2.3.1) (for $z \neq z_1$) as

$$\frac{|f(z_1) - f(z)|}{|z_1 - z|} \cdot \frac{1}{|1 - \overline{f(z_1)} f(z)|} \leq \frac{1}{|1 - \bar{z}_1 z|} .$$

[3] We shall use the words "hyperbolic" and "non-euclidean" synonymously, although there exist other geometries (of positive curvature) that deserve the appellation "non-euclidean" as well; see the remarks on elliptic geometry at the end of this chapter.

Letting z tend to z_1, we get (2.3.2); observe that

$$|1 - \overline{w}w| = |1 - |w|^2| = 1 - |w|^2 \text{ for } |w| < 1.$$

The assertion regarding equality in (2.3.1) or (2.3.2) also follows from Theorem 2.3.1. □

Analogously, one can prove

Theorem 2.3.3

$$\left|\frac{f(z_1) - f(z_2)}{f(z_1) - \overline{f(z_2)}}\right| \le \frac{|z_1 - z_2|}{|z_1 - \overline{z}_2|}, \qquad z_1, z_2 \in H, \tag{2.3.3}$$

and

$$\frac{|f'(z)|}{Imf(z)} \le \frac{1}{Im(z)}, \qquad z \in H. \tag{2.3.4}$$

Equality for some $z_1 \neq z_2$ in (2.3.3) or for some z in (2.3.4) holds if and only if f is a Möbius transformation (in which case both inequalities become identities).
(Here, in fact, f must have the form $z \mapsto \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{R},\ ad-bc > 0$.) □

Corollary 2.3.1 *Let $f : D \to D$ (or $H \to H$) be biholomorphic (i.e. conformal and bijective). Then f is a Möbius transformation.*

Proof. After composing with a Möbius transformation if necessary, we may suppose that we have $f : D \to D$ and $f(0) = 0$. Then, by Theorem 2.3.1, we have

$$|f'(0)| \le 1 \text{ and } |(f^{-1})'(0)| \le 1.$$

Hence $|f'(0)| = 1$, so that f must be a Möbius transformation. □

Let now

$$\mathrm{SL}(2, \mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{R},\ ad - bc = 1 \right\},$$

$$\mathrm{PSL}(2, \mathbb{R}) := \mathrm{SL}(2, \mathbb{R}) / \left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}.$$

Via $z \to (az + b)/(cz + d)$, an element of $\mathrm{SL}(2, \mathbb{R})$ defines a Möbius transformation which maps H onto itself. Any element of $\mathrm{PSL}(2, \mathbb{R})$ can be lifted to $\mathrm{SL}(2, \mathbb{R})$ and thus defines a Möbius transformation which is independent of the lift.

We recall a general definition:

Definition 2.3.7 A group G acts as a group of transformations or transformation group of a manifold E, if there is given a map

$$\begin{aligned} G \times E &\to E \\ (g, x) &\to gx \end{aligned}$$

with

$$(g_1 g_2)(x) = g_1(g_2 x) \qquad \text{for all } g_1, g_2 \in G,\ x \in E,$$

and

$$ex = x \qquad \text{for all } x \in E$$

where e is the identity element of G. (In particular, each $g : E \to E$ is a bijection, since the group inverse g^{-1} of g provides the inverse map).

Specially important for us is the case when E carries a metric, and all the maps $g : E \to E$ are isometries. It is easy to see that the isometries of a manifold always constitute a group of transformations.

Theorem 2.3.4 $\mathrm{PSL}(2, \mathbb{R})$ *is a transformation group of* H*. The operation is transitive (i.e. for any* $z_1, z_2 \in H$*, there is a* $g \in \mathrm{PSL}(2, \mathbb{R})$ *with* $gz_1 = z_2$*) and effective (or faithful, i.e. if* $gz = z$ *for all* $z \in H$*, then* $g = e$*).*
The isotropy group of a $z \in H$ *(which is by definition* $\{g \in \mathrm{PSL}(2, \mathbb{R}) : g(z) = z\}$*) is isomorphic to* $\mathrm{SO}(2)$.

Proof. The transformation group property is clear, and the faithfulness of the action is a consequence of the fact that we have normalised the determinant $ad - bc$ to 1.
To prove transitivity, we shall show that, given $z = u + \mathrm{i}v \in H$, we can find g with $g\mathrm{i} = z$. Thus we are looking for

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$$

with

$$\frac{a\mathrm{i} + b}{c\mathrm{i} + d} = u + \mathrm{i}v$$

or equivalently,

$$\frac{bd + ac}{c^2 + d^2} = u, \quad \frac{1}{c^2 + d^2} = v. \tag{2.3.5}$$

We can always solve (2.3.5) with $ad - bc = 1$. In particular, if

$$\frac{a\mathrm{i} + b}{c\mathrm{i} + d} = \mathrm{i},$$

we must have

$$\begin{aligned} bd + ac &= 0 \\ c^2 + d^2 &= 1 \\ ad - bc &= 1, \end{aligned}$$

so that (up to the freedom in the choice of the sign),

$$a = d = \cos\varphi, \quad b = -c = \sin\varphi.$$

Thus the isotropy group at i is isomorphic to SO(2). For any other $z \in H$, any $g \in \mathrm{PSL}(2,\mathbb{R})$ with $g\mathrm{i} = z$ provides an isomorphism between the isotropy groups at i and z. □

Definition 2.3.8 The hyperbolic metric on H is given by

$$\frac{1}{y^2}\,\mathrm{d}z\,\mathrm{d}\overline{z} \qquad (z = x + \mathrm{i}y).$$

Lemma 2.3.4 $\log\frac{1}{y}$ *is a potential for the hyperbolic metric. The hyperbolic metric has curvature* $K \equiv -1$*. Also, it is complete. In particular, every curve with an endpoint on the real axis and otherwise contained in* H *has infinite length.* □

Lemma 2.3.5 $\mathrm{PSL}(2,\mathbb{R})$ *is the isometry group of* H *for the hyperbolic metric.*

Proof. By Lemma 2.3.2, an isometry $h : H \to H$ is conformal. For any curve γ in H,

$$\begin{aligned} \int_{h(\gamma)} \frac{|\mathrm{d}h(z)|}{\mathrm{Im}h(z)} &= \int_{h(\gamma)} \frac{|h'(z)||\mathrm{d}z|}{\mathrm{Im}h(z)} \\ &\le \int_{\gamma} \frac{|\mathrm{d}z|}{\mathrm{Im}z}, \end{aligned}$$

and equality holds precisely when $h \in \mathrm{PSL}(2,\mathbb{R})$ (by Theorem 2.3.3). □

Lemma 2.3.6 *The hyperbolic metric on* D *is given by*

$$\frac{4}{(1-|z|^2)^2}\,\mathrm{d}z\,\mathrm{d}\overline{z},$$

and the isometries between H *and* D *are again Möbius transformations.*

Proof. This lemma again follows from Schwarz's lemma, just like Theorems 2.3.2 and 2.3.3. It has of course to be kept in mind that, up to a Möbius transformation, $w = (z - \mathrm{i})/(z + \mathrm{i})$ is the only transformation which carries the metric $\frac{1}{y^2}\mathrm{d}z\mathrm{d}\overline{z}$ to the metric $\frac{4}{(1-|w|^2)^2}\,\mathrm{d}w\mathrm{d}\overline{w}$. □

Consider now the map

$$z \to e^{iz} =: w$$

of H onto $D\backslash\{0\}$. This local homeomorphism (which is actually a covering) induces on $D\backslash\{0\}$ the metric

$$\frac{1}{|w|^2 \left(\log|w|^2\right)^2}\, dw\, d\overline{w}$$

with potential $\frac{1}{4}\log\log|w|^{-2}$, i.e. the map becomes a local isometry between H with its hyperbolic metric and $D\backslash\{0\}$ with this metric. This metric is complete; in particular, every curve going to 0 has infinite length. On the other hand, for every $r < 1$, $\{w : \ |w| \le r\}\backslash\{0\}$ has finite area.

Finally, we consider briefly the sphere

$$S^2 = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : \quad x_1^2 + x_2^2 + x_3^2 = 1\right\},$$

with the metric induced on it by the Euclidean metric $dx_1^2 + dx_2^2 + dx_3^2$ of $\mathbb{R}^3$.

If we map S^2 onto $\mathbb{C} \cup \{\infty\}$ by stereographic projection:

$$(x_1, x_2, x_3) \mapsto \frac{x_1 + i x_2}{1 - x_3} = z,$$

then the metric takes the form

$$\frac{4}{\left(1 + |z|^2\right)^2}\, dz\, d\overline{z},$$

as a computation shows.

We shall briefly state a few results concerning this case; we omit the necessary computations, which are straightforward: the curvature is $K \equiv 1$, Area$\,(S^2) = 4\pi$, the isometries are precisely the Möbius transformations of the form

$$z \mapsto \frac{az - \overline{c}}{cz + \overline{a}}, \quad |a|^2 + |c|^2 = 1.$$

We now wish to introduce the concept of geodesic lines.

Let

$$\gamma : [0,1] \to \Sigma$$

be a smooth curve. The length of γ then is

$$\ell(\gamma) = \int_0^1 \lambda(\gamma(t))\ |\dot{\gamma}(t)|\ dt.$$

We have

$$\frac{1}{2}\,\ell^2(\gamma) \le E(\gamma) = \frac{1}{2}\int_0^1 \lambda^2(\gamma(t))\,\dot{\gamma}(t)\,\dot{\overline{\gamma}}(t)\ dt. \tag{2.3.6}$$

($E(\gamma)$ is called the energy of γ), with equality precisely if

$$\lambda(\gamma(t))\ |\dot{\gamma}(t)| \equiv \text{const.} \tag{2.3.7}$$

in which case we say that γ is parametrized proportionally to arclength. Therefore, the minima of ℓ that satisfy (2.3.7) are precisely the minima of E. In other words, the energy functional E, when compared with ℓ, selects a distinguished parametrization for minimizers. We want to characterize the minimizers of E by a differential equation. In local coordinates, let

$$\gamma(t) + s\,\eta(t)$$

be a smooth variation of γ, $-s_0 \leq s \leq s_0$, for some $s_0 > 0$. If γ minimizes E, we must have

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}s} E(\gamma + s\eta)\Big|_{s=0} \\ &= \frac{1}{2}\int_0^1 \left\{ \lambda^2(\gamma)\ \left(\dot{\gamma}\dot{\overline{\eta}} + \dot{\overline{\gamma}}\dot{\eta}\right) + 2\lambda\ (\lambda_\gamma \eta + \lambda_{\overline{\gamma}}\overline{\eta})\ \dot{\gamma}\dot{\overline{\gamma}} \right\}\ \mathrm{d}t \text{ (here, } \lambda_\gamma = \tfrac{\partial\lambda}{\partial\gamma} \text{ etc.)} \\ &= \mathrm{Re}\int_0^1 \left\{ \lambda^2(\gamma)\dot{\gamma}\dot{\overline{\eta}} + 2\lambda\lambda_\gamma \dot{\gamma}\dot{\overline{\gamma}}\,\overline{\eta} \right\}\ \mathrm{d}t. \end{aligned}$$

If the variation fixes the end points of γ, i.e. $\eta(0) = \eta(1) = 0$, we may integrate by parts to obtain

$$0 = -\mathrm{Re}\int_0^1 \left\{ \lambda^2(\gamma)\ddot{\gamma} + 2\lambda\lambda_\gamma \dot{\gamma}^2 \right\}\ \overline{\eta}\ \mathrm{d}t.$$

If this holds for all such variations η, we must have

$$\ddot{\gamma}(t) + \frac{2\lambda_\gamma(\gamma(t))}{\lambda(\gamma(t))}\,\dot{\gamma}^2(t) = 0. \tag{2.3.8}$$

Definition 2.3.9 A curve γ satisfying (2.3.8) is called a geodesic.

We note that (2.3.8) implies (2.3.7) so that any geodesic is parametrized proportionally to arclength. Since the energy integral is invariant under coordinate chart transformations, so must be its critical points, the geodesics. Therefore (2.3.8) is also preserved under coordinate changes. Of course, this may also be verified by direct computation.

Lemma 2.3.7 *The geodesics for the hyperbolic metric on H are subarcs of Euclidean circles or lines intersecting the real axis orthogonally (up to parametrization).*

Proof. For the hyperbolic metric, (2.3.8) becomes

$$\ddot{z}(t) + \frac{2}{z - \overline{z}}\,\dot{z}^2(t) = 0 \tag{2.3.9}$$

for a curve $z(t)$ in H. Writing

$$z(t) = x(t) + \mathrm{i}\, y(t),$$

we obtain

$$\ddot{x} - \frac{2\dot{x}\dot{y}}{y} = 0, \quad \ddot{y} + \frac{\dot{x}^2 - \dot{y}^2}{y} = 0. \tag{2.3.10}$$

If $\dot{x} = 0$, then x is constant, and so we obtain a straight line intersecting the real axis orthogonally. If $\dot{x} \neq 0$, the first equations of (2.3.10) yields

$$\left(\frac{\dot{x}}{y^2}\right) = 0, \qquad \text{i.e. } \dot{x} = c_0 y^2, \quad (c_0 = \text{const. } \neq 0),$$

Since a geodesic is parametrized proportionally to arclength, we have

$$\frac{1}{y^2}\left(\dot{x}^2 + \dot{y}^2\right) = c_1^2 \qquad (c_1 = \text{ const.}).$$

We obtain

$$\left(\frac{\dot{y}}{\dot{x}}\right)^2 = \frac{c_1^2}{c_0^2\, y^2} - 1.$$

This equation is satisfied by the circle

$$(x - x_0)^2 + y^2 = \frac{c_1^2}{c_0^2}$$

that intersects the real axis orthogonally. A careful analysis of the preceding reasoning shows that we have thus obtained all geodesics of the hyperbolic metric. □

Correspondingly, the geodesics on the model D of hyperbolic geometry are the subarcs of circles and straight lines intersecting the unit circle orthogonally.

For our metric on the sphere S^2, the geodesics are the great circles on $S^2 \subset \mathbb{R}^3$ or (in our representation) their images under stereographic projection. Thus, any two geodesics have precisely two points of intersection (which are diametrically opposite to each other). We can pass to a new space $P(2, \mathbb{R})$ by identifying each point of S^2 with its diametrically opposite point. We then obtain the so-called elliptic geometry. In this space, two geodesics meet in exactly one point.

If we think of geodesics as the analogues of the straight lines of Euclidean geometry, we thus see that, in elliptic geometry, we cannot draw a parallel to a given straight line g through a point $p_0 \notin g$, since every straight line through p_0 does in fact meet g. In hyperbolic geometry on the other hand, there always exist, for every straight line g, infinitely many parallels to g (i.e. straight lines which do not meet g) passing through a prescribed point $p_0 \notin g$.

However, all the other axioms of Euclidean geometry, with the single exception of the parallel postulate, are valid in both geometries; this shows that the parallel postulate is independent of the remaining axioms of Euclidean geometry.
This discovery, which is of very great significance from a historical point of view, was made independently by Gauss, Bolyai and Lobačevsky at the beginning of the 19th century.

Exercises for § 2.3

1) Prove the results about S^2 stated at the end of § 2.3.
2) Let Λ be the group of covering transformations for a torus T. Let $\lambda^2 \mathrm{d}z\mathrm{d}\overline{z}$ be a metric on $\mathbb{C}$ which is invariant under all elements of Λ (i.e. each $\gamma \in \Lambda$ is an isometry for this metric). Then $\lambda^2 \mathrm{d}z\mathrm{d}\overline{z}$ induces a metric on T. Let K be its curvature.
   Show
   $$\int_T K = 0.$$
   Having read § 2.5, you will of course be able to deduce this from the Gauss-Bonnet theorem. The argument needed here actually is a crucial idea for proving the general Gauss-Bonnet theorem (cf. Cor. 2.5.6).

2.3.A Triangulations of Compact Riemann Surfaces

We let S be a compact surface, i.e. a compact manifold of dimension 2. A triangulation of S is a subdivision of S into triangles satisfying suitable properties:

Definition 2.3.A.1 A triangulation of a compact surface S consists of finitely many "triangles" $T_i,\ i = 1, \ldots, n$, with

$$\bigcup_{i=1}^{n} T_i = S.$$

Here, a "triangle" is a closed subset of S homeomorphic to a plane triangle Δ, i.e. a compact subset of the plane $\mathbb{R}^2$ bounded by three distinct straight lines. For each i, we fix a homeomorphism

$$\varphi_i : \Delta_i \to T_i$$

from a plane triangle Δ_i onto T_i, and we call the images of the vertices and edges of Δ_i vertices and edges, resp., of T_i. We require that any two triangles $T_i, T_j,\ i \neq j$, either be disjoint, or intersect in a single vertex, or intersect in a line that is an entire edge for each of them.

Remark. Similarly, one may define a "polygon" on S.

The notion of a triangulation is a topological one. The existence of triangulations may be proved by purely topological methods. This is somewhat tedious, however, although not principally difficult. For this reason, we shall use geometric constructions in order to triangulate compact Riemann surfaces. This will also allow us to study geodesics which will be useful later on as well. Only for the purpose of shortening our terminology, we say

Definition 2.3.A.2 A metric surface is a compact Riemann surface equipped with a conformal Riemannian metric.

