Automorphic Functions

Daniel Nicks

May 16, 2006

Contents

1 Discontinuous Groups **3**

2 Geometry of Γ **7**

2.1 The Isometric Circle 7

2.2 Fundamental Regions 7

2.3 The Modular Group $\Gamma(2)$ 9

3 Riemann Surfaces and Uniformisation **10**

3.1 Riemann Surfaces 10

3.2 Uniformisation 13

3.3 Riemann Surfaces with $\mathbb{C}$ or $\mathbb{C}_\infty$ as the Universal Covering Surface 14

4 Existence of Automorphic Functions on Fuchsian Groups **17**

4.1 Automorphic Forms 17

4.2 Existence 18

5 A Return to Geometry **22**

5.1 Fixed Points 22

5.2 Normal Polygons 23

6 Quotient Spaces as Riemann Surfaces **24**

Introduction

A Riemann surface is essentially a two-dimensional surface that locally looks like the complex plane. Many concepts of complex analysis generalise to Riemann surfaces. In particular, we can define analytic and meromorphic functions on them. The first aim of this essay is to show that every Riemann surface admits non-constant meromorphic functions. To achieve this we will appeal to the Uniformisation Theorem. This states that any Riemann surface can be expressed as a quotient of a complex domain by a group of conformal[1] self-maps of that domain. This group of maps is necessarily *discontinuous*—the orbit of any point can only accumulate on the boundary of the domain. The search for a non-constant meromorphic function on the Riemann surface is then reduced to finding an *automorphic* function on the domain. A meromorphic function is said to be automorphic with respect to a group of maps if the action of the maps leaves the value of the function unchanged: letting Γ denote the group, a meromorphic function $f(z)$ is automorphic if,

$$f(z) = f(Tz) \quad \text{for all} \quad T \in \Gamma.$$

An automorphic function assigns values to the orbits of the group and so gives a well-defined function on the quotient space—our original Riemann surface. The discontinuity of the group of conformal maps is immediately seen to be vital; without it the Isolated Values Theorem would doom any automorphic function to be constant.

The first two sections of this essay are devoted to building up the necessary results on discontinuous groups and their associated geometry. In the third section we discuss Riemann surfaces and see how the Uniformisation Theorem will lead us to our goal of constructing non-constant meromorphic functions on them. The end of this third section and the whole of the fourth contain the existence proofs for the automorphic functions required to complete the story.

In the penultimate section we develop further the underlying geometry of a discontinuous group. We put these results to use in the final section where we deal with the second aim of this essay: exposing how quotient spaces similar to those encountered in the Uniformisation Theorem can be given the structure of a Riemann surface.

The primary sources for this essay were the two books by Joseph Lehner. Sections 1, 2.3, 5 and 6 draw mainly from [1], while Sections 2.1, 2.2 and 4 are based more on [2]. Also consulted were Beardon's book [3] on discrete groups and Ford's book [4] in which he introduced the isometric circle. Acknowledgment is also due to Dr Kovalev and Dr Carne for their lecture courses on *Riemann Surfaces* and *Advanced Complex Variable*.

1 Discontinuous Groups

In this first section we will introduce discontinuous groups of conformal mappings and examine some of their properties.

Definition. Let Γ be a group of conformal maps of $\mathbb{C}_\infty$ onto itself[2]. A point $z \in \mathbb{C}_\infty$ is a *limit point* of Γ if there exists a sequence of distinct maps $T_n \in \Gamma$ and a point $w \in \mathbb{C}_\infty$ such that $T_n(w) \to z$. The *limit set* of Γ is simply the set of all limit points of Γ and will be denoted by L. The complement $\mathbb{C}_\infty \setminus L$ is the set of *ordinary points* of Γ which we denote by $\mathcal{O}$. The group Γ is said to be *discontinuous* if $\mathcal{O}$ is non-empty and is *discontinuous on U* if $U \subseteq \mathcal{O}$.

[1] A *conformal* map is simply an analytic bijection.

[2] By $\mathbb{C}_\infty$ we mean the *extended complex plane* $\mathbb{C} \cup \{\infty\}$, also called the *Riemann sphere* (see Section 3.1).

Since we are considering conformal self-maps of $\mathbb{C}_\infty$ we are in fact dealing with subgroups of the group of Möbius transformations [5, §2.1]. A general Möbius map takes the form,

$$z \longmapsto \frac{az+b}{cz+d}$$

for some complex coefficients a, b, c, d such that $ad - bc \neq 0$. We can always normalise so that $ad - bc = 1$ and throughout this essay we shall assume that this normalisation has been carried out. We adopt the usual conventions regarding the point ∞. For example, under the above map ∞ has image a/c and pre-image $-d/c$.

It is often useful to consider the map

$$z \longmapsto \frac{az+b}{cz+d}$$

as the complex 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

It is a simple exercise in algebra to see that this gives a group homomorphism with composition and inversion of maps corresponding to matrix multiplication and inversion. At this point we should note that the matrices A and $-A$ represent the same Möbius transformation. It is not hard to see that the set of all normalised Möbius transformations is isomorphic to $PSL(2, \mathbb{C})$, the quotient group of all 2×2 complex matrices with determinant 1 under the equivalence $A \sim -A$. Consequently, we may refer to a Möbius map in its matrix form, provided we bear in mind the $\pm A$ equivalence.

The group of Möbius maps naturally carries with it the topology of uniform convergence (with respect to the chordal metric on $\mathbb{C}_\infty$). The corresponding group of complex matrices can be viewed as a subset of $\mathbb{C}^4$ and so inherits the subspace topology. Intuitively, one would hope that the convergence of Möbius maps is related to the convergence of representing matrices. The topologies do indeed coincide appropriately, as is shown in [3, §4.5].

A group Γ of Möbius transformations naturally acts on $\mathbb{C}_\infty$. We say that two points $z, w \in \mathbb{C}_\infty$ are *equivalent* (under Γ) if there exists $T \in \Gamma$ such that $T(z) = w$. This defines an equivalence relation on $\mathbb{C}_\infty$, the equivalence classes being the *orbits* of Γ. We write the orbit of z as,

$$\Gamma z = \{Tz : T \in \Gamma\}.$$

Observe that an orbit of Γ can only accumulate at a limit point of Γ.

We shall now establish some results regarding the structure of discontinuous groups.

Definition. A group of matrices is *discrete* if it contains no elementwise convergent sequence of distinct matrices. The limit need not be a member of the group.

We shall soon see that the discontinuity of a group of Möbius maps is closely related to the discreteness of the corresponding group of matrices. As always, we consider groups of matrices with determinant 1.

Lemma 1.1. *A group Γ is discrete iff it contains no sequence of distinct matrices converging to the identity.*

Proof. If Γ contains a sequence of distinct matrices converging to I then by definition it is not discrete.

Conversely, suppose that Γ is not discrete and so contains a distinct sequence $T_n \to T$. Note that,

$$\det T = \lim \det T_n = 1$$

so the inverse T^{-1} exists and $T_n^{-1} \to T^{-1}$ implying that,

$$T_{n+1}^{-1} T_n \to T^{-1}T = I.$$

The latter sequence is in Γ so it simply remains to show that it has a subsequence of distinct elements. If this is not the case then eventually (i.e. for all sufficiently large n) $T_{n+1}^{-1}T_n = I$ and so $T_{n+1} = T_n$. This contradicts the fact that the T_n are all distinct. □

Theorem 1.2. *A discontinuous group is discrete.*

Proof. Recall that matrix convergence corresponds to the convergence of Möbius maps. Suppose that Γ is not discrete. By the previous lemma there exists a distinct sequence T_n in Γ such that $T_n \to I$. Then for all $z \in \mathbb{C}_\infty$ we have $T_n z \to z$ so that z is a limit point of Γ. Hence Γ is not discontinuous. □

Lemma 1.3. *A discontinuous group is countable.*

Proof. Let Γ be a discontinuous group. With each element $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ associate the point $(a, b, c, d) \in \mathbb{C}^4$. If Γ is uncountable then so is the set of such points in $\mathbb{C}^4$. Hence some closed ball in $\mathbb{C}^4$ must contain uncountably many points. Since such a ball is compact we can find a convergent sequence of these points. This corresponds to an elementwise convergent sequence of distinct matrices in Γ, which contradicts the preceding theorem that Γ is discrete. □

From this point on we shall largely restrict our attention to groups of mappings that preserve either the unit disc, $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, or the upper half plane, $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}\ z > 0\}$. Such discontinuous groups are called *Fuchsian groups*. The Riemann Mapping Theorem provides the motivation for making this restriction: the theorem states that any simply-connected domain in $\mathbb{C}$ (other than $\mathbb{C}$ itself) is conformally equivalent to $\mathbb{D}$. If we let U be such a domain, let $h : U \mapsto \mathbb{D}$ be a conformal map and take a discontinuous group Γ of conformal self-maps of U, then $\Gamma' = h\Gamma h^{-1}$ is an equivalent Fuchsian group. Therefore, by studying Fuchsian groups we study discontinuous groups of conformal maps on any simply-connected domain.

One consequence of this is the Uniformisation Theorem that allows us to express 'most' Riemann surfaces as a quotient of $\mathbb{D}$ by a discontinuous group. If we wish to consider meromorphic functions on these Riemann surfaces then we are led quite naturally to seek automorphic functions for Fuchsian groups. We shall return to these ideas later.

Note that we may freely move between the disc and the upper half plane since $\mathbb{D}$ is conformally mapped onto $\mathbb{H}$ by the Möbius transformation $z \mapsto \frac{i(1+z)}{1-z}$. This allows us to work in whichever space is most convenient.

We recall that the conformal maps fixing $\mathbb{H}$ are precisely the Möbius maps with real coefficients, while the conformal self-maps of $\mathbb{D}$ are those Möbius transformations that take the form,

$$\begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix}, \text{ where } |a|^2 - |b|^2 = 1.$$

Note that these maps act on the whole of $\mathbb{C}_\infty$ and preserve not only $\mathbb{D}$ (respectively $\mathbb{H}$) but also $\partial\mathbb{D}$ and $\mathbb{C}_\infty \setminus \overline{\mathbb{D}}$ (respectively $\partial\mathbb{H}$ and $\mathbb{C}_\infty \setminus \overline{\mathbb{H}}$).

We now establish a converse to Theorem 1.2 for Fuchsian groups.

Theorem 1.4. *A discrete group Γ of conformal self-maps of $\mathbb{H}$ is discontinuous on $\mathbb{H}$.*

Proof. Suppose that Γ has a limit point $z_0 \in \mathbb{H}$. Then there exists a distinct sequence $T_1, T_2, \ldots$ in Γ such that $T_n w \to z_0$ for some $w = x + iy$. Elements of Γ preserve $\mathbb{H}$ so $w \in \mathbb{H}$ (i.e. $y > 0$). By transforming by the normalised real Möbius map that sends i to w,

$$A : z \longmapsto \frac{y^{\frac{1}{2}} z + x y^{-\frac{1}{2}}}{y^{-\frac{1}{2}}}$$

we may instead consider $\Gamma' = A^{-1}\Gamma A$ which must also be discrete. The transformed group Γ' has limit point $z_1 = A^{-1} z_0$ and contains the distinct sequence $T'_n = A^{-1} T_n A$ such that,

$$T'_n i = A^{-1} T_n w \to z_1.$$

Now write $T'_n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}$. Since $z_1 \in \mathbb{H}$ we have that,

$$\text{Im } T'_n i = \frac{1}{c_n^2 + d_n^2} \to \text{Im } z_1 > 0$$

from which we deduce that the sequences c_n and d_n are bounded. We also have,

$$|T'_n i|^2 = \frac{a_n^2 + b_n^2}{c_n^2 + d_n^2} \to |z_1|^2$$

so the sequences a_n and b_n are also bounded. Therefore, we can find a convergent subsequence of the T'_n. This contradicts the fact that Γ' is discrete. $\square$

Corollary 1.5. *A discrete group of conformal maps that preserve $\mathbb{H}$ has $L \subseteq \mathbb{R} \cup \{\infty\}$. That is, all limit points lie on the boundary of $\mathbb{H}$.*

Proof. The proof of the theorem shows that $L \cap \mathbb{H} = \emptyset$. An almost identical argument applies for a limit point in the lower half plane. Hence $L \cap (\mathbb{C} \setminus \overline{\mathbb{H}}) = \emptyset$. Therefore $L \subseteq \partial\mathbb{H}$. $\square$

Remark 1.6. These results do of course apply to $\mathbb{D}$ as well as to $\mathbb{H}$. In particular, the limit set of a Fuchsian group lies on the boundary of the preserved domain.

Example. We close this section with a quick look at some Fuchsian groups on $\mathbb{H}$. A famous example is the *modular group*,

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

This is clearly a real discrete group, so by Theorem 1.4 it is discontinuous on $\mathbb{H}$. Any subgroup of $SL(2,\mathbb{Z})$ will also be Fuchsian, for example the grandly-titled *principal congruence subgroup of level n*,

$$\Gamma(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \;:\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \pm I \pmod{n} \right\}$$

We shall learn more about $\Gamma(2)$ in Section 2.3.

