INTRODUCTION TO KLEINIAN GROUPS

Abstract. Pekka Tukia has given this course in Helsinki during spring 2007. I have corrected some spelling mistakes.

Joensuu, October 15, 2017, Juha-Matti Huusko

An introduction to Kleinian groups

Spring 2007 (III period)

Pekka Tukia

1. Möbius transformations

We will work in the extended complex plane $\overline{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$. The extended complex plane has a concrete model which is the Riemann sphere S consisting of points $u = (u_1, u_2, u_3) \in \mathbf{R}^3$ such that $|u - (0, 0, 1/2)| = 1/2$. Thus S has diameter 1 and lies on the complex plane $\mathbf{C}$ touching $\mathbf{C}$ at 0 when $\mathbf{C}$ is identified with $\mathbf{R}^2 \subset \mathbf{R}^3$ so that $x + iy$ corresponds to $(x, y, 0)$. The "north pole" of S is the point $(0, 0, 1)$ and S touches $\mathbf{C}$ at the south pole $0 = (0, 0, 0)$ The stereographic projection s maps $\overline{\mathbf{C}}$ bijectively onto S. The point $z = x + iy$ would correspond to the point $s(z) \in S$ so the line through the north pole and z intersects S at $s(z)$.

We can define a metric in the extended complex plane by means of the Riemann sphere and the stereographic projection. We will call this metric the *spherical metric* of $\overline{\mathbf{C}}$ and denote it by q. The spherical distance of z and w is $q(z, w) = |s(z) - s(w)|$ = the euclidean distance of the points $s(z)$ and $s(w)$ in $\mathbf{R}^3$. It is calculated in the complex analysis course that

$$q(z, w) = \frac{|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}}$$

and $q(z, \infty) = 1/\sqrt{1 + |z|^2}$. The actual formula for q is not so important, what is enough is that ∞ has a neighborhood basis consisting of sets of the form $D(\infty, r) = \{\infty\} \cup \{w \in \mathbf{C} : |w| > 1/r\}$ and points $z \in \mathbf{C}$ have the usual neigborhoods.

A Möbius transformation is a bijective self-map of $\overline{\mathbf{C}}$ of the form

$$f(z) = \frac{az + b}{cz + d}$$

where the coefficients a, b, c, d are complex numbers such that $ad - bc \neq 0$. We will recapitulate a few facts about Möbius transformations. A Möbius transformation is a bijection of $\overline{\mathbf{C}}$, if we set $f(\infty) = a/c$ (if $c = 0$, then $f(\infty) = \infty$) and

Date: October 15, 2017.

$f(-d/c) = \infty$. This is how they must be defined if f is to be continuous. One can directly compute that the inverse of a Möbius transformation is a Möbius transformation and the composition of two Möbius transformations. Thus Möbius transformations form a group. We denote

$$M = \text{the group of Möbius transformations of } \overline{\mathbf{C}}.$$

There is an important connection between matrices and Möbius transformations. If A is the complex 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{0}$$

we can set

$$f_A(z) = \frac{az+b}{cz+d}$$

This is a Möbius transformation if the determinant of A, denoted $\det A = ad - bc$, is non-zero. One can show (exercise) that the matrix multiplication corresponds to composition of Möbius transformations, that is

$$f_{AB} = f_A \circ f_B.$$

The set of complex 2×2 matrices with non-vanishing determinant is denoted by $GL(2, \mathbf{C})$ (GL comes from "general linear group"). Thus the map $A \mapsto f_A$ is a homomorphism $GL(2, \mathbf{C}) \to M$.

We now recall some basic properties of Möbius transformations, proved in Complex Analysis I.

Conformality. The derivative is

$$f'(z) = \frac{ad - bc}{(cz+d)^2}.$$

This is non-zero at all finite points where $cz + d \neq 0$, i.e. $f(z \neq \infty)$, and so f is conformal at all such points, i.e. f is angle preserving. At the point $z = -d/c$ or at the point ∞ conformality is defined by means of the auxiliary map $\varrho(z) = 1/z$ which is used to transfer the situation so that the point is finite. For instance at $z = -d/c$, $\varrho \circ f(z) = 0$ and one easily checks that $(\varrho \circ f)'(-d/c) \neq 0$. The map f is conformal at ∞, if $f \circ \varrho$ is conformal at 0; if it should happen that $f(\infty) = \infty$, we have to check that $\varrho \circ f \circ \varrho$ is conformal at 0. We leave as an exercise to check that f is conformal at all points.

Thus Möbius transformations are conformal self-maps of the Riemann sphere $\overline{\mathbf{C}}$. One can show using the generalized Liouville theorem, that all conformal homeomorphisms of $\overline{\mathbf{C}}$ are Möbius transformations. This is one of the reasons why Möbius transformations are important.

Preservation of cross-ratio. We define the cross ratio of four points z_i as

$$(z_1, z_2, z_3, z_4) = \frac{z_1 - z_3}{z_1 - z_4} : \frac{z_2 - z_3}{z_2 - z_4}.$$

Thus the cross-ratio is the ratio of two ratios justifying the name crossratio (kaksoissuhde in Finnish). This is not the only way to define the cross ratio, sometimes one takes the crossratio as the quotient

$$\frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_3)(z_2 - z_4)}.$$

However, this can be obtained from the first crossratio by a permutation of z_i, namely permutate $(z_1, z_2, z_3, z_4) \mapsto (z_1, z_3, z_4, z_2)$. To begin with, z_i are distinct complex numbers but by continuity the cross-ratio can be extended to the case that $z_i \in \overline{\mathbf{C}}$ and that at most two of the numbers z_i coincide. If two z_i's coincide, then the crossratio assumes the value 0, 1 or ∞.

Three points determine a Möbius transformation. Let z_i and w_i $i \leq 3$, be two sets of three distinct points. Then there is a uniquely determined Möbius transformation f such that $f(z_i) = w_i$. The map f can be determined by solving $w = f(z)$ from the equation

$$(w, w_1, w_2, w_3) = (z, z_1, z_2, z_3).$$

In practice, it is often simpler to solve the coefficients from the three equations

$$w_i = \frac{az_i + b}{cz_i + d}.$$

The coefficients a, b, c, d are not well-defined and one can often fix that one of them is 1.

Preservation of circles. We say that $S \subset \overline{\mathbf{C}}$ is a Möbius circle, if either $S \subset \mathbf{C}$ and S is a usual euclidean circle or S is of the form $L \cup \{\infty\}$ where L is a line of $\mathbf{C}$. These are circles if mapped to the Riemann sphere by the stereographic projection. Möbius transformations preserve Möbius circles. If S is a Möbius circle, then the image fS is also a Möbius circle for any Möbius transformation f.

If S_1 and S_2 are Möbius circles, then there is a Möbius transformation f such that $fS_1 = S_2$. This can be found by taking 3 points $z_i \in S_1$ and $w_i \in S_2$ and finding the Möbius transformation such that $f(z_i) = w_i$.

If S_i is an euclidean circle, then $\overline{\mathbf{C}} \setminus S_i$ has two components, one of which is an ordinary euclidean disk and the other contain ∞. If $S_2 = L \cup \infty$, then $\overline{\mathbf{C}} \setminus S_2$ has two components, and both are half-spaces. For instance

$$f(z) = \frac{z - i}{z + i}$$

maps the upper half-space $U = \{z \in \mathbf{C}; \operatorname{Im} z > 0\}$ to the unit disk $|z| < 1$. (Check). There are often situation where need for this kind of mappings arises.

One sees easily (exercise) that if f is as in (1) and

$$g(z) = \frac{a'z + b'}{c'z + d'}$$

then $f = g$ if and only if the coefficients are proportional: that is, there is a complex number $\lambda \neq 0$ such that $a' = \alpha$, $b' = \lambda b$, etc. Thus we can multiply all coefficients by the same complex number $\lambda \neq 0$ without changing the complex number. If we multiply by λ the coefficients of f in (0), then $ad - bc$ is changed into $\lambda^2(ad - bc)$. Thus if λ^{-1} is one of the two values of $\sqrt{ad - bc}$, then λA, A as in (2), has determinant 1. Thus we can choose A so that $\det A = 1$. Since the square root has two values differing by sign, there are two matrices A and B such that $f = f_A = f_B$ and $B = -A$.

Since $\det AB = (\det A)(\det B)$, the complex 2×2 matrices with determinant 1 form a group, denoted $SL(2;\mathbf{C})$ (SL comes from the "special linear group"). Thus the map $A \mapsto f_A$ is a homomorphism $SL(2,\mathbf{C}) \to M$ so that to $f \in M$ corresponds two matrices $A, -A \in SL(2;\mathbf{C})$. So we obtain M from $SL(2,\mathbf{C})$ by identifying A and -A. The kernel of the map $SL(2,\mathbf{C})$ is $\{I, -I\}$ and M is isomorphic to the quotient $SL(2;\mathbf{C})/\{I,-I\}$.

The group $SL(2,\mathbf{C})$ has some advantages compared to $GL(2;\mathbf{C})$. For instance forming the inverse matrix is easy in $SL(2,\mathbf{C})$. One easily checks that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -b & a \end{pmatrix} = (\det A)I$$

where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is the unit matrix. Thus if

$$f(z) = \frac{az+b}{cz+d},$$

where $ad - bc = 1$, then

$$f^{-1}(z) = \frac{dz-b}{-cz+a}.$$

We will see that $SL(2,\mathbf{C})$ has other advantages as well.

Examples of Möbius transformations. Some very simple mappings are Möbius transformations. For instance, the identity map $\mathrm{id} : \overline{\mathbf{C}} \to \overline{\mathbf{C}}$ such that $z \mapsto z$ can be written as

$$\mathrm{id}(z) = \frac{1z+0}{0z+1}$$

and hence is a Möbius transformation. Similarly

$$\frac{1}{z} = \frac{0z+1}{1z+0}$$

A translation $f(z) = z + a$ can be written as

$$z + a = \frac{1z+a}{0z+1}$$

and finally $az + b = \frac{az+b}{0z+1}$. We will see that every Möbius transformation is topologically and conformally similar to one of these maps. We can divide Möbius transformations into types and each type can be exemplified by this kind of mapping.

A basic distinction between Möbius transformations is the number of fixpoints. A *fixpoint* of f is a point z such that $f(z) = z$. We will see that a Möbius transformation $f \neq \mathrm{id}$ has one or two fixpoints.

Translations. Translations are of the form $f(z) = z + a$ $(a \neq 0)$, or $f(z) = z+1$ if we want to be as simple as possible. It has one fixpoint which is ∞. These maps will exemplify *parabolic* Möbius transformation. One often characterizes Möbius transformations by means of circle families. Let

$$T_1 = \{L \cup \{\infty\} : L \text{ is a line parallel to } a\}. \tag{1}$$

Then f maps each $S \in T_1$ onto itself: $fS = S$. Actually, we can characterize T_1 as the family of Möbius circles preserved by f. Let T_2 be the family of Möbius circles orthogonal to the circles of T_1, i.e. if $L \cup \{\infty\} \in T_2$, then L intersects orthogonally L' if $L' \cup \{\infty\} \in T_2$. Now fL is the line $L+a = \{z+a : z \in L\}$ and so f preserves T_2 as a circle family: $fS \in T_2$ if $S \in P_2$ although $fS \neq S$.

We note also that $f^n = f \circ \ldots \circ f$ (f composed with itself n times) is the map $f(z) = z + na$ and f^n is the identity only if $n = 0$ (we set $f^{-n} = (f^{-1})^{-n} = f - nz$ if $n < 0$ and $f^0 = \text{id}$.)

Affine maps. A (complex) affine map is of the form $f(z) = \lambda z + b$, $\lambda \neq 0$, or more simply $f(z) = \lambda z$. These maps have two fixpoints. ∞ is one and $z = -b/(\lambda - 1)$ is the other. Thus the fixpoints of $f(z) = \lambda z$ are 0 and ∞. These maps exemplify several types of Möbius transformations, depending on the value of λ. We have

a) if $|\lambda| = 1$, then f is *elliptic*.

b) if $\lambda \in \mathbf{R}$ and $\lambda > 0$ and $\lambda \neq 1$, then f is *hyperbolic*

c) if $|\lambda| \neq 1$, then f is *loxodromic*.

We have above assumed that $\lambda \neq 1$; if $\lambda = 1$, then $f = \text{id}$ and we regard the identity mapping also elliptic. Sometimes one calls only such maps loxodromic where $|\lambda| \neq 1$ and $\lambda \notin \mathbf{R}$ so that the cases a), *b*) or *c*) are exclusive.

We will now study the behavior of these by means of two circle familes: Let S_r be the circle $\{z; |z| = r]$ and R_α the Möbius circle consisting ∞ and of the points $te^{i\alpha}$, $t \in \mathbf{R}$ and set

$$S_1 = \{R_\alpha : \alpha \in \mathbf{R}\} S_2 = \{S_r : r > 0\} \tag{2}$$

Thus the circle family S_1 is orthogonal to S_2. We easily see that S_1 and S_2 are preserved as circle families and have

$$fS_r = S_{|\lambda| r} f R_\alpha = R_{\alpha + \arg \lambda}.$$

If f we have case *a*) or b), If $|\lambda| = 1$ then $fS_r = S|\lambda|r = S_r$ and thus f preserves circles $S_r \in S_2$. If we have case b), then $\lambda > 0$, then $f(z) = \lambda z$ and we see that f preserves not only circles $S_\alpha \in S_1$, but that f preserves the two arcs to which the fixpoints 0 and ∞ divide S_α. We can think that when we apply f, we glide the points along the arcs. Note that if $\lambda < 0$, then f interchanges these two arcs.

Conjugation of maps. We will see that every Möbius transformations $f \neq \text{id}$ is similar to one of these types. This similarity is made precise by the notion of conjugation of maps. We explain this first generally, without reference to Möbius transformations.

Let A and B be sets and let $h : A \to B$ be a bijective map. We use h to transfer the features of A to B. In particular, we can transfer maps $A \to A$ to maps $B \to B$ by a process called conjugation (with h). If $f : A \to A$ is a map, we can transfer it to map of B conjugating with h; This is the map $g = h \circ f \circ h^{-1}$. We use usually the shorter notation $g = hfh^{-1}$ for the compositions of maps. Thus g

is a map $B \to B$; we say that g is obtained from f by conjugating with h. Note that we obtain f from g by conjugating with h^{-1}.

We note here some properties of the conjugated map $g = fgh^{-1}$.