The reader should be warned that this definition is not usually standard in the literature, and therefore, we shall employ it only in the present section. Let M be a metric surface with metric

$$\lambda^2(z)\,\mathrm{d}z\,\mathrm{d}\overline{z}.$$

We recall the equation (2.3.8) for geodesics in local coordinates

$$\ddot{\gamma}(t) + \frac{2\lambda_\gamma(\gamma(t))}{\lambda(\gamma(t))}\,\dot{\gamma}^2(t) = 0. \tag{2.3.A.1}$$

Splitting $\gamma(t)$ into its real and imaginary parts, we see that (2.3.A.1) constitutes a system of two ordinary differential equations satisfying the assumptions of the Picard-Lindelöf theorem. From that theorem, we therefore obtain

Lemma 2.3.A.1 *Let M be a metric surface with a coordinate chart $\varphi : U \to V \subset \mathbb{C}$. In this chart, let the metric be given by $\lambda^2(z)\,\mathrm{d}z\mathrm{d}\overline{z}$. Let $p \in V$, $v \in \mathbb{C}$. There exist $\varepsilon > 0$ and a unique geodesic (i.e. a solution of (2.3.A.1)) $\gamma : [0,\varepsilon] \to M$ with*

$$\begin{aligned}\gamma(0) &= p \\ \dot{\gamma}(0) &= v.\end{aligned} \tag{2.3.A.2}$$

γ depends smoothly on p and v. □

We denote this geodesic by $\gamma_{p,v}$.
If $\gamma(t)$ solves (2.3.A.1), so then does $\gamma(\lambda t)$ for constant $\lambda \in \mathbb{R}$. Thus

$$\gamma_{p,v}(t) = \gamma_{p,\lambda v}\Big(\frac{t}{\lambda}\Big) \qquad \text{for } \lambda > 0,\ t \in [0,\varepsilon]. \tag{2.3.A.3}$$

In particular, $\gamma_{p,\lambda v}$ is defined on the interval $\left[0, \frac{\varepsilon}{\lambda}\right]$. Since $\gamma_{p,v}$ depends smoothly on v as noted in the lemma and since $\left\{v \in \mathbb{C} : \|v\|_p^2 := \lambda^2(p)v\overline{v} = 1\right\}$ is compact, there exists $\varepsilon_0 > 0$ with the property that for any v with $\|v\|_p = 1$, $\gamma_{p,v}$ is defined on the interval $[0,\varepsilon_0]$. It follows that for any $w \in \mathbb{C}$ with $\|w\|_p \leq \varepsilon_0$, $\gamma_{p,w}$ is defined at least on $[0,1]$.

Let

$$V_p := \{v \in \mathbb{C} : \ \gamma_{p,v} \text{ is defined on } [0,1]\}\,.$$

Thus, V_p contains the ball

$$\{w \in \mathbb{C} : \ \|w\|_p \le \varepsilon_0\}\,.$$

We define the so-called exponential map

$$\begin{aligned} \exp_p : V &\to M \quad \text{(identifying points in } \varphi(U) = V \\ &\qquad\qquad \text{with the corresponding points in } M) \\ v &\mapsto \gamma_{p,v}(1). \end{aligned}$$

Lemma 2.3.A.2 $\exp_p$ *maps a neighbourhood of* $0 \in V_p$ *diffeomorphically onto some neighbourhood of* p.

Proof. The derivative of $\exp_p$ at $0 \in V_p$ applied to $v \in \mathbb{C}$ is

$$\begin{aligned} D\,\exp_p(0)(v) &= \frac{\mathrm{d}}{\mathrm{d}t}\gamma_{p,tv}(1)\bigg|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\gamma_{p,v}(t)\bigg|_{t=0} \\ &= \dot\gamma_{p,v}(0) \\ &= v \end{aligned}$$

by definition of $\gamma_{p,v}$.
Thus, the derivative of $\exp_p$ at $0 \in V_p$ is the identity. The inverse function theorem may therefore be applied to show the claim. □

In general, however, the map $\exp_p$ is not holomorphic. Thus, if we use $\exp_p^{-1}$ as a local chart, we preserve only the differentiable, but not the conformal structure. For that reason, we need to investigate how our geometric expressions transform under differentiable coordinate transformations. We start with the metric. We write

$$z = z^1 + \mathrm{i}\, z^2, \quad \mathrm{d}z = \mathrm{d}z^1 + \mathrm{i}\; \mathrm{d}z^2, \quad \mathrm{d}\overline{z} = \mathrm{d}z^1 - \mathrm{i}\; \mathrm{d}z^2.$$

Then

$$\lambda^2(z)\; \mathrm{d}z\; \mathrm{d}\overline{z} = \lambda^2(z)\; \big(\mathrm{d}z^1\; \mathrm{d}z^1 + \mathrm{d}z^2\; \mathrm{d}z^2\big).$$

If we now apply a general differentiable coordinate transformation

$$z = z(x), \text{ i.e. } z^1 = z^1(x^1, x^2), \quad z^2 = z^2(x^1, x^2) \qquad \text{with } x = (x^1, x^2)$$

the metric transforms to the form

$$\sum_{i,j,k=1}^{2} \lambda^2(z(x))\, \frac{\partial z^i}{\partial x^j} \frac{\partial z^i}{\partial x^k}\; \mathrm{d}x^j\; \mathrm{d}x^k.$$

We therefore consider metric tensors of the form

$$\sum_{j,k=1}^{2} g_{jk}(x)\ \mathrm{d}x^j\ \mathrm{d}x^k \tag{2.3.A.4}$$

with a positive definite, symmetric metric $(g_{jk})_{j,k=1,2}$. Again, we require that $g_{jk}(x)$ depends smoothly on x. The subsequent considerations will hold for any metric of this type, not necessarily conformal for some Riemann surface structure. W.r.t. such a metric, the length of a curve $\gamma(t)$ ($\gamma : [a, b] \to M$) is

$$\ell(\gamma) = \int_a^b \left(g_{jk}(\gamma(t))\, \dot\gamma^j(t)\, \dot\gamma^k(t) \right)^{\frac{1}{2}}\ \mathrm{d}t, \tag{2.3.A.5}$$

and its energy is

$$E(\gamma) = \frac{1}{2} \int_a^b g_{jk}(\gamma(t))\, \dot\gamma^j(t)\, \dot\gamma^k(t)\ \mathrm{d}t. \tag{2.3.A.6}$$

As before, one has

$$\ell^2(\gamma) \le 2(b-a)\, E(\gamma) \tag{2.3.A.7}$$

with equality iff

$$g_{jk}(\gamma(t))\, \dot\gamma^j(t)\, \dot\gamma^k(t) \equiv \text{const.}, \tag{2.3.A.8}$$

i.e. if γ is parametrized proportionally to arclength. The Euler-Lagrange equations for E, i.e. the equations for γ to be geodesic, now become

$$\ddot\gamma^i(t) + \sum_{j,k=1}^{2} \Gamma^i_{jk}(\gamma(t))\, \dot\gamma^j(t)\, \dot\gamma^k(t) = 0 \qquad \text{for } i = 1, 2, \tag{2.3.A.9}$$

with

$$\Gamma^i_{jk}(x) = \frac{1}{2} \sum_{l=1}^{2} g^{il}(x) \left(\frac{\partial}{\partial x^k} g_{jl}(x) + \frac{\partial}{\partial x^j} g_{kl}(x) - \frac{\partial}{\partial x^l} g_{jk}(x) \right) \tag{2.3.A.10}$$

where $\left(g^{jk}(x)\right)_{j,k=1,2}$ is the inverse matrix of $\left(g_{jk}(x)\right)_{j,k=1,2}$, i.e.

$$\sum_{k=1}^{2} g^{jk} g_{kl} = \begin{cases} 1 & \text{for } j = l \\ 0 & \text{for } j \neq l. \end{cases}$$

(The derivation of (2.3.A.9) needs the symmetry $g_{jk}(x) = g_{kj}(x)$ for all j, k.)
We now use the local coordinates $p \in M$ defined by $\exp_p^{-1}$. We introduce polar coordinates r, φ on V_p, ($x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$) on V_p, and call the resulting coordinates on M geodesic polar coordinates centered at p. By construction of $\exp_p$, in these coordinates the lines $r = t$, $\varphi = $ const. are geodesic.

We thus write the metric as $g_{11}\,\mathrm{d}r^2 + 2g_{12}\,\mathrm{d}r\mathrm{d}\varphi + g_{22}\,\mathrm{d}\varphi^2$. From (2.3.A.9), we infer

$$\Gamma^i_{11} = 0 \qquad \text{for } i = 1, 2$$

in these coordinates, i.e. by (2.3.A.10)

$$\sum_{l=1}^{2} g^{il}\left(2\,\frac{\partial}{\partial r} g_{1l} - \frac{\partial}{\partial l} g_{11}\right) = 0,$$

hence, since (g^{il}) is invertible,

$$2\,\frac{\partial}{\partial r} g_{1l} - \frac{\partial}{\partial l} g_{11} = 0 \qquad \text{for } l = 1, 2. \tag{2.3.A.11}$$

For $l = 1$, we obtain

$$\frac{\partial}{\partial r} g_{11} = 0. \tag{2.3.A.12}$$

Since by the properties of polar coordinates, φ is undetermined for $r = 0$,

$$g_{jk}(0, \varphi)$$

is independent of φ, and (2.3.A.12) implies

$$g_{11} \equiv \text{ const.} =: g. \tag{2.3.A.13}$$

(In fact, $g = 1$.) Inserting this into (2.3.A.11) yields

$$\frac{\partial}{\partial r} g_{12} = 0. \tag{2.3.A.14}$$

By the transformation rules for transforming Euclidean coordinates into polar coordinates, we have

$$g_{12}(0, \varphi) = 0$$

($x^1 = r\cos\varphi$, $x^2 = r\sin\varphi$, the metric in the coordinates x^1, x^2 written as $\sum_{j,k=1}^{2} \gamma_{jk}\,\mathrm{d}x^j\mathrm{d}x^k$, hence

$$g_{11} = \sum \gamma_{jk}\,\frac{\partial x^j}{\partial r}\frac{\partial x^k}{\partial r}, \quad g_{12} = \sum \gamma_{jk}\,\frac{\partial x^j}{\partial r}\frac{\partial x^k}{\partial \varphi},$$
$$g_{22} = \sum \gamma_{jk}\,\frac{\partial x^j}{\partial \varphi}\frac{\partial x^k}{\partial \varphi},$$

and $\frac{\partial x^j}{\partial \varphi} = 0$ at $r = 0$).
Thus, (2.3.A.14) implies

$$g_{12} = 0. \tag{2.3.A.15}$$

Since the metric is positive definite, we finally have

$$g_{22} > 0 \qquad \text{for } r > 0. \tag{2.3.A.16}$$

Lemma 2.3.A.3 *Let $\delta > 0$ be chosen such that*

$$\exp_p : \{v \in V_p : \ \|v\|_p < \delta\} \to M$$

is injective. Then for every $q = \exp_p(v)$ with $\|v\|_p < \delta$, the geodesic $\gamma_{p,v}$ is the unique shortest curve from p to q. In particular

$$d(p,q) = \|v\|_p.$$

Proof. Let $\gamma(t)$, $0 \leq t \leq T$ be any curve from p to q. Let

$$t_0 := \inf\bigl\{t \leq T : \ \gamma(t) \notin \exp_p\{\|v\|_p < \delta\}\bigr\},$$

$$\text{or } t_0 := T \text{ if no such } t \leq T \text{ can be found.}$$

We shall show that the curve $\gamma_{|[0,t_0]}$ is already longer than $\gamma_{p,v}$, unless it coincides with the latter one. For that purpose, we represent $\gamma(t)$ as $(r(t), \varphi(t))$ for $0 \leq t \leq t_0$ in our geodesic polar coordinates and compute

$$\begin{aligned}
\ell(\gamma_{|[0,t_0]}) &= \int_0^{t_0} \bigl(g_{11}\,\dot r^2(t) + 2g_{12}\,\dot r(t)\dot\varphi(t) + g_{22}\,\dot\varphi^2(t)\bigr)^{\frac12}\ \mathrm{d}t \\
&\geq \int_0^{t_0} \bigl(g\,\dot r^2(t)\bigr)^{\frac12}\ \mathrm{d}t \qquad \text{by } (2.3.\text{A}.13), (2.3.\text{A}.15), (2.3.\text{A}.16) \\
&= \int_0^{t_0} g^{\frac12}\ |\dot r(t)|\ \mathrm{d}t \\
&\geq \int_0^{t_0} g^{\frac12}\ \dot r(t)\ \mathrm{d}t \\
&= g^{\frac12}\, r(t_0) = \max\left(\delta, \ell(\gamma_{p,v})\right) \qquad \text{by definition of } t_0 \\
&\geq \ell(\gamma_{p,v}),
\end{aligned}$$

with equality only if $t_0 = T$ and $\varphi(t) = \text{const.}$, $\dot r(t) \geq 0$, i.e. if γ coincides with $\gamma_{p,v}$ up to parametrization. □

Corollary 2.3.A.1 *Let M be a compact metric surface. There exists $\varepsilon > 0$ with the property that any two points in M of distance $< \varepsilon$ can be connected by a unique shortest geodesic (of length $< \varepsilon$) (up to reparametrization).*

(Note, however, that the points may well be connected by further geodesics of length $> \varepsilon$.)

Proof. By the last sentence of Lemma 2.3.A.1, $\exp_p$ depends smoothly on p. Thus, if $\exp_p$ is injective on the open ball $\{\|v\|_p < \delta\}$, there exists a neighbourhood Ω of p such that for all $q \in \Omega$, $\exp_q$ is injective on $\{\|v\|_q < \delta\}$. Since M is compact, it may be covered by finitely many such neighbourhoods, and we then choose ε as the smallest such δ. Thus, for any $p \in M$, any point q in $\exp_p\{\|v\|_p < \varepsilon\}$ can be connected with p by a unique shortest geodesic, namely the geodesic $\gamma_{p,\exp_p^{-1}(q)}$, by Lemma 2.3.A.3. □

For our purposes, these geodesic arcs are useful because they do not depend on the choice of local coordinates. While the equation (2.3.A.9) is written in local coordinates, a solution satisfies it for any choice of local coordinates, as the equation preserves its structure under coordinate changes. This may be verified by direct computation. It can also be seen from the fact that these geodesics minimize the length and energy integrals, and these are readily seen to be coordinate independent.

Theorem 2.3.A.1 *Any compact metric surface - and hence by Lemma 2.3.3 any Riemann surface - can be triangulated.*

Proof. The idea of the proof is very simple. We select a couple of points and connect them by geodesics. More precisely, we choose them in such a manner that each of them has a certain number of other ones that are so close that the shortest geodesic connection is unique. Those geodesic connections then subdivide our surface into small pieces. One might now try to choose the points so carefully that these pieces are already triangles. It seems easier, however, to simply subdivide those pieces that happen not to be triangles. Such nontriangular pieces may arise because some of our geodesic connection may intersect. The subdivision presents no problem because at any such intersection point, the geodesics intersect at a nonvanishing angle.

We now provide the details.

Let Σ be a metric surface. Let ε be as in Corollary 2.3.A.1. We select finitely many points $p_1, \dots, p_n \in \Sigma$ with the following properties:
(i) $\forall\, p \in \Sigma\ \exists\, i \in \{1, \dots, n\}: \quad d(p, p_i) < \varepsilon$
(ii) $\forall\, i \in \{1, \dots, n\}\ \exists\, j, k \in \{1, \dots, n\}: \quad i \neq j,\ i \neq k,\ j \neq k$:

$$\begin{aligned} d(p_i, p_j) &< \frac{\varepsilon}{3} \\ d(p_i, p_k) &< \frac{\varepsilon}{3} \\ d(p_j, p_k) &< \frac{\varepsilon}{3}. \end{aligned}$$

Whenever $i, j \in \{1, \dots, n\}$ and $d(p_i, p_j) < \frac{\varepsilon}{3}$, we connect p_i and p_j by the unique shortest geodesic $\gamma_{i,j}$ of Corollary 2.3.A.1. By Lemma 2.3.A.2, any two such geodesics $\gamma_{i,j}$ and $\gamma_{k,l}$ intersect at most once. Namely, if there were two points q_1 and q_2 of intersection, then q_2 would have two different preimages under $\exp_{q_1}$, namely the two tangent vectors at the two geodesic subarcs of $\gamma_{i,j}$ and $\gamma_{k,l}$ from q_1 to q_2; since both these subarcs have length $< \varepsilon$ by construction, this would contradict the local injectivity of $\exp_{q_1}$. For any three points p_i, p_j, p_k as in (iii), the union of the geodesic arcs $\gamma_{i,j}, \gamma_{j,k}, \gamma_{i,k}$ subdivides Σ into a triangle T contained in $\{p : \ d(p, p_i) < \frac{2}{3}\}$ and its exterior.
This property may readily be deduced from the following observation. Any of the three geodesic arcs, say $\gamma_{i,j}$, may be extended as a geodesic up to a distance of length ε in both directions from either of its two endpoints, say

$\gamma_{i,j}$ by Lemma 2.3.A.1. By Lemma 2.3.A.2 this extended geodesic arc then divides $\exp_{p_i}\{\|v\|_{p_i} < \varepsilon\}$ into two subsets. T then is the intersection of three such sets.

We now enlarge the collection $\{p_1, \ldots, p_n\}$ to a collection $\{p_1, \ldots, p_N\}$ by including all points where any two such geodesics $\gamma_{i,j}, \gamma_{k,l}$ intersect $(i, j, k, l = 1, \ldots, n)$.