2 Geometry of Γ

In this section we shall introduce two geometric objects that arise from discontinuous groups: isometric circles and fundamental regions.

2.1 The Isometric Circle

Definition. For a transformation $T(z) = \frac{az+b}{cz+d}$ not fixing ∞, the *isometric circle* of T is defined as

$$\mathbf{I}(T) = \{z \in \mathbb{C} : |cz + d| = 1\}$$

We note immediately that $c \neq 0$ for a transformation that does not fix ∞, so the isometric circle is indeed a circle of radius $|c|^{-1}$ with centre $-d/c$. The isometric circle is the set of points for which $|T'(z)| = 1$. This explains the name: on the isometric circle the infinitesimal euclidean length is unchanged by T.

Lemma 2.1. *The transformation* T *maps*[3] $\left.\begin{array}{l} \mathbf{I}(T) \\ \textit{the interior of } \mathbf{I}(T) \\ \textit{the exterior of } \mathbf{I}(T) \end{array}\right\}$ *onto* $\left\{\begin{array}{l} \mathbf{I}(T^{-1}) \\ \textit{the exterior of } \mathbf{I}(T^{-1}) \\ \textit{the interior of } \mathbf{I}(T^{-1}) \end{array}\right.$

Proof. This is just a quick calculation. If $T(z) = \frac{az+b}{cz+d}$ then,

$$T^{-1}(z) = \frac{dz - b}{-cz + a} \qquad \text{and so} \qquad \mathbf{I}(T^{-1}) = \{z \in \mathbb{C} : |cz - a| = 1\}$$

$$|cT(z) - a| = \left| c\frac{az + b}{cz + d} - a \right| = \left| \frac{caz + bc - acz - ad}{cz + d} \right| = \frac{1}{|cz + d|}$$

Therefore,

$$\begin{array}{rlllllll} z \in \mathbf{I}(T) & \Rightarrow & |cz + d| = 1 & \Rightarrow & |cT(z) - a| = 1 & \Rightarrow & T(z) \in \mathbf{I}(T^{-1}) \\ z \in \text{Int } \mathbf{I}(T) & \Rightarrow & |cz + d| < 1 & \Rightarrow & |cT(z) - a| > 1 & \Rightarrow & T(z) \in \text{Ext } \mathbf{I}(T^{-1}) \\ z \in \text{Ext } \mathbf{I}(T) & \Rightarrow & |cz + d| > 1 & \Rightarrow & |cT(z) - a| < 1 & \Rightarrow & T(z) \in \text{Int } \mathbf{I}(T^{-1}) \end{array}$$

□

2.2 Fundamental Regions

We have already remarked that a discontinuous group Γ of Möbius self-maps of a domain D partitions D into orbits. We shall also be interested in the quotient space D/Γ. Therefore, it is natural to consider subsets of D obtained by selecting one point from each of the group's orbits. This section discusses how we can construct such sets possessing useful geometric and topological properties.

Definition. Let Γ be a discontinuous group of self-maps of D. A set $F \subseteq D$ is a *fundamental set* for Γ if F contains exactly one point of the orbit Γz for every $z \in D$.

Since it is often more convenient to work with open sets, we shall chiefly deal with a slightly modified concept.

[3] By the *interior* of a circle we mean the set of points enclosed within the circle, rather than the topological set interior.

Definition. An open set $R \subseteq D$ is a *fundamental region* for Γ if,

(i) $\overline{R}$ contains a point of Γz for all $z \in D$

(ii) No two points of R are equivalent under Γ

By forming the union of a fundamental region R with some of its boundary points we can obtain a fundamental set F such that $R \subset F \subset \overline{R}$.

So far there is no reason to believe that a fundamental region has any geometric structure. In fact, for Fuchsian groups, it is possible to construct simply-connected fundamental regions bounded by line segments and arcs of circles. The two primary ways of achieving this are the *normal polygons* of Poincaré and Ford's approach using isometric circles. We shall return to normal polygons in Section 5.2. For now, we follow Ford's method.

Definition. For Γ a Fuchsian group acting on $\mathbb{D}$ with no non-identity element fixing ∞, let the region R_0 be the set of points external to all isometric circles. That is,

$$R_0 = \{z \in \mathbb{C}_\infty : z \in \text{Ext } \mathbf{I}(T) \text{ for all } T \in \Gamma \setminus \{I\}\}$$

We note that all non-identity elements have isometric circles since they must not fix ∞.

It can be shown that R_0 is open and that $R_0 \cap \mathbb{D}$ is a fundamental region for Γ (see [2, p115]). We shall just establish those properties of R_0 that will be needed later when we come to prove the existence of non-constant automorphic functions. *For the rest of this section we assume that* Γ *is a Fuchsian group on* $\mathbb{D}$ *such that no non-identity element fixes* ∞.

Theorem 2.2. R_0 *does not contain two equivalent points.*

Proof. Let $z \in R_0$ and $T \in \Gamma \setminus \{I\}$. Since z lies outside the isometric circle $\mathbf{I}(T)$, the point $T(z)$ lies inside $\mathbf{I}(T^{-1})$ by Lemma 2.1. Hence T maps z outside R_0, so no two distinct points of R_0 can be equivalent under Γ. $\square$

Lemma 2.3. *There exists* $\tilde{b} > 0$ *such that for all* $\begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix} \in \Gamma \setminus \{I\}$ *we have* $|b| > \tilde{b}$.

Proof. We will show that the set $B = \left\{ b : \begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix} \in \Gamma \right\}$ has no finite accumulation points. We will then be done—zero cannot be an accumulation point of B, and an element of Γ with $b = 0$ fixes ∞, and so, by our assumption on Γ, it can only be the identity.

The proof is by contradiction. Suppose there exist distinct $\begin{pmatrix} a_n & \bar{b}_n \\ b_n & \bar{a}_n \end{pmatrix}$ in Γ such that $b_n \to b$. Then,

$$|a_n|^2 - |b_n|^2 = 1 \quad \Rightarrow \quad |a_n|^2 \to 1 + |b|^2$$

Hence the sequences a_n, b_n are bounded and so we can find a convergent subsequence, $\begin{pmatrix} a_{n_j} & \bar{b}_{n_j} \\ b_{n_j} & \bar{a}_{n_j} \end{pmatrix}$. This contradicts Theorem 1.2 that Γ is discrete. $\square$

Remark 2.4. Moreover, for a distinct sequence $\begin{pmatrix} a_n & \bar{b}_n \\ b_n & \bar{a}_n \end{pmatrix}$ in Γ, we must have $|b_n| \to \infty$. Otherwise b_n possesses a convergent subsequence in contradiction to the above.

Theorem 2.5. R_0 *contains a neighbourhood of* ∞. *That is, there exists* ρ *such that*

$$\{z : |z| > \rho\} \subseteq R_0.$$

Proof. This is equivalent to showing that all isometric circles lie within a bounded disc about the origin. Recall that the isometric circle for $\begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix}$ has radius $|b|^{-1}$ and centre $-\bar{a}/b$. By the preceding lemma we know that the radii of all the isometric circles is bounded above by $\tilde{b}^{-1}$. By Remark 1.6 we know that ∞ is an ordinary point—no orbit can accumulate at infinity. In particular, the set of pre-images of ∞ cannot accumulate at ∞. This set is simply,

$$\left\{ -\frac{\bar{a}}{b} : \begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix} \in \Gamma \right\}$$

which is seen to be the set of centres of the isometric circles of elements of Γ together with ∞ itself. We conclude that the set of the centres of the isometric circles is bounded. Hence all isometric circles lie in $\{z : |z| < \rho\}$ for some ρ. □

2.3 The Modular Group $\Gamma(2)$

We now construct an example of a fundamental region for the principal congruence subgroup of level 2 of the modular group. Recall from the end of Section 1 that $\Gamma(2)$ is the group of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where a, d are odd integers and b, c are even integers.

The group $\Gamma(2)$ contains a subgroup of translations generated by $z \mapsto z + 2$. Let Γ_∞ denote this subgroup. Any transformation in $\Gamma(2)$ which fixes ∞ must lie in this subgroup (since it must have $c = 0$, $ad = 1$ and b even). If we define the strip,

$$S = \{z \in \mathbb{H} : -1 < \operatorname{Re} z < 1\}$$

then we see immediately that any point of $\mathbb{H}$ can be mapped to a point of $\overline{S}$ by a translation in Γ_∞. In fact, S is a fundamental region for Γ_∞.

The elements of $\Gamma(2) \setminus \Gamma_\infty$ do not fix ∞ and so all have isometric circles. These are simply given by $|cz + d| = 1$ for all odd d and non-zero even c. The isometric circles and the strip S are shown in Figure 1. A quick check confirms that any isometric circle that meets S lies inside one of the circles $|2z \pm 1| = 1$. If we let,

$$T_\pm = \begin{pmatrix} \pm 1 & 0 \\ 2 & \pm 1 \end{pmatrix}$$

then these circles are $\mathbf{I}(T_\pm)$. We define the set R to be those points of S that lie outside $\mathbf{I}(T_+)$ and $\mathbf{I}(T_-)$, as shown in Figure 1.

The set R will be our fundamental region for $\Gamma(2)$. We first show that a point $z \in R$ is mapped outside of R by any transformation $T \in \Gamma(2)$. If $T \in \Gamma_\infty$ is a translation then $T(z)$ clearly lies outside S and so outside R. Otherwise, T has an isometric circle and by the construction of R the point z must lie outside $\mathbf{I}(T)$. Then Lemma 2.1 states that T maps z to a point inside $\mathbf{I}(T^{-1})$ and so outside R. Hence R cannot contain equivalent points.

We must now prove that any point $z_0 \in \mathbb{H}$ is equivalent under $\Gamma(2)$ to a point in $\overline{R}$. Some translation in Γ_∞ must map z_0 to a point $z_1 \in \overline{S}$. If $z_1 \notin \overline{R}$ then it must lie within either $\mathbf{I}(T_+)$ or $\mathbf{I}(T_-)$, i.e. $|2z_1 \pm 1| < 1$ for the appropriate choice of sign. Applying $T_\pm$ to z_1 then increases its imaginary part:

$$\operatorname{Im}(T_\pm z_1) = \operatorname{Im}\left(\frac{\pm z_1}{2z_1 \pm 1}\right) = \frac{\operatorname{Im} z_1}{|2z_1 \pm 1|^2} > \operatorname{Im} z_1$$

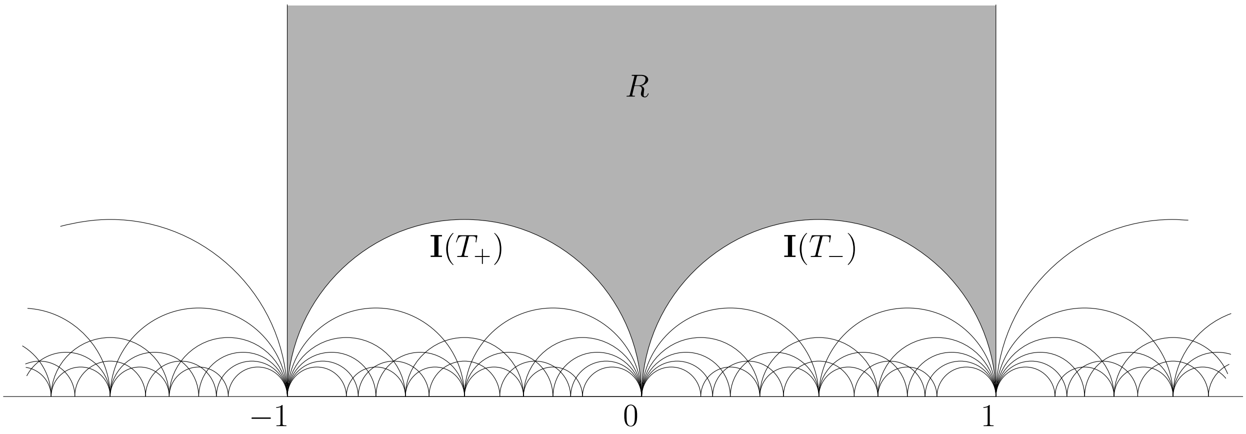

Figure 1: The region R and isometric circles for $\Gamma(2)$.

By applying a translation, let z_2 be a point of $\overline{S}$ equivalent to $T_\pm z_1$. Note that z_2 has the same imaginary part as $T_\pm z_1$ and is equivalent to z_0.