1° *$f(a) = b$ if and only if $g(h(a)) = h(b)$ $(a, b \in A)$. In particular, a is a fixpoint of f if and only if $h(a)$ is a fixpoint of g. Thus f is the identity map of A if and only if hfh^{-1} is the identity map of B.*

2°. *If $X, Y \subset A$, then $fX = Y$ if and only if $g(hY) = hY$. In particular, $fX = X$ if and only if $g(hX) = Y$.*

3°. *The map f is a bijection (injection, surjection) of A if and only if hfh^{-1} is a bijection (injection, surjection) of B.*

4° *If G is a group of bijections of A, then $hGh^{-1} = \{hgh^{-1} : g \in G\}$ is a group of bijections of B and if $\varphi(g) = hgh^{-1}$ is an isomorphism onto the group $H = \varphi G$ of bijections of B, the inverse being $\varphi^{-1}(g) = h^{-1}gh$.*

These are easy to check by using the definition. It is also useful to draw a diagram in the first two cases which shows graphically what happens.

We now return to Möbius transformations. We say that $f, g \in M$ are conjugate by a Möbius transformation (or simply conjugate) if there is $h \in M$ such that $g = hfh^{-1}$. Since M is a group, this is an equivalence relation (exercise).

We will consider here only conjugation with a Möbius transformation, but it would be reasonable to consider conjugations by a homeomorphism of $\overline{\mathbf{C}}$. This would be also an equivalence relation with larger equivalence classes than conjugation by a Möbius transformation.

We will now show that every Möbius transformation is conjugate either to a translation or to a map which fixes 0 and ∞.

Theorem 1.1. *If $f \neq \mathrm{id}$ is a Möbius transformation, then f has one or two fixpoints. We have that f has one fixpoint if and only if f is conjugate to a translation and we can take this translation be $z \mapsto z + 1$. The map f has two fixpoints if and only if f is conjugate to the map $z \mapsto \lambda z$; here $\lambda \neq 0, 1$ and λ is uniquely determined if we require that either $|\lambda| > 1$ or that $\lambda = e^{i\alpha}$ where $0 < \alpha \leq \pi$.*

Proof. Let
$$f(z) = \frac{az + b}{cz + d}.$$
If $c = 0$, then f is of the form $f(z) = \lambda z + \beta$ and has ∞ as one fixpoint. If $\lambda = 1$, then $\beta \neq 0$ and hence f is a translation. If $\lambda \neq 1$, then the only fixpoint $z \in \mathbf{C}$ of f is $-\beta/(\lambda - 1)$. (The case $\lambda = 1$ and $\beta = 0$ is excluded since $f \neq \mathrm{id}$. If $c \neq 0$, then $f(z) = z$ is equivalent to
$$cz^2 + (d - a)z - b = 0$$
and this equation has one or two complex solutions. Note that in this case ∞ is not a fixpoint. So ∞ is a fixpoint if and only if $c = 0$ and in this case f is of the form $f(z) = \lambda z + \beta$.

So f has 1 or 2 fixpoints. Suppose that z_0 is the only fixpoint of f. We choose a Möbius transformation h such that $h(z_0) = \infty$. Thus if $g = hgh^{-1}$, then ∞ is the only fixpoint of g. Thus f is of the form $\lambda z+\beta, \beta \neq 0$, and since ∞ is the only fixpoint, $\lambda=1$. We can conjugate further by the map $k(z)=\beta^{-1}$ and then

$$kgk^{-1}(z)=(kh)f(kh)^{-1}(z)=\beta^{-1}(\beta z+\beta)=z+1.$$

It follows that f is conjugate to the map $z\mapsto z+1$ if and only if f has only one fixpoint.

Suppose then that f has fixpoints z_1 and z_2. Then we can find $h\in M$ such that $f(z_1)=0$ and $f(z_2)=\infty$. Thus the map hfh^{-1} has fixpoints 0 and ∞ and hence is of the form $h(z)=\lambda z$ where $\lambda\neq 0,1$. If $k(z)=1/z$, then $k(z)$ interchanges the fixpoints of g and so kgk^{-1} has also fixpoints 0 and ∞ but

$$kgk^{-1}(z)=\lambda^{-1}z$$

and so, possibly by replacing h with kh, we can assume that $|\lambda|\geq 1$ and if $|\lambda|=1$, we can obtain that $\lambda=e^{-i\alpha}$ where $0<\alpha\leq\pi$.

So to conclude the proof we must show that λ is unique if it is chosen as indicated in the theorem. Suppose that also $\tilde{h}$ conjugates f to a map with fixpoints 0 and ∞. Thus $\tilde{h}(z_1,z_2\}=\{0,\infty\}$ and hence if $k=\tilde{h}h^{-1}$, then $k\{0,\infty\}=\{0,\infty\}$. If $k(0)=0$ and $k(\infty)=\infty$, then $k(z)=\mu z$ for some $\mu\in\mathbf{C}\setminus\{0\}$ and one easily checks $\tilde{h}f\tilde{h}^{-1}(z)=kgk^{-1}(z)=\lambda z$.

The other case is that $k(0)=\infty$ and $k(\infty)=0$. Let $\sigma(z)=1/z$. Then σk preserves 0 and ∞ and hence $(\sigma k)g(\sigma k)^{-1}(z)=\lambda z$ and so $kgk^{-1}(z)=\sigma k\sigma^{-1}(z)=\lambda^{-1}z$ (note that $\sigma^2=\mathrm{id}$). Since $kgk^{-1}=\tilde{h}f\tilde{h}^{-1}$, we see that by changing the conjugating map, we can only obtain that λ is replaced by λ^{-1} and λ is unique if chosen so as indicated in the theorem.

A consequence is that if $f\in M$ has two fixpoints and we conjugate f to a map of the form $z\mapsto\lambda z$, then if $|\lambda|=1$ for one conjugating map, then this is true for any conjugating map and the same is true of the positivity of λ. Hence we can define.

Definition. Let $f\in M$. Then

(a) f is parabolic if f can be conjugated to a translation.
(b) f is elliptic if f can be conjugated to a map $z\to\lambda z$ where $|\lambda|=1$.
(c) f is hyperbolic if f can be conjugated to a map $z\mapsto\lambda z$ where $\lambda>0$ and $\lambda\neq 1$.
(d) f is loxodromic if f can be conjugated to a map $z\mapsto\lambda z$ where $|\lambda|\neq 1$.

Remark. We regard he identity map also elliptic. This fits to the above definition although f has infinite number of fixpoints.

We can now transfer the properties of translations and maps $z\mapsto\lambda z$ to the general case using the conjugating map.

If f is parabolic, then there is $h\in M$ such that hfh^{-1} is the translation $g(z)=z+1$. Thus $f=h^{-1}gh$, and so h^{-1} conjugates g back to f. If T_i are as defined above (cf. (1)), and if $P_i=h^{-1}T_i=\{h^{-1}S : S\in T_i\}$, then P_1 and P_2 are two families of circles such that if $S_i\in P_i$, then S_1 and S_2 intersect orthogonally and one of the points of intersection is the fixpoint of f. We can note that if S is a Möbius circle invariant under $z\mapsto z+1$, then $S\in S_1$. We can conclude that R_1 is the family of Möbius circles preserved by f.

If f has two fixpoints, then there is $h \in M$ conjugating f to the map $g(z) = \lambda z$ and h conjugates g back to h. Let S_i be as in (2) and set $R_i = h^{-1} S_i$. Then R_1 and R_2 are families of Möbius circles so that R_1 is orthogonal to R_2. Since the situation regarding S_i and g is transferred to the situation of R_i are and g, we see that f preserves the family R_i as a family. Furthermore, if f is elliptic, we see that f preserves S if $S \in S_2$ and if f is hyperbolic, f preserves the arcs into which the fixpoints of f divide circles $S \in R_1$. If f is hyperbolic, we can describe f so that if $z \in S \in R_1$, then f glides z along on arc of $S \setminus \{\text{fixpoints of } f\}$ containing z.

There is a fundamental difference between elliptic and loxodromic maps. If f is elliptic and z is not a fixpoint of f, then $z \in S \in R_2$. Since S is preserved by f, f stays always on S which is a circle seprating the fixpoints of f. On the other hand, if f is loxodromic and z is not fixed by f, then $f^n(z)$ tends toward a fixpoint of f. This can be seen easily if $f(z) = \lambda z$ where $|\lambda| \neq 1$, and since every loxodromic f is conjugate to such a map, this follows for every loxodromic f. We leave the checking of this as an exercise.

There is also another difference between elliptic and other types of Möbius transformations. If $f \in M$, we say that the *order* of f is the smallest number $n > 0$ such that $f^n = \text{id}$; if $f^n \neq \text{id}$ for all $n > 0$, we say that f is of *infinite order* otherwise f is of *finite order*. A map $f \in M$ can be of finite order only if it is elliptic. This can be easily checked from the basic form of different types. If f is elliptic, then f is conjugate to $z \mapsto e^{i\alpha} z$ and $f^n(z) = e^{in\alpha} z$ and so f^n is the identity if $n\alpha$ is a multiple of 2π. This happens for some $n > 0$ if and only if α is a rational multiple of π.

Remark. If f is loxodromic, R_1 consists of Möbius circles passing through the fixpoints of f. To given an intrinsic for R_2 is a little less straightforward but it can be show that R_2 consists of Möbius circles

$$R_t = \{z \in \overline{\mathbf{C}} : \frac{|z - z_1|}{|z - z_2|} = t\}, \quad t > 0,$$

when z_1 and z_2 are the fixpoints of f so that $R_2 = \{R_t : t > 0\}$. This can be seen by using the cross-ratio and noting that S_r in (2) is the set

$$\{z \in \overline{\mathbf{C}} : |(0, \infty, z, z_1)| = 1\}$$

where $|z_1| = r$. Details are left as an exercise.

Trace and the type of a Möbius transformation. If

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

is a 2×2 matrix, the *trace* (in Finnish “jälki”) $\operatorname{tr} A$ is

$$\operatorname{tr} A = a + d.$$

A basic property of the trace is that $\operatorname{tr} AB = \operatorname{tr} BA$ (exercise). Using this one easily sees that trace is conjugation invariant

$$\operatorname{tr} BAB^{-1} = \operatorname{tr} A. \tag{3}$$

If $f \in M$, then $f = f_A = f_{-A}$ for some $A \in SL(2, \mathbf{C})$. The matrix A is well-determined up to sign. Thus we can define the trace of $f \in M$ up to sign by the formula

$$\operatorname{tr} f = \operatorname{tr} A$$

if A is one of the two matrices such that $f_A = f$. We could fix $\operatorname{tr} f$ for instance by requiring that that $\mathrm{Re}(\backslash tr; f) \geq 0$ but we prefer to think that $\operatorname{tr} f$ is two-valued and if we write $t = \operatorname{tr} f$, we mean that t is one of the two values of $\operatorname{tr} f$ and equation

$$\operatorname{tr} f = \operatorname{tr} g$$

means that we can choose the traces to be equal. We see from (3) that conjugacy invariancy extends to traces of Möbius transformations: $\operatorname{tr} gfg^{-1} = \operatorname{tr} f$.

We can characterize the type of a Möbius transformation by means of the trace.

Theorem 1.2. *Let $f \in M$, $f \neq \mathrm{id}$. Then*

Remark. If f is the identity map, then $\operatorname{tr} f = 2$. Thus the trace cannot distinguish a parabolic map from the identity. *Proof.* In view of the conjugacy invariancy, it suffices to consider the basic case that f is a translation (if f is parabolic) or that 0 and ∞ are the fixpoints. If f is translation, then $f = f_A$ for $A \in SL(2, \mathbf{C})$ of the form

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

and hence $\operatorname{tr} A = 2$. If $f(z) = \lambda z$, then $f = f_A$ if

$$A = \begin{pmatrix} \mu & 0 \\ 0 & \mu \end{pmatrix}$$

where $\mu = \sqrt{\lambda}$. Thus, if $\lambda = re^{i\alpha}$, then

$$\operatorname{tr} A = (r^{1/2} + r^{-1/2}) \cos \alpha/2 + i(r^{1/2} - r^{-1/2}) \sin \alpha/2.$$

The elliptic case corresponds to $r = 1$ and the hyperbolic case to $\alpha = 0$. If f is loxodromic, then $r > 0$ and we see that our claim is valid.

2. The convergence properties of Möbius transformations

The notion of uniform convergence (*"tasainen suppeneminen"* suomeksi) is important for us. If f_i and f are mappings of $\overline{\mathbf{C}}$ onto itself we say that f_i converge uniformly (in the spherical metric) toward f, if given $\varepsilon > 0$, there is n_0 such that spherical distance

$$q(f_i(z)), f(z)) < \varepsilon$$

for *all* $z \in \overline{\mathbf{C}}$ whenever $i \geq n_0$. In particular, $f_i(z)$ converges pointwise to $f(z)$ for every $z \in \overline{\mathbf{C}}$ but the converse is not true. Uniform convergence is a much stronger notion. Uniform convergence depends on the metric used but we will use the spherical metric except if not otherwise stated. Thus when we say that $f_i \to f$ uniformly, uniform convergence with respect to the spherical metric is meant.

A general property of uniform convergence is that uniform limits of continuous functions are continuous. Since $\overline{\mathbf{C}}$ is compact, we have the following useful characterization of uniform convergence.

Lemma 2.1. *Let f_i and f be continuous maps of $\overline{\mathbf{C}}$ onto itself. Then $f_i \to f$ uniformly if and only if $f_i(z_i) \to f(z)$ whenever $z_i \in \overline{\mathbf{C}}$ is a sequence such that $z_i \to z$ as $i \to \infty$.*

The proof is left as an exercise. We can use it to have a characterization of uniform convergence using the matrix representation of Möbius transformations. Let

$$A_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \text{ and } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{1}$$

be complex matrices. We say that A_i converge to A (and write $A_i \to A$) if we have the convergences $a_i \to a$, $b_i \to b$, etc. for the entries of matrices.

Lemma 2.2. *If A_i and A are matrices of $GL(2, \mathbf{C})$ such that $A_i \to A$, then $f_{A_i} \to f_A$ uniformly.*

Proof. This follows easily using Lemma 2.1. Let $z_i \in \overline{\mathbf{C}}$ be points converging toward z. It is easy to see that if A_i and A are as in (1), then

$$\frac{a_i z_i + b_i}{c_i z_i + d_i} \to \frac{az + b}{cz + d}.$$

If $z \neq \infty$ and $cz + d \neq 0$, this follows from standard convergence properties of complex numbers and the remaining cases are easily dealt with. For instance, if $cz + d = 0$ (and $c \neq 0$), then one checks that $az + b \neq 0$ because of the determinant condition and hence $f_{A_i}(z_i) \to \infty = f_A(z)$. Other cases are dealt similarly.