This subdivides Σ into finitely many "polygons" with geodesic sides. All angles at the vertices are different from 0, because by the uniqueness statement of Lemma 2.3.A.1 any two geodesics with the same initial direction coincide. Similarly, by Lemma 2.3.A.2, any polygon has at least three vertices. We now want to subdivide any such polygon P with more than three vertices into triangles in the sense of Def. 2.3.A.1. By construction, any two vertices of P have distance $< \frac{\varepsilon}{3}$. In order to carry out the subdivision, we always have to find two vertices of any such polygon P whose shortest geodesic connection is contained in the interior of P. Thus, let us suppose that p_0 is a vertex of P that cannot be connected in such manner with any other vertex of P. Let p_1 and p_2 be the two vertices adjacent to p_0, i.e. connected to p_0 by an edge of P. Let $\gamma_{0,1}$ be the edge from p_0 to p_1, $\gamma_{0,2}$ the one from p_0 to p_2, and let $\gamma_{1,2}$ be the shortest geodesic from p_1 to p_2. $\gamma_{0,1}, \gamma_{0,2}$ and $\gamma_{1,2}$ form a geodesic triangle T. We claim that some such T does not contain any vertices of P in its interior. Otherwise, let p_3 be a point on ∂P in the interior of T closest to p_0, with geodesic connection $\gamma_{0,3}$. The geodesic arcs $\gamma_{0,1}, \gamma_{0,2}$ and $\gamma_{1,2}$ can be continued beyond their endpoints up to a length of at least $\frac{\varepsilon}{3}$ in each direction, by choice of ε. By uniqueness of short geodesics, $\gamma_{0,3}$ cannot intersect any of these extended arcs, and $\gamma_{0,3}$ therefore is contained in the interior of T. Assume that $\gamma_{0,3}$ is contained in the interior of P, except for its endpoints. By choice of p_3, $\gamma_{0,3}$ does not contain points of ∂P besides p_0 and p_3. If p_3 is a vertex, it can thus be connected to p_0 in the interior of P contrary to the choice of p_0. Thus, let p_3 be contained in some edge ℓ of P. Let p_4 be one of the vertices of ℓ. We represent the part of ℓ between p_3 and p_4 as a (geodesic) smooth curve $\gamma : [0, 1] \to M$, with $\gamma(0) = p_3$. Let γ_t be the geodesic arc from p_0 to $\gamma(t)$. By Lemma 2.3.A.1, γ_t depends smoothly on t. Let t_0 be the smallest value of t for which the interior of t_0 is not disjoint to ∂P. If no such t_0 exists, then p_0 can be connected to the vertex p_4 in the interior of P, contrary to our assumption. If γ_{t_0} contains a vertex p_5, p_5 can be connected to p_0 in the interior of P, again a contradiction. Otherwise, however, γ_{t_0} is tangent to some edge of P. By the uniqueness statement of Lemma 2.3.A.1, it then has to coincide with that edge. This is only possible if that edge is $\gamma_{0,1}$ or $\gamma_{0,2}$. In that case, we perform the same construction with the other vertex of ℓ, reaching the same conclusion. Thus, ℓ has to coincide with $\gamma_{1,2}$. P thus is the triangle T, and there is nothing to prove. If $\gamma_{0,3}$ is contained in the exterior of P, we perform the same construction at another vertex that cannot be connected with any other vertex in the discussed manner, until we reach the desired conclusion, because it is impossible that for all vertices

the corresponding triangle T lies in the exterior of P, since P is contained in some geodesic triangle with vertices as in (ii).
Thus, we may always construct a subdivision of Σ into triangles. □

Exercises for § 2.3.A:

1) Show that for the unit disk D with the hyperbolic metric and $p \in D$, the exponential map $\exp_p$ is not holomorphic. Same for S^2.

2.4 Discrete Groups of Hyperbolic Isometries. Fundamental Polygons. Some Basic Concepts of Surface Topology and Geometry.

Definition 2.4.1 An action of the transformation group G on the manifold E is said to be properly discontinuous if every $z \in E$ has a neighbourhood U such that
$\{g \in G : \ gU \cap U \neq \emptyset\}$ is finite, and if z_1, z_2 are not in the same orbit, i.e. there is no $g \in G$ with $gz_1 = z_2$, they have neighbourhoods U_1 and U_2, resp., with $gU_1 \cap U_2 = \emptyset$ for all $g \in G$.

We have obviously:

Lemma 2.4.1 *If G acts properly discontinuously, then the orbit $\{gp : \ g \in G\}$ of every $p \in E$ is discrete (i.e. has no accumulation point in E).*

We now wish to study properly discontinuous subgroups Γ of $\mathrm{PSL}(2,\mathbb{R})$; Γ acts on H as a group of isometries. Being properly discontinuous, Γ has to be a discrete subgroup of $\mathrm{PSL}(2,\mathbb{R})$.[4] Indeed, if $g_n \to g$ for some sequence (g_n) in Γ, then $g_n z_0 \to g z_0$ for every $z_0 \in H$, in contradiction to Lemma 2.4.1. In particular, Γ is countable, because every uncountable set in $\mathbb{R}^4$, and hence also any such subset of $\mathrm{SL}(2,\mathbb{R})$ or $\mathrm{PSL}(2,\mathbb{R})$, has an accumulation point.

We now form the quotient H/Γ:

Definition 2.4.2 Two points z_1, z_2 of H are said to be equivalent with respect to the action of Γ if there exists $g \in \Gamma$ with $gz_1 = z_2$.
H/Γ is the space of equivalence classes, equipped with the quotient topology. This means that $(q_n)_{n\in\mathbb{N}} \subset H/\Gamma$ converges to $q \in H/\Gamma$ if and only if it is

---

[4] $\mathrm{SL}(2,\mathbb{R}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1\}$ is a subset of $\mathbb{R}^4$ in a natural way, and thus is equipped with a topology: $\begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ if and only if $a_n \to a, \ldots, d_n \to d$. This then also induces a topology on $\mathrm{PSL}(2,\mathbb{R})$. For every $z \in H$, the map $g \mapsto gz$ from $\mathrm{PSL}(2,\mathbb{R})$ to H is continuous.

possible to represent each q_n by an element $z_n \in H$ in the equivalence class defined by q_n such that $(z_n)_{n\in\mathbb{N}}$ converges to some $z \in H$ in the equivalence class of q.

If the action of Γ is free from fixed points, i.e. $gz \neq z$ for all $z \in H$ and all $g \neq \mathrm{id}$ in Γ, then H/Γ becomes a Riemann surface in a natural way. For $p_0 \in H/\Gamma$, choose $z_0 \in \pi^{-1}(p_0)$; since Γ is fixed-point-free and properly discontinuous, z_0 has a neighbourhood U such that $g(U) \cap U = \emptyset$ for $g \neq \mathrm{id}$ in Γ, so that $\pi : U \to \pi(U)$ is a homeomorphism. But this procedure provides H/Γ not only with a Riemann surface structure, but also a hyperbolic metric, because $\mathrm{PSL}(2,\mathbb{R})$ acts by isometries on H.

In order to develop some geometric understanding of such surfaces H/Γ, we start by establishing some elementary results in hyperbolic geometry. Let $\mathrm{SL}(2,\mathbb{R})$ operate as before on the upper half plane H via

$$z \mapsto \frac{a\,z+b}{c\,z+d} \quad (a,b,c,d \in \mathbb{R},\ ad-bc=1).$$

Lemma 2.4.2 *Each* $\gamma \in \mathrm{SL}(2,\mathbb{R})$, $\gamma \neq$ identity*, either has one fixed point in* H*, one fixed point on the extended real line* $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\} = \partial H$*, or two fixed points on* ∂H*. If* $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$*, this corresponds to* $|\mathrm{tr}\,\gamma| < 2$, $|\mathrm{tr}\,\gamma| = 2$, or $|\mathrm{tr}\,\gamma| > 2$*, resp., with* $\mathrm{tr}\,\gamma := a+d$.

Proof. If z is a fixed point of γ, then

$$c\,z^2 + (d-a)\,z - b = 0,$$

i.e.

$$z = \frac{a-d}{2\,c} \pm \sqrt{\frac{(a-d)^2 + 4\,b\,c}{4\,c^2}} = \frac{a-d}{2\,c} \pm \frac{1}{2\,c}\sqrt{(a+d)^2 - 4},$$

using $a\,d - b\,c = 1$, and the conclusion easily follows. □

Definition 2.4.3 An element of $\mathrm{SL}(2,\mathbb{R})$ with one fixed point in H is called elliptic, an element with one fixed point on $\overline{\mathbb{R}}$ parabolic, and one with two fixed points on $\overline{\mathbb{R}}$ hyperbolic.[5]

In order to see the geometric relevance of the distinction between elliptic, parabolic and hyperbolic automorphisms of H, let us discuss some examples:

– In the proof of Theorem 2.3.4, we have already determined all the transformations that fix i,

[5] This use of the word "hyperbolic" is not quite compatible with its use in "hyperbolic" geometry as now only certain isometries of hyperbolic space are called "hyperbolic". This is unfortunate, but we are following customary terminology here.

$$\frac{a\mathrm{i}+b}{c\mathrm{i}+d}=\mathrm{i}$$

An example is the transformation $z \mapsto -\frac{1}{z}$ that also maps 0 to ∞ and, in fact, is a reflection in the sense that every geodesic through i gets mapped onto itself, but with the direction reversed. Other such elliptic elements do not leave geodesics through i invariant. (The elliptic elements are perhaps even more easily understood when we let them operate on the unit disk D in place of H as our model space. In fact, the elliptic elements leaving the origin 0 of D fixed are simply the rotations $z \mapsto e^{\mathrm{i}\alpha} z$ with real α.)

– The transformation $\gamma : z \mapsto z+1$ is parabolic as it has a single fixed point on the boundary, namely ∞. This is already the typical case: If ∞ is a fixed point of γ, with $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $c = 0$, and, if γ is parabolic, by Lemma 2.4.1 $a+d=2$. Since also $a\,d=1$, it follows that

$$\gamma = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

Likewise, $z \mapsto \frac{z}{z+1}$ has a single fixed point on the boundary, namely 0, and is thus parabolic.

– The transformation of the form

$$\gamma = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

i.e. $z \to \lambda^2 z$, with real $\lambda \neq 1$, is hyperbolic. It fixes 0 and ∞. Again, this is the typical case as by applying an automorphism of H, we may assume that the two fixed points of our hyperbolic transformation are 0 and ∞. Such a γ leaves the geodesic connected these two points, namely the imaginary axis, invariant. This can be trivially seen by direct inspection of the operation of γ. A more abstract reason is that automorphisms of H are isometries with respect to the hyperbolic metric and therefore map geodesics to geodesics, and geodesics are uniquely determined by their endpoints on the boundary of H (see Lemma 2.3.6). In particular, since γ has no other fixed points, the imaginary axis is the only geodesic left invariant. On that geodesic, it operates as a translation, that is, shifts points along it by the distance

$$\int_{v\mathrm{i}}^{\lambda^2 v\mathrm{i}} \frac{1}{y}\mathrm{d}y = \log \lambda^2.$$

Lemma 2.4.3 *Let H/Γ be a compact Riemann surface for a subgroup Γ of* $\mathrm{PSL}(2,\mathbb{R})$*, as described in 2.4. Then all elements of Γ are hyperbolic.*

Proof. Γ cannot contain elliptic elements, because it has to operate without fixed points in H.

Let $\gamma \in \Gamma$. Since H/Γ is compact, there exists some z_0 in a fundamental region with

$$d(z_0, \gamma z_0) \leq d(z, \gamma z) \quad \text{for all } z \in H, \tag{2.4.1}$$

where d denotes hyperbolic distance.

Assume now that γ is parabolic. By applying an automorphism of H, we may assume that ∞ is the unique fixed point of γ, and as we have seen above, before the statement of the present lemma, γ is of the form

$$\gamma = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}.$$

Then for each $z \in H$,

$$d(z, \gamma z) = d(z, z + b),$$

and this goes to zero as $\operatorname{Im} z \to \infty$. Thus, if γ is parabolic there can be no z_0 satisfying (2.4.1), as γ has no fixed point in H. Therefore, γ cannot be parabolic. □

Lemma 2.4.4 *Let again H/Γ be a compact Riemann surface. Then for each $\gamma \in \Gamma, \gamma \neq$ identity, the free homotopy class of loops determined by γ contains precisely one closed geodesic (w.r.t. the hyperbolic metric).*

Proof. By applying an automorphism of H, we may assume that γ has 0 and ∞ as its fixed points (recall that γ is hyperbolic by Lemma 2.4.3). Thus, as we have already seen before Lemma 2.4.3, γ is of the form

$$\gamma = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

i.e. $z \to \lambda^2 z$, and there is precisely one geodesic of H which is invariant under the action of γ, namely the imaginary axis. A moment's reflection shows that the closed geodesics on H/Γ are precisely the projections of geodesics on H which are invariant under some nontrivial element of Γ, and this element of Γ of course determines the homotopy class of the geodesic. □

From the preceding proof and the discussion before Lemma 2.4.3, we also observe that the length of that closed geodesic on H/Γ is $\log \lambda^2$ because γ identifies points that distance apart on the imaginary axis.

Definition 2.4.4 An open subset F of H is called a fundamental domain for Γ if every $z \in H$ is equivalent under the action of Γ to a point z' in the closure of F, whereas no two points of F are equivalent.

Definition 2.4.5 A fundamental domain F is said to be a fundamental polygon if ∂F is a finite or countable union of geodesic arcs (together with their limit points in the latter case), the intersection of two such arcs being a single common end-point if non-empty.

We shall now construct a fundamental polygon for a given group Γ. For simplicity, we shall restrict ourselves to the case when Γ is fixed-point-free and H/Γ is compact; this is also the case we shall be mainly concerned with in the rest of this book.

Theorem 2.4.1 *Suppose $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ acts properly discontinuously and without fixed points on H, and that H/Γ is compact. Let $z_0 \in H$ be arbitrary. Then $F := \{z \in H : \ d(z, z_0) < d(z, gz_0)$ for all $g \in \Gamma\}$ [6] is a convex [7] fundamental polygon for Γ with finitely many sides. For every side σ of F, there exists precisely one other side σ' of F such that $g\sigma = \sigma'$ for a $g \in \Gamma$. Different pairs of such sides are carried to each other by different elements of Γ.*

Definition 2.4.6 Such a fundamental polygon is called the metric (or Dirichlet) fundamental polygon with respect to z_0.

Proof of Theorem 2.4.1. Since H/Γ is compact, F is bounded. Indeed, if

$$\mathrm{diam}\,(H/\Gamma) := \sup\{d(z_1, z_2) : \ z_1, z_2 \in H/\Gamma\},$$

where $d(\cdot,\cdot)$ now denotes the induced hyperbolic metric on H/Γ, then

$$\mathrm{diam}\, F \le \mathrm{diam}\,(H/\Gamma).$$

For each $g \in \Gamma$, the line

$$\{z \in H : \ d(z, z_0) = d(z, gz_o)\}$$

is a geodesic. (Actually, for every two $z_1, z_2 \in H$, $L = \{z \in H : \ d(z, z_1) = d(z, z_2)\}$ is geodesic. In order to see this, we first apply an isometry of H so that we can assume that z_1, z_2 are symmetric to the imaginary axis. Then L is the imaginary axis, hence geodesic.) Since F is bounded (so that $\overline{F}$ is compact), there can exist only finitely many $g \in \Gamma$ such that

$$d(z, z_0) = d(z, gz_0) \qquad \text{for some } z \in \overline{F};$$

indeed, since Γ operates properly discontinuously, Lemma 2.4.1 ensures that, for every $K > 0$, there are only finitely many $g \in \Gamma$ with $d(z_0, gz_0) \le K$. Thus F is the intersection of finitely many half-planes (with respect to hyperbolic geometry) of the form

$$\{z \in H : \ d(z, z_0) < d(z, gz_0)\};$$

[6] $d(\cdot\,,\cdot)$ denotes the distance with respect to the hyperbolic metric.

[7] "Convex" means that the geodesic segment joining any two points of F is entirely contained in F.

let $g_1, \dots, g_m$ be the elements thus occuring. In particular, F is convex and has finitely many sides, all of which are geodesic arcs. The intersection of two of these arcs (when not empty) is a common end-point, and the interior angle of F at the vertex is less than π.

To prove that every $z \in H$ has an equivalent point in $\overline{F}$, we first determine a $g \in \Gamma$ such that

$$d(z, gz_0) \le d(z, g'z_0) \qquad \text{for all } g' \in \Gamma.$$

Then $g^{-1}z$ is equivalent to z, and it lies in $\overline{F}$ since the action of Γ, being via isometries, preserves distances:

$$d(g^{-1}z, z_0) \le d(g^{-1}z, g^{-1}g'z_0) \qquad \text{for all } g' \in \Gamma,$$

and $g^{-1}g'$ runs through all of Γ along with g'.

Conversely, if z is a point of F, then

$$d(z, z_0) < d(z, gz_0) = d(g^{-1}z, z_0)$$

for all $g \in \Gamma$, so that no other point equivalent to z lies in F.

Thus F is a fundamental polygon.

A side σ_i of F is given by

$$\sigma_i = \{z : \; d(z, z_0) = d(z, g_i z_0),\; d(z, z_0) \le d(z, gz_0) \text{ for all } g \in \Gamma\}. \tag{2.4.2}$$

Since $d(g_i^{-1}z, gz_0) = d(z, g_i g z_0)$, we have for $z \in \sigma_i$

$$d(g_i^{-1}z, z_0) \le d(g_i^{-1}z, gz_0) \qquad \text{for all } g \in \Gamma,$$

with equality for $g = g_i^{-1}$. Thus g_i^{-1} carries σ_i to a different side. Since F is a convex polygon (with interior angles $< \pi$), g_i^{-1} carries all other sides of F outside F. Thus, different pairs of sides are carried to each other by different transformations. Further, the transformation which carries σ_i to another side is uniquely determined since, for an interior point of σ_i, equality holds in the inequality sign in (2.4.2) only for $g = g_i$. Thus all the assertions of the theorem have been proved. □

Corollary 2.4.1 *The transformations $g_1, \dots, g_m$ (defined in the proof of Theorem 2.4.1) generate Γ.*

Proof. For any $g \in \Gamma$, we consider the metric fundamental polygon $F(g)$ with respect to gz_0. Among the $F(g')$, only the $F(g_i)$ have a side in common with F, and g_i^{-1} carries $F(g_i)$ to F. If now $F(g')$ has a side in common with $F(g_i)$ say, then $g_i^{-1}F(g_i)$ has a side in common with F, so that there exists $j \in \{1, \dots, m\}$ with $g_j^{-1}g_i^{-1}F(g') = F$. Now, any $F(g_0)$ can be joined to F by a chain of the $F(g)$ in which two successive elements have a common side; hence, by what we have seen above, $F(g_0)$ can be carried to F by a product of the g_i^{-1}. Hence g_0 is a product of the g_i. □

Let us emphasize once again that, for a fundamental domain F of Γ,

$$H = \bigcup_{g \in \Gamma} g\overline{F}.$$

Thus the hyperbolic plane is covered without gaps by the closure of the fundamental domains gF, and these fundamental domains are pairwise disjoint. For the fundamental domain F of Theorem 2.4.1, the adjacent ones are precisely the $g_i F$, the g_i being as in Corollary 2.4.1.
To help visualisation, we shall now discuss some examples, though they are rather simple compared to the situation considered in Theorem 2.4.1.
Suppose first that Γ is a cyclic group. If Γ is to be fixed-point-free, then its generator must have its two fixed points (distinct or coincident) on the real axis. Of course H/Γ is not compact in this case, but a metric fundamental polygon for Γ can be constructed exactly as in Theorem 2.4.1.
We consider first the parabolic case, when Γ has only one fixed point on $\mathbb{R} \cup \{\infty\}$. As explained above, by conjugating with a Möbius transformation, we may assume that the fixed point is ∞, so that Γ is generated by a transformation of the form $z \to z + b$ ($b \in \mathbb{R}$). Thus, for any $z_0 \in H$, all the points gz_0 ($g \in \Gamma$) are of the form $z_0 + nb\,(n \in \mathbb{Z})$, i.e. lie on a line parallel to the real axis (which can be thought of as a circle with centre at infinity). The F of Theorem 2.4.1 is given in this case by

$$F = \{z = x + \mathrm{i}y : \ |x - \mathrm{Re}\, z_0| < \frac{b}{2}\},$$

see Fig.2.4.1.
Similarly, for the group Γ generated by the $z \mapsto \frac{z}{z+1}$, the fixed point of Γ is the point $p = 0$ on the real axis. In this case, the gz_0 lie on a Euclidean circle around 0, and the sides of the metric fundamental polygon are again geodesics orthogonal to these circles, see Fig.2.4.2. More generally, given any $p \in \mathbb{R}$, the map $z \mapsto \frac{z(p+1)-p^2}{z-p+1}$ has that point p as its unique fixed point and generates a parabolic Γ.