If $z_2 \notin \overline{R}$ then we again find its image under T_+ or T_- and translate to obtain an equivalent point $z_3 \in \overline{S}$ such that $\operatorname{Im} z_3 > \operatorname{Im} z_2$. By induction, we define a sequence of points $\{z_1, z_2, \ldots\} \subseteq \overline{S}$ of strictly increasing imaginary part, all of which are equivalent to z_0. If for some N we have $z_N \in \overline{R}$ then we are done. Otherwise, we would have an infinite sequence of points in $\overline{S}$, the imaginary part of each point lying between $\operatorname{Im} z_1$ and 1 (if $z_n \in \overline{S}$ and $\operatorname{Im} z_n > 1$ then $z_n \in \overline{R}$). This sequence of equivalent points would therefore have an accumulation point in $\mathbb{H}$—a contradiction to the discontinuity of $\Gamma(2)$. This completes the proof that R is a fundamental region for $\Gamma(2)$.

We conclude by noting that the vertical sides of R are equivalent under the translations $z \mapsto z \pm 2$, while the two curved sides of R (the semi-circles of $\mathbf{I}(T_+)$ and $\mathbf{I}(T_-)$) are equivalent under $T_\pm$.

3 Riemann Surfaces and Uniformisation

Fundamental to our proof of the existence of non-constant meromorphic functions on any Riemann surface will be the Uniformisation Theorem. This allows us to classify all Riemann surfaces and express them as quotient spaces of subsets of $\mathbb{C}_\infty$. We can then 'lift' the question of existence to that of finding non-constant automorphic functions for discontinuous groups.

3.1 Riemann Surfaces

Definition. A connected Hausdorff topological space S is a *Riemann surface* if

1. S is a surface (or 'two-dimensional manifold'). That is, S has an open cover of *co-ordinate neighbourhoods* $\{\mathcal{U}_\alpha\}$ together with homeomorphisms, $\varphi_\alpha : \mathcal{U}_\alpha \to \varphi_\alpha(\mathcal{U}_\alpha) \subseteq \mathbb{C}$, called *charts*.

2. All the *transition functions* $\varphi_\beta \circ \varphi_\alpha^{-1}$ are analytic where defined, i.e. on $\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$.

See Figure 2.

Observe that any domain in $\mathbb{C}$ is trivially a Riemann surface. Furthermore, the Riemann sphere $\mathbb{C}_\infty$ can be seen to be a Riemann surface by taking co-ordinate neighbourhoods $\mathbb{C}_\infty \setminus \{\infty\}$ and $\mathbb{C}_\infty \setminus \{0\}$. As charts we use respectively stereographic projection from the north pole and

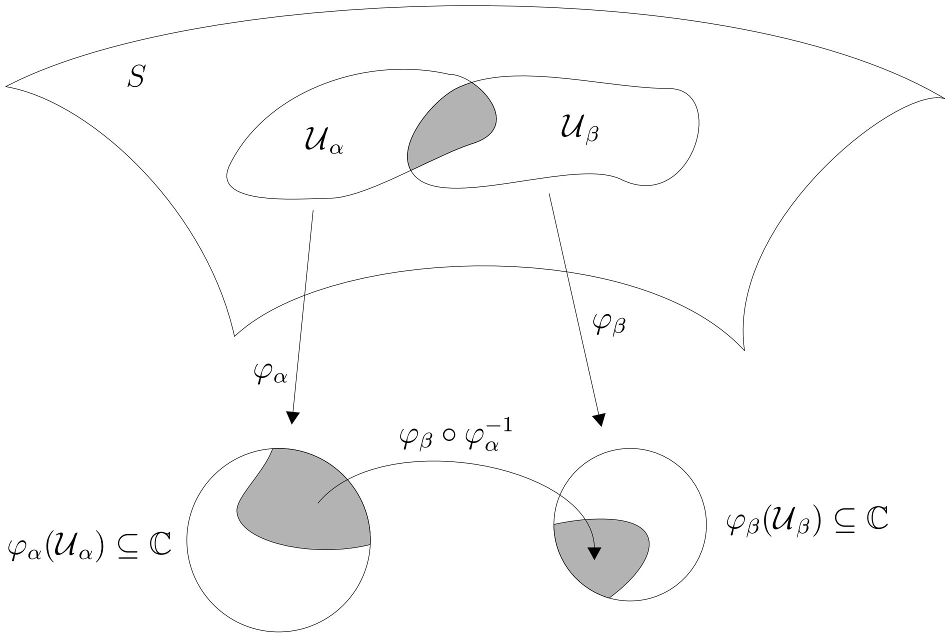

Figure 2: Charts and transition functions for a Riemann surface

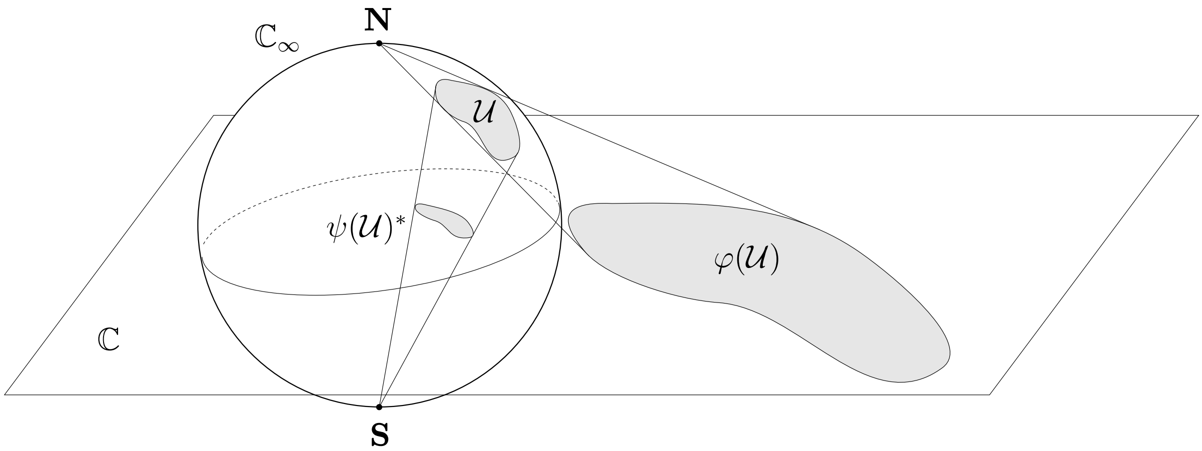

Figure 3: Possible charts for the Riemann sphere

stereographic projection from the south pole followed by complex conjugation. We shall denote these charts by φ, ψ as shown in Figure 3.

The transition functions, $\psi \circ \varphi^{-1}(z) = \varphi \circ \psi^{-1}(z) = \frac{1}{z}$, are then analytic on their domain $\mathbb{C} \setminus \{0\}$.

Definition. Let R and S be Riemann surfaces. A continuous function $f : R \to S$ is *analytic* if, for all charts φ_α on R and ψ_β on S, the composition $\psi_\beta \circ f \circ \varphi_\alpha^{-1}$ is analytic. See Figure 4.

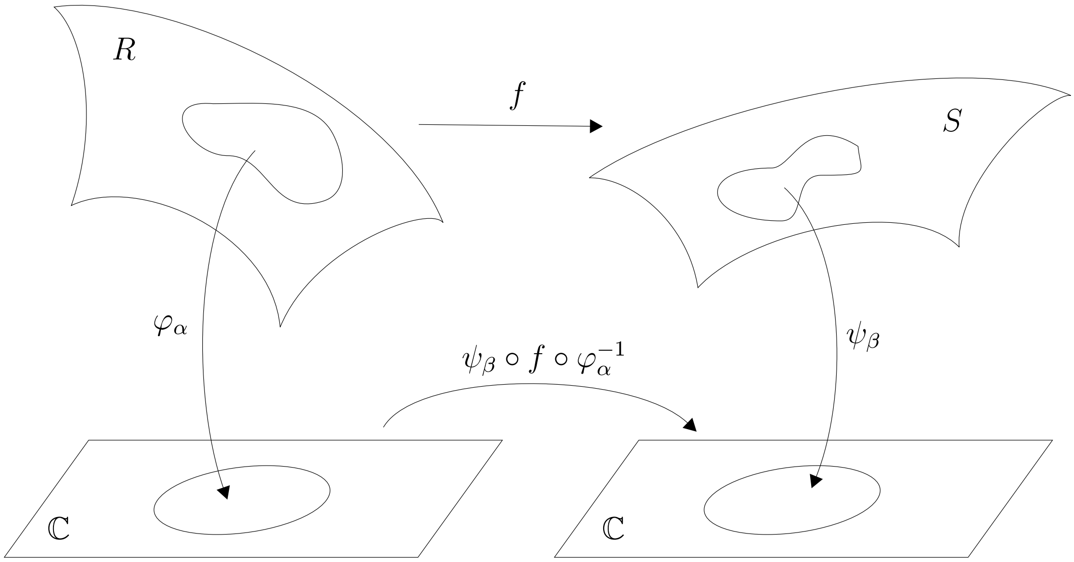

Figure 4: An analytic function between Riemann surfaces

In particular, a function $f : R \to \mathbb{C}$ is analytic if $f \circ \varphi_\alpha^{-1}$ is analytic on its domain in $\mathbb{C}$. By extension, we say that $f : R \to \mathbb{C}$ is *meromorphic* if $f \circ \varphi_\alpha^{-1}$ is meromorphic (i.e. analytic except at poles).

In fact, $f : R \to \mathbb{C}$ is meromorphic if and only if $f : R \to \mathbb{C}_\infty$ is analytic.

Definition. If $f : R \to S$ is analytic and a bijection then it is a theorem (see [5, §4.17]) that f^{-1} is also analytic and f is called a *conformal equivalence*. The Riemann surfaces R and S are said to be *conformally equivalent*.

Conformal equivalence is a notion of two Riemann surfaces being 'essentially the same'. We do not generally distinguish between conformally equivalent Riemann surfaces.

Our abstract definition of a Riemann surface differs from that used by Riemann himself. In his day, Riemann surfaces were defined as the most general domains on which analytic functions could be considered. A function such as $\sqrt{z}$ or $\log z$, initially defined on a small neighbourhood, can be analytically continued to a larger domain. If we are restricted to the complex plane this continuation may be 'multi-valued'. Riemann surfaces were conceived as multi-sheeted coverings of $\mathbb{C}$ upon which these analytic continuations become single-valued functions. Taking $\sqrt{z}$ as an example, the corresponding Riemann surface is a two-sheeted covering of $\mathbb{C}$ *branched* over the origin—each time a path encircles the origin it moves from one sheet of the covering to the other. Whether we take 2 or -2 as the square root of 4 depends upon which sheet of the covering we are on. In this way $\sqrt{z}$, considered on a Riemann surface, is rendered single-valued. Similarly, the Riemann surface that is the most general domain for $\sqrt[n]{z}$ is an n-sheeted covering of $\mathbb{C}$, while $\log z$ leads to an infinitely-sheeted covering.

Taking Riemann's viewpoint there is no question about the existence of non-constant meromorphic functions on Riemann surfaces. However, using the modern, complex manifold definition the existence question is certainly non-trivial. By answering it, we demonstrate a consistency between modern Riemann surfaces and the spirit of Riemann's original definition.

3.2 Uniformisation

Given a domain $D \subseteq \mathbb{C}$ and a discontinuous group Γ acting on D we define the quotient D/Γ to be the set of orbits of Γ, that is,

$$D/\Gamma = \{\Gamma z : z \in D\}.$$

The *quotient topology* on D/Γ is the coarsest topology which makes the *projection map*,

$$\begin{array}{rcl} \pi : D & \longrightarrow & D/\Gamma \\ z & \longmapsto & \Gamma z \end{array}$$

continuous. Explicitly, $U \subseteq D/\Gamma$ is open iff $\pi^{-1}(U)$ is open.

Observe that D/Γ is in one-to-one correspondence with any fundamental set for Γ. In Section 6 we exploit this for a Fuchsian group Γ in order to give $\mathbb{H}/\Gamma$ the structure of a Riemann surface obtained by identifying equivalent boundary points of a normal polygon.

Theorem 3.1 (Uniformisation)**.** *For any Riemann surface S there exists a domain D equal to either $\mathbb{C}$, $\mathbb{C}_\infty$ or $\mathbb{D}$ and a discontinuous group Γ acting on D such that the quotient space D/Γ can be given a Riemann surface structure conformally equivalent to S. Moreover, the projection map π is locally conformal—every point of D has a neighbourhood U such that the restriction $\pi|_U$ is conformal.*

No non-identity element of Γ has any fixed points in D.

The domain D is called the *universal covering surface* of the Riemann surface.

A complete proof of this profound result was eventually given by Koebe in 1912. A modern account may be found in Sections 9.1 and 9.2 of [6].

At this point automorphic functions re-enter the story. If S is a Riemann surface conformally equivalent to D/Γ then the automorphic functions with respect to Γ correspond precisely to the meromorphic functions on S.