The converse to Lemma 2.2 is more difficult since to each $f \in M$ corresponds several matrices but we have:

Lemma 2.3. *Let f_i and f be elements of M such that $f_i \to f$ uniformly. Then there are $A_i \in SL(2, \mathbf{C})$ and $A \in SL(2, C)$ such that $f_i = f_{A_i}$ and $f = f_A$ and such that $A_i \to A$.*

Proof. Pick three distinct points $z_k \in \mathbf{C}$, $k \leq 3$, such that $w_k = f(z_k) \in \mathbf{C}$. If we set $w_{ki} = f_i(z_k)$, then $w_{ki} \to z_k$ as $i \to \infty$ and hence it is enough to consider the case that $w_{ki} \in \mathbf{C}$. By the Möbius invariance of the crossratio, we have

$$(f_i(z), w_{1i}, w_{2i}, w_{3i}) = (z, z_1, z_2, z_3)$$

from which $f_i(z)$ can be solved in the form

$$f_i(z) = \frac{a_i z + b_i}{c_i z + d_i}$$

so that the coefficient a_i, b_i, c_i and d_i are rational functions of the points z_k and w_{ki} not depending on i. Thus $a_i = g(z_1, z_2.z_3, w_{1i}, w_{2i}, w_{3i})$. Since

$$(f(z), w_1, w_2, w_3) = (z, z_1, z_2, z_3)$$

we see that if we solve $f(z)$ from this equation in the form

$$f(z) = \frac{az + b}{cz + d}$$

then $a = g(z_1, z_2, z_3, w_1, w_2, w_3)$. Since $w_{ki} \to w_i$, it follows that $a_i \to z$. The same argument is valid for other coefficients as well and hence we have found $A_i \in GL(2, \mathbf{C})$ and $A \in GL(2; \mathbf{C})$ so that $A_i \to A$.

In order to obtain that the matrices are in $SL(2, \mathbf{C})$ we have to divide the entries by $\sqrt{\det A_i}$ and $\sqrt{\det A}$, respectively. There is the problem that the square root does not have a global branch in all of $\mathbf{C}$. However, it is enough that we find a branch in neighborhood U of $\det A$ and use this branch whenever $\det A_i \in U$. Since $\det A_i \in U$ for large i, we can ignore other A_i.

Remark. Actually, the lemma is true as soon as $f_i(z) \to f(z)$ for three points; for instance if $f_i(z) \to f(z)$ pointwise. This is a consequence of Theorem 2.6 (cf. corollary 2.6). Actually, this is what we need and in this manner we need not treat the cases where one z_i or w_i is ∞.

Lemmas 2.2 and 2.3 have the following consequence.

Lemma 2.4. *Let f_i, g_i, f and g be Möbius transformations such that $f_i \to f$ uniformly and $g_i \to g$ uniformly. Then the $f_i \circ g_i \to f \circ g$ uniformly and $f^{-1} \to f^{-1}$ uniformly.*

Proof. We find matrix representations for the maps and obviously the matrix product and taking the inverse of the matrix depend continuously on the entries and the theorem follows. (Remember that taking the matrix inverse is especially easy in $SL(2, \mathbf{C})$.

Convergence sequences. The following situation occurs often in the situation of Kleinian groups. Let $(g_i)_{i>0}$ be a sequence of Möbius transformations. We say that (g_i) is a *convergence sequence* if there are points $a, b \in \overline{\mathbf{C}}$ such that

$$g_i(z) \to a$$

if $z \neq b$ and if the convergence is uniform outside neighborhoods of b. That is, given a neighborhood U of A and V of b, there is n_0 such that

$$g_i(\overline{\mathbf{C}} \setminus V) \subset U \tag{1}$$

if $i \geq n_0$. We can think that a attracts the points $g_i(z)$, $z \neq b$, and so a is called the *attracting point* of the sequence. Similarly, we can think that b repels these points and hence b is called the *repelling point* of the sequence. We can also think that g_i blow up neighborhoods of b so that they eventually fill $\overline{\mathbf{C}} \setminus \{a\}$.

Examples. There are two simple examples of convergence sequences which provide good models. The first $g_i(z) = a_i z$ where $a_i \to \infty$ as $i \to \infty$. In this case ∞ is the attracting and 0 the repelling point. If $a_i \to 0$, the attracting and repelling point are interchanged. The second is $g_i(z) = z + b_i$ where $b_i \to \infty$. In this case ∞ is both the attracting and repelling point. Thus the attracting and the repelling point need not be distinct.

The following is an easy consequence of the definitions.

Theorem 2.5. *If (g_i) is a convergence sequence, then (g_i^{-1}) is also a convergence sequence so that the attracting point of (g_i^{-1}) is the repelling point of (g_i) and the repelling point of $(g_i)^{-1}$ is the attracting point of (g_i).*

For the proof, we need simply to note that (1) can be written also as $g_i^{-1}(\overline{\mathbf{C}} \setminus U) \subset V$, by applying g_i^{-1} onto both sides and taking complements.

The following lemma is used in our main theorem in this section:

Lemma 2.6. *Suppose that (g_i) is a convergence sequence whose attracting point is a and the repelling point is b. Let $h_i, h \in M$ be mappings such that $h_i \to h$ uniformly. Then*

a) $h_i \circ g_i$ is a convergence sequence whose attracting point is $h(a)$ and the repelling point b.

b) $g_i \circ h_i$ is a convergence sequence whose attracting point is a and the repelling point $h^{-1}(b)$.

Proof. To prove a), let U be a neighborhood of $h(a)$ and V a neighborhood of b. Because of uniform convergence, there exists a neighborhood U' of a and n_0 such that $h_i U' \subset U$ if $i \geq n_0$. Because of the convergence property of (g_i), we can find n_i such that $g_i(\overline{\mathbf{C}} \setminus V) \subset U'$ if $i \geq n_i$. Thus $h_i g_i(\overline{\mathbf{C}} \setminus V) \subset U$ if $i \geq \max(n_0, n_1)$.

The proofs of other points are similar and left as exercise. It is enough to prove *a*) as this will imply *b*) in view of Lemma 2.1.

The following theorem expresses the general convergence property of Möbius transformations.

Theorem 2.7. *A sequence g_i of Möbius transformations has a subsequence g_{n_i} such that g_{n_i} is either a convergence sequence or converges uniformly in the spherical metric toward a Möbius transformation g.*

Proof. With these lemmas at our disposal, we can easily prove Theorem 2.2. We will pass several times to subsequences of g_i and to avoid complicated notation, we will still denote the subsequence as g_i. We can also always replace g_i by $g_i \circ h_i$ or by $h_i \circ g_i$ and the uniform convergence or convergence sequence property of the original sequence does not change as can be seen from Lemma 2.4 and 2.6.

We will first reduce the proof to the case that $g_i(\infty) = \infty$ for all i. In any case, $\overline{\mathbf{C}}$ being compact, we can pass to a subsequence so that $g_i(\infty) \to a \in \overline{\mathbf{C}}$. If $a = \infty$, we replace g_i by $\varrho \circ g_i$ where $\varrho(z) = 1/z$ (i.e. $g_i(z)$ is replaced by $1/g_i(z)$). Lemmas 2.3 and 2.4 say that we can do this replacement. If we do this replacement, $g_i(\infty)$ will tend to 0. Thus we can assume that $g_i(\infty) \neq \infty$, possibly by passing to a subsequence. Let $h_i(z) = 1/(z - g_i(\infty))$. Then $h_i \circ g_i(\infty) = \infty$ and it is not difficult to see that $h_i \to h$ uniformly when $h(z) = 1/(z-a)$, $a = \lim_{i\to\infty} g_i(\infty)$.

So, replacing g_i by $h_i \circ g_i$, we can assume that $g_i(\infty) = \infty$. If it is now the case that given $R > 0$, there is n_0 such that

$$|g_i(z)| \geq R, \quad |z| \leq R, \quad i \geq n_0, \tag{2}$$

whenever $|z| \leq R$, then (g_i) is a convergence sequence such that both attracting and the repelling point coincide of (g_i) is ∞. Every neighborhood of ∞ contains a set of the form $D(\infty, r) = \{\infty\} \cup \{z \in \mathbf{C} : |z| > 1/r\}$ and we see from (2) that beginning from some i, g_i maps the complement of $D(\infty, r)$ into $D(\infty, r)$.

Thus we can assume that there is a sequence of points z_i such that $|z_i| \leq R$ and $|g_i(z_i)| \leq R$. Thus, passing to a subsequence we can assume that $z_i \to a \in \mathbf{C}$ and $f(z_i) \to b \in \mathbf{C}$.

Let $h_i(z) = z + z_i$ and $f_i(z) = z - f(z_i)$. Using Lemma 2.2, we see that $h \to h$ and $f_i \to f$ uniformly if $h(z) = z + a$ and $f(z) = z - b$. Using Lemma 2.4 twice, we can replace g_i by $f_i g_i h_i$ and obtain a map that fixes 0 and ∞. Thus, after this replacement

$$g_i(z) = \lambda_i z.$$

If now there are $m > 0$ and $M > 0$ such that $m \leq |\lambda_i| \leq M$, we can pass to a subsequence so that $\lambda_i \to \lambda$. Thus if $g(z) = \lambda z$, $g_i \to g$ uniformly. Otherwise there is a subsequence so that either $\lambda_i \to 0$ or $\lambda_i \to \infty$. In both cases g_i is a convergence sequence so that $\{0, \infty\}$ is the set of attracting and repelling points.

Corollary 2.7. Let $f_i \in M$ be a sequence such that there are three distinct points $z_k \in \overline{\mathbf{C}}$, $k \leq 2$, such that $f_i(z_k) \to w_k$ where w_k are distinct. Then there is $f \in M$ such that $f_i \to f$ uniformly

If $f_i \in M$ tend pointwise toward a Möbius transformation, then the convergence is uniform.

The proof is left as an exercise. Use Lemma 2.1 and and Theorem 2.6 and note that no subsequence of (f_i) can be a convergence sequence.

A consequence is that a sequence of Möbius transformations converges uniformly as soon as it converges pointwise. Thus we can simply say that $f_i \to f$ since the convergence will be uniform.

3. Kleinian and Fuchsian groups.

We can now present the definition of Kleinian groups. Kleinian groups are subgroups of M whose action on $\overline{\mathbf{C}}$ satisfies a certain condition. We call subgroups of M Möbius groups. Thus elements of a Möbius group G are Möbius transformations, G contains the identity map, and if $f, g \in G$ both f^{-1} and $f \circ g$ are in M. In the sequel we usually denote fg for $f \circ g$ for brevity.

The *action* of G on $\overline{\mathbf{C}}$ is simply the rule $(g, x) \mapsto g(x)$ which assigns to each $g \in G$ and $x \in \overline{\mathbf{C}}$ the point $g(x)$ to which g moves x. Here this notion is so natural as to be obvious but sometimes one considers abstract groups which may have different actions on the same space.

Definition. The action of a Möbius group G is discontinuous at a point $x \in \overline{\mathbf{C}}$ if x has a neighborhood U such that $gU \cap U \neq \emptyset$ for only finitely many $g \in G$.

Definition. The group G is Kleinian if G is discontinuous at some point $x \in \overline{\mathbf{C}}$.

We denote

$$\Omega(G) = \{x \in \overline{\mathbf{C}} : G \text{ is discontinuous at some point } x \in \overline{\mathbf{C}}\}, L(G) = \overline{\mathbf{C}} \setminus \Omega(G).$$

The set $\Omega(G)$ is the *set of discontinuity* for or the *ordinary set* for G. The set $L(G)$ is *the limit set* of G. Points of $\Omega(G)$ are *ordinary points* of G and points of $L(G)$ are *limit points* of G. A direct characterization of $L(G)$ would be that $z \in L(G)$ if, given any neighborhood U of z, then $U \cap gU \neq \emptyset$ for infinitely many $g \in G$. We will see that the action of G is very different in these two sets.

The first rather obvious theorem for Kleinian groups is

Theorem 3.1. *If G is Kleinian, then every $z \in \overline{\mathbf{C}}$ is either an ordinary point or a limit point of G (but not both). The set of discontinuity $\Omega(G)$ is non-empty and open and the limit set $L(G)$ is closed.*

Proof. We need only to observe that if $z \in \Omega(G)$ and U is a neighborhood of z such that $U \cap gU \neq \emptyset$ for only finitely many $g \in G$ and since U is a neighborhood of $w \in U$, it follows that $U \subset \Omega(G)$. Thus $\Omega(G)$ is open and hence $L(G)$ is closed. By the definition of Kleinian groups $\Omega(G) \neq \emptyset$.

Examples. Obviously, the full Möbius group M is not Kleinian. Equally obviously, the trivial group G consisting of the identity mapping is Kleinian and $\Omega(G) = \overline{\mathbf{C}}$. Simple non-trivial examples are cyclic groups generated by a parabolic or a loxodromic element. If for instance $g(z) = 2z$, then the fixpoints 0 and ∞ are obviously limit points. If $z \neq 0, \infty$, then $U = \{w : |z|/\sqrt{2} < |w| < \sqrt{2}|z|\}$ is a neighborhood of z such that $U \cap gU \neq \emptyset$ only for $g = \mathrm{id}$. Similarly if g is parabolic with fixpoint ∞.

A slightly more complicated Kleinian group is

$$G = \{T_{n+mi} : n, m \in \mathbf{Z}\}$$

and where T_a is the translation $T_a(z) = z + a$. Its only limit point is ∞ and hence $\Omega(G) = \mathbf{C}$. We leave the details as an exercise.

These are all groups of the type called *elementary*. The construction of more complicated Kleinian groups requires more work.

G-invariant sets. A subset $A \subset \overline{\mathbf{C}}$ is G-invariant if $gA = A$ for every $g \in G$. A basic example of a G-invariant set are the sets $Gz = \{g(z) : g \in G\}$; such sets Gz are called *orbits* and Gz is the orbit of z. The ordinary and the limit set are G-invariant.