If the generator g_1 of Γ is hyperbolic so that it has two fixed points on $\mathbb{R} \cup \{\infty\}$, we recall from our above discussion of hyperbolic transformations

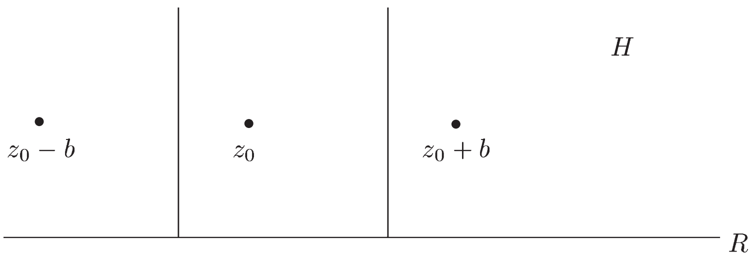

Fig. 2.4.1.

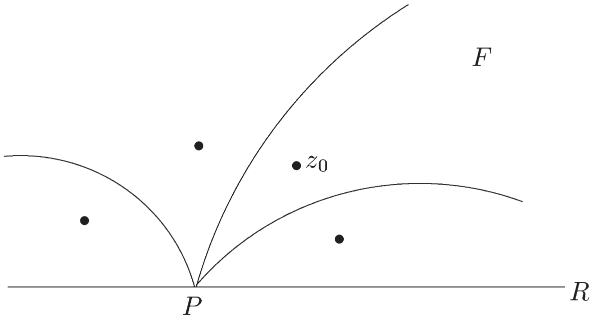

Fig. 2.4.2.

that wecan by conjugation send the fixed points e.g. to 0 and ∞. Then $g_1 z = \lambda z,\ \lambda > 0$. Hence the points equivalent to z_0 lie on the ray from the origin through z_0, and F will be bounded by two circles orthogonal to these rays and the real axis, see Fig.2.4.3.

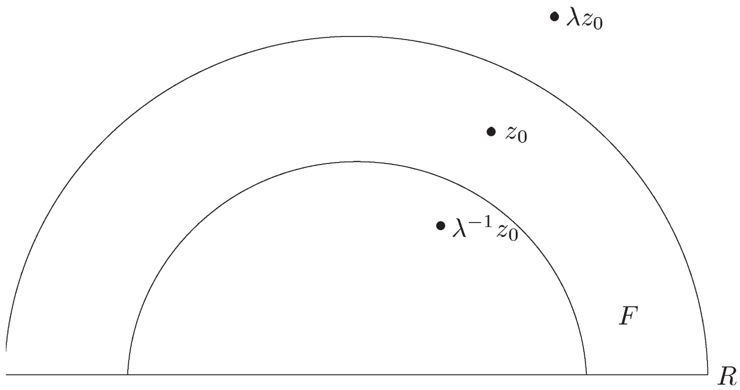

Fig. 2.4.3.

Correspondingly, for a generator g_1 with fixed points 0 and $p \in \mathbb{R}$, the gz_0 lie on the circle through 0, z_0 and p, and the sides of F are orthogonal to this circle, see Fig.2.4.4.

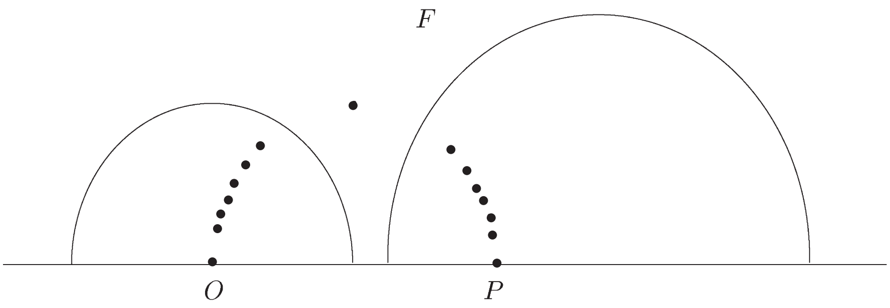

Fig. 2.4.4.

Finally, we consider groups Γ of Euclidean motions. In the compact cases $\mathbb{C}/\Gamma$ is a torus as we shall see later. In this case, a metric fundamental polygon is in general not a fundamental parallelogram, but a hexagon. If e.g. $\Gamma = \{z \mapsto z + n + m e^{\frac{2\pi i}{3}}, \quad n, m \in \mathbb{Z}\}$, then one obtains a regular hexagon, see Fig.2.4.5.

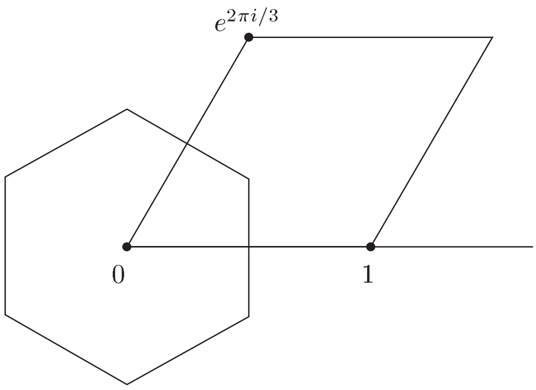

Fig. 2.4.5.

Theorem 2.4.2 *Under the assumptions of Theorem 2.4.1, there exists a fundamental polygon with finitely many sides, all of whose vertices are equivalent. Here again, every side a is carried by precisely one element of Γ to another side a', and the transformations corresponding in this way to distinct pairs of equivalent sides are distinct. The sides will be described in the order*

$$a_1 b_1 a_1' b_1' a_2 b_2 \cdots a_p b_p a_p' b_p';$$

in particular, the number of sides is divisible by 4.

The proof will be carried out in several steps. We start from the fundamental polygon F of Theorem 2.4.1.

1) Construction of a fundamental domain with finitely many sides, all of whose vertices are equivalent.
During this step, we shall denote equivalent vertices by the same letter. We choose some vertex p of F. If F has a vertex not equivalent to p, then F has also a side a with p as one end-point and $q \neq p$ as another. We join p to the other adjacent vertex of q, say r, by a curve d in F. Let g be the element of Γ which carries the side b between q and r to another side b' of F. We then get a new fundamental domain by replacing the triangle abd by its image under g; this fundamental domain has one p-vertex more, and one q-vertex less, than F. After repeating this process finitely many times, we finally get a fundamental domain with only p-vertices.

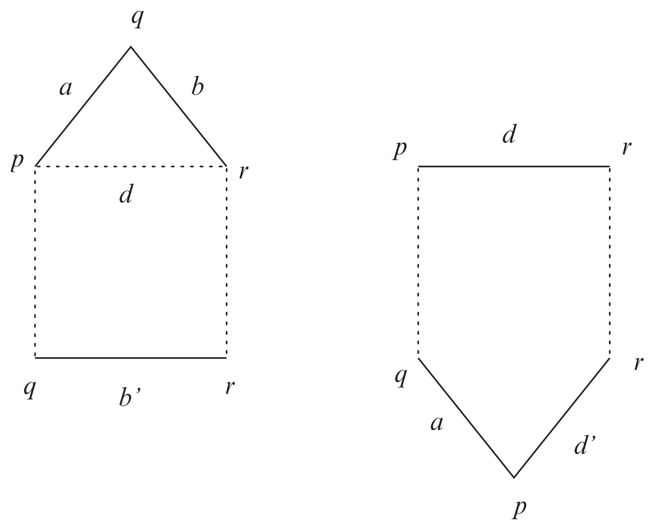

Fig. 2.4.6.

(For the curve d above, we could have chosen the geodesic arc from p to r the first time, since F was convex. But the modification made could destroy the convexity, so that it may not be possible to choose a geodesic diagonal inside the modified fundamental domain. We have therefore taken an arbitrary Jordan curve for d; thus the resulting fundamental domain is in general not a polygon. This defect will be rectified only at the very end of our construction.)

2)

Lemma 2.4.5 *With the above notation, a and a' cannot be adjacent, i.e. cannot have a common vertex.*

Proof. Let g be the transformation carrying a to a'. If F is the fundamental domain under consideration, then F and $g(F)$ are disjoint. Thus, if a and a'

have a common end-point, this point must be fixed by g (since g preserves orientation), in contradiction to our assumption that Γ acts without fixed points. □

3)

Lemma 2.4.6 *Let F be a fundamental domain all of whose vertices are equivalent. Then each $g \neq \mathrm{Id}$ in Γ carries at the most two vertices of F to vertices of F. In such a case, these two vertices are adjacent, and g carries the side of F between them to the side of F between their g-images.*

Proof. Let p_0 be a vertex of F, which is carried by $g \in \Gamma$ to another vertex p_0'; let a_1 be a side of F with p_0 as an end-point. Then either

$$g(a_1) \subset \partial F$$

or

$$g(a_1) \cap \partial F = p_0'.$$

In the first case, let p_1 be the other end-point of a_1, and a_2 the side of F adjacent to a_1 at p_1. Since $g(p_1) \in \partial F$, there are again the same two possibilities for a_2. Continuing in this manner, we arrive at a first vertex p_{j-1} and side a_j such that

$$g(a_j) \cap \partial F = g(p_{j-1}).$$

Then we have, for p_j and a_{j+1}, either

$$g(a_{j+1}) \cap \partial F = \emptyset$$

or

$$g(a_{j+1}) \cap \partial F = g(p_{j+1}).$$

In the first case, we again continue, till we arrive at the first vertex p_{k-1} and side a_k with

$$g(a_k) \cap \partial F = g(p_k)\ (k > j).$$

Continuing cyclically, we must return after finitely many steps to the vertex $p_m = p_0$ we started with. We now want to show that the whole chain $a_{k+1}, a_{k+2}, \ldots, a_m, a_1, \ldots, a_{j-1}$ is mapped by g into ∂F.
Thus, let F' be the domain bounded by $g(a_{k+1}), \ldots, g(a_{j-1})$ and the subarc of ∂F from $g(p_k)$ to $g(p_{j-1})$; here the latter is to be so chosen that F' and F are disjoint. Similarly, let F'' be the domain bounded by $g(a_j), \ldots, g(a_k)$ and the subarc of ∂F from $g(p_{j-1})$ to $g(p_k)$. We must show that either F' or F'' has empty interior.
Now,

$$\partial F' \subset F \cup g(F),\ \partial F'' \subset F \cup g(F),$$

and $\bigcup_{g\in\Gamma} g(F)$ is the whole of H, while $g_1(F)$ and $g_2(F)$ have no interior point in common if $g_1 \neq g_2$. Thus if F' and F'' had non-empty interior, we would have

$$F' = g'(F) \text{ and } F'' = g''(F)$$

for some $g', g'' \neq g$; in particular the interiors of F' and F'' would be fundamental domains. But $g(a_1) \subset \partial F$, hence either F' or F'' would have at least two sides fewer than $g(F)$. This is clearly impossible, since all images of F by elements of Γ of course have the same number of sides. Hence $F' = \emptyset$ or $F'' = \emptyset$, as asserted. Without loss of generality, let $F'' = \emptyset$. Then the chain $a_{k+1}, a_{k+2}, \ldots, a_{j-1}$ is mapped by g into ∂F.
If $j = 1$ and $k = m$, this chain is empty, and

$$F \cap g(F) = g(p_0)$$

in this case.
The important point to understand from the above considerations is however the following: if g maps two vertices of F into ∂F, then it also maps one of the two chains of sides between these two vertices into ∂F.
(By the way, we have not so far made use of the assumption that all vertices of F are equivalent; hence the above statement holds even if there are several equivalence classes of vertices).
We shall now show that the assumption that $g \neq \text{id}$ carries more than two vertices into ∂F leads to a contradiction. Indeed, what we have proved above shows that, in such a situation, we can find three successive vertices p_1, p_2, p_3 which, along with the sides a_1 and a_2 between them, are mapped by g into ∂F.

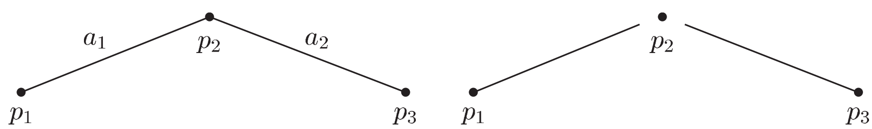

Fig. 2.4.7.

We now modify F slightly: instead of joining a_1 and a_2 at the intermediate vertex p_2, we connect them by means of a small arc going around p_2. We modify $g(a_1 \cup a_2)$ correspondingly. We then obtain a new fundamental domain with p_2 or $g(p_2)$ as an interior point. But this is the desired contradiction, since the closure of this fundamental region contains points equivalent to p_2 e.g. as boundary points.
This proves the lemma. □

4) In what follows, we need to use a modification of the fundamental domain which generalises the one we have already used in 1):

Let a and a' be equivalent sides of ∂F, so that $a' = ga$ for some $g \in \Gamma$. Let d be a diagonal curve which joins two vertices of F but otherwise lies in the interior of F. Then d divides F into two regions F_1 and F_2. Let $a \subset \overline{F}_1$, $a' \subset \overline{F}_2$. Then $g(F_1)$ has precisely the side a' in common with F_1 and $F_2 \cup g(F_1)$ is therefore again a fundamental domain (when the common sides are adjoined to it).

5) We now bring the sides to the desired order. We first choose the ordering of the sides such that c' is always to the right of c. Let c_1 be a side for which the number of sides between c_1 and c_1' is minimal: this number is positive by Lemma 2.4.5. Then the arrangement of the sides looks like

$$c_1 c_2 \cdots c_1' \ldots c_2' \tag{2.4.3}$$

where the dots indicate the possible presence of other sides. If there are no such intermediate sides, we look among the remaining sides for a c with the distance between c and c' minimal. Continuing this way, we must arrive at the situation of (2.4.3) with intermediate sides present (unless the sides are already in the desired order and there is nothing to prove). We now join the end-point of c_1 with the initial point of c_1' (end-point and initial point with respect to the chosen orientation of ∂F) by a diagonal, say b_1, and apply the modification of 4) to the pair c_1, c_1' and the diagonal b_1. We then obtain a fundamental domain with sides in the order

$$c_1 b_1 c_1' \cdots b_1'.$$

Without loss of generality, we may suppose that there are again some other sides between c_1' and b_1'. We now join the end-point of b_1 with the initial point of b_1' by a diagonal a_1 and again apply the modification of 4) to c_1, c_1' and a_1 and obtain the order

$$a_1 b_1 a_1' b_1' \cdots$$

for sides of the new fundamental domain.

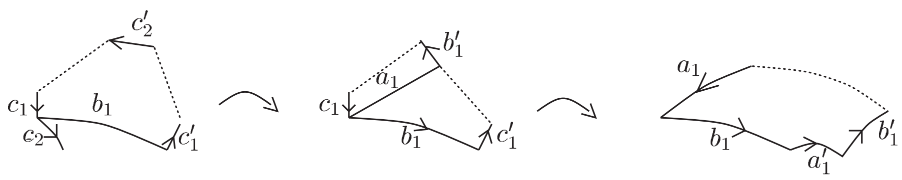

Fig. 2.4.8.

We repeat the above procedure for the remaining sides; this does not disturb the portion $a_1 b_1 a_1' b_1'$. After finitely many steps, we thus reach the desired order

$$a_1 b_1 a_1' b_1' \cdots a_p b_p a_p' b_p'.$$

6) In this last step we move the sides of the fundamental domain F constructed in 5) to geodesic arcs:
We fix a side a of F, and consider the geodesic $\widetilde{a}$ with the same end-points. Let $A(t,s)$ be a homotopy with $A(\cdot,0) = a$ and $A(\cdot,1) = \widetilde{a}$ such that none of the curves $A(\cdot,s)$ has self-intersections.[8]
We now deform the side a by the homotopy $A(\cdot,s)$ and the equivalent side $a' = g(a)$ by the homotopy $g(A(\cdot,s))$. We wish to say that we obtain in this way a new fundamental domain F_s.
Suppose first that, as s increases from 0 to 1, the curve $A(\cdot\,,s)$ meets another side b without crossing any vertex. Then the domain acquires for example some points which were previously exterior to the fundamental domain; but these points (or points equivalent to them) will be taken away at the side b' equivalent to b. Thus we will always be left with a fundamental domain.