This correspondence is given by $F = f \circ \pi$ where F is an automorphic function on Γ and f is a meromorphic function on D/Γ.

$$\begin{tikzcd}
 & D \arrow[r, "F", "\text{automorphic}"'] \arrow[d, "\pi", "\substack{\text{locally}\\\text{conformal}}"'] & \mathbb{C}_\infty \\
S \arrow[r, leftrightarrow, "\substack{\text{conformal}\\\text{equivalence}}"'] & D/\Gamma \arrow[r, "f", "\text{meromorphic}"'] & \mathbb{C}_\infty
\end{tikzcd}$$

- Given f meromorphic, we see by composition that $F(z) = (f \circ \pi)(z)$ is meromorphic. For $T \in \Gamma$ we have that $F(Tz) = f(\pi(Tz)) = f(\pi(z)) = F(z)$. Therefore F is automorphic.

- Given F automorphic, $f(z) = F(\pi^{-1}(z))$ is well-defined since F takes the same value at all points of the orbit $\pi^{-1}(z)$. The local conformality of π ensures that f is meromorphic.

This correspondence is actually an isomorphism between the fields of meromorphic functions on D/Γ and automorphic functions on Γ.

This result is of paramount importance in our attempt to prove the existence of non-constant meromorphic functions on any Riemann surface. We are now well placed to show this for Riemann surfaces with universal covering surface $\mathbb{D}$—by the above discussion this is equivalent to the existence of non-constant automorphic functions on a Fuchsian group. This will be done in Section 4 but we delay the proof for a moment so that we may deal with the other possible universal covering surfaces.

3.3 Riemann Surfaces with $\mathbb{C}$ or $\mathbb{C}_\infty$ as the Universal Covering Surface

By classifying those Riemann surfaces which have either $\mathbb{C}$ or $\mathbb{C}_\infty$ as their universal covering surface we shall show that they admit non-constant meromorphic functions.

All Möbius transformations have fixed points in $\mathbb{C}_\infty$. Hence, referring back to the Uniformisation Theorem, we see that (up to conformal equivalence) the only Riemann surface with $\mathbb{C}_\infty$ as the universal covering surface is $\mathbb{C}_\infty$ itself. It is well known that the rational functions are the meromorphic functions on $\mathbb{C}_\infty$.

We now turn our attention to Riemann surfaces conformally equivalent to $\mathbb{C}/\Gamma$, where Γ is a discontinuous group of Möbius maps having no fixed points in $\mathbb{C}$ (except the identity map). Hence Γ consists of transformations of the form $z \mapsto z + \lambda$.

Theorem 3.2. *If $\Gamma = \{z \mapsto z + \lambda : \lambda \in \Lambda\}$ is a non-trivial discontinuous group of translations of $\mathbb{C}$ then either,*

(i) $\Lambda = \{n\alpha : n \in \mathbb{Z}\} = \mathbb{Z}\alpha$ *for some* $\alpha \in \mathbb{C} \setminus \{0\}$, *or*

(ii) $\Lambda = \{n\alpha + m\beta : n, m \in \mathbb{Z}\} = \mathbb{Z}\alpha + \mathbb{Z}\beta$ *for some* $\alpha, \beta \in \mathbb{C} \setminus \{0\}, \quad \alpha/\beta \notin \mathbb{R}$.

Proof. The lattice Λ is the orbit of the origin so it cannot accumulate in $\mathbb{C}$, i.e. Λ is a discrete set. Note also that Λ is an additive subgroup of the complex plane.

Since Γ is non-trivial and Λ is discrete we may pick $\alpha \in \Lambda \setminus \{0\}$ with $|\alpha|$ minimal. If $\Lambda = \mathbb{Z}\alpha$ then we have case (i).

Otherwise, by again appealing to the discreteness of Λ, pick $\beta \in \Lambda \setminus \mathbb{Z}\alpha$ with $|\beta|$ minimal. We are required to show that we now have case (ii).

Suppose that $\alpha/\beta \in \mathbb{R}$. Let $k = \lfloor \beta/\alpha \rfloor$ be the greatest integer less than or equal to β/α. Then,

$$\beta - k\alpha \in \mathbb{Z}\alpha + \mathbb{Z}\beta$$

but

$$|\beta - k\alpha| = |\beta/\alpha - k||\alpha| < |\alpha|.$$

So $|\alpha|$ minimal $\Rightarrow$ $\beta - k\alpha = 0$ $\Rightarrow$ $\beta \in \mathbb{Z}\alpha$. A contradiction, therefore $\alpha/\beta \notin \mathbb{R}$.

Suppose now that $\gamma \in \Lambda \setminus (\mathbb{Z}\alpha + \mathbb{Z}\beta)$. Let $n\alpha + m\beta$ be a closest point of $\mathbb{Z}\alpha + \mathbb{Z}\beta$ to γ. The point γ lies in some parallelogram with vertices in $\mathbb{Z}\alpha + \mathbb{Z}\beta$. One possibility is shown in Figure 5.

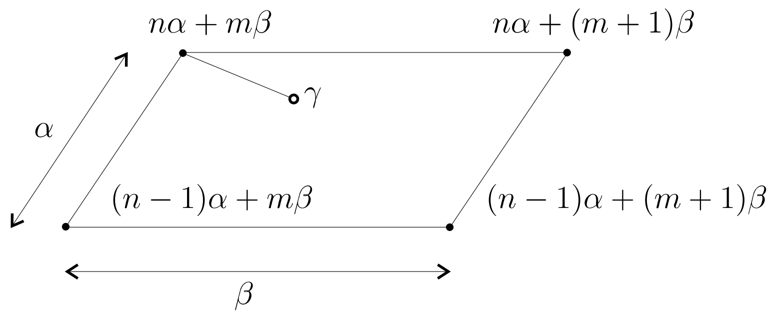

Figure 5: The $\alpha/\beta \notin \mathbb{R}$ condition ensures that the parallelogram is not degenerate

Straightforward geometry gives that,

$$|(n\alpha + m\beta) - \gamma| < |\beta|,$$

so our choice of β implies that either $\gamma \in \mathbb{Z}\alpha$ or $\gamma - (n\alpha + m\beta) = 0$. Both cases are a contradiction.
Therefore $\Lambda = \mathbb{Z}\alpha + \mathbb{Z}\beta$. □

The problem of finding a non-constant meromorphic function on a Riemann surface with universal covering surface $\mathbb{C}$ is reduced to that of finding an automorphic function on (i) $\mathbb{C}/\mathbb{Z}\alpha$ and (ii) $\mathbb{C}/(\mathbb{Z}\alpha + \mathbb{Z}\beta)$. The first case is easily settled by taking, for example, the function $z \mapsto e^{\frac{2\pi i z}{\alpha}}$. If α were real or pure imaginary we could take trigonometric or hyperbolic functions respectively.

To tackle the second case we must find a meromorphic function on $\mathbb{C}$ that is invariant under translations $z \mapsto z + \lambda$, where $\lambda \in \mathbb{Z}\alpha + \mathbb{Z}\beta$. Such a function is called a *doubly-periodic* or *elliptic* function with *periods* α and β.

We now take the periods to be 1 and τ where $\text{Im } \tau > 0$. To see that this incurs no loss of generality, note that $\alpha/\beta \notin \mathbb{R}$ implies that either $\text{Im}(\alpha/\beta) > 0$ or $\text{Im}(\beta/\alpha) > 0$. Take $\tau = \alpha/\beta$ or β/α appropriately and let $f(z)$ be elliptic with periods 1 and τ. Then $f(\beta z)$ or $f(\alpha z)$ is the required elliptic function with periods α and β.

We begin our search for elliptic functions by introducing theta functions. While not doubly-periodic themselves they have some of the properties we are interested in and we shall later build an elliptic function from them.

Definition. For $\text{Im } \tau > 0$ define the *theta function*,

$$\vartheta(z) = \sum_{n=-\infty}^{\infty} \exp 2\pi i \left(\frac{1}{2} n^2 \tau + nz \right)$$

Lemma 3.3. *The function $\vartheta(z)$ is analytic.*

Proof. We use the Weierstrass M-test to show that the series converges uniformly on

$$S_C = \{z \in \mathbb{C} : |\text{Im } z| \leq C\} \quad \text{for} \quad C > 0$$

Write $z = x + iy$ and $\tau = \tau_R + i\tau_I$.

$$\begin{aligned}
\left| \exp 2\pi i \left(\frac{1}{2} n^2 \tau + nz \right) \right| &= \exp\left(-\pi \tau_I n^2 - 2n\pi y\right) \\
&\leq \exp\left(-\pi \tau_I n^2 + 2n\pi C\right) \\
&= \exp\left(-\pi \tau_I \left(n - \frac{C}{\tau_I} \right)^2 + \pi \frac{C^2}{\tau_I} \right) \\
&= M_C \exp\left(-\pi \tau_I \left(n - \frac{C}{\tau_I} \right)^2 \right)
\end{aligned}$$

Where the constant M_C is independent of z and n. By comparison with a geometric series we see that,

$$\sum_{n=-\infty}^{\infty} M_C \exp\left(-\pi\tau_I\left(n-\frac{C}{\tau_I}\right)^2\right)$$

converges since $\tau_I > 0$. Therefore, the M-test tells us that the series defining $\vartheta(z)$ converges uniformly on S_C. A uniformly convergent series of analytic functions converges to an analytic function. Hence $\vartheta(z)$ is analytic on $\mathbb{C}$ since C was arbitrary. $\square$

Lemma 3.4.

(i) $\vartheta(z+1)=\vartheta(z)$. *The theta function is 1-periodic.*

(ii) $\vartheta(z+\tau)=\exp\left(2\pi i\left(-\frac{1}{2}\tau-z\right)\right)\vartheta(z)$

Proof. (i) Clear from the series definition of $\vartheta(z)$.

(ii)
$$\begin{aligned}\vartheta(z+\tau) &= \sum_{n=-\infty}^{\infty}\exp 2\pi i\left(\frac{1}{2}n^2\tau+n(z+\tau)\right)\\ &= \sum_{n=-\infty}^{\infty}\exp 2\pi i\left(\frac{1}{2}(n+1)^2\tau-\frac{1}{2}\tau+nz\right)\\ &= \exp 2\pi i\left(-\frac{1}{2}\tau-z\right)\sum_{n=-\infty}^{\infty}\exp 2\pi i\left(\frac{1}{2}(n+1)^2\tau+(n+1)z\right)\\ &= \exp 2\pi i\left(-\frac{1}{2}\tau-z\right)\vartheta(z)\end{aligned}$$
$\square$

We finish this section by writing down a meromorphic function invariant under $z\mapsto z+\lambda$, $\lambda\in\mathbb{Z}+\mathbb{Z}\tau$.

Theorem 3.5. *There exists an elliptic function.*

Proof. Let

$$f(z)=\left(\frac{\vartheta\left(z-\frac{\tau}{2}\right)}{\vartheta\left(z-\frac{\tau}{2}-\frac{1}{2}\right)}\right)^2$$

Since $\vartheta(z)$ is 1-periodic and analytic, f is 1-periodic and meromorphic.

Finally, using Lemma 3.4 (ii),

$$\begin{aligned}f(z+\tau) &= \left(\frac{\exp 2\pi i\left(-\frac{\tau}{2}-\left(z-\frac{\tau}{2}\right)\right)\vartheta\left(z-\frac{\tau}{2}\right)}{\exp 2\pi i\left(-\frac{\tau}{2}-\left(z-\frac{\tau}{2}-\frac{1}{2}\right)\right)\vartheta\left(z-\frac{\tau}{2}-\frac{1}{2}\right)}\right)^2\\ &= \frac{\exp 2\pi i(-2z)}{\exp 2\pi i(1-2z)}\left(\frac{\vartheta\left(z-\frac{\tau}{2}\right)}{\vartheta\left(z-\frac{\tau}{2}-\frac{1}{2}\right)}\right)^2\\ &= f(z)\end{aligned}$$
$\square$

4 Existence of Automorphic Functions on Fuchsian Groups

We now complete the proof of the existence of non-constant meromorphic functions on a Riemann surface by constructing automorphic functions on Fuchsian groups. This covers the one remaining case of Riemann surfaces with $\mathbb{D}$ as the universal covering surface.

4.1 Automorphic Forms

We first introduce automorphic forms, whose role in this proof is analogous to that of the theta function in the construction of an elliptic function.

Definition. Let Γ be a Fuchsian group on $\mathbb{D}$. A meromorphic function $F(z)$ on $\mathbb{D}$ satisfying,

$$F(Tz) = (bz + \bar{a})^r F(z) \quad \text{for all} \quad T = \begin{pmatrix} a & \bar{b} \\ b & \bar{a} \end{pmatrix} \in \Gamma$$

is an *automorphic form of dimension* $-r$.

Since $\frac{dTz}{dz} = (bz + \bar{a})^{-2}$ this condition has the equivalent form,

$$F(Tz)(dTz)^{\frac{r}{2}} = F(z)(dz)^{\frac{r}{2}}$$

Note that an automorphic form of zero dimension is an automorphic function. Furthermore, if F_1, F_2 are two automorphic forms of equal dimension and $F_2(z) \not\equiv 0$ then $\frac{F_1(z)}{F_2(z)}$ is an automorphic function.