Theorem 3.2. *The ordinary set $\Omega(G)$ and the limit $L(G)$ of G are G-invariant.*

Proof. We need to prove this only for $\Omega(G)$ and it follows for $L(G)$ by taking the complements. Suppose that $z \in \Omega(G)$ and $\gamma \in G$. We claim that also $\gamma(z) \in\in \Omega(G)$. Since $z \in \Omega(G)$, there is a neighborhood U of z elements $g_1, \ldots, g_n \in G$ such that if $U \cap gU \neq \emptyset$ where $g \in G$, then g is some g_i. Let $V = \gamma U$ which is a neighborhood of $\gamma(z)$. Now, $V \cap gV \neq \emptyset$, $g \in G$, is equivalent to

$$\emptyset \neq \gamma^{-1}(V \cap gV) = \gamma^{-1}V \cap \gamma^{-1}g\gamma\gamma^{-1}V = U \cap \gamma^{-1}g\gamma U.$$

Thus $\gamma^{-1}g\gamma$ is some g_i, or equivalently, g is some $\gamma g_i \gamma^{-1}$ and hence $V \cap gV \neq \emptyset$, $g \in G$, only if $g \in \{\gamma g_1 \gamma^{-1}, \ldots, \gamma g_n \gamma^{-1}\}$

Conjugate groups. Let $\gamma \in M$. We denote $\gamma G \gamma^{-1} = \{\gamma g \gamma^{-1} : g \in G\}$. If $\Gamma = \gamma G \gamma^{-1}$, we say that Γ is obtained from G by conjugating with γ and Γ is *conjugate* to G. The map $\varphi : G \to \Gamma$, $\varphi(g) = \gamma h \gamma^{-1}$ is an isomorphism with inverse $g \mapsto \gamma^{-1}g\gamma$. If $\gamma \in G$, then $\gamma G \gamma^{-1} = G$ and so φ is in this case an isomorphism of G.

If $z \in \Omega(G)$, and U is a neighborhood of z such that $U \cap gU \neq \emptyset$ only if $g \in \{g_1, \ldots, g_n\}$, then we see as above that γU is a neighborhood of $\gamma(z)$ such that $\gamma U \cap g(\gamma U) \neq \emptyset$ only if $g \in \{\gamma g_1 \gamma^{-1}, \ldots, \gamma g_n \gamma^{-1}\}$. Thus $\gamma(z) \in \Omega(\Gamma)$. We have

Theorem 3.3. *If $\Gamma = \gamma G \gamma^{-1}$, then $\Omega(\Gamma) = \gamma\Omega(G)$ and $L(\Gamma) = \gamma L(G)$.*

Cyclic groups. The group generated by an $g \in M$ is denoted by $\langle g \rangle$ and it consists of elements of the form g^n, $n \in \mathbf{Z}$. If $g^n = \mathrm{id}$ for some $n > 0$, then the group is finite and $\langle g \rangle = \{g, g^2, ^n\}$. Otherwise, all g^n are distinct amd in this case the group is called *infinite cyclic*. Examples are groups generated by a parabolic or loxodromic g. These are conjugate to the group generated by $z \mapsto \lambda z$ or by $z \mapsto z + 1$. We have seen that these groups are Kleinian and the limit set is the fixpoint set of the generator and hence this is true for all infinite cyclic groups generated b a parabolic or loxodromic element.

On the other hand, we will see that, if g is elliptic, then g generates a Kleinian groups if and only if g is of finite order, i.e. $g^n = \mathrm{id}$ for some $n > 0$. Obviously, if g is of finite order $\langle g \rangle$ is finite and hence Kleinian.

Discrete groups. A Möbius group G is *discrete* if it is a discrete subset of M in the sense that if $f \in M$, then there are no sequences of distinct $g_i \in G$ such that $g_i \to f$ uniformly. We could define a metric d in M so that

$$d(f, g) = \sup_{z \in \overline{\mathbf{C}}} q(f(z), g(z))$$

and then discreteness of G would have the usual meaning that G has no accumulation points in M. However, the above definition is enough for us.

Theorem 3.4. *A Möbius group G is non-discrete if and only if there is a sequence of distinct elements $g_i \in G$ such that $g_i \to \mathrm{id}$.*

Proof. We need only to prove that if G is non-discrete, then there are distinct $g_i \in G$ such that $g_i \to \mathrm{id}$ uniformly. Since G is non-discrete, there are $f \in M$ and distinct $g_i \in G$ such that $g_i \to f$. Thus $g_i^{-1} \to f^{-1}$ (Lemma 2.4). Using again Lemma 2.4, have that

$$g_i \circ g_{i+1}^{-1} \to f \circ f^{-1} = \mathrm{id}$$

uniformly. Since g_i are distinct, $g_i \circ g_{i+1}^{-1} \neq \mathrm{id}$ and hence we can pick a subsequence so that g_i are distinct.

Theorem 3.5. *A Kleinian group G is discrete.*

Proof. If G is not discrete, then there is a sequence of distinct $g_i \in G$ such that $g_i \to \mathrm{id}$ uniformly. It follows that if $z \in \overline{\mathbf{C}}$ and U is a neighborhood of z, then $g_i U \cap U \neq \emptyset$ beginning from some n_0 (which depends on U). Hence $g_i \cap U \neq \emptyset$ for infinitely manu g_i and so the discontinuity set of G is empty and G is not Kleinian.

Remark. The converse is not true: there are discrete groups which are not Kleinian, i.e. the discontinuity set is empty. Sometimes one defines that a Kleinian group is a discrete Möbius group but then it may happen that $L(G) = \overline{\mathbf{C}}$ and $\Omega(G) = \emptyset$.

We can also characterize discreteness using convergence sequences:

Theorem 3.6. *A Möbius group is discrete if and only if every sequence of distinct elements $g_i \in G$ has a subsequence which is a convergence sequence.*

Proof. If $g_i \to g \in M$, then no subsequence of g_i is a convergence sequence and hence non-discreteness of G implies that there are sequence of distinct elements with no convergence subsequences.

Suppose that G is discrete and suppose that $g_i \in G$ are distinct. Theorem 2.7 says that g_i has a subsequence (denoted in the same manner) such that either $g_i \to g \in M$ uniformly or (g_i) is a convergence sequence. By non-discreteness the first case is impossible and hence there are convergence subsequences.

The following corollary will be important for us.

Corollary 3.7. *If G is Kleinian and $g_i \in G$ are distinct, then (g_i) has a convergence subsequence.*

It is possible to use matrices to characterize discreteness. This will help us later to find some non-trivial Kleinian groups.

We can identify a complex 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $(a, b, c, d) \in \mathbf{C}^4$. $\mathbf{C}^4$ has the usual euclidean metric so that the distance of z and w is

$$|z - w| = \sqrt{\sum_{i=1}^{4} |z_i - w_i|^2}$$

when $z = (z_1, z_2, z_3, z_4)$ and similarly w. Thus we can regard $SL(2, \mathbf{C})$ as a subset of $\mathbf{C}^4$ and this gives a topology to $SL(2, \mathbf{C})$. This is the topology what we used earlier in the convergence $A_i \to A$ of matrices.

If G is a Möbius group, we define

$$\hat{G} = \{A \in SL(2, \mathbf{C}) : f_A \in G\}.$$

It is easy to check that $\hat{G}$ is a subgroup of $SL(2, \mathbf{C})$. It is discrete if it has no accumulation points when regarded as a subset of $\mathbf{C}^4$.

Theorem 3.8. *A Möbius groups G is discrete if and only if $\hat{G}$ is discrete.*

Proof. We note that if A is an accumulation point of $\hat{G}$, then $A \in SL(2, \mathbf{C})$ since $\det A$ is a continuous function of A and $\det A = 1$ in $SL(2, \mathbf{C})$. Suppose that A is such an accumulation point. Thus there are distinct $A_i \in \hat{G}$ so that $A_i \to A$, implying that $f_{A_i} \to f_A \in M$ uniformly. Since $f_{A_i} = f_{A_k}$ for at most one $k \neq i$, we can find a subsequence so that f_{A_i} are distinct and hence G is not discrete if A is not discrete.

On the other hand, suppose that $\hat{G}$ is discrete but G is not. Thus there are distinct $g_i \in G$ such that $g_i \to g \in M$ uniformly. By Lemma 2.3, there matrices A_i and A in $SL(2, \mathbf{C})$ such that $A_i \to A$ and $f_{A_i} = g_i$ and $f_A = f$. Thus $A_i \in \hat{G}$ and so $\hat{G}$ is not discrete.

We have the following corollary, analogous to Theorem 3.4 and proved like it. The unit matrix $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is the neutral element of $SL(2, \mathbf{C})$.

Corollary 3.9. *A Möbius group G is discrete if and only if the unit matrix I has a neighborhood U such that $U \cap \hat{G} \setminus \{I\}$.*

Corollary 3.10. *A Kleinian group is countable.*

Proof. The group $\hat{G}$ is a discrete subset $\mathbf{C}^4 = \mathbf{R}^8$ and hence is countable as proved in topology. This implies the countability of G.

We can use these results to characterize elliptic elements in a Kleinian group.

Theorem 3.11. *If g is an elliptic element of a Kleinian group, then g is of finite order.*

Proof. We can conjugate the group so that the fixpoints of g are 0 and ∞, that is $g(z) = e^{i\alpha} z$ for some $\alpha \in \mathbf{R}$. Let

$$A = \{e^{in\alpha} : \alpha \in \mathbf{Z}\} \subset S^1$$

Thus g^n is of the form $e^{i\beta}$ where $\beta \in A$. If A is finite, then there are n and m such that $e^{i\alpha n} = e^{i\alpha m}$ where $n \neq m$, that is $g^n = g^m$ and hence $g^{n-m} \neq \text{id}$. Since $n - m \neq 0$, g is of finite order.

If A is not finite. then it a subset of the circle S^1 which is compact. Hence, there is a sequence $e^{in_k\alpha} \to e^{i\beta}$ where $e^{in_k\alpha}$ are distinct. Hence

$$\beta_k = e^{i(n_{k+1}\alpha)}/\alpha^{in_k\alpha} = e^{i(n_{k+1}-n_k)\alpha} \neq 0$$

and $\beta_k \to 0$ as $k \to \infty$. But if $\gamma_k = g^{n_{k+1}} g^{-n_k}$, then $\gamma_k = \beta_k(z)$ and hence $\gamma_k \to \text{id}$ and $\gamma_k \neq \text{id}$. Thus G cannot be discrete and hence G is not Kleinian.

Corollary 3.12. If $H = \langle g \rangle$ is a cyclic subgroup of a Kleinian group and H is infinite, then g is parabolic or loxodromic, otherwise elliptic.

The limit set and convergence sequences. There is a connection between the limit set and convergence sequence given by the following theorem. A convergence sequence of G is a convergence sequence g_i such that $g_i \in G$.

The following notation is handy

$$\#A = \text{ the number of elements of the set } A.$$

Theorem 3.12. *If g_i is a convergence sequence of a Kleinian group G, then the attractive and repelling points of (g_i) are in $L(G)$. Conversely, every $z \in L(G)$ is the attractive (or repelling) point of convergence sequence of G.*

Proof. Let a be the attractive point of such a convegence sequence g_i. Let b be the repelling point. Let $z \neq B$. Since the orbit Ga is countable by the countability of G, there exists $z \neq b$ such that $z \notin Ga$. Thus $g_i(z) \to a$ but $g_i(z) \neq a$ for every i. It follows that if U is a neighborhood of z, then $g_i(z) \in U$ for infinitely g_i. Thus

$$\#Gz \cap U = \infty$$

for infinitely many $g \in G$. This is possible only if $a \in L(G)$.

If a is the repelling point of g_i, then a is the attractive point of g_i^{-1}. Hence $a \in L(G)$ also in this case.

Suppose then that $z \in L(G)$, i.e. if U is a neighborhood of z, then $g(z) \cap U \neq \emptyset$ for infinitely many $g \in G$. Thus we can inductively define $z_i \in \overline{\mathbf{C}}$ and $g_i \in G$ such that

$$q(z, z_i) < 1/n \text{ and } q(z, g_i(z_i) < 1/n$$

and that g_i are distinct. Since $q(z, w) \leq 1$ for all z, w we can choose z_1 and g_1 freely. Suppose then that we have chosen z_i and g_i for $i \leq n$. Let $U = \{w : q(z, w) < 1/n\}$. Thus $U \cap gU \neq \emptyset$ for infinitely many $g \in G$, and hence we can find g_{n+1} distinct from earlier g_i and $z_i \in U$ so that $g_i(z_i) \in U$.

So we have such a sequence as claimed. Since g_i are distinct, we can pass to a convergence subsequence. Let a be its attractve and b the repelling point. If $z \neq a, b$, then $g_i(z_i) \to a$ which is impossible since $g_i(z_i) \to z$. Hence either $z = a$ or $z = b$. If $z = a$, we are done. If $z = b$, then z is attractive point of g_i^{-1} and we are again done.

Corollary 3.13. *The limit set of a Kleinian group G is empty if and only if G is finite.*

Proof.. Obviously. $L(G) = \emptyset$ is G is finite. If G is infinite, then there is a sequence g_i of distinct elements of G. This has a convergence subsequence (Corollary 3.7). The attractive point of this convergence sequence is in $L(G)$.

We can give as a consequence another characterization of the limit set and have some results on topological character of the limit set.

Theorem 3.14. *The limit set $L(G)$ of a Kleinian group is the set of accumulation points of any orbit Gz where $z \in \Omega(G)$.*

Proof. Let $z \in \Omega(G)$ and let $a \in L(G)$. Thus there is a convergence sequence (g_i) of G whose attractive point is a. Since the repelling point is a limit point, z cannot be the repelling point and hence $g_i(z) \to a$. Since $g_i(z) \in \Omega(G)$, $g_i(z) \neq a$ and hence a is an accumulation point of $\{g_i(z)\}$ and hence of Gz.

Conversely, if a is an accumulation point of Gz, then every neighborhood U of a contains infinitely many points of the form $g(z)$, $g \in G$, and hence $U \cap gU \neq \emptyset$ for infinitely many $g \in G$.

This theorem gives a reason for the name "limit point". Such a point is the limit of points in a fixed orbit Gz. An ordinary point cannot be such a limit.

Kleinian groups are divided into two types depending on the number of points in the limit set:

Definition. A Kleinian group G is elementary if $\#L(G)$ is at most 2. Otherwise G is non-elementary.

In topology, one calls closed sets which do not contain isolated points "perfect sets". Thus every point of the set is an accumulation point of the set. Limit sets of non-elementary Kleinian groups are of this type and hence if there are more than two points in $L(G)$, $L(G)$ is actually an infinite set. One can show that $L(G)$ is even uncountable in this case.

Theorem 3.15. *The limit set of a non-elementary Kleinian group is a perfect set.*

Proof. The basic fact is that since G is non-elemetary, the limit set contains at least 3 points. Thus, given a convergence sequence there is a point of $L(G)$ which is not the attractive nor the repelling point of the sequence.