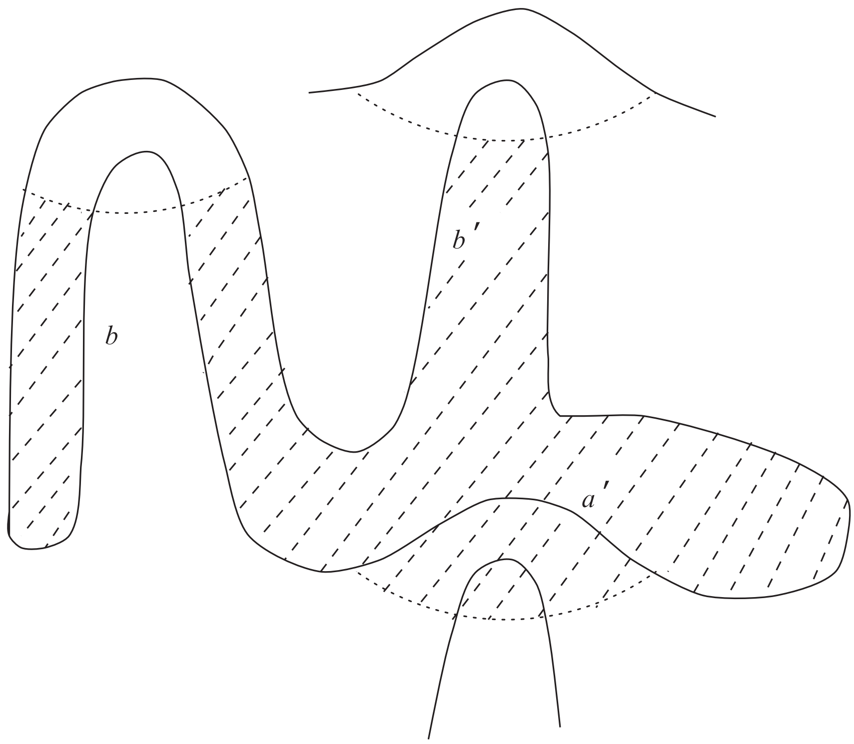

Fig. 2.4.9.

[8] The existence of such a homotopy is easy to prove. We would like to remark however that, in our constructions, the sides a may in any case be taken to lie in a suitably restricted class of curves (e.g. piecewise geodesic), and this makes the proof even simpler.

But the homotopy $A(\cdot\,,\cdot)$ never passes through a vertex. For suppose for an $s \in [0,1]$ that the vertex p is an interior point of the curve $A(\cdot\,,s)$. Then g would map both the end points of a (and thus also of $A(\cdot\,,s)$) as well as p into ∂F_s. But this is excluded by Lemma 2.4.6.
After performing the above homotopies for all pairs of equivalent sides (a,a'), we end up with a fundamental polygon with all the desired properties. This concludes the proof of Theorem 2.4.2. □

Finally, we wish to discuss briefly the structure of the fundamental group of a surface H/Γ.

Theorem 2.4.3 *Let H/Γ be a compact Riemann surface of genus $p\,(>1)$. Then the fundamental group $\pi_1(H/\Gamma,p_0)$ has $2p$ generators $a_1,b_1,a_2,\cdots,a_p,b_p$ with the single defining relation*

$$a_1b_1a_1^{-1}b_1^{-1}a_2b_2\cdots a_pb_pa_p^{-1}b_p^{-1}=1.$$

Proof. We represent H/Γ by the fundamental domain given in Theorem 2.4.2. Let p_0 be a vertex. Suppose e.g. that $g \in \Gamma$ carries the side a_1 to a_1'. Then, since $F \cap gF = \emptyset$, and g preserves orientations, $g(a_1)$ is a_1' described in the opposite direction, i.e. $a_1' = a_1^{-1}$ in H/Γ, and similarly for the other sides. ¿From Corollary 2.4.1 and Theorem 1.3.4, it follows that $a_1,b_1,\cdots,a_p,b_p$ generate $\pi_1(H/\Gamma,p_0)$.
By Theorem 1.3.2, a path is trivial in $\pi_1(H/\Gamma,p_0)$ if and only if its lift to H is closed. It follows that

$$a_1b_1a_1^{-1}\cdots a_pb_pa_p^{-1}b_p^{-1}=1$$

is the only relation among the given generators. That this is indeed a relation is clear. We show that there are no other ones (apart from trivial ones like $a_1a_1^{-1}=1$).
This is not hard to see. Let

$$c_1\cdots c_k=1$$

be any such relation. It is then represented by a closed loop based at p_0. Since each c_j, $j=1,\ldots,k$, is equivalent to a side of our fundamental domain, the loop is disjoint to the interiors of all translates of this fundamental domain. We claim that the loop is a multiple of

$$a_1b_1\cdots a_p^{-1}b_p^{-1}.$$

By what we have just said, it encloses a certain number of fundamental domains, and we shall see the claim by induction on this number. Let F be any such domain whose boundary contains part of the loop. Let c be such a boundary geodesic forming part of the loop. Replacing c by the remainder of this boundary, traversed in the opposite direction, yields a homotopic loop as the boundary represents the trivial loop $a_1b_1\cdots a_p^{-1}b_p^{-1}$. We observe that we can always choose F in such a way that this replacement decreases the number of enclosed fundamental domains by one. This completes the induction step and the proof of the claim. □

Corollary 2.4.2 *Every compact Riemann surface of the form H/Γ has a non-abelian fundamental group.* □

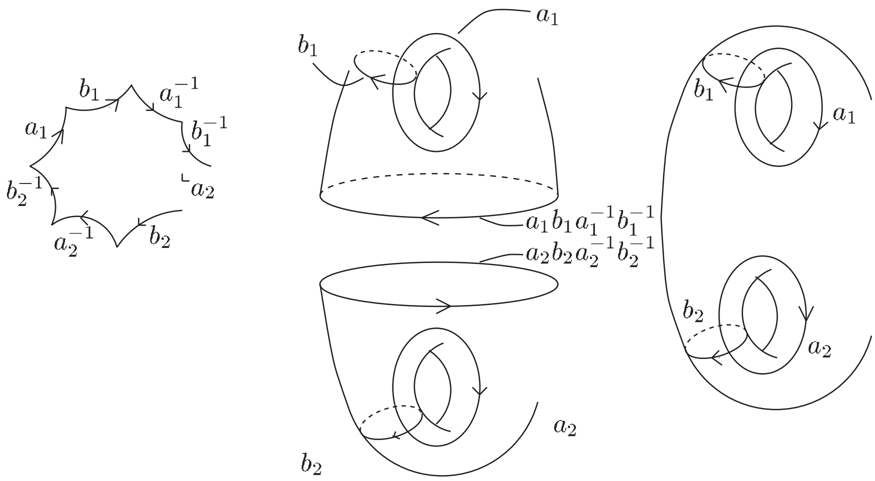

Fig. 2.4.10.

Exercises for § 2.4

1) Let H/Γ be a compact Riemann surface. Show that each nontrivial abelian subgroup of Γ is infinite cyclic.
2) Provide the details of the construction of a metric fundamental polygon for a group of Euclidean motions.

2.4.A The Topological Classification of Compact Riemann Surfaces

We start with

Definition 2.4.A.1 A differentiable manifold M is called orientable if it possesses an atlas whose chart transitions all have positive functional determinant. An orientation of M consists in the choice of such an atlas.

Corollary 2.4.A.1 *Any Riemann surface is orientable, and a conformal atlas provides an orientation.*

Proof. All transition maps of a conformal atlas are holomorphic and therefore have positive functional determinant. □

In this section, we shall classify the possible topological types of two-dimensional differentiable, orientable, triangulated, compact surfaces S. By Theorem 2.3.A.1 and Corollary 2.4.A.1, we shall therefore obtain a topological classification of compact Riemann surfaces.
Let $(T_i)_{i=1,\ldots,f}$ be a triangulation of S as in Def. 2.3.A.1. We choose an orientation of S. That orientation then determines an orientation of each triangle T_i, i.e. an ordering of the vertices and we then orient ∂T_i in the manner induced by the ordering of the vertices. We note that if ℓ is a common edge of two adjacent triangle T_i and T_j, then the orientations of ℓ induced by T_i and T_j, resp., are opposite to each other.
Let

$$\varphi_1 : \Delta_1 \to T_1$$

be a homeomorphism of a Euclidean triangle $\Delta_1 \subset \mathbb{R}^2$ onto the triangle T_1 as in Def. 2.3.A.1. We now number the triangles $T_2, \ldots, T_n$ so that T_2 has an edge in common with T_1. We then choose a triangle $\Delta_2 \subset \mathbb{R}^2$ that has an edge in common with Δ_1 so that $\Delta_1 \cup \Delta_2$ is a convex quadrilateral and a homeomorphism

$$\varphi_2 : \Delta_2 \to T_2$$

satisfying the requirements of Def. 2.4.A.1.
Renumbering again, T_3 has an edge in common with either T_1 or T_2, and we choose an Euclidean triangle Δ_3 so that $\Delta_1 \cup \Delta_2 \cup \Delta_3$ forms a regular convex pentagon and a homeomorphism

$$\varphi_3 : \Delta_3 \to T_3$$

as before. We iterate this process in such a manner that each new triangle Δ_j shares an edge with precisely one of the preceding ones and is disjoint to all the other ones. We obtain a regular convex polygon Π. The orientations of the triangles Δ_j induced by the homeomorphism $\varphi_j^{-1} : T_j \to \Delta_j$ in term induce an orientation of $\partial\Pi$. This orientation will be called positive. Points in $\partial\Pi$ that correspond to the same point in S will be called equivalent. Π has $f + 2$ edges, where f is the number of triangles T_i.
Since each edge of Π belongs to precisely two of the T_i, precisely two edges of Π correspond to the same edge of the triangulation. Let a be an edge of Π. The induced orientation of a allows us to distinguish between an initial point p and a terminal point q of a. The other edge of Π equivalent to a then has initial point q and terminal point p because of our convention on orienting the edges of Π. Therefore, we denote that edge by a^{-1}.
With this convention, the edges of Π are now labeled $a, b, c, \ldots, a^{-1}, b^{-1}, c^{-1}, \ldots$. By writing down these letters in the order in which the corresponding edges of $\partial\Pi$ are traversed, we obtain the so-called symbol of Π (cf. Thm. 2.4.3.).
The following process will repeatedly be applied below:
We dissect Π by an interior straight line connecting two edges into two subgroups Π_1, Π_2, and we glue Π_1 and Π_2 along a pair of edges a and a^{-1} by

identifying equivalent points. We thus obtain a new polygon Π' that again constitutes a model for S. After possibly applying a homeomorphism, we may again assume that Π' is convex.
This process will now be used in order to transform the symbol of Π into a particularly simple form as in Thm. 2.4.2. In fact what follows will essentially be the construction of steps 1), 2), 3), 4), 5) of the proof of that theorem.

1) (Corresponding to step 2 of the proof of Thm. 2.4.2)

In case the symbol of Π contains the sequence aa^{-1} plus some other letters, these edges a and a^{-1} are now eliminated by identifying, i.e. glueing equivalent points on them. This step is repeated until either no such sequence aa^{-1} occurs any more, or the entire symbol is given by aa^{-1}. In the latter case, we have reached the desired form already.

2) (Corresponding to step 1 of the proof of Thm. 2.4.2)

(Construction of a polygon with finitely many sides, all of whose vertices are equivalent).

Let p be a vertex of Π. If Π possesses another vertex not equivalent to p, then it has an edge a with initial point p and terminal point $q \neq p$. We join p to the terminal vertex r of the edge b with initial point q, by a line d in Π. Note that by 1), $b \neq a^{-1}$. We obtain a triangle T with edges a, b, d. This triangle is now cut off along d, and its edge b is then glued to the edge b^{-1} of Π. The resulting Π' then has one more vertex equivalent to p, while the number of vertices equivalent to q is decreased by one. After finitely many repetitions, we obtain a polygon with the desired property.

3) (Corresponding to step 4 of the proof of Thm. 2.4.2)

Subsequently, we shall need the following type of modification of Π that generalizes the one employed in 1):
Let a, a^{-1} be edges of Π, and use an interior line d of Π connecting two of the vertices, in order to dissect Π into two parts Π_1, Π_2, with $a \subset \Pi_1,\ a^{-1} \subset \Pi_2$. We glue Π_1 and Π_2 along the edges a, a^{-1}.

4) (Corresponding to step 5 of the proof of Thm. 2.4.2)

We first label the edges of Π in such a manner that c^{-1} is always to the right of c. Let c_1 be an edge for which the number of edges between c_1 and c_1^{-1} is minimal.
By 2), this number is positive. Thus, the arrangement looks like

$$c_1 c_2 \cdots c_1^{-1} \cdots c_2^{-1}.$$

If there are no intermediate edges in the places denoted by dots, i.e. if we already have the sequence $c_1 c_2 c_1^{-1} c_2^{-1}$ in our symbol, we look among the remaining edges for an edge c with minimal number of intermediate edges. When we arrive at the above situation with intermediate edges present, we connect the terminal point of c_1 with the initial point of c_1^{-1} by a line b_1 in F and apply the modification of 3) to the pair c_1, c_1^{-1} and the diagonal b_1.

The resulting symbol is

$$c_1 b_1 c_1^{-1} \cdots b_1^{-1}.$$

If there are intermediate edges between c_1^{-1} and b_1^{-1}, we join the terminal point of b_1 with the initial point of b_1^{-1} by a line a_1 in Π and apply the modification of 4) to c_1, c_1^{-1} and the diagonal a_1^{-1}. We obtain the symbol

$$a_1 b_1 a_1^{-1} b_1^{-1} \cdots .$$

Repeating this process finitely many times, we conclude

Theorem 2.4.A.1 *The symbol of a polygon representing the differentiable, orientable, compact, triangulated surface S may be brought into either the form*

(i)
$$aa^{-1}$$

or

(ii)
$$a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \cdots a_p b_p a_p^{-1} b_p^{-1}.$$

In case (ii), all vertices are equivalent.
In particular, the number of edges is either 2 *or a multiple of* 4.

Definition 2.4.A.2 The genus of S as in Theorem 2.4.A.1 is 0 in case (i), p in case (ii), and the Euler characteristic is

$$\chi := 2 - 2p.$$

Corollary 2.4.A.2 *Two differentiable, orientable, compact, triangulated surfaces are homeomorphic iff they have the same genus.*

Proof. A homeomorphism between two surfaces with the same symbol is produced by a vertex preserving homeomorphism between the corresponding polygons. That surfaces of different genus are not homeomorphic follows for example from Thm. 2.4.3, noting that homeomorphic surfaces must have isomorphic fundamental groups by Lemma 1.2.3. □

2.5 The Theorems of Gauss-Bonnet and Riemann-Hurwitz

We now proceed to the Gauss-Bonnet formula for hyperbolic triangles.

Theorem 2.5.1 *Let B be a hyperbolic triangle in H (so that the sides of B are geodesic arcs) with interior angles $\alpha_1, \alpha_2, \alpha_3$. Let K be the curvature of the hyperbolic metric (thus $K \equiv -1$). Then*

$$\int_B K \frac{1}{y^2} \frac{\mathrm{i}}{2} \, \mathrm{d}z \, \mathrm{d}\overline{z} = \sum_{i=1}^{3} \alpha_i - \pi. \tag{2.5.1}$$

Proof. Quite generally, we have

$$\begin{aligned}\int_B K \cdot \lambda^2 \frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} &= -\int_B \frac{4\partial^2}{\partial z \partial \overline{z}} \log \lambda\, \frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \\ &= -\int_{\partial B} \frac{\partial}{\partial n} \log \lambda\ |\mathrm{d}z|\end{aligned}$$

($\frac{\partial}{\partial n}$ denotes differentiation in the direction of the outward normal of ∂B), hence in our situation

$$\int_B K \frac{1}{y^2} \frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} = -\int_{\partial B} \frac{\partial}{\partial n} \log \frac{1}{y}\ |\mathrm{d}z|.$$

Now, ∂B consists of three geodesic arcs a_1, a_2, a_3. Thus each a_i is either a Euclidean line segment perpendicular to the real axis, or an arc of a Euclidean circle with centre on the real axis. In the former case, $\frac{\partial}{\partial n} \log y = 0$ on a_i; in the latter case, we can write $y = r \sin\varphi$ in polar coordinates, so that

$$-\int_{a_i} \frac{\partial}{\partial n} \log \frac{1}{y}\ |\mathrm{d}z| = -\int_{\varphi_1}^{\varphi_2} \frac{\partial}{\partial r} \log \frac{1}{r\sin\varphi}\ r\ \mathrm{d}\varphi = \int_{\varphi_1}^{\varphi_2} \frac{1}{r}\ r\ \mathrm{d}\varphi = \varphi_2 - \varphi_1,$$

where the angles φ_1 and φ_2 correspond of course to the end-points of a_i.
Now an elementary geometric argument keeping in mind the correct orientations of the sides of B yields (2.5.1).
(By a hyperbolic isometry, we can always assume in the above that one of the sides is an interval on the imaginary axis, so that $\frac{\partial}{\partial n} \log y = 0$ on this side.) □

Since $K \equiv -1$, we get:

Corollary 2.5.1 *hyperbolic area of the hyperbolic triangle with interior angles* $\alpha_1, \alpha_2, \alpha_3$, *we have the formula*

$$\mathrm{Area}(B) = \pi - \sum_{i=1}^{3} \alpha_i. \tag{2.5.2}$$

Although K is constant in our case, there are various reasons for giving the formula (2.5.1) the more prominent place. One reason is of course that our proof of (2.5.2) uses (2.5.1). But the most important reason is rather that (2.5.1) is valid for quite arbitrary metrics. Using the differential equation for the geodesics with respect to an arbitrary metric (see § 2.3.A), one can prove the general statement. We shall anyway prove the general Gauss-Bonnet formula for compact surfaces without boundary later on (Corollary 2.5.6). The Euclidean case is trivial, and the case of the spherical metric can be treated exactly in the same way as the hyperbolic case by means of the formulae given above. For this reason, we shall make use of the Gauss-Bonnet formula for geodesic triangles in all the three geometries.

Corollary 2.5.2 *Let P be a geodesic polygon in H with k vertices, of interior angles $\alpha_1, \cdots \alpha_k$. Then*

$$\int_P K \frac{1}{y^2} \frac{\mathrm{i}}{2} \, \mathrm{d}z \, \mathrm{d}\overline{z} = \sum_{i=1}^{k} \alpha_i + (2-k)\,\pi \tag{2.5.3}$$

and

$$\mathrm{Area}(P) = (k-2)\,\pi - \sum \alpha_i. \tag{2.5.4}$$

Proof. The proof is by dividing P into geodesic triangles: in every geodesic polygon which has more than 3 sides, we can find two vertices which can be joined by a geodesic running in the interior of P. In this way, P will be divided into two sub-polygons with fewer sides. We repeat this process till P has been decomposed into $(k-2)$ triangles. The corollary now follows from the corresponding assertions for triangles, since the sum of the interior angles of the triangles at a vertex of P is precisely the interior angle of P at that vertex. □

Corollary 2.5.3 *Suppose $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ operates without fixed points and properly discontinuously on H and that H/Γ is compact. Then*

$$\int_{H/\Gamma} K \frac{1}{y^2} \frac{\mathrm{i}}{2} \, \mathrm{d}z \, \mathrm{d}\overline{z} = 2\pi\,(2-2p) \tag{2.5.5}$$

and

$$\mathrm{Area}(H/\Gamma) = 2\pi\,(2p-2) \tag{2.5.6}$$

where $4p$ is the number of sides of the fundamental polygon for Γ constructed in Theorem 2.4.2.