As a brief aside we now explore how automorphic forms correspond to *differentials* on the Riemann surface $\mathbb{D}/\Gamma$.

Definition. Let S be a Riemann surface. Let $\mathcal{U} \subseteq \mathcal{S}/\Gamma$ be a co-ordinate neighbourhood and let φ be a chart on $\mathcal{U}$. Then $t = \varphi(q)$ is called a *local variable* at $q \in \mathcal{U}$. It clearly depends on the choice of chart.

Definition. Let S be a Riemann surface. A *meromorphic differential of weight* m assigns to each co-ordinate neighbourhood of S, and each choice of local variable t, a meromorphic function $\eta(t)$ obeying the transformation law,

$$\eta_1(t_1)(dt_1)^m = \eta_2(t_2)(dt_2)^m$$

where η_i is the function corresponding to the local variable t_i for $i = 1, 2$.

If $t_1 = \varphi_1(q)$ and $t_2 = \varphi_2(q)$ then this can be expressed in terms of the transition function $t_1 = \varphi_1 \circ \varphi_2^{-1}(t_2)$:

$$\eta_2(t_2) = \left(\frac{d}{dt_2} (\varphi_1 \circ \varphi_2^{-1})(t_2) \right)^m \eta_1 \left(\varphi_1 \circ \varphi_2^{-1}(t_2) \right).$$

Theorem 4.1. *Let $\pi : \mathbb{D} \to \mathbb{D}/\Gamma$ be the projection map and let t be a local variable at $q = \pi(z)$. An automorphic form $F(z)$ of dimension $-2m$ on $\mathbb{D}$ corresponds to a meromorphic differential $\eta(t)(dt)^m$ on $\mathbb{D}/\Gamma$, of weight m, where*

$$F(z)(dz)^m = \eta(t)(dt)^m.$$

Proof. The meromorphicity of one implies it for the other.

Suppose first that the differential is given. Under a transformation $T \in \Gamma$, the projection $\pi(Tz) = \pi(z) = q$ is unchanged, so t is unchanged. Hence,

$$F(Tz)(dTz)^m = F(z)(dz)^m$$

and so $F(z)$ is an automorphic form of dimension $-2m$.

Now suppose instead that the automorphic form F is given. The choice of inverse $z = \pi^{-1}(q)$ is immaterial since $F(z)(dz)^m$ takes the same value for any choice. Since z depends only on q and not on the choice of local variable,

$$\eta_1(t_1)(dt_1)^m = F(z)(dz)^m = \eta_2(t_2)(dt_2)^m$$

so $\eta(t)(dt)^m$ is indeed a meromorphic differential of weight m. $\square$

In fact, the sets of automorphic forms and meromorphic differentials are complex vector spaces and the above correspondence is an isomorphism.

4.2 Existence

We shall now construct a family of automorphic forms from which the required automorphic functions can be built.

Let $T_0 = I, T_1, T_2, \ldots$ where $T_n = \begin{pmatrix} a_n & \bar{b}_n \\ b_n & \bar{a}_n \end{pmatrix}$ be the elements of a Fuchsian group Γ on $\mathbb{D}$ (any such group is countable by Lemma 1.3). We impose the condition that ∞ is only fixed by $T_0 = I$. If H is a rational function with no poles on $\partial\mathbb{D}$ then the Poincaré series,

$$F(z) = \sum_{n=0}^{\infty} H(T_n z)(b_n z + \bar{a}_n)^{-r} \qquad r \geq 4, \quad r \text{ is an integer}$$

is an automorphic form of dimension $-r$. Temporarily assuming appropriate convergence, we see that F transforms correctly under $V = \begin{pmatrix} \alpha & \bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix} \in \Gamma$;

$$\begin{aligned} F(Vz) &= \sum_{n=0}^{\infty} H(T_n V z)\left(b_n \frac{\alpha z + \bar{\beta}}{\beta z + \bar{\alpha}} + \bar{a}_n\right)^{-r} \\ &= (\beta z + \bar{\alpha})^r \sum_{n=0}^{\infty} H(T_n V z)((b_n \alpha + \bar{a}_n \beta)z + b_n \bar{\beta} + \bar{a}_n \bar{\alpha}) \\ &= (\beta z + \bar{\alpha})^r F(z) \end{aligned}$$

The final equality following from the observation that $T_n V = \begin{pmatrix} \cdot & \cdot \\ b_n \alpha + \bar{a}_n \beta & b_n \bar{\beta} + \bar{a}_n \bar{\alpha} \end{pmatrix}$ and that $\Gamma = \Gamma V$.

The next theorem will be of crucial importance in the proof of the convergence of the Poincaré series.

Theorem 4.2. *If $\Gamma = \{I, T_1, T_2, \ldots\}$ as above and $T_n(\infty) \neq \infty$ for $n \geq 1$ then the series,*

$$\sum_{n=0}^{\infty} |b_n z + \bar{a}_n|^{-r} \qquad r \geq 4, \quad r \text{ is an integer}$$

converges locally uniformly on $\mathbb{D}$.

Proof. Since no non-identity element of Γ fixes ∞, we can appeal to the results of Section 2.2. Recall that we defined R_0 as the region external to all of the isometric circles $\mathbf{I}(T_n)$ for $n \geq 1$. By Theorem 2.5 there exists $\rho > 0$ such that $\{z : |z| > \rho\} \subseteq R_0$.

Let $R_n = T_n(R_0)$. By Theorem 2.2, R_0 does not contain equivalent points—hence the R_n are disjoint. (If $\zeta \in R_i \cap R_j$ then $T_i^{-1}\zeta$ and $T_j^{-1}\zeta$ are equivalent points of R_0.)

As R_0 lies outside $\mathbf{I}(T_n)$, Lemma 2.1 tells us that R_n lies within $\mathbf{I}(T_n^{-1})$.

Let K be a finite disc in R_0 and, for $n \geq 1$, let $K_n = T_n(K) \subseteq R_n$. The K_n are disjoint and lie within isometric circles. Therefore,

$$\sum_{n=1}^{\infty} \text{Area } (K_n) \leq \pi\rho^2.$$

Writing $u + iv = T_n(x + iy)$ and recalling the Cauchy-Riemann relations we calculate the Jacobian,

$$\begin{aligned} \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} &= \begin{vmatrix} \frac{\partial u}{\partial x} & -\frac{\partial v}{\partial x} \\ \frac{\partial v}{\partial x} & \frac{\partial u}{\partial x} \end{vmatrix} = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 \\ &= \left|\frac{\partial T_n(x+iy)}{\partial x}\right|^2 = |T_n'(z)|^2 \\ &= |b_n z + \bar{a}_n|^{-4} \end{aligned}$$

So that,

$$\begin{aligned} \text{Area}(K_n) &= \iint_{K_n} du\, dv = \iint_K \frac{dx\, dy}{|b_n z + \bar{a}_n|^4} \\ &= |b_n|^{-4} \iint_K \left| z + \frac{\bar{a}_n}{b_n} \right|^{-4} dx\, dy \\ &\geq |b_n|^{-4} \iint_K \frac{dx\, dy}{(|z| + \rho)^4} \\ &= |b_n|^{-4} M \end{aligned}$$

Where M, the value of the integral, is finite since K is finite. The inequality is just the triangle rule together with the observation that $-\bar{a}_n/b_n$ is the centre of $\mathbf{I}(T_n)$, so that $|\bar{a}_n/b_n| \leq \rho$. Therefore,

$$\sum_{n=1}^{\infty} |b_n|^{-4} \leq \frac{1}{M} \sum_{n=1}^{\infty} \text{Area}(K_n) \leq \frac{\pi\rho^2}{M}$$

By Remark 2.4, $|b_n| \to \infty$ so that for $r \geq 4$ we have $|b_n|^{-r} \leq |b_n|^{-4}$ for sufficiently large n. Therefore, by comparison with the above we have established the convergence of the sequence,

$$\sum_{n=1}^{\infty} |b_n|^{-r}$$

We now prove the theorem. The $n = 0$ term is $|b_0 z + \bar{a}_0|^{-r} = 1$ so it may safely be ignored. The condition that $|a_n|^2 - |b_n|^2 = 1$ implies that $|a_n| > |b_n|$. This means that the points $-\bar{a}_n/b_n$ lie outside the unit disc. Let Q be a compact subset of $\mathbb{D}$. Let $\delta > 0$ be the minimum distance between Q and $\partial\mathbb{D}$. Then for $z \in Q$,

$$\left| z + \frac{\bar{a}_n}{b_n} \right| > \delta$$

$$\Rightarrow \quad \sum_{n=1}^{\infty} |b_n z + \bar{a}_n|^{-r} = \sum_{n=1}^{\infty} |b_n|^{-r} \left| z + \frac{\bar{a}_n}{b_n} \right|^{-r} \leq \delta^{-r} \sum_{n=1}^{\infty} |b_n|^{-r}$$

So the series converges uniformly on Q. $\square$

By proving the convergence assumed in the discussion at the beginning of this section, the following theorem establishes the existence of an automorphic form on Γ.

Theorem 4.3. *For Γ as in the previous theorem and H a rational function whose set of poles P does not meet $\partial\mathbb{D}$, the Poincaré series,*

$$F(z) = \sum_{n=0}^{\infty} H(T_n z)(b_n z + \bar{a}_n)^{-r} \qquad r \geq 4, \quad r \text{ is an integer}$$

(i) converges absolutely locally uniformly on $\mathbb{D} \setminus \Gamma(P)$

(ii) is meromorphic on $\mathbb{D}$.

Proof.

(i) Let Q be a compact subset of $\mathbb{D} \setminus \Gamma(P)$.

$$Q \cap \Gamma(P) = \emptyset \quad \Rightarrow \quad \Gamma(Q) \cap P = \emptyset$$

$\Gamma(Q)$ can only accumulate at limit points of Γ, which by Remark 1.6 lie in $\partial\mathbb{D}$ and so outside P. Therefore $\overline{\Gamma(Q)}$ is at a positive distance from P. Hence $H(T_n z)$ is uniformly bounded on Q, with the bound, say C, independent of n. Therefore, for $z \in Q$,

$$\sum_{n=0}^{\infty} |H(T_n z)(b_n z + \bar{a}_n)^{-r}| \leq C \sum_{n=0}^{\infty} |b_n z + \bar{a}_n|^{-r}$$

which converges uniformly on Q by the previous theorem.

(ii) If $z_0 \in \mathbb{D} \setminus \Gamma(P)$ then $\Gamma(P)$ cannot accumulate at z_0 since it only accumulates on $\partial\mathbb{D}$. Therefore, z_0 lies in a compact neighbourhood Q as in (i), and the Poincaré series converges uniformly on Q. Since each term in the series is analytic on Q (recall that $-\bar{a}_n/b_n \notin \mathbb{D}$) the function $F(z)$ is analytic at $z_0 \in Q$.

Now suppose that $z_0 \in \mathbb{D} \cap \Gamma(P)$. Then $T_n z_0 \in P$ for at most finitely many n, since P is a finite set disjoint from the limit set of Γ. Let N be maximal such that $T_N z_0 \in P$. Then the finite sum,

$$\sum_{n=0}^{N} H(T_n z)(b_n z + \bar{a}_n)^{-r}$$

is meromorphic on $\mathbb{C}$. The remainder of the series converges uniformly on a compact neighbourhood of z_0. To see this, let $Q \subseteq \mathbb{D}$ be a compact neighbourhood of z_0 that does not meet

$\Gamma(P) \setminus \{z_0\}$. (As before, this is possible because $\Gamma(P)$ can only accumulate on $\partial\mathbb{D}$.) Then the argument of (i) applies, so that,

$$\sum_{n=N+1}^{\infty} H(T_n z)(b_n z + \bar{a}_n)^{-r}$$

is analytic at z_0.

Therefore, $F(z)$ is meromorphic at z_0, and so on all of $\mathbb{D}$. □

We have now proved that the Poincaré series defines an automorphic form. To see that this may be non-constant pick $\alpha \in \mathbb{D}$ that is not fixed by any non-identity element of Γ. Setting $H(z) = (z-\alpha)^{-k}$,

$$F_k(z) = \sum_{n=0}^{\infty} (T_n z - \alpha)^{-k} (b_n z + \bar{a}_n)^{-r} \qquad r \geq 4, \quad r \text{ is an integer}$$

must have a pole of order k at $z = \alpha$, since the first term has such a pole, yet the sum of the other terms is analytic at α. We deduce that,

$$f(z) = \frac{F_2(z)}{F_1(z)}$$

is our long-sought automorphic function. The simple pole at α implies that f is non-constant.

We have in fact proved more. A quick exercise shows that the quotient of an automorphic form of dimension r_1 by one of dimension r_2 gives an automorphic form of dimension $r_1 - r_2$. Hence there exist non-constant automorphic forms of all positive and negative dimensions.