Let $a \in L(G)$. Thus there is a convergence sequence (g_i) of G whose attractive point is a. Let b be the repelling point. Since $\#L(G) > 2$, there is $z \in L(G)$ distinct from a and b. Thus $g_i(z) \to a$. If $\{g_i(z)\}_{i>0}$ is infinite, then a is an accumulation point of $\{g_i(z)\}$ and hence of $L(G)$ since $g_i(z) \in L(G)$.

If $\{g_i(z)\}_{i>0}$ is finite, then $g_i(z) = a$ for infinitely many g_i. Thus we pass to a subsequence so that $g_i(z) = a$. Let $h_i = g_i g_1^{-1}$. Then $h_i(a) = a$ and h_i is still a convergence sequence with attracting point a. The repelling point b may have changed but still there is $z \in L(G) \setminus \{a, b\}$. Since $h_i(a) = a$, $h_i(z) \neq a$ and we see that a is an accumulation point of $\{h_i(z)\}$ and of $L(G)$.

Another theorem in the same vein is

Theorem 3.16. *Let G be a discrete Möbius group and let A be a G-invariant closed set such that $A \neq \overline{\mathbf{C}}$ and that A contains at least 2 points. Then G is Kleinian and $L(G) \subset A$.*

Proof. We is enough to show that G acts discontinuously at every point $z \in \overline{\mathbf{C}} \setminus A$. If G is not discontinuous at such a point, we can find like in the proof of Theorem 3.10 points $z_i \in \overline{\mathbf{C}}$ and $g_i \in G$ such that $z_i \to a$ and $g_i(z) \to a$ and that g_i are distinct. Thus we can pass to a convergence subsequence. Let a be the attractive point *and* b the repelling point.

We claim that $a \in A$. To see this, we can find a point $w \in A \setminus \{b\}$ since $\#A \geq 2$. Thus $g_i(z) \to a$. If $g_i(z) = a$ for some i, then $a \in A$ since A is G-invariant. If $g_i(z) \neq a$ for all i, then a is an accumulation point of $\{g_i(z)\}_{i>0}$ and hence of A. However, A is closed and so $a \in A$.

Similarly, b is the attractive point of (g_i^{-1}) and a similar reasoning shows that $b \in A$.

Thus $a, b \in A$. Since $z \notin A$, z is distinct from a and b and we can find a neighborhood W for z, a neighborhood V of b and a neighborhood U of a such that $W \cap V = W \cap U = \emptyset$. Now, g_i is a convergence sequence such that $g_i(\overline{\mathbf{C}} \setminus V) \subset U$ beginning from some $i = n_0$. Thus $g_i W \subset U$ if $i \geq \cap n_0$ and hence $W \cap \setminus g_i W = \emptyset$ if $i \geq n_0$. This contradicts the fact that $z_i \to z$ and $g_i(z) \to z$.

So every $z \in \overline{\mathbf{C}} \setminus A$ is an ordinary point and hence G is Kleinian and $L(G) \subset A$.

Remark. It is essential that $\#A \geq 2$. Obviously, $\emptyset$ is G-invariant but $L(G) \not\subset \emptyset$ unless G is finite. A more non-trivial counterexample is give *by* $G = \langle g \rangle$ where g is loxodromi. Now, $L(G)$ is the set of fixpoints of g. If $A = \{v\}$ where v is a fixpoint of g, then A is G-invariant but $L(G) \not\subset \{v\}$.

Example. Let $SL(2, \mathbf{Z})$ consists of matrices of $SL(2, \mathbf{C})$ so that all the entries of matrices are integers. It is obviously a discrete subset of $\mathbf{C}^4$ when we identify $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $(a, b, c, d) \in \mathbf{C}^4$. It follows that the group

$$M = \{f_A : A \in SL(2, \mathbf{Z})\}$$

is discrete (Theorem 3.8). If

$$g(z) = \frac{az+b}{cz+d}$$

where the coefficients are real, then obviously $gA = A$ for $A = \mathbf{R} \cap \{\infty\}$. It follows that M is Kleinian and $L(M) \subset \mathbf{R} \cup \{\infty\}$. This group is called the *modular group*. It is important in number theory and one can show that $L(M) = \mathbf{R} \cup \{\infty\}$.

4. The quotient surface of a Kleinian group

A very important notion in the study of a Kleinian group is the quotient surface obtained from $\Omega(G)$ as follows. We note first that the orbits $Gz = \{g(z) : g \in G\}$ form a partition of $\overline{\mathbf{C}}$, that is, every point of $w \in \overline{\mathbf{C}}$ is in one and only one orbit Gz. That is the relation $\sim$ such that

$$z \sim w \Leftrightarrow \text{ there is } g \in G \text{ such that } g(z) = w \tag{1}$$

is an equivalence relation and its equivalence classes are the orbits Gz, $z \in \overline{\mathbf{C}}$. This follows from the group properties of G and we leave the proof of this fact as an (easy) exercise.

It turns out that the orbits Gz where $z \in L(G)$ behave very differently from the orbits Gz, $z \in \Omega(G)$. In the first case Gz is usually dense in $L(G)$ (proved like Theorem 3.16 if $\#Gz \geq 2$). On the other hand, we know by Corollary 3.15 that Gz is a discrete subset of $\Omega(G)$. This is the fact that makes it very useful to consider the space of orbits Gz, $z \in \Omega(G)$. Note that since $\Omega(G)$ is G-invariant, $Gz \subset \Omega(G)$ if $z \in \Omega(G)$, and thus $\{Gz : z \in \Omega(G)\}$ is a partition of $\Omega(G)$, We denote

$$\Omega(G)/G = \{Gz : z \in L(G)\}$$

that is, $\Omega(G)/G$ is the set of equivalence classes of $\sim$ when restricted to $\Omega(G)$. We call $\Omega(G)/G$ the *quotient of* $\Omega(G)$ *by* G or more simply the *quotient surface* since it will turn out that $\Omega(G)/G$ is a surface in the topology that it inherits from $\Omega(G)$, in the so-called factor topology.

We will show in this section that $\Omega(G)/G$ is a surface, that is every point has a neighborhood homeomorphic to an open subset of the complex plane. We will even show that $\Omega(G)/G$ has a natural conformal structure so that it is a Riemann surface. We will define this notion later.

We start with some preliminary results. If $z \in \Omega(G)$, we set

$$G_z = \{g \in G : g(z) = z\}.$$

G_z is the *stabilizer* of z (in G). It is easily seen to be a subgroup of G. More generally, if $A \subset \Omega(G)$, we set

$$G_A = \{g \in G : gA = A\}$$

which is also a subgroup of G.

The following theorem will be important for the quotient surface. The letter G will denote a Kleinian group during this section.

Theorem 4.1. *If $a \in \Omega(G)$, then G_a is either the trivial group $\{\mathrm{id}\}$ or is a finite cyclic group generated by an elliptic element g which is conjugate to a map of the form $z \mapsto e^{2\pi/n}$ where n is the order of g, that is the smallest number $n > 0$ such that $g^n = \mathrm{id}$.*

Proof. Suppose that $h \in G_z \setminus \{\mathrm{id}\}$. We note that G_a must be a finite group since if U is any neighborhood of a, then

$$(*) \qquad gU \cap U \neq \emptyset$$

for any $g \in G_a$. However, we know that there are U such that the number of $g \in G$ satisfying (*) is finite. Hence G_a is finite. It follows that G_a cannot contain parabolic or loxodromic g since $g \in G_a$ implies all powers g^n are in G_a. Hence G_a is all elliptic group (we regard the identity element also as elliptic).

Let now $h, g \in G_a$ be two non-identity elements. Thus both fix a and let b the other fixpoint of h and c the other fixpoint of g. We claim that $b = c$. This follow since if $b \neq c$, there are parabolic elements in the group G_z (this was exercise 3 in the 3rd exercise set) which is seen to be *impossible*. The existence of parabolic elements is easiest to see if we conjugate the group so that $a = \infty$ and $b = 0$ so that $h(z) = \lambda z$ and $g(z) = \nu z + \beta$. An easy calculation shows that if $k(z) = hg^{-1}h^{-1}$, then $k(z) = \nu^{-1}z + \lambda\beta$ and $kh(z) = z + \alpha$ where $\alpha \neq 0$ since the fixpoint z_0 of g is not the fixpoint of k and hence $kg(z_0) \neq z_0$ and hence $kg \neq \mathrm{id}$ and so $\alpha \neq 0$. $k(z) = z + \alpha'$

Thus there is $b \neq a$ so that all elements of G_a are elliptic with fixpoints a and b. Let $\gamma \in M$ be an element such that $\gamma(a) = 0$ and $\gamma(b) = \infty$. If we conjugate G by γ, we obtain a group Γ such that $\Gamma_0 = \gamma G_a \gamma^{-1}$ and all elements of Γ_0 fix 0 and ∞. It suffices to prove the theorem for Γ and Γ_0 and we can assume that $a = 0$ and $b = \infty$ (and return to original notation). Thus every $g \in G_a = G_0$ is of the form $g(z) = e^{i\alpha_g}z$ where $0 \leq \alpha_g < 2\pi$. Since G_a is finite, we can enumerate $G_a \setminus \{\mathrm{id}\}$ as $\{h_1, \dots, h_n\}$ so that if $\beta_i = \alpha_{h_i}$, then

$$0 < \beta_1 < \cdots < \beta_n < 2\pi.$$

We claim that if we set $g = h_1$ then $h_k = g^k$ or equivalently $b_k = k\beta_1$. Obviously $g^1 = h_1$ and we are done if $n = 1$.

If $n > 1$ and if $\beta_1 \neq 2\beta_1$, then we would have $\beta_1 < \beta_2 < 2\beta_1$ and thus $0 < \beta_2 - \beta 1 < \beta_1$. However $h_2h_1^{-1} \in G_a$ and $h_2h_1^{-1}(z) = (\beta_2 - \beta_1)z$. This contradicts the definition of β_i and so $g^2 = h_2$. If $n > 2$ and $\beta_3 \neq 3\beta_1$ we would obtain a similar contradiction from $0 < (k-1)\beta_1 = \beta_{k-1} < \beta_k < k\beta_1$.

Corollary 4.2. *If $g, h \in G \setminus \{\mathrm{id}\}$ have a common fixpoint $a \in \Omega(G)$, then they fix the same points and are in a cyclic subgroup generated by an elliptic element of G.*

Remark. This need not be true $a \in L(G)$ as is shown by Ex. 2 of 3rd set of exercises.

Lemma 4.3. *The set of $a \in \Omega(G)$ such that $G_a \neq \{\mathrm{id}\}$ is a discrete subset of $\Omega(G)$.*

Proof. Let $A = \{a \in \Omega(G) : G_a \neq \{\text{id}\}\}$ and suppose that $b \in \Omega(G)$ is an accumulation point of A. Thus there is a sequence $a_i \in A$ of distinct points such that $G_i = G_{a_i}$ is not trivial. Let g_i be the generator of G_i. Thus g_i is elliptic and has fixpoint a_i. We can pass to a subsequence so that $a_i \neq c_k$ for all k. Choose $g_i \in G_i \setminus \{\text{id}\}$. If U is a neighborhood of a, then $a_i \neq U \cap g_i U \neq \emptyset$ for infinitely many i. If $g_i = g_k$ where $g_i \neq g_k$, then the fixpoint set of $g_i = g_k$ is $\{a_i, a_k\}$. Thus we see that $g_i U \cap U \neq \emptyset$ actually for infinitely many g_i's (and not only for infinitely many i's).

Theorem 4.4. *Let $z \in \Omega(G)$. Then z has arbitrarily small neighborhoods U such that $gU = U$ if $g \in G_z$ and $gU \cap U = \emptyset$ if $g \in G \setminus G_z$.*

Proof. Since $z \in \Omega(G)$, z has a neighborhood U such that the set

$$\{g \in G : gU \cap U \neq \emptyset\} = H$$

is finite. Obviously, $H \supset G_z$. Suppose that $g \in H$ and $g(z) \neq z$. Choose a neighborhood W for $g(z)$ and a neighborhood V for z such that $W \cap U = \emptyset$. Since g is continuous, there is $U_0 \subset W$ such that $gU_0 \subset V$. Thus $U_0 \cap gU_0 = \emptyset$. Thus if we replace U by U_0, the number of elements of H is decreased by 1. If we continue in this manner, we see that we can replace U by a smaller set, still denoted U, so that $H = G_z$. Since G_z is finite,

$$V = \bigcap_{g \in G_z} gU$$

is a finite intersection of neighborhoods of z and hence a neighborhood of z. If $h \in G_z$,

$$gV = \bigcap_{g \in G_z} hgV = \bigcap_{g \in G_z} gV$$

since $g \mapsto hg$ is a bijection of G_z onto itself.

The quotient topology. We recall the quotient topology (this is treated in Topology II) what we will use to topologize $\Omega(G)/G$. A topology of a set X is given by giving its open sets, that is a family T of subsets of X. In order to be a topology of X, T needs to satisfy

(a) $\emptyset \in T$ and $X \in T$.
(b) If U_i, $i \in I$, are in T, then $\bigcup_{i \in I} U_i$ is in T.
(c) If $U_1, \dots, U_n \in T$, then $U_1 \cap \ldots \cap U_n \in T$.

The elements $U \in T$ are the open subsets of X and a set is closed if and only if its complement is open. The conditions $a)$ - $b)$ is everything that is needed to define topological notions like continuity etc.

Let now $\sim$ be an equivalence relation of X. Let $X/\sim$ be the set of equivalence classes of $\sim$ and let $p : X \to X/\sim$ be the canonical projection so that

$$p(x) = \text{the equivalence class of } x.$$

The quotient topology is defined so that $U \subset X/\sim$ is open if and only if $p^{-1}U$ is open. It is easy to see that the open subsets of $X/\sim$ defined in this manner satisfy conditions a)-b).

The following theorem is useful in connection with the quotien topology. Here Y is another topological space. Recall that an open map is a map such that fU is open whenever U is open. Actually, we do not need this notion but state it as some kind of characterization of the quotient topology.

Theorem 4.5. *A map $f : X/\sim \to Y$ is continuous if and only if $f \circ p : X \to Y$ is continuous. In particular, the canonical projection is continuous.*

A bijection $f : X/\sim \to Y$ is a homeomorphism if $f \circ p$ is continuous and open.

We refer to the topology for the course. We do not use it here but it gives a characterization of the quotient topology.

The quotient surface. A topological space X is a surface if it is a Hausdorff space such that any point has a neighborhood which is homeomorphic to an open subset of the plane $\mathbf{C} = \mathbf{R}^2$. A space is Hausdorff if any two points x and y, $x \neq y$, have disjoint neighborhoods.