Proof. Since all the vertices of the fundamental polygon for Γ constructed in Theorem 2.4.2 are equivalent under Γ, it follows that the sum of the interior angles of the polygon is exactly 2π. Indeed, if we draw a circle around any of the vertices of the polygon, we see that each point of the circle is equivalent to precisely one interior point or to some boundary point of the polygon. But the second alternative occurs only for finitely many points of the circle, so we conclude that the sum of the interior angles is indeed 2π.
The assertions of the corollary now follow from Corollary 2.5.2. □

Definition 2.5.1 The p of Theorem 2.4.2 is called the genus, and $\chi := 2-2p$ the Euler characteristic of the Riemann surface H/Γ.
(Note that this coincides with the purely topological Def. 2.4.A.2.)

It follows from Corollary 2.5.3 that p and χ are well-defined, since they are invariants of the surface H/Γ and do not depend on the fundamental domain of Theorem 2.4.2.
(In particular, every fundamental domain with the properties established in

Theorem 2.4.2 must have the same number of sides.) On the other hand, it follows from Theorem 2.4.2 that surfaces with the same genus or Euler characteristic are mutually homeomorphic, since one can directly produce a homeomorphism between the fundamental polygons given by the theorem which respects the boundary identifications.
From Corollary 2.5.3 we also get:

Corollary 2.5.4 *For a Riemann surface of the form $H/\Gamma, p>1$, i.e. $\chi<0$.*

As a consequence of the uniformization theorem (Thm. 4.4.1 below), in fact every compact Riemann surface is conformally equivalent to S^2, or a torus $\mathbb{C}/M$ (M a module over $\mathbb{Z}$ of rank two, cf. § 2.7) or a surface H/Γ, since the universal covering is S^2, $\mathbb{C}$ or H. But S^2 admits no non-trivial quotients; and the compact quotients of $\mathbb{C}$ are the tori. We shall have $p=1$ for the torus and $p=0$ for S^2 (so that $\chi=0$ and $\chi=2$ respectively) if we define the genus and Euler characteristic for these surfaces in the analogous way. Thus, the genus of a Riemann surface already determines the conformal type of the universal covering. Further, we have thus obtained a complete list of the topological types of compact Riemann surfaces, since the topological type is already determined by the genus. We recall that in the appendix to Sec. 2.4, we obtained the topological classification directly by topological methods. For this, it was necessary to triangulate the surface, i.e. decompose it into triangles (see 2.3.A), and then dissect the surface to get an abstract polygon from which the surface could be reconstructed by boundary identifications. This polygon then was brought to a normal form as in Theorem 2.4.2 - for this, the steps 1), 2), 4) and 5) of the proof of Theorem 2.4.2 sufficed.

Corollary 2.5.5 *Suppose given a decomposition of the Riemann surface Σ into polygons (i.e. Σ is represented as the union of closed polygons with disjoint interiors and boundaries consisting of finitely many geodesic arcs) and suppose the number of polygons occuring is f, the number of sides k, and the number of vertices e. Then*

$$\chi(\Sigma)=f-k+e. \tag{2.5.7}$$

Proof. This again follows from the Gauss-Bonnet formula. If we sum (2.5.3) over all the polygons of the decomposition, then the first summand on the right-hand side contributes the sum of all the interior angles, i.e. $2\pi e$; the summand 2π occurs f times, i.e. contributes $2\pi f$, and each edge occurs twice, hence the contribution from the edges in $-2\pi k$. By (2.5.5), the left side of the sum is $2\pi\chi$. (The argument is the same if the universal covering is S^2 or $\mathbb{C}$.) □

The relation (2.5.7) is valid even if the sides of the polygons of the decomposition are not necessarily geodesics; it is easy and elementary to reduce the general case to the case considered above. But the general case also follows from the general Gauss-Bonnet formula, in which there will be additional boundary integrals in general; however, these additional terms cancel on addition since every edge appears twice in the sum with opposite orientations. Finally, the topological invariance of $f-k+e$ can also be proved directly by a combinatorial argument.

The principle of the proof presented above, which consists in representing a topological quantity as an integral of an analytically defined expression (here the curvature), so that the invariant nature of the topological quantity on the one hand and the integrality of the integral on the other follow simultaneously, is of considerable importance in mathematics. A higher dimensional generalisation leads to Chern classes. And in the principal index theorems of mathematics (e.g. that of Atiyah and Singer) one proves likewise the equality of two expressions, one of which is defined topologically and the other analytically.

We shall now prove the **Gauss-Bonnet theorem** for compact surfaces without boundary with respect to an arbitrary metric.

Corollary 2.5.6 *Let Σ be a compact Riemann surface without boundary, of genus p* [9]*, with a metric $\rho^2(z)\mathrm{d}z\mathrm{d}\overline{z}$ of curvature K_ρ. Then*

$$\int_\Sigma K_\rho \rho^2(z) \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z} = 2\pi\,(2-2p).$$

Proof. We put another metric $\lambda^2\mathrm{d}z\mathrm{d}\overline{z}$ on Σ, of constant curvature K. For this second metric, we know by Corollary 2.5.3 that

$$\int_\Sigma K\lambda^2 \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z} = 2\pi\,(2-2p).$$

Now the quotient $\rho^2(z)/\lambda^2(z)$ is invariant under coordinate transformations, i.e. behaves like a function, since by Def. 2.3.1 both $\rho^2(z)$ and $\lambda^2(z)$ get multiplied by the same factor. We compute now

$$\begin{aligned}
&\int K\,\lambda^2 \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z} - \int K_\rho \rho^2 \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z}v \\
&= -4 \int \frac{\partial^2}{\partial z \partial \overline{z}} \log\lambda\, \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z} + 4 \int \frac{\partial^2}{\partial z \partial \overline{z}} \log\rho\, \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z} \\
&= 4 \int \frac{\partial^2}{\partial z \partial \overline{z}} \log\frac{\rho}{\lambda}\, \frac{\mathrm{i}}{2}\, \mathrm{d}z\, \mathrm{d}\overline{z}
\end{aligned}$$

which vanishes by Gauss' Divergence Theorem (note that ρ and λ are everywhere positive), so that our assertion follows. □

[9] Since we shall prove the Uniformization Theorem only in § 4.4 below, we should strictly assume at this stage that Σ is diffeomorphic to S^2 or a torus, or is a quotient of H. It follows from the Uniformization Theorem that this assumption is automatically satisfied.

It is worthwhile to reflect briefly once again on the above statement and its proof. We consider an arbitrary metric on a compact surface, and construct from it a quantity, namely the curvature integral, which now no longer depends on the particular metric, but is determined by the topological type of the surface. Thus, to compute the Euler characteristic of the surface, we may choose an arbitrary metric.
For the proof of Corollary 2.5.6, we had only to observe that, for any two metrics, the integrands differ only by a divergence expression, which integrates to zero. In the terminology introduced later on in Chapter 5, if K_ϱ is the curvature of the metric $\varrho^2 \, \mathrm{d}z\mathrm{d}\overline{z}$,

$$K_\rho \rho^2 \, \mathrm{d}z \wedge \mathrm{d}\overline{z} = -4 \frac{\partial^2}{\partial z \partial \overline{z}} \log \rho \, \mathrm{d}z \wedge \mathrm{d}\overline{z}$$

defines a cohomology class of Σ which does not depend on the special choice of ϱ (namely the so-called first Chern class of Σ up to a factor), cf. § 5.6.

We next consider a (non-constant) holomorphic map $f : \Sigma_1 \to \Sigma_2$ between compact Riemann surfaces.
According to the local representation theorems for holomorphic functions, we can find for each $p \in \Sigma_1$ local charts around p and $f(p)$ in which (assuming without loss of generality that $p = 0 = f(p)$) f can be written as

$$f = z^n. \tag{2.5.8}$$

(First, we can write $\zeta = f(w) = \sum_{k \geq n} a_k w^k$ with $n > 0$ and $a_n \neq 0$. Since a non-vanishing function has a logarithm locally, we have $\zeta = w^n g(w)^n$, with g holomorphic and $g(0) \neq 0$. Set $z = wg(w)$.)

Definition 2.5.2 p is called a branch point or ramification point of f if $n > 1$ in (2.5.8). We call $n - 1$ the order of ramification of f at p (in symbols: $v_f(p) := n - 1$).

Since Σ_1 is compact, there are only finitely many points of ramification.

Lemma 2.5.1 *Let $f : \Sigma_1 \to \Sigma_2$ be a non-constant holomorphic map of compact Riemann surfaces. Then there exists $m \in \mathbb{N}$ such that*

$$\sum_{p \in f^{-1}(q)} (\, v_f(p) + 1 \,) = m$$

for all $q \in \Sigma_2$. Thus f takes every value in Σ_2 precisely m times, multiplicities being taken into account.

Definition 2.5.3 We call m the (mapping) degree of f. If f is constant, we set $m = 0$.

The *proof* of Lemma 2.5.1 follows by a simple open-and-closed argument. □

We now prove the **Riemann-Hurwitz formula**:

Theorem 2.5.2 *Let $f : \Sigma_1 \to \Sigma_2$ be a non-constant holomorphic map of degree m between compact Riemann surfaces of genera g_1 and g_2 respectively. Let $v_f := \sum_{p\in\Sigma_1} v_f(p)$ be the total order of ramification of f. Then*

$$2 - 2g_1 = m\,(2 - 2g_2) - v_f. \tag{2.5.9}$$

Proof. Let $\lambda^2\, \mathrm{d}w\mathrm{d}\overline{w}$ be a metric on Σ_2. Then

$$\lambda^2(w(z))\frac{\partial w}{\partial z}\frac{\partial \overline{w}}{\partial \overline{z}}\ \mathrm{d}z\ \mathrm{d}\overline{z}$$

(where $f = w(z)$ in local coordinates) defines a metric on Σ_1 outside the ramification points of f, and f is a local isometry with respect to these two metrics).
Let $p_1, \cdots, p_k$ be the ramification points. Suppose f is given in a local chart near p_j by $w = z^{v_f}$, and let $B_j(r)$ be a disc of radius r around p_j in this chart.
Since f is a local isometry, we will have, as $r \to 0$,

$$\begin{aligned} &-\frac{1}{2\pi}\int_{\Sigma_1\setminus\bigcup_{j=1}^k B_j(r)} 4\frac{\partial^2}{\partial z\partial\overline{z}}\log\left(\lambda\,(w_z\overline{w}_{\overline{z}})^{\frac{1}{2}}\right)\frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \\ &= -\frac{m}{2\pi}\int_{\Sigma_1\setminus\bigcup B_j(r)} 4\frac{\partial^2}{\partial w\partial\overline{w}}(\log\lambda)\,\frac{\mathrm{i}}{2}\ \mathrm{d}w\ \mathrm{d}\overline{w} \\ &\to m\,(2-2g_2) \qquad \text{by Cor. 2.5.6.} \end{aligned}$$

On the other hand, as $r \to 0$,

$$-\frac{1}{2\pi}\int_{\Sigma_1\setminus\bigcup B_j(r)} 4\frac{\partial^2}{\partial z\partial\overline{z}}(\log\lambda)\,\frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \to 2-2g_1$$

$(\lambda^2(w(z))\,\mathrm{d}z\mathrm{d}\overline{z}$ transforms like a metric except for a factor which is the square of the absolute value of a non-vanishing holomorphic function; when we form $\frac{\partial^2}{\partial z\partial\overline{z}}\log\lambda$, this factor plays no role, hence Cor. 2.5.6 provides the value of the limit of the integral).

$$\begin{aligned} &-\frac{1}{2\pi}\int_{\Sigma_1\setminus\bigcup B_j(r)} 4\frac{\partial^2}{\partial z\partial\overline{z}}\left(\log w_z^{\frac{1}{2}}\right)\frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \\ &= \frac{1}{2\pi}\sum\int_{\partial B_j(r)}\frac{\partial}{\partial r}\log w_z^{\frac{1}{2}}\ r\ \mathrm{d}\varphi \\ &= \frac{1}{2}\sum_j (n_j-1), \text{ since } w = z^{n_j} \text{ in } B_j(r), \end{aligned}$$

and similarly for the integral involving $\overline{w}_{\overline{z}}^{\frac{1}{2}}$. These formulæ imply (2.5.9). $\square$

We collect some consequences of (2.5.9) in the following:

Corollary 2.5.7
(i) v_f *is always even;*
(ii) $g_1 \geq g_2$*;*
(iii) $g_2 = 0,\ f$ *unramified* $\Rightarrow g_1 = 0, \quad m = 1$*;*
(iv) $g_2 = 1,\ f$ *unramified* $\Rightarrow g_1 = 1 \quad$ *(m arbitrary);*
(v) $g_2 > 1,\ f$ *unramified* $\Rightarrow g_1 = g_2$ and $m = 1$ or $g_1 > g_2,\ m > 1$;
(vi) $g_2 = g_1 = 1 \Rightarrow f$ *unramified;*
(vii) $g_2 = g_1 > 1 \Rightarrow m = 1,\ f$ *unramified.* □

Exercises for § 2.5

1) State and prove the Gauss-Bonnet formula for spherical polygons.
2) We have defined the degree of a holomorphic map between compact Riemann surfaces in Def. 2.5.3. However, a degree can also be defined for a continuous map between compact surfaces, and such a definition can be found in most textbooks on algebraic topology. For a differentiable map $g : \Sigma_1 \to \Sigma_2$ between compact Riemann surfaces, the degree $d(g)$ is characterized by the following property:
   If $\lambda^2(g)\mathrm{d}g\mathrm{d}\overline{g}$ is a metric on Σ_2, and if $\varphi : \Sigma_2 \to \mathbb{R}$ is integrable, then
$$\int_{\Sigma_1} \varphi(g(z))\ (g_z\overline{g}_{\overline{z}} - \overline{g}_z g_{\overline{z}})\ \lambda^2(g(z))\,\frac{\mathrm{i}}{2}\ \mathrm{d}z\ \mathrm{d}\overline{z}$$
$$= d(g) \int_{\Sigma_2} \varphi(g)\,\lambda^2(g)\,\frac{\mathrm{i}}{2}\ \mathrm{d}g\ \mathrm{d}\overline{g}$$
   Show that the degree of a holomorphic map as defined in Def. 2.5.3 satisfies this property.

2.6 A General Schwarz Lemma

We begin with the Ahlfors-Schwarz lemma:

Theorem 2.6.1 *hyperbolic metric*

$$\lambda^2(z)\ \mathrm{d}z\mathrm{d}\overline{z} := \frac{4}{(1-|z|^2)^2}\ \mathrm{d}z\ \mathrm{d}\overline{z} \qquad \text{(cf. Lemma 2.3.6).}$$

Let Σ *be a Riemann surface with a metric*

$$\rho^2(w)\ \mathrm{d}w\ \mathrm{d}\overline{w}$$

whose curvature K *satisfies*

$$K \leq -\kappa < 0 \tag{2.6.1}$$

(for some positive constant κ). Then, for any holomorphic map $f : D \to \Sigma$, we have

$$\rho^2(f(z))\, f_z\overline{f}_{\overline{z}} \le \frac{1}{\kappa}\lambda^2(z) \qquad \left(f_z := \frac{\partial f}{\partial z} \text{ etc.}\right). \tag{2.6.2}$$

Proof. We recall the curvature formulæ

$$-\frac{4}{\lambda^2}\frac{\partial^2}{\partial z\partial\overline{z}}\log\lambda = -1 \tag{2.6.3}$$

and

$$-\frac{4}{\rho^2(f(z))f_z\overline{f}_{\overline{z}}}\frac{\partial^2}{\partial z\partial\overline{z}}\log\left(\rho^2(f(z))f_z\overline{f}_{\overline{z}}\right)^{\frac{1}{2}} \le -\kappa \tag{2.6.4}$$

at all points where $f_z \neq 0$, by Lemma 2.3.7 and (2.6.1). We put

$$u := \frac{1}{2}\log\left(\rho^2(f(z))f_z\,\overline{f}_{\overline{z}}\right)$$

so that

$$4\frac{\partial^2}{\partial z\partial\overline{z}}u \ge \kappa e^{2u} \tag{2.6.5}$$

wherever u is defined, i.e. $f_z \neq 0$. For any $0 < R < 1$, we also put

$$v_R(z) := \log\frac{2R}{\kappa^{\frac{1}{2}}(R^2 - |z|^2)}, \qquad |z| < R$$

and compute

$$4\frac{\partial^2}{\partial z\partial\overline{z}}v_R = \kappa e^{2v_R}. \tag{2.6.6}$$

From (2.6.5) and (2.6.6) we get

$$4\frac{\partial^2}{\partial z\partial\overline{z}}(u - v_R) \ge \kappa\left(e^{2u} - e^{2v_R}\right) \tag{2.6.7}$$

wherever $f_z \neq 0$. Let

$$S := \{|z| < R : \; u(z) > v_R(z)\}.$$

Since u tends to $-\infty$ as f_z tends to zero, S cannot contain any zeros of $f(z)$. Hence (2.6.7) is valid in S. Therefore, by the maximum principle, $u - v_R$ cannot attain an interior maximum in S. But the boundary of S (in $\mathbb{C}$) is contained in $|z| < R$, since $v_R(z) \to -\infty$ as $|z| \to R$. Hence $u - v_R = 0$ on ∂S, by continuity. This means that the maximum of $u - v_R$, which has to be attained in ∂S, is zero, i.e. that S is empty.
We conclude:

$$u(z) \le v_R(z), \; |z| < R,$$

and letting R tend to 1, we get

$$u(z) \le \log\frac{2}{\kappa^{\frac{1}{2}}(1 - |z|^2)}$$

which is equivalent to (2.6.2). □

Theorem 2.6.1, of which Theorems 2.3.1 and 2.3.2 are special cases, shows the importance of negatively curved metrics on Riemann surfaces. In this section, we shall exploit the strong connection between the conformal structure of a Riemann surface and the curvature properties of the metrics which can be put on it. Often one can construct a metric with suitable properties on a Riemann surface and deduce consequences for the holomorphic structure of the surface. Such techniques are of even greater importance in higher-dimensional complex geometry. And, although it is not necessary for our present applications, we also want to introduce a concept that abstracts the assertion of the Ahlfors-Schwarz lemma, because it again illustrates ideas that are useful in the higher dimensional case.
Thus, let Σ be a Riemann surface. For any $p, q \in \Sigma$, we define

$$d_H(p,q) := \inf \Bigg\{ \sum_{i=1}^{n} d(z_i, w_i) : \ n \in \mathbb{N},\ p_0, p_1, \cdots, p_n \in \Sigma,\ p_0 = p,\ p_n = q, \\ f_i : D \to \Sigma \text{ holomorphic}, \quad f_i(z_i) = p_{i-1},\ f_i(w_i) = p_i \Bigg\}.$$

Here, $d(\cdot\,,\cdot)$ is the distance on D defined by the hyperbolic metric.
It is easily seen that d_H satisfies the triangle inequality

$$d_H(p,q) \le d_H(p,r) + d_H(r,q), \qquad p,q,r \in \Sigma,$$

and is symmetric and non-negative.