We must now lift the restriction that no non-identity element of Γ fixes ∞. There must exist some point $\zeta \in \mathbb{C} \setminus \overline{\mathbb{D}}$ such that $T_n \zeta \neq \zeta$ for all $n \geq 1$ (i.e. by a countability argument). Let A be a Möbius map fixing $\mathbb{D}$ and sending ζ to ∞. Then ∞ is not fixed by any non-identity element of the transformed group $A\Gamma A^{-1}$, so there exist non-constant automorphic functions and forms with respect to $A\Gamma A^{-1}$.

Definition. If F is an automorphic form of dimension $-2m$ with respect to $A\Gamma A^{-1}$, then the A^{-1}-*transform* of F is,

$$F_{A^{-1}}(z) = F(Az) \left(\frac{dAz}{dz} \right)^m$$

Theorem 4.4. *Let F be an automorphic form of dimension $-2m$ with respect to $A\Gamma A^{-1}$. Then $F_{A^{-1}}(z)$ is an automorphic form of dimension $-2m$ with respect to Γ.*

Proof. $F_{A^{-1}}(z)$ is meromorphic on $\mathbb{D}$ since F and A are.

We now verify the transformation condition,

$$\begin{array}{rcll} F_{A^{-1}}(Tz)(dTz)^m & = & F(ATz)(dATz)^m & \text{by definition of } F_{A^{-1}} \\ & = & F((ATA^{-1})Az)(d(ATA^{-1})(Az))^m & \text{inserting } A^{-1}A = I \\ & = & F(Az)(dAz)^m & \text{since } F \text{ automorphic on } A\Gamma A^{-1} \\ & = & F_{A^{-1}}(z)(dz)^m & \end{array}$$

□

This means that given any Fuchsian group Γ on $\mathbb{D}$ we can find non-constant automorphic functions and forms of any dimension on a transformed group $A\Gamma A^{-1}$. By taking the A^{-1}-transform we get non-constant automorphic functions and forms on Γ. Hence, given a Riemann surface $S = \mathbb{D}/\Gamma$, we have established the existence of non-constant meromorphic functions and differentials on S.

We have now dealt with all the cases presented by the Uniformisation Theorem and so have completed our proof of the existence of non-constant meromorphic functions on an arbitrary Riemann surface.

5 A Return to Geometry

The Uniformisation Theorem realises all Riemann surfaces as quotients by discontinuous groups. The goal of the remainder of this essay is to show how a Riemann surface structure may be put on a certain class of quotient spaces. From now on, we shall concern ourselves with Fuchsian groups on $\mathbb{H}$, that is, discrete subgroups of the group of real Möbius maps (i.e. maps with real coefficients). To get us started quickly, several preliminary results will be stated without proof.

5.1 Fixed Points

A point $z \in \mathbb{C}_\infty$ is a *fixed point* of a Möbius map T if $Tz = z$. It is clear that a non-identity map has at most two fixed points. If $T : z \mapsto \frac{az+b}{cz+d}$ has exactly one fixed point then it is called a *parabolic* fixed point. A quick calculation shows that if a, b, c, d are real then a parabolic fixed point is either ∞ or is real. If T has real coefficients and two distinct fixed points then it is conjugate[4] to one of the following:

(i) $z \mapsto e^{i\theta} z$ for some $\theta \in (0, 2\pi)$. In this case the fixed points form a complex conjugate pair and are said to be *elliptic*.

(ii) $z \mapsto kz$ for some $k \in (0, 1)$. In this case the fixed points are real and are said to be *hyperbolic*.

Theorem 5.1. *If Γ is a Fuchsian group on $\mathbb{H}$ and w is an elliptic fixed point of an element of Γ then the* stabiliser, *$\Gamma_w = \{T \in \Gamma : Tw = w\}$, is a finite cyclic subgroup of Γ. The* order *of w is defined to be the order of Γ_w. If p is a parabolic fixed point then Γ_p is an infinite cyclic subgroup of Γ.*

See [1, p15] for a proof.

Remark 5.2. It follows that when w is an elliptic fixed point of order l, with E the generator of Γ_w and $A(z) = \frac{z-w}{z-\bar{w}}$,

$$AEA^{-1}(z) = e^{\pm \frac{2\pi i}{l}} z$$

Definition. A *fixed circle* of a Möbius transformation is a circle or straight line that is mapped onto itself.

Remark 5.3. If C is a fixed circle of T then $A(C)$ is a fixed circle of $T' = ATA^{-1}$.

[4] Two maps T and S are *conjugate* if $S = ATA^{-1}$ for some A.

If T is a real Möbius map with a parabolic fixed point then it is conjugate to a translation (with unique fixed point at ∞). The fixed circles of a translation are clearly straight lines in the direction of translation. The fixed circles of T are therefore circles through the parabolic fixed point. Since T maps $\mathbb{H}$ to itself and the parabolic fixed point is real, the fixed circles must be tangent to $\mathbb{R}$ at the fixed point.

If T has an elliptic fixed point w then it is conjugate (by a map sending $w \mapsto 0, \bar{w} \mapsto \infty$) to a rotation about the origin. The fixed circles of this rotation are clearly circles centred on the origin. It follows that the fixed circles of T are orthogonal to circular arcs joining w and $\bar{w}$ and each fixed circle encloses either w or $\bar{w}$.

5.2 Normal Polygons

By constructing a normal polygon we obtain a fundamental region with many important properties. In this section we state a variety of results that will be useful later. Essentially, a normal polygon is the set of points that, under the hyperbolic metric, are strictly nearer to some w_0 than to any other point of Γw_0.

Definition. For Γ a Fuchsian group on $\mathbb{H}$ we define the *normal polygon* N_0 with *centre* $w_0 \in \mathbb{H}$ to be,

$$\{z \in \mathbb{H} : \rho(z, w_0) < \rho(z, Tw_0) \text{ for all } T \in \Gamma \setminus \{I\}\}$$

where the hyperbolic metric ρ is derived from the Riemannian metric $ds = \dfrac{|dz|}{\text{Im } z}$ on $\mathbb{H}$. We require that w_0 is not fixed by any element of Γ.

The geodesics of hyperbolic geometry are called *hyperbolic lines*. They are circles and straight lines orthogonal to the real axis.

Normal polygons are similarly defined for Fuchsian groups on $\mathbb{D}$. The intersection of $\mathbb{D}$ with the region R_0 defined using isometric circles in Section 2.2 is actually the normal polygon with the origin as the centre (see [2, p151]).

This raises an interesting historical point: Poincaré introduced the use of hyperbolic geometry in this subject in the 1880s. Considering conformal mappings as hyperbolic isometries is a powerful technique for studying all types of discontinuous groups. Indeed, hyperbolic geometry is the natural geometry to use in many areas of complex analysis and its applications are widespread. However, some decades after Poincaré, Ford introduced the isometric circle in an attempt to simplify the subject by removing all reference to non-Euclidean geometry. There is some small merit to this: we have set up the basics of the theory without (yet) using hyperbolic geometry and Poincaré himself, despite arriving at his results through hyperbolic geometry, avoided any use of it in his papers, for fear of causing confusion. However, to proceed any further the use of non-Euclidean geometry is virtually essential. The above remark about R_0 demonstrates how Ford's approach for Fuchsian groups is a very special case of what can be achieved using hyperbolic geometry.

Definition. A subset N of $\mathbb{H}$ is *H-convex* if the arc of a hyperbolic line joining any two points of N lies entirely within N.

It is apparent that a H-convex set is connected and, in fact, simply-connected.

Theorem 5.4. *A normal polygon is a fundamental region for Γ and is H-convex. Its boundary consists of hyperbolic line segments and possibly also points of $\mathbb{R}$.* [1, §I.4C–4D]

Definition. A maximal hyperbolic line segment lying on the boundary of a normal polygon N_0 is called a *side*. If two sides of N_0 meet at a parabolic fixed point $p \in \mathbb{R} \cup \{\infty\}$ then p is called a *parabolic vertex* of N_0.

Theorem 5.5. *Every parabolic fixed point is a parabolic vertex of some normal polygon N_0. Furthermore, if K is the interior of a fixed circle of p and $\Delta = \overline{N_0} \cap K$, then the images of Δ under Γ cover K. See Figure 6.* [1, p42–47]

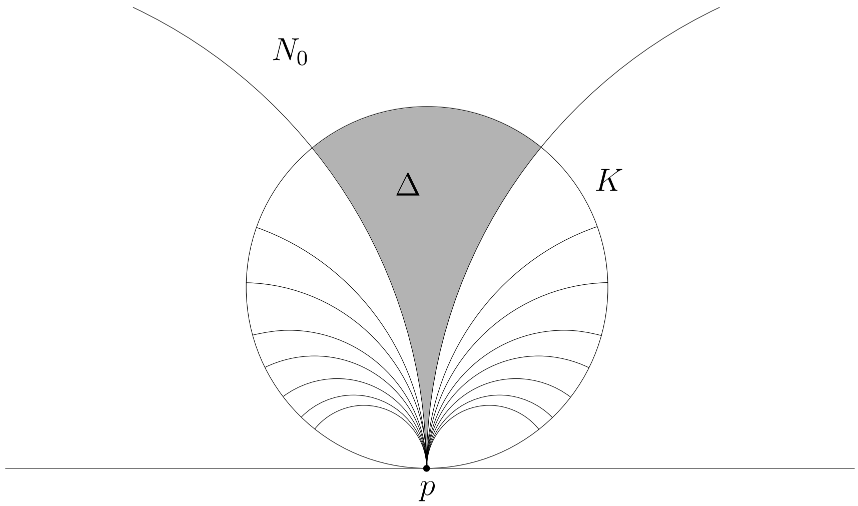

Figure 6: $\Gamma(\Delta)$ covers K

Remark 5.6. The image of a parabolic vertex is also a parabolic vertex: if p is a parabolic fixed point of $T \in \Gamma$ then for any $V \in \Gamma$ the point $V(p)$ is a parabolic fixed point of VTV^{-1}.

Definition. A fundamental region R for Γ is said to be *locally finite* if each compact subset of $\mathbb{H}$ intersects only finitely many of the images of $\overline{R}$ under Γ.

Theorem 5.7. *A normal polygon is locally finite.* [1, p30–31]

6 Quotient Spaces as Riemann Surfaces

In this final section we describe in detail how the quotient space of a Fuchsian group can be made into a Riemann surface.

Definition. For Γ a Fuchsian group on $\mathbb{H}$, let P denote its set of parabolic vertices. We then define $\mathbb{H}^+ = \mathbb{H} \cup P$.

Remark 5.6 implies that $\Gamma P = P$, so the group Γ acts on the space $\mathbb{H}^+$. We shall soon show how $\mathbb{H}^+/\Gamma$, the set of orbits of Γ in $\mathbb{H}^+$, can be endowed with the structure of a Riemann surface. First we look at a consequence of this result that is analogous to our earlier work.

Definition. We shall call an automorphic function $f(z)$ *simple* if, for each parabolic vertex p, $f(z)$ tends to a definite value (possibly ∞) as $z \to p$ from within a normal polygon.

Our discussion of Section 3.2 can be extended to give an isomorphism between the field of simple automorphic functions and the field of meromorphic functions on $\mathbb{H}^+/\Gamma$. In fact, by defining *simple* *automorphic forms* in the obvious way, Theorem 4.1 generalises to relate these to meromorphic differentials on $\mathbb{H}^+/\Gamma$. See [1, p131–132].

The Riemann surface $\mathbb{H}/\Gamma$ can be viewed as a 'punctured' version of $\mathbb{H}^+/\Gamma$ with the points corresponding to parabolic vertices removed. Then a meromorphic function on $\mathbb{H}/\Gamma$ can be extended to one on $\mathbb{H}^+/\Gamma$ if we can assign values to the punctured points in such a way that the function remains meromorphic. This can be done if and only if the original function corresponds to a *simple* automorphic function on Γ. If this is not the case, then the function has an 'essential singularity' at the punctured point.

The set $\mathbb{H}^+/\Gamma$ is clearly in one-to-one correspondence with any fundamental set for Γ relative to $\mathbb{H}^+$. Hence, we may picture the space $\mathbb{H}^+/\Gamma$ as a closed normal polygon of Γ with equivalent boundary points identified.

Example. Recall the fundamental region for the modular subgroup $\Gamma(2)$ of Section 2.3. Figure 1 shows that this fundamental region has parabolic vertices at 0, -1, 1 and ∞. The vertices at -1, 1 are equivalent under $\Gamma(2)$, as are the two vertical sides and the two semi-circular sides. If we were to glue together these identified sides we would obtain a sphere. The quotient $\mathbb{H}^+/\Gamma(2)$ is conformally equivalent to the Riemann sphere. Likewise, the punctured surface $\mathbb{H}/\Gamma(2)$ is a 3-punctured sphere—the punctured points corresponding to the distinct orbits of the parabolic vertices 0, 1 and ∞.