An example of a surface is $\overline{\mathbf{C}}$. If $z \in \mathbf{C}$, then $\mathbf{C}$ is a neighborhood of z homeomorphic to an open subset of $\mathbf{C}$. If $z = \infty$, then $\overline{\mathbf{C}} \setminus \{0\}$ is homeomorphic to $\mathbf{C}$ and the homeomorphism between them is $f(z) = 1/z$.

Another example is the sphere $S^2 = \{z \in \mathbf{R}^3 : |z| = 1$. We leave it as an exercise to check that it is a surface. One can use the projections like $(x, y, z) \to (x, z)$ (there are 3 different projections of this kind), restricted to suitable subsets to have neighborhoods of the kind required.

Still another is the product of two circles, called torus. The typical torus is $S^1 \times S^1$. Again, it is an exercise to show that this is a surface.

Now, we come to our main theorem. We use the following notation to denote $p(z)$ and $p(A)$: We set

$$p(z) = Gz = \tilde{z}.$$
$$pA = \{Gz : z \in A\} = \tilde{A}.$$

We start with the following lemma.

Lemma 4.6. *a) If $U \subset \Omega(G)$ is open, then $\tilde{U}$ is open and hence the canonical projection is an open mapping..*

b) If $z \in \Omega(G)$ and U_i, $i \in I$, form a basis of neighborhoods of z, then $\tilde{U}_i$ $i \in I$ form a basis of neighborhoods for $\tilde{z}$.

Proof. To prove a), we need only to note that $p^{-1}\tilde{U} = \bigcup_{g \in G} gU$ is open as a union of open sets.

To prove b), let W be a neighborhood of $\tilde{x}$, $x \in \Omega(G)$. Thus $U = p^{-1}(\tilde{x})$ is an open set containing x. Thus there is some $U_i \subset U$. Obviously, $Gz \subset U$ if $z \in U$ and hence $GU_i = \bigcup_{z \in_i} Gz \subset U$, showing that $\tilde{U}_i \subset \tilde{U}$.

A notion that is important in the study of the quotient is that of a *covering map*. A map $p : X \to Y$ is called a covering map or a covering projection if the following is true. Any point $x \in X$ has a neighborhood

U such that $p^{-1}U = \bigcup_{i \in I} V_i$ where V_i are disjoint subsets of X such that $p|V_i$ is a homemorphism $V_i \to U$.

Theorem 4.7. The quotient $\Omega(G)/G$ is a surface in the quotient topology. If G does not contain elliptic elements, the canonical projection $p : \Omega(G) \to \Omega(G)/G$ is a covering map.

Proof. We first show that $\Omega(G)/G$ is Hausdorff. Let $x, y \in \Omega(G)$ be points such that $\tilde{x} \neq \tilde{y}$. If $\tilde{x}$ and $\tilde{y}$ do not have disjoint neighborhoods, we derive a contradiction as follows. Let $U_n = B(x, 1/n)$ and $V_n = B(y, 1/n)$ when $B(z, r) = \{w \in \overline{\mathbf{C}} : q(w, z) < r\}$. Thus $\tilde{U}_n$ is a neighborhood of $\tilde{x}$ and $\tilde{V}_n$ a neighborhood of $\tilde{y}$. If $\tilde{U}_n \cap \tilde{V}_n \neq \emptyset$, there is $x_n \in U_n$ and $y_n \in V_n$ such that $\tilde{x}_n = \tilde{y}_n$, that is $Gx_n = Gy_n$. Thus there is $g_n \in G$ such that $g_n(x_n) = y_n$. Thus

$$x_n \to x \text{ and } y_n = g_n(x_n) \to y.(*)$$

There are two cases depending on whether $\{g_n : n > 0\}$ is finite or infinite. If it is finite, then it is possible to pass to a subsequence so that $g_n = g$ for some fixed g. If this is the case, it would follow from the convergences (*) that $g(x) = y$, contrary to the assumption that $\tilde{x} = Gx \neq \tilde{y} = Gy$.

If $\{g_n : n > 0\}$ is infinite, it is possible to pass to convergence subsequence. Its repelling point is a limit point and hence distinct from x. Thus $g_n(x_n)$ tend toward the attracting point which is also a limit point and so distinct from y. This contradicts (*). We have proved that $\Omega(G)/G$ is Hausdorff.

We then show that $\tilde{x}$ has a neighborhood homeomorphic with an open subset of $\mathbf{C}$. This is easier if $G_x = \{\mathrm{id}\}$ and so we first assume this. In this case x has a neighborhood U such that $gU \cap U = \emptyset$ if $g \in G \setminus \{\mathrm{id}\}$. We claim that in this case $p|U$ is a homeomorphism $U \to \tilde{U}$.

Note that it may be that $\infty \in U$ and hence U is not a subset of $\mathbf{C}$. This does not matter, since we can assume that $U \neq \overline{\mathbf{C}}$ and in this case U can be mapped by a homeomorphism, for instance by a Möbius transformation onto an open subset of $\mathbf{C}$.

The map $p|U$ is surjective by the definition of $\tilde{U}$. It is injective, since $Gz \cap U$ contains at most one point.

The continuity of $p|U$ follows from the fact that the canonical projection is continuous. Thus it suffices to show that if $V \subset U$ is open, then $pV = \tilde{V}$ is open. But this is the case by Lemma 4.6

It now follows easily that if G does not contain elliptic, elements, and U is a neighborhood of x such that the sets gU, $g \in G$, are disjoint, then $p^{-1}\tilde{U} = \bigcup_{g \in G} gU$ and $p|gU$ is a homeomorphism $gU \to \tilde{U}$ for every $g \in G$. Thus p is a covering projection.

If G_x is not trivial, then it is more difficult to prove that $\tilde{x}$ has a neighborhood homeomorphic to an open subset of the plane. In any case G_x is a cyclic group generated by an element conjugate to $h(z) = e^{2\pi i/n}z$.

We assume first that $x = 0$ and that G_x is generated by the above map $h(z) = e^{2\pi i/n}z$. Thus ∞ is the other fixpoint of non-trivial elements of G_x.

Again, we choose a neighborhood U of x such that $gU \cap U = \emptyset$ if $g \in G \setminus G_x$ and $gU = U$ if $g \in G_x$. By making U smaller if necessary we can assume that

$$U = \{z \in \mathbf{C} : |z| < R\}.$$

Let $f(z) = z^n$. Thus f maps U onto $V = \{z : |z| < R^n\}$. The map f is not injective, but the points $re^{i(\varphi+2\pi k/n)}$ are mapped onto the same point $w = r^n e^{in\varphi}$. Thus $f^{-1}\{w\}$ consists of these n points if $w \neq 0$, but $f^{-1}\{0\} = \{0\}$. The map f is useful to us since if $z = re^{i\varphi}$, then

$$(**) \qquad Gz \cap U = \{re^{i(\varphi+2\pi k/n)} : k = 0, \dots, n-1\} = f^{-1}\{w\}.$$

Thus we can define a map $k : \tilde{U} \to V$ so that $k(\tilde{z}) = f(z) = z^n$. Obviously, k is surjective and it is injective by (**). Hence k is bijective and we will show that it is a homeomorphism.

The continuity of k follows from the general properties of the quotient topology (Lemma 4.6) since $k \circ p(z) = f(z) = z^n$ is a continuous map $U \to V$. So we need only to show that k is open. Suppose that $W \subset V$ is open, then $kW = kp(p^{-1}W \cap U)$. Here $p^{-1}W$ is open by the definition of the quotient topology and henced so is $W' = p^{-1}W \cap U$. However, $f(z) = kp(z) = z^n$ is an open mapping and so $fW' = kW$ is open. We leave it as exercise to show that z^n is open. This is a special case of the more general fact that non-constant analytic functions are open.

We still need to justify that we can assume that $x = 0$ and that G_x is generated by $h(z) = e^{2\pi i/n}$. This follows by the next theorem which we will need to apply separately for each elliptic fixpoint $a \in \Omega(G)$ and find a suitable conjugation for each such a.

Theorem 4.7. *Let $\Gamma = hGh^{-1}$. Then $h(Gz) = \Gamma h(z)$ and the map $\tilde{h} : \Omega(G)/G \to \Omega(\Gamma)/\Gamma$ defined $\tilde{h}(Gz) = \Gamma h(z) = h(Gz)$ is a homeomorphism in the quotient topology.*

Proof. We have

$$\Gamma h(z) = \{\gamma(z) : \gamma \in \Gamma\} = \{hgh^{-1}h(z) : g \in G\} = \{hg(z) : g \in G\} = h(Gz)$$

and so the map $\tilde{h}$ is well-defined. Thus h maps the partition $\{Gz : z \in \Omega(G)\}$ of G onto the partition $\{\Gamma z : z \in \Omega(\Gamma)\}$ of $\Omega(\Gamma)$. Since h is a homeomorphism, i.e. sends the topology (that is open subsets) of $\Omega(G)$ bijectively onto the topology of $\Omega(\Gamma)$, it is obvious that the induced map is also a homeomorphism. If one needs a formal proof, it is easy.

Example. Let G be the parabolic group generated by $g(z) = z+1$. Thus $G = \{g^n : n \in \mathbf{Z}\}$. The point ∞ is a parabolic fixpoint and hence a limit point and it is not difficult to see that $\Omega(G) = \mathbf{C}$. The orbit of a point $z \in \mathbf{C}$ is

$$Gz = \{z + n : n \in \mathbf{Z}\}.$$

We claim that $\Omega(G)/G = \mathbf{C}/G$ is homeomorphic to $X = S^1 \times \mathbf{R}$. Here S^1 is the unit circle whose elements are of the form e^{iz}. Define $h : \mathbf{C} \to X$ so that

$$h(x+iy) = (e^{2\pi ix}, y)$$

We note that $f(z) = f(w)$ if and only if $w = z + n$ for some $n \in \mathbf{Z}$. Thus

$$(*) \qquad h^{-1}(x,y) = Gz$$

if $z = x + iy$. We can define $f : \Omega(G)/G \to X$ so that $h(Gz) = f(z)$. Thus $h = f \circ p$ if $p : \Omega(G) \to \Omega(G)/G$ is the canonical projection. Obviously, h and f are surjective and in view of (*), h is a injective and hence a bijection. Since f is continuous, it follows from Theorem 4.5 that h is continuous. The map $h = f \circ p$ is open (this is a general fact for non-constant analytic mappings and easy to see directly). Hence the second part of this theorem implies that f is open. We have proved that f is a homeomorphism.

The method of the above example is difficult to apply to more complicated groups, but there are other methods to find the homeomorphism class of the quotient surface.

Riemann surfaces. It is possible to give a complex analytic structure to the quotient surface. These structure are called conformal structure and surface provided with this kind of structure is called a Riemann surface. If S is a Riemann surface we can speak of analytic and conformal maps of S.

A *Riemann surface* is a surface such that there are given a family U_i, $i \in I$, of open subsets of S as well homeomorphisms $\Phi_i : U_i \to V_j$ of U_i onto an open subset V_i of $\mathbf{C}$. If $U_i \cap U_j \neq \emptyset$, then we denote by

$$(2) \qquad \Phi_{ij} = \Phi_j \circ \Phi_i^{-1} | \Phi_i(U_i \cap U_j)$$

which is a hoemomorphism $\Phi_i(U_i \cap U_j) \to \Phi_j(U_i \cap U_j)$ and we require that these maps are conformal. That is they are analytic and the derivative is non-zero at all points.

The family $(U_i, \Phi_i)_{i \in I}$ is a called a *conformal atlas* of S. The maps Φ_i are often called chart mappings or charts. Often one uses the name local coordinates or coordinate maps since one can transfer considerations of analyticity on S by means of the coordinate maps to $\mathbf{C}$. The maps Φ_{ij} in (2) are called coordinate change maps.

The coordinate maps allow defining analyticity and conformality on Riemann surfaces. For instance, a map $f : S \to \mathbf{C}$ is analytic if all the maps $f \circ \Phi_i^{-1} : V_j \to \mathbf{C}$ are analytic. Note that if $z \in U_i \cap U_j$ and we set $z_i = \Phi_i(z)$ and $z_j = \Phi_j(z)$, then $a = (f \circ \Phi_i^{-1})'(z_i)$ may differ from $b = (f \circ \Phi_j^{-1})'(z_j)$ but $a \neq 0$ implies $b \neq 0$.

In the other direction, a map $f : V \to S$, $V \subset S$ open is analytic, if Φf, Φ a chart mapping, is analytic at all points where it is defined.

We define that a map $\Phi : U \to V$ between an open subset of S and an open subset of $\mathbf{C}$ is conformal if $(\Phi \circ \Phi_i^{-1})'(z) \neq 0$ for all $z \in \Phi_i(U \cap U_i)$. A map $\Psi : V \to U$ is conformal if $(\Phi_i \circ \Phi)'(z) \neq 0$ if $\Phi(z) \in U_i$. In particular, all *chart mappings are conformal.*

More generally, if S is a Riemann surface with atlas A and S' is another Riemann surface with atlas A', then $f : S \to S'$ is conformal if $\Psi_j \circ f \circ \Phi_i^{-1}$,

$\Phi_i \in A$ and $\Psi_j \in A'$, is conformal at points where it is defined. Note that we do not require that f is a homeomorphism. It is useful to note that the composition of conformal mappings is still conformal.

Often, if given an atlas $A = (U_i, \Phi_i)$ one adds to it all maps $\Phi : U \to V$ (U open subset of S and V an open subset of $\mathbf{C}$) which are conformal in the above sense. It satisfies the compatibility condition and is called the maximal conformal atlas of S compatible with A.

Example. 1. Any open subset U of $\mathbf{C}$ is a Riemann surface in the above sense. In this case the conformal atlas contains just the pair (U, id).

2. The extended complex plane $\overline{\mathbf{C}}$ can be given a conformal structure as follows. The conformal atlas consists of two coordinate maps $\Phi_i : U_i \to V_1$, $i = 1, 2$, such that $U_1 = V_1$ and Φ_1 is the identity map $U_1 \to V_1$ and $\Phi_2 = 1/z$ and is a map $U_2 = \overline{\mathbf{C}} \setminus \{0\} \to V_2 = \mathbf{C}$. Now, $U_1 \cap U_2 = \mathbf{C} \setminus \{0\}$ and the map Φ_{12} (and also the map Φ_{21}) is the map $1/z$ regarded as a map $\mathbf{C} \setminus \{0\}$ onto itself.

Theorem 4.8. *The quotient surface $\Omega(G)/G$ has a conformal structure in which the canonical projection p is analytic and is conformal at each point z not fixed by an elliptic element.*

Remark. The fact that p is analytic and in addition conformal outside elliptic fixpoints implies that this conformal structure is uniquely determined. The verification is not difficult and left as an exercise.