Definition 2.6.1 Σ is said to be hyperbolic if d_H defines a distance function on Σ, i.e.

$$d_H(p,q) > 0 \qquad \text{if } p \neq q.$$

Important Note. This usage of the term "hyperbolic"[10] is obviously different from its usage in other parts of this book. It has been adopted here because the same definition is used in the higher-dimensional case. This usage is restricted to the present section; in all other parts of the book "hyperbolic" has a different meaning.

Remark. d_H is continuous in q for fixed p and if Σ is hyperbolic the topology on Σ defined by the distance function coincides with the original one. If d_H is complete, then bounded sets are relatively compact. We leave it as an exercise to the reader to check these assertions.

Corollary 2.6.1 *Suppose Σ carries a metric $\rho^2(w)\,\mathrm{d}w\mathrm{d}\overline{w}$ with curvature K bounded above by a negative constant. Then Σ is hyperbolic (in the sense of Definition 2.6.1.).*

[10] In the literature, it is sometimes called "Kobayashi-hyperbolic".

Proof. Let $p, q \in \Sigma$, and let $f : D \to \Sigma$ be a holomorphic map with $f(z_1) = p$, $f(z_2) = q$ for some $z_1, z_2 \in D$. Let Γ be the geodesic arc in D joining z_1 to z_2. Then

$$\begin{aligned} d(z_1, z_2) &= \int_\gamma \lambda(z)\ |\mathrm{d}z| \qquad (\text{where } \lambda^2(z) = \frac{4}{(1-|z|^2)^2}) \\ &\geq C \int_\gamma \rho(f(z))|f_z|\ |\mathrm{d}z| \qquad \text{by (2.6.2)} \\ &= C \int_{f(\gamma)} \rho(w)\ |\mathrm{d}w| \\ &\geq C d_\rho(p, q) \end{aligned}$$

where $C > 0$ is a constant and d_ρ denotes the distance on Σ defined by the metric $\rho^2(w)\, \mathrm{d}w \mathrm{d}\overline{w}$. The corollary follows easily. □

From the proof of Corollary 2.6.1, we see that, under the assumptions of Theorem 2.6.1, any holomorphic map $f : D \to \Sigma$ is distance-decreasing (up to a fixed factor determined by the curvature of the metric on Σ). On the other hand, this is essentially the content of Definition 2.6.1.

Examples. 1) On the unit disc, d_H coincides with the distance function defined by the hyperbolic metric. This is again a consequence of the Schwarz lemma.
2) $\mathbb{C}$ is not hyperbolic. Namely, if $p, q \in \mathbb{C},\ p \neq q$, there exist holomorphic maps $f_n : D \to \mathbb{C}$ with $f_n(0) = p,\ f_n(\frac{1}{n}) = q$ $(n \in \mathbb{N})$. Hence $d_H(p, q) = 0$.
In view of Corollary 2.6.1, it follows that $\mathbb{C}$ cannot carry any metric with curvature bounded above by a negative constant. Thus the conformal structure puts restrictions on the possible metrics on a Riemann surface even in the non-compact case.

Lemma 2.6.1 *d_H is non-increasing under holomorphic maps: If $h : \Sigma_1 \to \Sigma_2$ is a holomorphic map, then*

$$d_H(h(p), h(q)) \leq d_H(p, q)$$

for all $p, q \in \Sigma_1$. In particular, d_H is invariant under biholomorphic maps:

$$d_H(h(p), h(q)) = d_H(p, q)$$

for all $p, q \in \Sigma_1$ if h is bijective and holomorphic.

Proof. If $f_i : D \to \Sigma_1$ is holomorphic with $f_i(z_i) = p_{i-1}$ and $f_i(w_i) = p_i$, then $h \circ f_i : D \to \Sigma$ is holomorphic with $h \circ f_i(z_i) = h(p_{i-1})$ and $h \circ f_i(w_i) = h(p_i)$.
The lemma follows easily from this. □

Lemma 2.6.2 *Let Σ be a Riemann surface and $\widetilde{\Sigma}$ its universal covering. Then Σ is hyperbolic if and only if $\widetilde{\Sigma}$ is.*

Proof. First, suppose Σ is hyperbolic. Let $\pi : \widetilde{\Sigma} \to \Sigma$ be the covering projection, and $p, q \in \widetilde{\Sigma}$, $p \neq q$. Then, by Lemma 2.6.1,

$$d_H(p,q) \geq d_H(\pi(p), \pi(q)) > 0 \qquad \text{if } \pi(p) \neq \pi(q). \tag{2.6.8}$$

To handle the case $\pi(p) = \pi(q)$, we make a geometric observation. Let f_i, z_i, w_i be as in the definition of $d_H(p,q)$, and c_i the geodesic in D from z_i to w_i. Then

$$\gamma := \bigcup_{i=1}^{n} f_i(c_i)$$

is a curve joining p to q, and

$$d_H(p,r) \leq \sum_{i=1}^{n} d(z_i, w_i)$$

for every $r \in \gamma$. Thus, if $d_H(p,q) = 0$ for $p \neq q$, we can find a sequence γ_ν, $\nu \in \mathbb{N}$, of such curves such that the sums of the lengths of the corresponding c_i tends to zero. And for every point r which is a limit point of points on the γ_ν, we would have

$$d_H(p,r) \leq d_H(p,q) = 0.$$

In particular, on every sufficiently small circle around p, there would be an r with $d_H(p,r) = 0$. But in our situation, in view of (2.6.9) and the fact that the fibres of π are discrete, this is impossible.
Now suppose conversely that $\widetilde{\Sigma}$ is hyperbolic. Let $p, q \in \Sigma$, $p \neq q$. Then, arguing as above, one shows that, for any $\widetilde{p} \in \pi^{-1}(p)$

$$\inf \left\{ d_H(\widetilde{p}, \widetilde{q}) : \ \widetilde{q} \in \pi^{-1}(q) \right\} > 0$$

using the fact that $\pi^{-1}(q)$ is a closed set containing $\widetilde{q}$. Moreover, this infimum is independent of the choice of $\widetilde{p} \in \pi^{-1}(p)$ since covering transformations act transitively on the fibres of π (Corollary 1.3.3), and are isometries with respect to d_H (Lemma 2.6.1) since they are biholomorphic (cf. the end of § 2.1). If now π_i, f_i, z_i, w_i are as in the definition of $d_H(p,q)$, we know by Theorem 1.3.1 that there exist holomorphic maps $g_i : D \to \widetilde{\Sigma}$ with $g_i(z_i) = g_{i-1}(w_{i-1})$ for $i > 1$ ($g_1(z_0) = \widetilde{p}_0 \in \pi^{-1}(p)$ arbitrary). Consequently

$$d_H(p,q) \geq \inf \left\{ d_H(\widetilde{p}, \widetilde{q}) : \ \widetilde{p} \in \pi^{-1}(p), \ \widetilde{q} \in \pi^{-1}(q) \right\}.$$

Combined with the earlier observations, this proves that Σ is hyperbolic. □

Theorem 2.6.2 *Let S, Σ be Riemann surfaces, and $z_0 \in S$. Assume that Σ is hyperbolic in the sense of Definition 2.6.1 and complete with respect to d_H. Then any bounded holomorphic map $f : S \backslash \{z_0\} \to \Sigma$ extends to a holomorphic map $\overline{f} : S \to \Sigma$.*

Proof. The problem is local near z_0, hence it suffices to consider the case when S is the unit disk D. Then f lifts to a holomorphic map $\overline{f} : D \to \widetilde{\Sigma}$ of the universal coverings of $D\backslash\{0\}$ and Σ (cf. § 2.3). By Lemma (2.6.2), $\widetilde{\Sigma}$ is also hyperbolic . As always, we equip D with its standard hyperbolic metric and induced distance d. Then, by Lemma (2.6.1),

$$d_H(\widetilde{f}(w_1), \widetilde{f}(w_2)) \leq d(w_1, w_2), \qquad w_1, w_2 \in D.$$

Hence we also have for f:

$$d_H(f(z_1), f(z_2)) \leq d(z_1, z_2), \qquad z_1, z_2 \in D\backslash\{0\}, \tag{2.6.9}$$

where d now denotes the distance on $D\backslash\{0\}$ induced by the hyperbolic metric

$$\frac{1}{|z|^2(\log |z|^2)^2} \, \mathrm{d}z \, \mathrm{d}\overline{z}$$

(cf. § 2.3).
Let now

$$S_\delta := \{|z| = \delta\}$$

for $0 < \delta < 1$. The length of S_δ in the hyperbolic metric of $D\backslash\{0\}$ tends to zero as δ tends to zero, hence the diameter of $f(S_\delta)$ with respect to d_H tends to zero by (2.6.10). Since f is bounded, and Σ is complete, there exists for every sequence $\delta_n \to 0$ a subsequence δ'_n such that $f(S_{\delta'_n})$ converges to a point in Σ. We must show that this limit point is independent of the choice of (δ_n) and (δ'_n).
Suppose this is not the case. Then we argue as follows. Let p_0 be the limit point for some sequence $f(S_{\delta_n})$. Choose a holomorphic coordinate $h : D \to \Sigma$ with $h(0) = p_0$, and choose $\varepsilon > 0$ so small that

$$\{\, p \in \Sigma : \ d_H(p, p_0) < 5\varepsilon \,\} \subset h(D).$$

Now choose $\delta_0 > 0$ such that

$$\operatorname{diam}(f(S_\delta)) < \varepsilon \tag{2.6.10}$$

for $0 < \delta \leq \delta_0$. Since we are assuming that the limit point of the $f(S_\delta)$ is not unique, we can find

$$0 < \delta_1 < \delta_2 < \delta_3 < \delta_0$$

such that, if

$$K_\eta := \{p \in \Sigma : \ d_H(p, p_0) < \eta\}\,,$$

then
(i) $f(S_{\delta_2}) \subset K_{2\varepsilon}$,
(ii) $f(S_\delta) \subset K_{3\varepsilon}$ for $\delta_1 < \delta < \delta_3$,
(iii) $f(S_{\delta_1})$ and $f(S_{\delta_3})$ are not contained in $K_{2\varepsilon}$.

We now identify D and $h(D)$ via h; in particular, we regard $f \mid \{\delta_1 \leq |z| \leq \delta_3\}$ as a holomorphic function. Choose a point $p_1 \in f(S_{\delta_2}) \subset K_{2\varepsilon}$. By (2.6.11) and (iii), p_1 does not lie on the curves $f(S_{\delta_1})$ and $f(S_{\delta_3})$. On the other hand, p_1 is attained by f at least in $\{\delta_1 \leq |z| \leq \delta_3\}$, namely on $|z| = \delta_2$. Hence

$$\int_{S_{\delta_3}} \frac{f'(z)}{f(z) - p_1}\,\mathrm{d}z - \int_{S_{\delta_1}} \frac{f'(z)}{f(z) - p_1}\,\mathrm{d}z \neq 0. \tag{2.6.11}$$

But $f(S_{\delta_1})$ and $f(S_{\delta_3})$ are contained in simply connected regions not containing p_1. Hence the integrand in (2.6.12) can be written as $\frac{\mathrm{d}}{\mathrm{d}z}\log(f(z) - p_1)$. Thus both integrals in (2.6.12) must vanish. This contradiction shows that the limit point of $f(S_\delta)$ as $\delta \to 0$ is unique. Hence f extends to a continuous map $\overline{f} : D \to \Sigma$. The proof can now be completed by an application of the removability of isolated singularities of bounded harmonic functions, which is recalled in the lemma below. □

Lemma 2.6.3 *Let $f : D\backslash\{0\} \to \mathbb{R}$ be a bounded harmonic function. Then f can be extended to a harmonic function on D.*

Proof. Let $D' = \{ z \in \mathbb{C} : |z| < \frac{1}{2}\}$ and let $h : D' \to \mathbb{R}$ be harmonic with boundary values $f_{|\partial D'}$ (the existence of h is guaranteed by the Poisson integral formula). For $\lambda \in \mathbb{R}$, let

$$h_\lambda(z) = h(z) + \lambda \log 2|z|.$$

Then h_λ is a harmonic function on $D'\backslash\{0\}$, with $h_\lambda\ |_{\partial D'} = f_{|\partial D'}$; also for $\lambda < 0$ (resp. $\lambda > 0$), $h_\lambda(z) \to +\infty$ (resp. $-\infty$) as $z \to 0$. Since f is bounded, it follows that $h_\lambda - f$, which is a harmonic function on $D'\backslash\{0\}$, has boundary values 0 on $\partial D'$ and $+\infty$ at 0, for all $\lambda < 0$. Hence $h_\lambda - f \geq 0$ on D' for all $\lambda < 0$, by the maximum principle. Similarly $h_\lambda - f \leq 0$ on D' for all $\lambda > 0$. Letting $\lambda \to 0$, we conclude

$$f \equiv h \text{ in } D'\backslash\{0\},$$

hence f extends through 0. □

Theorem 2.6.3 *If Σ is hyperbolic, then any holomorphic map $f : \mathbb{C} \to \Sigma$ is constant.*

Proof. As already observed, $d_H \equiv 0$ on $\mathbb{C}$. Hence the theorem follows from the non-increasing property of d_H under holomorphic maps (Lemma 2.6.1). □

Corollary 2.6.2 *An entire holomorphic function omitting two values is constant.*

Proof. Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic with $f(z) \neq a, b$ for all z. To conclude from Theorem 2.6.3 that f is constant, we must show that $\mathbb{C}\backslash\{a, b\}$ is hyperbolic. For that purpose, we construct a metric on $\mathbb{C}\backslash\{a, b\}$ with curvature bounded above by a negative constant: the metric

$$|z-a|^{\mu}\ |z-b|^{\mu}\ (\,|z-a|^{\mu}+1\,)\,(\,|z-b|^{\mu}+1\,)\ \mathrm{d}z\ \mathrm{d}\overline{z}$$

has curvature

$$-\frac{\mu^2}{2}\Big\{(\,|z-b|^{\mu}+1\,)^{-3}\ |z-a|^{2-\mu}\ (\,|z-a|^{\mu}+1\,)^{-1} + (\,|z-a|^{\mu}+1\,)^{-3}\ |z-b|^{2-\mu}\ (\,|z-b|^{\mu}+1\,)^{-1}\Big\},$$

which is bounded above by a negative constant if $0 < \mu < \frac{2}{5}$. Hence the result follows from Corollary 2.6.1. □

To prove the "big" Picard theorem, we need a slight extension of Theorem 2.6.2.

Theorem 2.6.4 *Let $\overline{\Sigma}$ be a compact surface, and $\Sigma := \overline{\Sigma}\backslash\{w_1, \cdots w_k\}$ for a finite number of points in $\overline{\Sigma}$. Assume that Σ is hyperbolic. Let S be a Riemann surface, and $z_0 \in S$. Then any holomorphic map $f : S\backslash\{z_0\} \to \Sigma$ extends to a holomorphic map $\overline{f} : S \to \overline{\Sigma}$.*

Proof. As in Theorem 2.6.2, we may assume $S = D$, $z_0 = 0$. Now we observe that, in Theorem 2.5.2, the boundedness of f and the completeness of d_H on Σ were only used to ensure that, for some sequence $z_n \to 0$, $f(z_n)$ converged in Σ. In the present situation, the set of limiting values of $f(z)$ as $z \to 0$, being the intersection of the closures in $\overline{\Sigma}$ of the $f(0 < |z| < r)$, $0 < r < 1$, is a connected compact set, hence must reduce to one of the w_i if contained entirely in $\overline{\Sigma}\backslash\Sigma$. Hence f extends continuously to D in any case, and the rest of the argument is the same as in Theorem 2.6.2. □

The "big" Picard theorem follows:

Corollary 2.6.3 *Let $f(z)$ be holomorphic in the punctured disc $0 < |z| < R$, and have an essential singularity at $z = 0$. Then there is at the most one value a for which $f(z) = a$ has only finitely many solutions in $0 < |z| < R$.*

Proof. If $f(z) = a$ has only finitely many solutions in $0 < |z| < R$, then there is also an r, $0 < r \leq R$, such that $f(z) = a$ has no solutions at all in $0 < |z| < r$. Hence it suffices to prove that if $f(z)$ is holomorphic in $0 < |z| < r\,(r > 0)$ and omits two finite values a and b, then it has a removable singularity or a pole at 0 (in other words, that it can be extended to a meromorphic function on $|z| < r$. Hence the result follows from Theorem 2.6.4, with $\overline{\Sigma} = S^2$ and $\Sigma = \mathbb{C}\backslash\{a, b\}$ (which was shown to be hyperbolic in the proof of Corollary 2.6.2). □

Exercises for § 2.6

1) Which of the following Riemann surfaces are hyperbolic in the sense of Def. 2.6.1?
S^2, a torus T, $T\backslash\{z_0\}$ for some $z_0 \in T$, an annulus $\{r_1 < |z| < r_2\}$, $\mathbb{C}\backslash\{0\}$.
2) Let S, Σ be Riemann surfaces, and suppose Σ is hyperbolic. Show that the family of all holomorphic maps $f : S \to \Sigma$ which are uniformly bounded is normal. (One needs to use the fact that S as a Riemann surface has countable topology.) If Σ is complete w.r.t. the hyperbolic distance d_H, then the family of all holomorphic maps $f : S \to \Sigma$ - whether bounded or not - is normal.
*3) Write down a čomplete metric on $\mathbb{C}\backslash\{a, b\}$ with curvature bounded from above by a negative constant. (Hint: In punctured neighbourhoods of a, b, ∞, add a suitable multiple of the hyperbolic metric on the punctured disk $D\backslash\{0\}$, multiplied by a cut-off function. If you are familiar with elliptic curves, you can also use the modular function $\lambda : H \to \mathbb{C}\backslash\{0, 1\}$, where H is the upper half plane, to get a metric with constant curvature -1 on $\mathbb{C}\backslash\{0, 1\}$.)
Using the result of 2), conclude Montel's theorem that the family of all holomorphic functions $f : \Omega \to \mathbb{C}$ that omit two values a, b is normal ($\Omega \subset \mathbb{C}$).