The first step towards making $\mathbb{H}^+/\Gamma$ into a Riemann surface is to define a basis, $\mathcal{B}$, for a topology on $\mathbb{H}^+$. Let $\mathcal{B}$ contain:

- All open discs in $\mathbb{H}$. That is, all sets of the form $\{z \in \mathbb{H} : |z - z_0| < r\}$ for some $z_0 \in \mathbb{H}$ and $0 < r \leq \text{Im } z$
- For each finite $p \in P$, and for each $K \subseteq \mathbb{H}$ a fixed circle of p, the set $\{p\} \cup \text{Int } K$
- If $\infty \in P$, all sets of the form $\{z : \text{Im } z > h\} \cup \{\infty\}$ for some $h > 0$

It is not hard to check that the basis $\mathcal{B}$ gives rise to a Hausdorff topology on $\mathbb{H}^+$. Note that the resulting subspace topology on $\mathbb{H}$ coincides with its usual topology. Hence the subspace $\mathbb{H}$ is connected. Any open set in $\mathbb{H}^+$ containing a point of P necessarily intersects $\mathbb{H}$, so it is not possible to write $\mathbb{H}^+$ as a disjoint union of two open sets. Therefore, $\mathbb{H}^+$ is a connected space.

An element of Γ, being a Möbius map, will carry one member of $\mathcal{B}$ onto another (recalling that $\Gamma P = P$). This implies that it is an open map under the topology for $\mathbb{H}^+$. It is in fact a homeomorphism, since the same is true of the inverse map.

The projection mapping,

$$\begin{aligned} \pi \ : \ \mathbb{H}^+ &\longrightarrow \mathbb{H}^+/\Gamma \\ z &\longmapsto \Gamma z \end{aligned}$$

is clearly surjective. As before, we can use this projection map to induce a topology on $\mathbb{H}^+/\Gamma$ from that of $\mathbb{H}^+$. We simply define a set $\mathcal{U} \subseteq \mathbb{H}^+/\Gamma$ to be open iff $\pi^{-1}(\mathcal{U})$ is open. The projection mapping is then automatically continuous and so $\mathbb{H}^+/\Gamma$ becomes a connected topological space.

The projection map is open. To prove this we must show that if $A \subseteq \mathbb{H}^+$ is open then $\pi(A) \subseteq \mathbb{H}^+/\Gamma$ is open. Under the induced topology this image is open iff $\pi^{-1}\{\pi(A)\}$ is open. This is just the set ΓA of images of A under elements of Γ, but these maps are homeomorphisms and so ΓA is indeed open.

Lemma 6.1. $\mathbb{H}^+/\Gamma$ *is a Hausdorff space.*

Proof. Let Γx and Γy be distinct points of $\mathbb{H}^+/\Gamma$. Let N_0 be a normal polygon for Γ. We can assume that $x, y \in \overline{N_0}$ since $\Gamma(Tx) = \Gamma x$ for all $T \in \Gamma$. We work through the possible cases:

1. Suppose first that $x, y \in \mathbb{H}$. Since x is an ordinary point of Γ there exists a closed disc B in $\mathbb{H}$ that contains x but no points of Γy. Let K be any compact neighbourhood of y. By the local finiteness of N_0 (Theorem 5.7), B meets only $V_1\overline{N_0}, \dots, V_n\overline{N_0}$ and K meets only $W_1\overline{N_0}, \dots, W_m\overline{N_0}$ for some $V_i, W_j \in \Gamma$. Now, if $T \in \Gamma$ and $T(B)$ meets K then it must meet some $W_j\overline{N_0}$ (since the $W_j\overline{N_0}$ cover K). Then B must meet $T^{-1}W_j\overline{N_0}$. Hence $T = W_jV_i^{-1}$ for some i, but there are only finitely many such T. We have thus shown that any such K meets only finitely many images of B. Therefore, we can choose K sufficiently small so that it meets no image of B, i.e. $\Gamma K \cap \Gamma B = \emptyset$. Let X and Y be the interiors of B and K respectively. Then, since π is an open map, $\pi(X)$ and $\pi(Y)$ are the required disjoint open neighbourhoods of Γx and Γy.

2. Now suppose that $x, y \in P$. Let S_x and S_y be disjoint sets in $\mathcal{B}$ such that $x \in S_x$ and $y \in S_y$. Let $\Delta_x = S_x \cap N_0$ and $\Delta_y = S_y \cap N_0$. By Theorem 5.5, $\Gamma\Delta_x$ is dense in S_x and $\Gamma\Delta_y$ is dense in S_y. This implies that if ΓS_x intersects ΓS_y, then for some $V, W \in \Gamma$ the set $V\Delta_x$ meets $W\Delta_y$. However, Δ_x and Δ_y both lie in N_0, which does not contain equivalent points. Hence $V = W$ and Δ_x meets Δ_y, so that S_x meets S_y—a contradiction. Therefore, $\Gamma S_x \cap \Gamma S_y$ is empty and $\pi(S_x)$, $\pi(S_y)$ are disjoint open neighbourhoods of Γx and Γy.

3. The final case is that $x \in P,\ y \in \mathbb{H}$. We use the same ideas as before. Let B be an open disc in $\mathbb{H}$ containing y. By local finiteness, $\overline{B}$ meets only $V_1\overline{N_0}, \dots, V_n\overline{N_0}$. Letting $B_i = N_0 \cap V_i^{-1}B$ we see that the sets $V_i\overline{B_i}$ cover B, so $\Gamma\left(\bigcup_i B_i\right)$ is dense in B. See Figure 7.

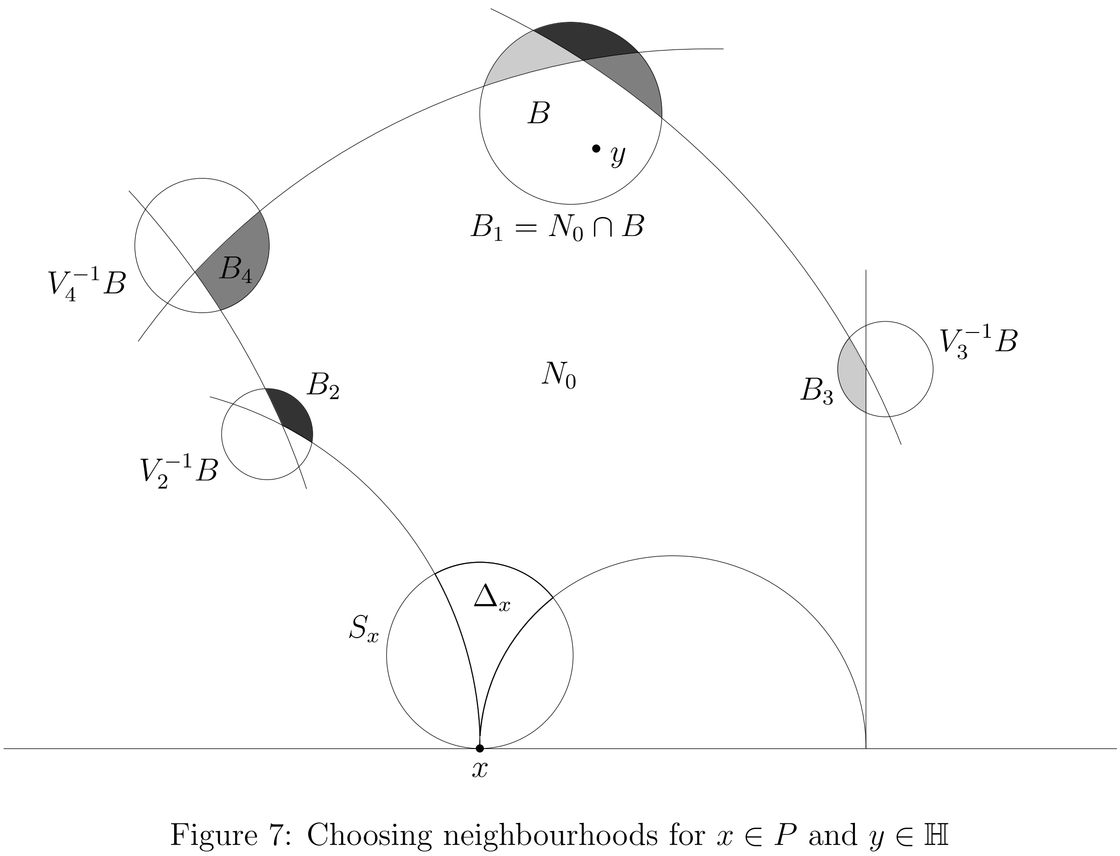

Figure 7: Choosing neighbourhoods for $x \in P$ and $y \in \mathbb{H}$

Choose $S_x \in \mathcal{B}$, a neighbourhood of x, such that S_x is disjoint from $B_1, \ldots, B_n$. Following (2) let $\Delta_x = S_x \cap N_0$ then $\Gamma\Delta_x$ is dense in S_x. The argument of (2) now shows that ΓS_x and ΓB are disjoint, and we are done as before. $\square$

We shall now describe a set of co-ordinate neighbourhoods and charts that make $\mathbb{H}^+/\Gamma$ a surface.

For each $\Gamma x \in \mathbb{H}^+/\Gamma$ define the co-ordinate neighbourhood $\mathcal{U}_x$ of Γx to be $\pi(U_x)$, where $U_x \in \mathcal{B}$ is chosen such that:

- $x \in U_x$
- $U_x \setminus \{x\}$ contains no fixed points of elements of Γ
- $TU_x \cap U_x = \emptyset$ for all $T \in \Gamma$ that do not fix x
- If x is a fixed point then the boundary of U_x is a fixed circle of some element fixing x (see Section 5.1)

Since the projection mapping is open, $\mathcal{U}_x$ is an open neighbourhood of Γx.

Define π_x to be the restriction of π to U_x. The map π_x is continuous because π is continuous. Moreover, if $A \subseteq U_x$ is open (in the subspace topology on U_x) then A is open in $\mathbb{H}^+$ so that $\pi_x(A) = \pi(A)$ is open. Therefore, π_x is an open mapping.

We now seek maps $\tau_x : U_x \to \mathbb{C}$ such that the composition,

$$\varphi_x = \tau_x \circ \pi_x^{-1}$$

is a chart on $\mathcal{U}_x$. Since π_x^{-1} is continuous and open, it will suffice to choose τ_x continuous and open such that $\tau_x \circ \pi_x^{-1}$ is well-defined[5] and injective. Then φ_x will be a homeomorphism from $\mathcal{U}_x$ onto its image. These compositions are shown in Figure 10.

To choose τ_x appropriately we must work in cases:

1. x is not fixed by any element of Γ.

   Our choice of U_x makes π_x^{-1} a well-defined and injective function from $\mathcal{U}_x$ to the disc U_x. Taking $\tau_x(z) = z - x$ will do.

2. x is an elliptic fixed point of order l.

   Let E generate the stabiliser Γ_x (see Theorem 5.1). The transformation $A(z) = \frac{z-x}{z-\bar{x}}$ carries the fixed points of E to 0 and ∞. Remarks 5.2 and 5.3 imply that $AE(z) = e^{\pm\frac{2\pi i}{l}}A(z)$ and that, since ∂U_x is a fixed circle of E, A maps U_x onto an open disc centred on the origin. We define,

$$\tau_x(z) = [A(z)]^l = \left(\frac{z-x}{z-\bar{x}}\right)^l$$

   so that τ_x maps U_x to an open disc D, as shown in Figure 8.

[5] i.e. single-valued

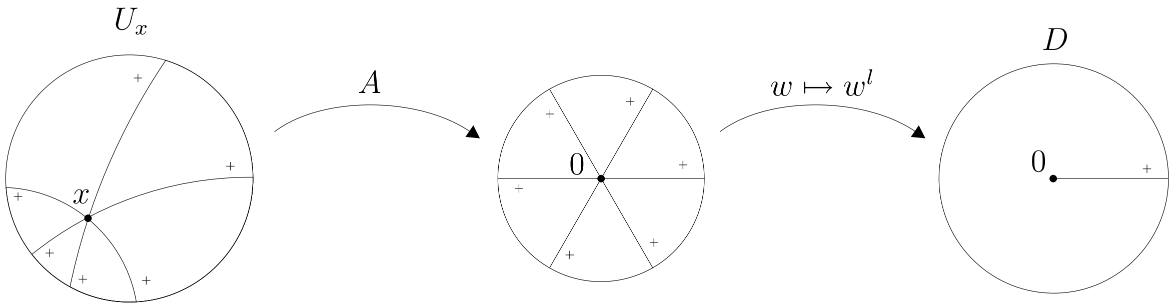

Figure 8: Defining τ_x when x is an elliptic fixed point. The + points in U_x are equivalent.