Helpful in the proof. In the construction of the conformal structure, it is natural to allow that the coordinate maps Φ_i are homeomorphisms of open subsets of $\Omega(G)/G$ onto open subsets of $\overline{\mathbf{C}}$. Conformality is defined between maps of open subsets of $\overline{\mathbf{C}}$ using the auxiliary map $\varrho(z) = 1/z$ and we require that the coordinate change maps are conformal in this sense. We can always go back to the original definition so that Φ_i is replaced by the map $\varrho \circ \Phi_i = 1/\Phi_i$ and possibly by making the domain of definition of Φ_i smaller.

Proof if there are no elliptic elements. Assume first that there are no elliptic elements. In this we define the conformal atlas A so that it consists of charts obtained as follows. Actually, the charts are the maps constructed in the proof that $\Omega(G)/G$ is a surface. Thus if $x \in \Omega(G)$, then, since there are no elliptic elements, x has a neighborhood V such that $gV \cap V = \emptyset$ if $g \in G \setminus \{\mathrm{id}\}$. Set $U = pV$ and since $Gz \cap V$ consists of at most one point, there is a well-defined homeomorphism $\Phi : U \to V$ such that $\Phi \circ p(z) = z$. This is a homeomorphism as we have seen.

Since p is open $V = pU$ is open. Obviously, every point of $\Omega(G)/G$ is in a set of this form and hence they form an open cover of $\Omega(G)/G$. Thus we must only show the conformality condition for two charts. Let $\Psi : U' \to V'$ be another chart obtained in this manner. We claim that $\Phi_0 = \Psi \circ \Phi^{-1}$ is conformal at $z \in \Phi(U \cap U')$. Let $w = \Phi_0(z)$. Thus $\Psi^{-1}(w) = p(w) = \Phi^{-1}(z) = p(z)$. Thus $w = g(z)$ for some well-defined $g \in G$; note that g may depend on z. However, z has a neighborhood W

such that $gW \subset U'$. If $\zeta \in W$, then $g(\zeta) \in U'$ and hence $g(\zeta) = \Phi_0(\zeta)$ since $g(\zeta)$ is the only point of $G\zeta \cap U'$. Since Möbius transformations are conformal, Φ_0 is indeed conformal at z.

Thus the chart mappings are local inverses of the canonical projection p. A *local inverse* of p is a a homeomorphism map $q : U \to V$ where $U \subset \Omega(G)/G$ and $V \subset \Omega(G)$ are open and $pq(z) = z$ on U. If $q : U \to V$ is such a local inverse, then p is injective on V and hence U is an open subset of $\Omega(G)$ such that $\#(Gz \cap V) \leq 1$. Hence $(U, q) \in A$.

It is immediate from the definitions that the canonical projection p is conformal. Let $x \in \Omega(G)$ and $y = p(x)$. Suppose that $y \in U$ such that $\Phi : U \to V$ is a chart mapping. Thus $\Phi p(x) = g(x)$ for some $g \in G$ and we see as above that this is true also in some neighborhood of x. Hence Φp is conformal at points where it is defined.

Proof if there are elliptic elements. Let $E \subset \Omega(G)$ be the set of elliptic fixpoints of G, that is $x \in E$ if there is elliptic $g \in G \setminus \{\mathrm{id}\}$ such that $g(x) = x$ and set

$$\Omega'(G) = \Omega(G) \setminus E.$$

We note that $\Omega'(G)$ is G-invariant and hence we can consider the quotient $\Omega'(G)/G = \{Gx : x \in \Omega'(G)\} \subset \Omega(G)/G$. Since every $x \in E$ has a neighborhood U such that $U \cap gU = \emptyset$ if $g \in G \setminus G_x$, the set pE is a discrete subset of $\Omega(G)/G$ and hence

$$\Omega/G)/G = \Omega'(G)/G \cup pE$$

is obtained from $\Omega'(G)/G$ by adding a discrete subset.

The first part of the proof applies if we replace $\Omega(G)$ by $\Omega'(G)$ and also $p|\Omega'(G)$ is conformal. Hence we have an atlas A of $\Omega'(G)/G$ whose elements are local inverses of the canonical projection. We will extend this conformal structure to $\Omega(G)/G$ so that if $u \in pE$, we find a neighborhood $U \subset \Omega'(G)/G \cup \{u\}$ and a homeomorphism $\Phi : U \to V$, $V \subset \mathbf{C}$ open so that $\Phi|U \setminus \{u\}$ is conformal in the conformal structure of $\Omega'(G)/G$ we have defined. Thus adding the chart (U, Φ) to A, elements of A satisfy the compatibility condition (2) and hence we have extended the conformal structure to $\Omega'(G)/G \cup \{u\}$. We add such a chart for every element of pE and this extends the conformal structure to $\Omega(G)/G$.

The actual construction was done when we showed that $\Omega(G)/G$ is a surface also if there are elliptic elements. We only show that we get in this manner a conformal atlas.

So, let $x \in E$. Let $\{x, y\}$ be the fixpoint set of elements of G_x. Find $\gamma \in M$ so that $\gamma(x) = 0$ and $\gamma(y) = \infty$. Thus γ conjugates G_x to a group Γ_0 generated by $h(z) = e^{2\pi i/n}$. We also know that there is a neighborhood U of x such that

$$gU \cap U = \emptyset \text{ if } g \in G \setminus G_x$$

and that $\gamma U = \{z \in \mathbf{C} : |z| < R\}$. Now, if $z \in U$,

$$\gamma(G_x z) = \gamma(Gz \cap U) = \Gamma_0 \gamma(z)$$

and we know that $f(z) = z^n$ maps orbits of Γ_0 onto the same point so that $f^{-1}\{f(w)\} = \Gamma_0 w$. Hence $f\gamma$ send orbits $G_x z$ onto the same point and

$$G_x w = (f\gamma)^{-1}\{f\gamma(w)\}.$$

It follows that

$$\Phi(p(z)) = f\gamma(z) \tag{3}$$

maps pU bijectively onto $V = \{|z| < R^n\}$.

We claim that $\Phi|pU \setminus \{p(x)\}$ is conformal. To see this, let $\zeta = p(z) \in pU$ where $z \in U$. We can use the fact that the canonical projection is conformal in $\Omega'(G)$ and hence so are the local inverses of p. By (3), $\Phi p = f\gamma$ is conformal in $U \setminus \{x\}$. If $q : W \to U$ is a local inverse of p in $W \subset U$, then $\Phi|W = \Phi pq = f\gamma q$ and hence is conformal.

Finally, the analyticity of the canonical projection p needs to be checked only at elliptic fixpoints since we already know that it is conformal at other points. This follows from (3) since $f\gamma$ is analytic. Recall that analyticity of p means that Φp is analytic; note that, given elliptic fixpoint x, there is only one chart mapping Φ so that $\Phi p(x)$ is defined. It is the one constructed above and is analytic by (2).

5. Fuchsian groups and the hyperbolic metric

A *Möbius disk* is the image of an ordinary euclidean disk, by a Möbius transformation. It is either a euclidean disk a half-space or the complement of the closure of a euclidean disk.

Definition. A Fuchsian group is a Kleinian group G such that G has a G-invariant Möbius disk..

If D is a Möbius disk such that $gD = D$ for $g \in G$, then we say that D is the invariant disk for G or that G is a (Fuchsian) group of D. Note that $\overline{\mathbf{C}} \setminus \overline{D}$ is also an invariant disk for G but except for this ambiguity the invariant disk is well-defined for non-elementary groups, as follows from Theorem 5.1. Usually the invariant disk is either

$$\Delta = \{z \in \mathbf{C} : |z| < 1\} = \text{ the unit disk, or}$$
$$U = \{z \in \mathbf{C} : \operatorname{Im} z > 0\} = \text{ the upper half-space.}$$

Some tasks are easier to do in Δ, others in U. One can change between them by a conjugating with a Möbius transformation f such that $gU = \Delta$, for example by

$$f(z) = \frac{z - i}{z + i}.$$

Let D be a Möbius disk. We denote the group of Möbius transformations leaving a Möbius disk D invariant by $M(D)$. Elements of $M(D)$ are also called Möbius transformations of D.

Theorem 5.1. *A subgroup G of $M(D)$ is a Fuchsian group of D as soon as it is discrete and in this case $L(G) \subset \partial D$.*

Proof. This is a special case of Theorem 3.16.

Thus $L(G)$ is a subset of ∂D and it is customary to say that G is of the *first kind* if $L(G) = \partial D$ and of the *second kind* if $L(G) \neq \partial D$.

An advantage of the upper half plane U is that elements $M(U)$ can be represented by elements of $GL(2,\mathbf{R})$, that is by 2×2 matrices with real entries and non-vanishing determinant. $SL(2,\mathbf{R})$ is the subgroup of $GL(2,\mathbf{R})$ consisting of matrices with determinant 1.

Lemma 5.2. If $A \in GL(2,\mathbf{R})$ and $\det A > 0$, then $f_A \in M(U)$. Conversely, every $g \in M(U)$ can be represented by a matrix of $SL(2,\mathbf{R})$.

Proof. Note that $\partial U = \mathbf{R} \cup \{\infty\}$ and hence $f_A(\partial U) = \partial U$ if A has real entries. Thus it either preserves or interchanges components of $\overline{\mathbf{C}} \setminus \partial U$ which are U and the lower half-space L. It preserves the components if $f_A(i) \in U$. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. We compute

$$f_A(i) = \frac{ai+b}{ci+d} = \frac{c^2+bd+i(ad-bc)}{c^2+d^2};$$

and the imaginary part of this is postive if and only if $ad - bc > 0$.

Conversely, choose distinct real numbers z_i, $i \leq 3$, and let $w_i = f(z_i)$. We can assume that z_i are so chosen that w_i are real. Now, $w = f(z)$ can be solved from

$$(w, w_1, w_2, w_3) = (z, z_1, z_2, z_3)$$

as

$$w = f(z) = \frac{az+b}{cz+d}$$

where coefficients are rational functions of z_i and w_i and hence real. We have seen that the condition $fU = U$ implies that $ad - bc > 0$. We can obtain that $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL(2,\mathbf{R})$ by dividing the entries with $\sqrt{ad-bc}$ which is a real number.

It follows by Corollary 3.9 that if H is a discrete subgroup of $SL(2,\mathbf{R})$, then the group $G = \{f_A : A \in H\}$ defined by the matrices of H is discrete and hence Fuchsian. In particular, the modular group consisting of Möbius transformations f_A such that $A \in SL(2,\mathbf{Z})$ which was found to be discrete in the end of Section 3, is a Fuchsian group. It is of the first kind since $L(M) = \mathbf{R} \cup \{\infty\}$ (an exercise in the 4th set.) The group generated by hyperbolic $g(z) = \lambda z$, $\lambda > 1$ would be an example of an elementary group of the second kind. In this case, $L(G) = \{0, 1\}$. One can show that if G is a non-elementary group of the second kind, then the limit set is homeomorphic to the Cantor set.

The mirror point. Mirror point are taken with respect to a Möbius circle, that is the Möbius image of a euclidean circle S. If $S = \mathbf{R} \cup \{\infty\}$, then the mirror point of z is $\bar{z}$ (and the mirror point of ∞ is ∞). Intuitively, we think that $\mathbf{R}$ is the mirror and we think that z is reflected on $\mathbf{R}$ to obtain the mirror point. We usually denote the mirror point by z^* though it depends on also on the Möbius circle S.

Given $z_1, z_2, z_3 \in \mathbf{R} \cup \{\infty\}$, the map $z \mapsto (z, z_1, z_2, z_3)$ is a Möbius transformation and hence a bijection of $\overline{\mathbf{C}}$. Since $\bar{z}_i = z_i$, we have

$$\overline{(z, z_1, z_2, z_3)} = (\bar{z}, z_1, z_2, z_3).$$

Thus we can characterize $z^* = \bar{z}$, the mirror point of z with respect to $\mathbf{R} \cup \{\infty\}$ as the unique point z^* such that

(0) $$(z^*, z_1, z_2, z_3) = \overline{(z, z_1, z_2, z_3)};$$

this is indepedent of the choice of the points $z_i \in \mathbf{R} \cup \{\infty\}$.

Equation (0) makes it possible to define the mirror point with respect to *any* Möbius circle S as the point z^* which satisfies (0), provided that $z_i \in S$ are distinct points. The independence of the definition of z^* from the points $z_i \in S$ was clear if $S = \mathbf{R} \cup \{\infty\}$ but for general Möbius circles this must be checked. This follow from the following lemma which gives also the rule for calculating z^*.

Lemma 5.2. *If $S = \{z : |z - a| = r]$, then the mirror point with respect to S is given by*

(1) $$z^* = a + r^2/\overline{(z - a)}$$

and if $S = L \cup \{\infty\}$, L a euclidean line, then the mirror point is the point z^* such that L intersects orthogonally the line segment J with endpoints z and z^* as the midpoint of J.

Note that if S is the unit circle $|z| = 1$, then we have the simple formula

(2) $$z^* = 1/\bar{z}$$

for the mirror point. In the general situation, if $z = a + te^{i\alpha}$, we have

(3) $$(a + te^{i\alpha})^* = a + r^2 e^{i\alpha}/t;$$

thus z and z^* are on the same ray starting from the center a of S but on different components of $\overline{\mathbf{C}} \setminus S$. Using (3) we see that

(4) $$z^* = z \text{ on } S$$

and that

(5) $$(z^*)^* = z.$$

These are valid also if $S = L \cup \{\infty\}$.

Proof. Let $f(z) = \bar{a} + r^2/(z - a)$ which is a Möbius transformation. Then (1) is equivalent to $z^* = \overline{f(z)}$. We need only to check that if $z_i \in S$ are distinct, then $\overline{f(z)}$ is the point z^* which satisfies (1). Let $z_i \in S$ be distinct and note that $\overline{f(z)} = z$ if $z \in S$.

$$\overline{(f(z), z_1, z_2, z_3)} = \{\overline{f(z)}, \overline{f(z_1)}, \overline{f(z_2)}, \overline{f(z_3)})0\} = \overline{(z, z_1, z_2, z_3)}.$$

We leave the case that $S = L \cup \{\infty\}$ as an exercise.

Lemma 5.3. *Let S be a Möbius circle. Then a Möbius transformation g preserves the mirror point:*

$$g(z^*) = g(z)^* \tag{2}$$

where on the left hand side the mirror point is taken with respect to S and on the right hand side with respect to gS.