2.7 Conformal Structures on Tori

We begin by recalling some facts from [A1] (p. 257).
Let f be a meromorphic function on $\mathbb{C}$. An ω in $\mathbb{C}$ is said to be a period of f if

$$f(z + \omega) = f(z) \qquad \text{for all } z \in \mathbb{C}. \tag{2.7.1}$$

The periods of f form a module M over $\mathbb{Z}$ (in fact an additive subgroup of $\mathbb{C}$). If f is non-constant, then M is discrete.
The possible discrete subgroups of $\mathbb{C}$ are

$$\begin{aligned} M &= \{0\}, \\ M &= \{n\omega : \ n \in \mathbb{Z}\}, \\ M &= \{n_1\omega_1 + n_2\omega_2 : \ n_1, n_2 \in \mathbb{Z}\}, \quad \frac{\omega_2}{\omega_1} \notin \mathbb{R}. \end{aligned}$$

Here the third case is the interesting one. A module of that form is also called a lattice.
As we have already seen, such a module defines a torus $T = T_M$ if we identify the points z and $z + n_1\omega_1 + n_2\omega_2$; let $\pi : \mathbb{C} \to T$ be as before the projection. The parallelogram in $\mathbb{C}$ defined by ω_1 and ω_2 (with vertices $0, \omega_1, \omega_2, \omega_1 + \omega_2$) is a fundamental domain for T.

By (2.7.1), f becomes a meromorphic function on T.
If (ω_1', ω_2') is another basis for the same module, the change of basis is described by

$$\begin{pmatrix} \omega_2' & \overline{\omega_2'} \\ \omega_1' & \overline{\omega_1'} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_2 & \overline{\omega_2} \\ \omega_1 & \overline{\omega_1} \end{pmatrix}$$

with $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ belonging to

$$\mathrm{GL}(2,\mathbb{Z}) := \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : \ \alpha, \beta, \gamma, \delta \in \mathbb{Z}, \ \alpha\delta - \beta\gamma = \pm 1 \right\}.$$

Its subgroup $\mathrm{SL}(2,\mathbb{Z})$ consisting of matrices of determinant $+1$ is called the modular group; and the elements of $\mathrm{SL}(2,\mathbb{Z})$ are called unimodular transformations.
As in § 2.3, we define $\mathrm{PSL}(2,\mathbb{Z}) := \mathrm{SL}(2,\mathbb{Z}) / \left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$. As a subgroup of $\mathrm{PSL}(2,\mathbb{R})$, it acts by isometries on H.

Theorem 2.7.1 *There is a basis (ω_1, ω_2) for M such that, if $\tau := \frac{\omega_2}{\omega_1}$, we have*
(i) $\mathrm{Im}\,\tau > 0$,
(ii) $-\frac{1}{2} < \mathrm{Re}\,\tau \leq \frac{1}{2}$,
(iii) $|\tau| \geq 1$,
(iv) $\mathrm{Re}\,\tau \geq 0$ *if* $|\tau| = 1$.
τ is uniquely determined by these conditions, and the number of such bases for a given module is $2, 4$ or 6.

Thus τ lies in the region sketched in Fig. 2.7.1. Theorem 2.7.1 can also be interpreted as saying that the interior of the region decribed by (i)–(iv) is a fundamental polygon for the action of $\mathrm{PSL}(2,\mathbb{Z})$ on the upper half-plane $\{\mathrm{Im}\,\tau > 0\}$, as in § 2.4.
That there are in general two such bases for a given M is simply because we can replace (ω_1, ω_2) by $(-\omega_1, -\omega_2)$. If $\tau = \mathrm{i}$, then there are 4 bases as in the theorem; namely we can also replace (ω_1, ω_2) by $(\mathrm{i}\omega_1, \mathrm{i}\omega_2)$. Finally we get 6 bases when $\tau = \mathrm{e}^{\frac{\pi \mathrm{i}}{3}}$, because we can in this case replace (ω_1, ω_2) by $(\tau\omega_1, \tau\omega_2)$ (hence also by $(\tau^2\omega_1, \tau^2\omega_2)$). We remark that $\tau = \mathrm{i}$ and $\tau = \mathrm{e}^{\frac{\pi \mathrm{i}}{3}}$ are precisely the fixed points of (non-trivial) elements of $\mathrm{PSL}(2,\mathbb{Z})$ (in the closure of the fundamental domain).
The normalisation in Theorem 2.7.1 can also be interpreted as follows: we choose $\omega_1 = 1$, and then ω_2 lies in the region described by the inequalities (i)–(iv).
In the sequel, we may always make this normalisation, since multiplication of the basis of the module by a fixed factor always leads to a conformally equivalent torus, and we are classifying the different conformal equivalence classes.

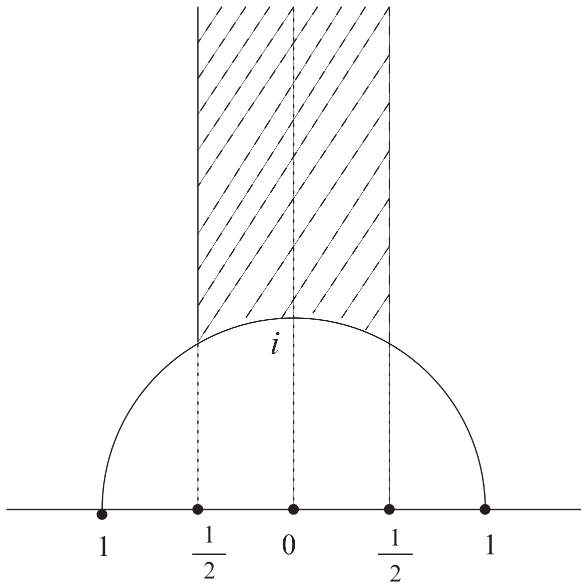

Fig. 2.7.1.

Let us also mention that by the Uniformization Theorem, every Riemann surface which is homeomorphic to a torus is in fact conformally equivalent to a quotient of $\mathbb{C}$, and hence of the form considered here.
As follows from Corollary 1.3.3, and as was explained in § 1.3, $\pi_1(T) = \mathbb{Z}\oplus\mathbb{Z}$; indeed, the group of covering transformations of $\pi : \mathbb{C} \to T$ is $\mathbb{Z}\oplus\mathbb{Z}$, generated by the maps

$$z \to z + \omega_1$$

and

$$z \to z + \omega_2.$$

Thus the fundamental group of T is canonically isomorphic to the module $\{n_1\omega_1 + n_2\omega_2 : \ n_1, n_2 \in \mathbb{Z}\}$.

Lemma 2.7.1 *Let $f_1, f_2 : T \to T'$ be continuous maps between tori. Then f_1 and f_2 are homotopic if and only if the induced maps*

$$f_{i*} : \pi_1(T) \to \pi_1(T') \qquad (i = 1, 2)$$

coincide.

Remark. We do not need to choose base points in this case, since the fundamental groups are abelian (so that all conjugations are the identity map). (Recall the discussion in §1.3.)

Proof of Lemma 2.7.1. "⇒" follows from Lemma 1.2.3, applied to a homotopy between f_1 and f_2.
"⇐": We consider lifts $\widetilde{f_i} : \mathbb{C} \to \mathbb{C}$ of the f_i (cf. Theorem 1.3.3). Let ω_1, ω_2 be a basis of $\pi_1(T)$. Then we have by assumption

$$\begin{aligned}&\widetilde{f}_1(z + n_1\omega_1 + n_2\omega_2) - \widetilde{f}_2(z + n_1\omega_1 + n_2\omega_2)\\ &\quad = \widetilde{f}_1(z) - \widetilde{f}_2(z) \quad \text{for all } z \in \mathbb{C},\, n_1, n_1 \in \mathbb{N}. \end{aligned} \tag{2.7.2}$$

It follows that $\widetilde{F}(z, s) := (1 - s)\widetilde{f}_1(z) + s\widetilde{f}_2(z)$ satisfies

$$\widetilde{F}(z + n_1\omega_1 + n_2\omega_2, s) = \widetilde{f}_1(z + n_1\omega_1 + n_2\omega_2) + s(\widetilde{f}_2(z) - \widetilde{f}_1(z)).$$

Hence each $\widetilde{F}(\cdot, s)$ induces a map

$$F(\cdot, s) : T \to T'.$$

This provides the desired homotopy between f_1 and f_2. □

We now proceed to the classification of conformal structures on tori. Actually, we shall only be giving a new interpretation of results already discussed. But it gives us an opportunity to illustrate in this simple case some concepts which we shall later have to discuss more precisely in the general case (which is much more difficult).
We shall make use of the normalisation discussed above, according to which the basis of a torus can be taken in the form 1, τ (τ as in Theorem 2.7.1). We denote the corresponding torus by $T(\tau)$.

Definition 2.7.1 The moduli space $\mathcal{M}_1$ is the space of equivalence classes of tori, two tori being regarded as equivalent if there exists a bijective conformal map between them. We say that a sequence of equivalence classes, represented by tori T^n ($n \in \mathbb{N}$) converges to the equivalence class of T if we can find bases (ω_1^n, ω_2^n) for T^n and (ω_1, ω_1) for T such that $\frac{\omega_2^n}{\omega_1^n}$ converges to $\frac{\omega_2}{\omega_1}$.

Definition 2.7.2 The Teichmüller space $\mathcal{T}_1$ is the space of equivalence classes of pairs $(T, (\omega_1, \omega_2))$ where T is a torus, and (ω_1, ω_2) is a basis of T (i.e. of the module M defining T); here, $(T, (\omega_1, \omega_2))$ and $(T', ({\omega_1}', {\omega_2}'))$ are equivalent if there exists a bijective conformal map

$$f : T \to T'$$

with

$$f_*(\omega_i) = \omega_i'.$$

(Here as before, (ω_1, ω_2) has been canonically identified with a basis of $\pi_1(T)$, and similarly (ω_1', ω_2'), f_* is the map of fundamental groups induced by f.) We say that $(T^n, (\omega_1^n, \omega_2^n))$ converges to $(T, (\omega_1, \omega_2))$ if $\frac{\omega_2^n}{\omega_1^n}$ converges to $\frac{\omega_2}{\omega_1}$.

We shall also call a pair $(T, (\omega_1, \omega_2))$ as above a marked torus.
The space $\mathcal{T}_1$ can also be interpreted as follows. We choose a fixed marked torus, e.g. $T(\mathrm{i})$ with basis $(1, \mathrm{i})$. We denote it by T_{top}, since it serves us as the underlying topological model. By Lemma 2.7.2, (ω_1, ω_2) defines a homotopy class $\alpha(\omega_1, \omega_2)$ of maps $T \to T_{\mathrm{top}}$. Namely, $\alpha(\omega_1, \omega_2)$ is that homotopy class for which the induced map of fundamental groups sends (ω_1, ω_2) to the given basis of T_{top} (ω_1 to 1 and ω_2 to i in our case). The existence of a map $T \to T_{\mathrm{top}}$ which induces the above map on fundamental groups is clear: the $\mathbb{R}$-linear map $g : \mathbb{C} \to \mathbb{C}$ with $g(\omega_1) = 1$, $g(\omega_2) = \mathrm{i}$ gives rise to one such map $T \to T_{\mathrm{top}}$.
Thus, instead of pairs $(T, (\omega_1, \omega_2))$, we can also consider pairs (T, α), where α is a homotopy class of maps $T \to T_{\mathrm{top}}$ which induces an isomorphism of fundamental groups (thus α should contain a homeomorphism). (T, α) and (T', α') are now to be regarded as equivalent if the homotopy class $(\alpha')^{-1} \circ \alpha$ of maps $T \to T'$ contains a conformal map. $\mathcal{T}_1$ is then the space of equivalence classes of such pairs.

Theorem 2.7.2 $\mathcal{T}_1 = H$; $\mathcal{M}_1 = H/\mathrm{PSL}(2, \mathbb{Z})$.

We have already seen that every torus is conformally equivalent to a $T(\tau)$ with τ in the fundamental domain of $\mathrm{PSL}(2, \mathbb{Z})$ (Theorem 2.7.1). Similarly, every marked torus can be identified with an element of H; just normalise so that $\omega_1 = 1$. Thus we must show that two distinct elements of $H/\mathrm{PSL}(2, \mathbb{Z})$ (resp. H) are not conformally equivalent (resp. equivalent as marked tori). There are many ways of doing this. We shall follow a method which illustrates by a simple example some considerations of great importance in the sequel.

Definition 2.7.3 A map $h : T \to T'$ is said to be harmonic if its lift $\widetilde{h} : \mathbb{C} \to \mathbb{C}$
(cf. Theorem 1.3.3) is harmonic.

Equivalently, the local expression of h in the charts induced by the projections $\mathbb{C} \to T$, $\mathbb{C} \to T'$ should be harmonic, i.e. have harmonic real and imaginary parts. Here, it is important to observe that the transition functions of such charts are linear, so that a change of charts in the target torus also preserves the harmonicity of the map; arbitrary changes of charts in the target do not preserve harmonicity.

Lemma 2.7.2 *Let T, T' be tori, $z_0 \in T$, $z_0' \in T'$. Then, in every homotopy class of maps $T \to T'$, there exists a harmonic map h; h is uniquely determined by requiring that $h(z_0) = z_0'$. The lift $\widetilde{h} : \mathbb{C} \to \mathbb{C}$ of a harmonic map h is affine linear (as a map $\mathbb{R}^2 \to \mathbb{R}^2$).*
If normalised by $\widetilde{h}(0) = 0$ (instead of $h(z_0) = z_0'$), it is therefore linear. h is conformal if and only if $\widetilde{h}$ (normalised by $\widetilde{h}(0) = 0$) is of the form $z \to \lambda z$, $\lambda \in \mathbb{C}$.

Proof. Let (ω_1, ω_2) be a basis of T, and ω_1', ω_2' the images of ω_1 and ω_2 determined by the given homotopy class (cf. Lemma 2.7.2). Then the $\mathbb{R}$-linear map $\widetilde{h} : \mathbb{C} \to \mathbb{C}$ with $\widetilde{h}(\omega_i) = \omega_i'$ induces a harmonic map $h : T \to T'$ in the given homotopy class.
Now suppose $\widetilde{f}$ is the lift of any map $f : T \to T'$ in the given homotopy class. Then

$$\widetilde{f}(z + n_1\omega_1 + n_2\omega_2) = \widetilde{f}(z) + n_1\omega_1' + n_2\omega_2', \tag{2.7.3}$$

hence

$$\frac{\partial \widetilde{f}}{\partial x}(z + n_1\omega_1 + n_2\omega_2) = \frac{\partial \widetilde{f}}{\partial x}(z) \tag{2.7.4}$$

and similarly for $\frac{\partial \widetilde{f}}{\partial y}$. Thus if f (hence $\widetilde{f}$) is harmonic, then so are $\frac{\partial \widetilde{f}}{\partial x}, \frac{\partial \widetilde{f}}{\partial y}$. But $\frac{\partial \widetilde{f}}{\partial x}, \frac{\partial \widetilde{f}}{\partial y}$ are then complex-valued harmonic functions on T by (2.7.4), hence constant by Lemma 2.2.1. Thus $\widetilde{f}$ is affine linear. It also follows that the harmonic map in a given homotopy class is uniquely determined by the requirement $h(z_0) = z_0'$. Another way of seeing this is to observe that, by (2.7.3), the difference between the lifts of two homotopic harmonic maps becomes a harmonic function on T, and is therefore constant.
The last assertion is clear. □

The *proof* of Theorem 2.7.2 is now immediate:
A conformal map is harmonic, hence has an affine linear lift $\widetilde{h}$ by Lemma 2.7.2; we may assume $\widetilde{h}(0) = 0$. We may also assume that the markings have been normalised by
$\omega_1 = 1 = \omega_1' \,(= \widetilde{h}(\omega_1))$. But if $\widetilde{h}$ is conformal, $\widetilde{h}(1) = 1$ implies $\widetilde{h}$ is the identity.
It follows that $\mathcal{T}_1 = H$.
To see that $\mathcal{M}_1 = H/\mathrm{PSL}(2, \mathbb{Z})$, we observe that we may now choose arbitrary markings. Thus we need only be able to say when a torus $T(\tau)$ with basis $(1, \tau)$ is equivalent to the torus $T(\tau')$ for some choice of a marking on it. But this, by what has been proved above, is the case precisely when $(1, \tau')$ is also a basis for $T(\tau)$.

The rest of the proof is straightforward and left to the reader as an exercise. □

Exercises for § 2.7

1) Compute the area of a fundamental domain for $\mathrm{PSL}\,(2, \mathbb{Z})$.
2) Determine a fundamental region for the congruence subgroup mod 2 of $\mathrm{PSL}\,(2, \mathbb{Z})$, namely

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}\,(2,\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod 2 \right\}.$$

Show that it is a normal subgroup of $\mathrm{PSL}\,(2,\mathbb{Z})$ and compute the number of elements of the quotient group.

3) Determine the set of conformal equivalence classes of annuli.