Since τ_x is analytic on U_x it is continuous and open. For $\Gamma z \in \mathcal{U}_x$, the set $\pi_x^{-1}\{\Gamma z\}$ equals $\{z, E(z), E^2(z), \ldots, E^{l-1}(z)\}$ but τ_x maps these equivalent points to a single point of D:

$$\tau_x(E^m(z)) = [A(E^m(z))]^l = [e^{\pm\frac{2\pi mi}{l}} A(z)]^l = [A(z)]^l = \tau_x(z)$$

so that $\tau_x \circ \pi_x^{-1}$ is single-valued.

Furthermore, π_x^{-1} maps distinct points of $\mathcal{U}_x$ to inequivalent sets in U_x. These have distinct images under τ_x, showing that $\tau_x \circ \pi_x^{-1}$ is injective.

3. $x \in \mathbb{R}$ is a parabolic fixed point.

We take a similar approach to that of the previous case. Let P generate the stabiliser Γ_x (again Theorem 5.1). Since $A(z) = \frac{1}{z-x}$ sends the fixed point x to ∞, the conjugated map APA^{-1} is a translation,

$$APA^{-1}(z) = z + c \quad \text{for some} \quad c \in \mathbb{R}$$

Equivalently, $AP(z) = A(z) + c$.

Since ∂U_x is a fixed circle of P, the region $A(U_x)$ is a half-plane as shown below (since by Remark 5.3, $\partial A(U_x)$ must be a fixed circle of the translation).

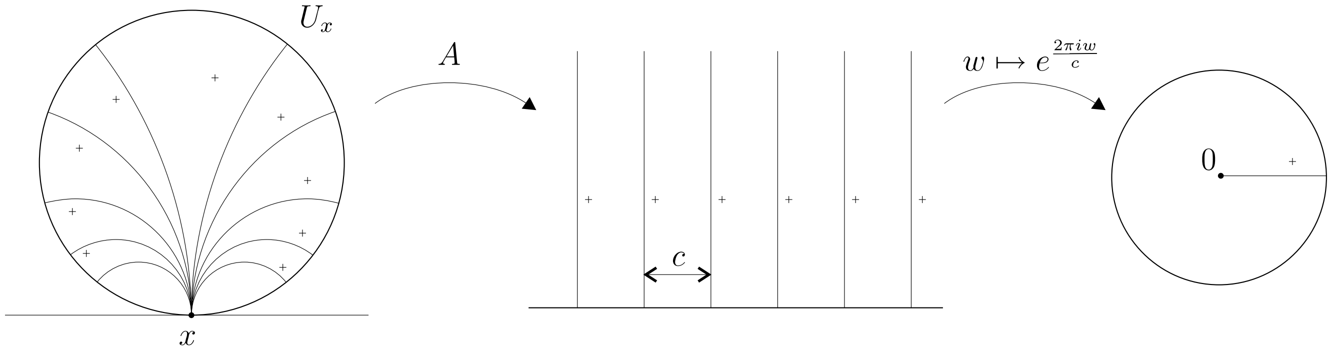

Figure 9: Defining τ_x when x is a parabolic fixed point. The + points in U_x are equivalent.

Now define $\tau_x(z) = \begin{cases} \exp\left(\frac{2\pi i A(z)}{c}\right) = \exp\left(\frac{2\pi i}{c(z-x)}\right), & z \in U_x \setminus \{x\} \\ 0, & z = x \end{cases}$

As shown in Figure 9, τ_x maps U_x to an open disc in $\mathbb{C}$.

For $\Gamma z \in \mathcal{U}_x$, the set $\pi_x^{-1}\{\Gamma z\}$ equals $\{P^m(z) : m \in \mathbb{Z}\}$, but τ_x maps this set to a single point:

$$\tau_x(P^m z) = \exp\left(\frac{2\pi i A(P^m z)}{c}\right) = \exp\left(\frac{2\pi i A(z)}{c} + 2\pi m i\right) = \tau_x(z)$$

So $\tau_x \circ \pi_x^{-1}$ is well-defined. As before, distinct points of $\mathcal{U}_x$ correspond to inequivalent sets in U_x so that $\tau_x \circ \pi_x^{-1}$ is injective. All that remains is to show that τ_x is continuous and open on U_x. This is clearly the case on $U_x \setminus \{x\}$, for τ_x is analytic there. To show that τ_x is open on all of U_x it will suffice that, for $x \in Q \in \mathcal{B}$, the set $\tau_x(Q)$ is an open disc about the origin. To see this, follow the maps described above—∂Q is a fixed circle of P, so $A(Q)$ is a half-plane and $\tau_x(Q)$ is a disc. Conversely, the pre-image under τ_x of any open disc centred on the origin is a member of $\mathcal{B}$. Hence, τ_x is continuous on all of U_x.

4. $x = \infty$ is a parabolic fixed point.

   In this case Γ_∞ is generated by $z \mapsto z + c$ for some $c \in \mathbb{R}$ and

$$\tau_x(z) = \exp\left(\frac{2\pi i}{c}\right)$$

   will do. The proof is as for (3), omitting the first steps.

We have now defined charts $\varphi_x = \tau_x \circ \pi_x^{-1}$ for all the co-ordinate neighbourhoods $\mathcal{U}_x$ and so $\mathbb{H}^+/\Gamma$ is a surface. To show that it is in fact a Riemann surface we must verify that the transition functions $\varphi_x \circ \varphi_y^{-1}$ are analytic.

The function $\varphi_x \circ \varphi_y^{-1}$ has domain $D = \varphi_y(\mathcal{U}_x \cap \mathcal{U}_y)$. If this is non-empty then we may assume that the representative x of Γx has been chosen so that U_x intersects U_y. See Figure 10.

Lemma 6.2. *We claim that on D,*

$$\varphi_x \circ \varphi_y^{-1} = \left(\tau_x \circ \pi_x^{-1}\right) \circ \left(\pi_y \circ \tau_y^{-1}\right) = \tau_x \circ r_{U_x} \circ \tau_y^{-1}$$

where r_{U_x} is just restriction to U_x (this is the domain of τ_x).

Proof. Let $t \in D$. Then $t = \varphi_y(\Gamma z)$ for some $\Gamma z \in \mathcal{U}_x \cap \mathcal{U}_y \subseteq \mathbb{H}^+/\Gamma$, as shown in Figure 10.

Our choices for U_y and τ_y ensure that τ_y maps equivalent points of U_y to a single point and inequivalent points to distinct points. Hence,

$$r_{U_x} \circ \tau_y^{-1}\{t\} \;=\; r_{U_x}(\Gamma z \cap U_y) \;=\; \Gamma z \cap U_x \cap U_y \tag{1}$$

$$\pi_x^{-1} \circ \left(\pi_y \circ \tau_y^{-1}\right)\{t\} \;=\; \left(\pi_x^{-1} \circ \pi_y\right)(\Gamma z \cap U_y) \;=\; \pi_x^{-1}(\Gamma z) \;=\; \Gamma z \cap U_x \tag{2}$$

The sets (1) and (2) are non-empty (since $\Gamma z \in \mathcal{U}_x \cap \mathcal{U}_y \;\Rightarrow\; \Gamma z \cap U_x \cap U_y \neq \emptyset$) and have the same image under τ_x. This proves the claim. □

We now tackle the question of analyticity.

Lemma 6.3. *The composition $\tau_x \circ r_{U_x} \circ \tau_y^{-1}(t)$ is analytic on D.*

Proof. If x is not a fixed point then $\tau_x(z) = z - x$ is analytic. Similarly, if y is not a fixed point then $\tau_y^{-1}(t) = t + y$ is analytic. Otherwise, by checking the possible definitions we see that $\tau_x(z)$ is locally conformal for $z \neq x$. Also, since $\tau_y(y) = 0$, the inverse $\tau_y^{-1}(t)$ is locally conformal for $t \neq 0$.

The sets U_x and U_y have been chosen so that $U_x \cap U_y$ contains no fixed points. Therefore, if x is a fixed point, then $x \notin U_x \cap U_y$ so τ_x is analytic on $r_{U_x} \circ \tau_y^{-1}(D) = U_x \cap U_y$. If y is a fixed point then $D = \tau_y(U_x \cap U_y)$ does not contain the origin, so τ_y^{-1} is analytic on D.

In all cases $\tau_x \circ r_{U_x} \circ \tau_y^{-1}(t)$ is analytic on D. □

Since the transition functions $\varphi_x \circ \varphi_y^{-1}$ are analytic, we have now seen that our system of charts and co-ordinate neighbourhoods makes $\mathbb{H}^+/\Gamma$ a Riemann surface.

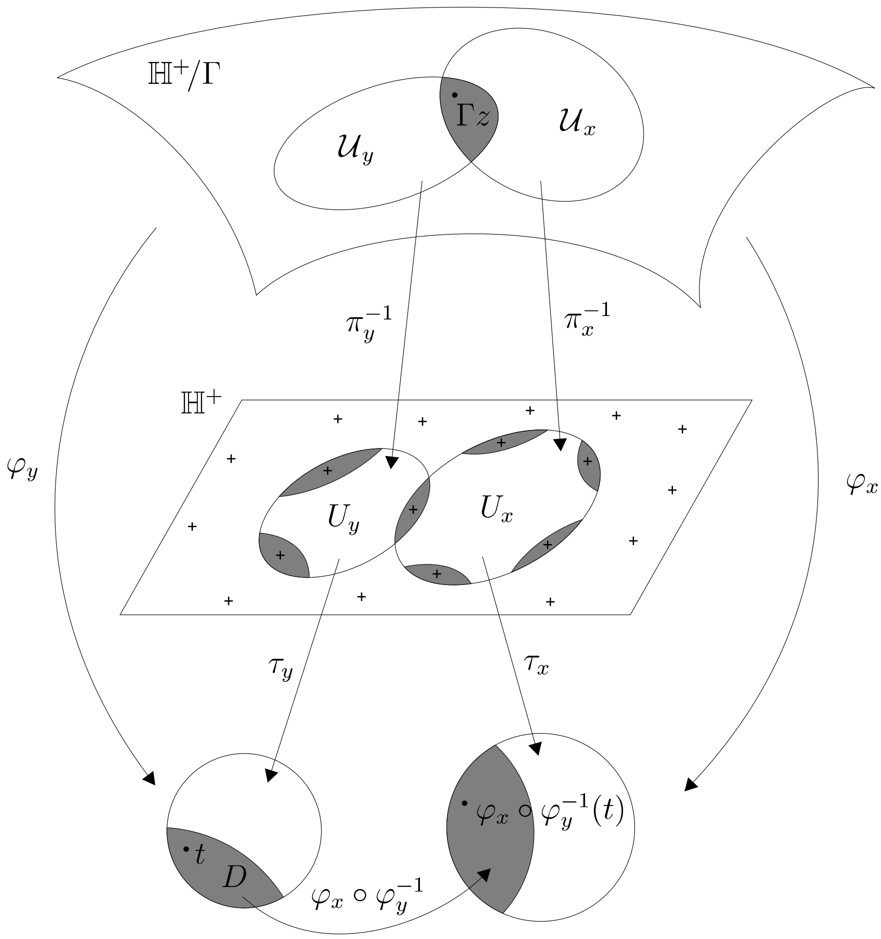

Figure 10: The points of Γz in $\mathbb{H}^+$ are marked $+$

Remarks

By restricting the argument of this section to $\mathbb{H}$ (with its usual topology) instead of $\mathbb{H}^+$, we see how $\mathbb{H}/\Gamma$ can be given the structure of a Riemann surface. Recalling that $\mathbb{D}$ and $\mathbb{H}$ are conformally equivalent, this is a pleasing complement to the result of Uniformisation Theorem that 'most' Riemann surfaces are conformally equivalent to $\mathbb{D}/\Gamma$ for some discontinuous group Γ.

We earlier exploited the link between discontinuous groups and Riemann surfaces in order to study the automorphic and meromorphic functions that naturally exist on them. In fact, this link can be used to discover something about the nature of the spaces themselves, as in the following result, a proof of which may be found in [1, p121–123].

Theorem 6.4. *Let Γ be a Fuchsian group on $\mathbb{H}$ with normal polygon N_0.*
The Riemann surface $\mathbb{H}^+/\Gamma$ is compact iff $\overline{N_0} \cap \mathbb{H}^+$ is compact in the topology of $\mathbb{H}^+$.
The Riemann surface $\mathbb{H}/\Gamma$ is compact iff $\overline{N_0} \cap \mathbb{H}$ is compact.

References

[1] Lehner, J., *A Short Course in Automorphic Functions*, Holt, Rinehart and Winston, 1966

[2] Lehner, J., *Discontinuous Groups and Automorphic Functions*, Amer. Math. Soc., 1964

[3] Beardon, A.F., *The Geometry of Discrete Groups*, Springer-Verlag, 1983

[4] Ford, L.R., *Automorphic Functions*, McGraw-Hill, 1929; 2nd ed., Chelsea, 1951

[5] Jones, G.A. & Singerman, D., *Complex Functions*, CUP, 1987

[6] Beardon, A.F., *A Primer on Riemann Surfaces*, LMS Lecture Note Series 78, CUP, 1984