Proof. Let $z_i \in S$, $i \leq 3$, be distinct. Then $w_i = g(z_i) \in gS$ are distinct. Now

$$(g(z^*), w_1, w_2, w_3) = (z^*, z_1, z_2, z_3) = \overline{(z, z_1.z_2, z_3)} = \overline{(g(z), w_1, w_2, w_3)}$$

and hence $g(z)^* = g(z^*)$.

The action of elements of a Fuchsian group. We can use these results to describe elements of a Fuchsian group.

In particular, (2) is valid if G is a Fuchsian group of D and z^* and $g(z)^*$ are mirror points with respect to ∂D, (2) is valid. Thus the actions of G in the two components of $\overline{\mathbf{C}} \setminus \partial D$ are mirror images of each other.

Theorem 5.4. *Let G be a Fuchsian group of the Möbius disk D. Let $g \in M(D) \setminus \{\mathrm{id}\}$. Then g is either hyperbolic, parabolic or elliptic. If g is hyperbolic or parabolic, the fixpoints of g are on ∂D. If g is elliptic, then one fixpoint of g is in D and the other outside $\overline{D}$ and they are mirror points of each other with respect to ∂D.*

Remark. Thus $M(D)$ does not contain loxodromic elements.

Proof. Let $g \in M(D)$, $g \neq \mathrm{id}$. If z is a fixpoint of g, then $g(z^*) = g(z)^* = z^*$ is also a fixpoint of D. Thus if $z \notin \partial D$, then z and z^* are the fixpoints of g. If g is not elliptic, it would follow that $g^n(z)$, $z \in \partial D$, would tend toward either z or z^* as $n \to \infty$. Thus if g has a fixpoint $z \notin \partial D$, then g is elliptic and z, z^* is the fixpoint pair of g.

So we need only t show that if g has fixpoint $z \in \partial D$, then g is either parabolic or hyperbolic. If g is parabolic, we are done. If this is not the case, then g has two fixpoints a and b on ∂D. We can conjugate the situation so that $D = U$, the upper half-space and $a = 0$ and $b = \infty$. Thus $g(z) = \lambda z$, $\lambda \neq 1$. Since $g(\partial D) = \partial D$, $g\mathbf{R} = \mathbf{R}$. Thus $\lambda \in \mathbf{R}$. If $\lambda < 0$, then g interchanges the upper and half-space which is impossible since g preserves D.

The hyperbolic metric. A very important notion for Fuchsian groups is the hyperbolic metric. It is a metric d on the invariant disk of the group which invariant for $g \in M(D)$. This means that

$$d(g(z), g(w)) = d(z, w) \tag{6}$$

if $g \in M(D)$ and $z, w \in D$. These metrics are usually constructed using a metric density ϱ which is a positive real function on D. If $\gamma : [a, b] \to d$ is path, then the ϱ-length of γ is

$$|\gamma|_\varrho = \int_\gamma \varrho |dz| = \int_a^b \varrho(\gamma(t)) |\gamma'(t)| dt.$$

The metric density is invariant for $g \in M(D)$ if

$$\varrho(g(z))|g'(z)| = \varrho(z) \tag{7}$$

for all $z \in D$. If ϱ is invariant, then ϱ-length of paths is unchanged under g

$$|g \circ \gamma|_\varrho = \int_a^b \varrho(g \circ \gamma(t))|(g \circ \gamma)'(t)|dt(8) = \int_a^b \varrho(g(\gamma(t))|g'(\gamma(t)||\gamma'(t)|dt = \int_a^b \varrho(\gamma(t))|\gamma'(t)dt = |\gamma|_\tau.$$

We see the reason for invariancy better if we write $w = g(z)$ and $dw = g'(z)dz$. If we multiply both sides of (7) by $|dz|$ we obtain

$$\varrho(w)|dw| = \varrho(z)|dz|; \tag{9}$$

here $\varrho(z)|dz|$ would be the infinitesimal ϱ-length of the infinitesimal euclidean length $|dz|$ and $\varrho(w)|dw|$ would be the corresponding ϱ-length of the induced infinitesimal dw.

Given a metric density ϱ, one defines a metric d_ϱ on D by

$$d_\varrho(x,y) = \inf_\gamma |\gamma|_\varrho$$

where the infimum is taken over all regular paths of D joining x and y. Under some reasonable condition, for instance ϱ is continuous, this is indeed a metric on D. If ϱ is invariant for $g \in M(D)$, also d_ϱ is invariant.

Thus the trick is to construct a Möbius-invariant metric is to find a Möbius invariant metric density. We use the Möbius invariance of the crossratio to derive such an invariant metric density and metric. Let $g \in M$ be a fixed Möbius transformation and let $z, \zeta \in \mathbf{C}$, $z \neq \zeta$, so that $w = g(z)$ and $\nu = g(z)$ are in $\mathbf{C}$. Further we consider two numbers h and k which are thought to be small and tend toward 0. Denote $\tilde{h} = g(z+h) - g(z)$ and $\tilde{k} = g(\zeta + h) - g(\zeta)$ so that $g(z+h) = w + \tilde{h}$ and $g(\zeta + k) = \nu + \tilde{k}$. Thus $\tilde{h}/h \to g'(z)$ and $\tilde{k}/k \to g'(\zeta)$ as $h \to 0$ and $k \to 0$. The crossratio

$$(z+h, \zeta, z, \zeta + k) = \frac{(z+h) - z}{(z+h) - (\zeta + k)} : \frac{\zeta - z}{\zeta - (\zeta + k)} = \frac{hk}{[(z-\zeta) + (h-k)](\zeta - z)}$$

is by the Möbius-invariance the same as

$$(w + \tilde{h}, \nu, w, \nu + \tilde{k}) = \frac{\tilde{h}\tilde{k}}{[(w-\nu) + (\tilde{h} - \tilde{k})](\nu - w)}$$

and hence

$$\frac{(\tilde{h}/h)(\tilde{k}/k)}{[(w-\nu) + (\tilde{h} - \tilde{k})](\nu - w)} = \frac{1}{[(z-\zeta) + (h-k)](\zeta - z)}.$$

If $h \to 0$ and $k \to 0$, then also $\tilde{h} \to 0$ and $\tilde{k} \to 0$, and $\tilde{h}/h \to g'(z)$ and $\tilde{k}/k \to h'(\zeta)$ and hence we have in the limit the following equality

$$\frac{g'(z)g'(\zeta)}{(w-\nu)^2} = \frac{g'(z)g'(\zeta)}{(g(z) - g(z))^2} = \frac{1}{(z-\zeta)^2} \tag{10}$$

and this is valid for all distinct $z, \zeta \in \mathbf{C}$ such that $g(z), g(\zeta) \in \mathbf{C}$.

Sometimes a more useful way to think of this equality is to multiply both sides by $dz\,d\zeta$ and write $dw = g'(z)dz$, $d\nu = g'(\zeta)d\zeta$ so that we obtain

$$\frac{dw\,d\nu}{(w-u)^2} = \frac{dz\,d\zeta}{(z-\zeta)^2} \tag{11}$$

this can be thought as a relation between the product of the infinitesimal chance dz of z and $d\zeta$ of ζ and of the corresponding infinitesimal changes of $w = g(z)$ and $\nu = g(\zeta)$.

Although formula (11) is an expression of Möbius-invariancy, it involves two variable, z and ζ. We have to get rid of one of them. We use the mirror point z^* and let $\zeta = z^*$. To be more precise, we assume that $z \in D_1$ and let $\zeta = z^*$ be the mirror point of z with respect to ∂D_1. We need a more precise notation and denote

$$z^* = \sigma_1(z)$$

for the mirror point with respect to ∂D_1. Let $D_2 = gD_1$. and thus $\nu = g(\zeta) = g(z^*) = w^*$ where now w^* is the mirror point with respect to ∂D_2. We let σ_2 be the mirror point with respect to σ_2 so that $w^* = \sigma_2(w)$. By Lemma 5.3, we have the following commutativity relation

$$g\sigma_1 = \sigma_2 g. \tag{12}$$

and so $w^* = g\sigma_1(z) = \sigma_2 g(z)$. Thus we can write (10) as

$$\frac{g'(z)g'(z^*)}{(w-w^*)^2} = \frac{1}{(z-z^*)^2}. \tag{13}$$

Before we continue the general case, we study the special case that $D_1 = D_2 = U =$ the upper half-plane. In this case $z^* = \bar{z}$ and $w^* = \bar{w}$. Since $g(\bar{z}) = \overline{g(z)}$, we obtain that

$$g'(\bar{z}) = \lim_{h\to 0} \frac{g(\bar{z}+\bar{h}) - g(\bar{z})}{\bar{h}} = \lim_{h\to 0} \frac{\overline{g(z+\bar{h}) - g(z)}}{\bar{h}} = \overline{g'(z)}.$$

Write $z = x + iy$ and $w = u + iv$ so that $z - z^* = 2y$ and $w - w^* = 2v$. We can now write (13) in the following form, after taking square roots of the moduli and multiplying by 2,

$$\frac{|g'(z)|}{v} = \frac{1}{y}. \tag{14}$$

In the infinitesimal form this is

$$\frac{|dw|}{v} = \frac{|dz|}{y}.$$

Equation (14) shows that the metric density $\varrho(z) = 1/\operatorname{Im} y = 1/y$ is invariant for $g \in M(U)$ and thus g does not change the ϱ-length of a path γ of U. The ϱ-length will be called the hyperbolic length of γ and ϱ is the hyperbolic metric density though sometimes ϱ is divided by 2 reflecting the fact that we multiplied by 2 to obtain (14).

The general case is a little more complicated. We need to calculate $g'(z^*)$. Here the problem is that the map $z \mapsto z^*$ is not analytic and so does not have complex derivative since it involves taking the complex conjugate and the map $z \mapsto \bar{z}$ is not analytic. However, the absolute value $|dz^*|/|dz|$ still exists, defined as the limit

$$\lim_{h \to 0} \frac{|(z+h)^* - z^*|}{|h|} \tag{15}$$

and this is all that is needed. If $z^* = \bar{z}$ as above, then obviously

$$\frac{|dz^*|}{|dz|} = 1. \tag{16}$$

In the general case, w^* and z^* could be mirror points with respect to different circles. Hence it might be better to use the notation $\sigma_2(w)$ and $\sigma_1(z)$ as above for the mirror points. Since $w^* = g(z)^* = g(z^*)$, we obtain by the appropriate form of the chain rule from $\sigma_2 g = g\sigma_1$ that

$$|\sigma_2'(w)||g'(z)| = |g'(z^*)||\sigma_1'(z)|$$

giving

$$|g'(z^*)| = |g'(z)||\sigma_2'(w)||\sigma_1'(z)|^{-1}. \tag{18}$$

Note that since σ_i are not analytic, we have to modify the chain rule so that we take absolute values. We leave it as an exercise to show that the chain rule is in the above form valid. Easiest, if unexact, way to see it, is to note that if ζ is a function of w and w a function of z, then $\frac{|d\zeta|}{dz|} = \frac{|d\zeta|}{|dw|}\frac{|dw|}{|dz|}$. Formally, easiest proof goes by noting that $\overline{\sigma_2(z)}$ and $\sigma_1(\bar{z})$ are analytic functions of z.

Now, we substitute (18) in (13) and take square root of the moduli and obtain

$$\frac{|g'(z)|^2|\sigma_2'(w)|}{|w-w^*|^2} = \frac{|\sigma_1'(z)|}{|z-z^*|^2} \tag{19}$$

Define

$$\varrho_{D_i}(z) = \frac{2\sqrt{|\sigma_i'(z)|}}{|z-z^*|}. \tag{20}$$

Here the constant 2 is added in order to get simplest possible form for $D_i = U$. We obtain from (19)

$$\varrho_{D_2}(g(z))|g'(z)| = \varrho_{D_1}(z). \tag{21}$$

In other words,

$$\varrho_{D_2}(w)|dw| = \varrho_{D_1}(z)|dz| \tag{22}$$

and we would obtain as in (8) that the ϱ_{D_2}-length of $g\gamma$ is the same as the ϱ_{D_1}-length of the path γ of D_1.

So we need to calculate

$$\varrho_D = \frac{2\sqrt{|dz^*|/|dz|}}{|z-z^*|} \tag{23}$$

where the mirror point is taken with respect to ∂D. We haved already seen that if $D = U$, then

$$\varrho_U(z) = \frac{1}{y}$$

if $z = x + iy$. If $D = \Delta =$ the unit disk, then $z^* = 1/\bar{z}$ and hence

$$\frac{|dz^*|}{|dz|} = |d(1/z)/dz| = \frac{1}{|z|^2}.$$

Hence, remembering that $|z| < 1$ in Δ,

$$\varrho_\Delta(z) = \frac{2}{|z - 1/\bar{z}||z|} = \frac{2}{1 - |z|^2}.$$

Conclusion, if $g \in M$ maps the Möbius disk D_1 onto D_2, then the hyperbolic metric density is preserved, i.e. (22) is true.

Construction of the metric from the metric density. So we now assume that we have a disk D on which we have a metric density ϱ_D given by (23). We will now define a metric using ϱ_D. This metric is the *hyperbolic metric* of D and it is defined so that the distance of two points $z, w \in D$ is

$$d(z, w) = \inf_\gamma |\gamma|_{\varrho_D} \tag{24}$$

where the infimum is taken over all regular paths γ such that z is the inital point w the endpoint and as above,

$$|\gamma|_{\varrho_D} = \int_\gamma \varrho_D |dz|$$

is the ϱ_D-length of γ.

We have to check that d is indeed a metric of D, that is we need to check that is it satisfies

(i) $d(z, z) \geq 0$ and $d(z, w) = 0$ only if $z = w$.
(ii) $d(z, w) = d(w, z)$
(iii) $d(z, w) \leq d(z, u) + d(u, z)$.

We check these points. Obviously, $d(z, w) \geq 0$ and $d(z, z) = 0$ since the ϱ_D-length of the constant path $\gamma(t) = z$ is 0. It would not be difficult to prove that $d(z, w) > 0$ but we postpone it and prove it in connection with geodesics.

To prove ii), we note that we can always parametrize paths so that the parameter interval is [0,1]. Then the reverse path of γ is $\sigma(t) = \gamma(1 - t)$ whose initial point is the endpoint of γ and vice verse. One easily checks (by performing the change of variable $t = 1 - s$ in the integral) that $|\sigma|_{\varrho_D} = |\gamma|_{\varrho_D}$. It follows that the of the ϱ_D-lengths in (24) is the same for $d(z, w)$ and for $d(w, z)$. Thus ii) is true.

To prove iii) we choose paths γ with the initial point z and endpoint u and another path with initial point u and endpoint z such that

$$|\gamma|_{\varrho_D} \leq d(z, u) + \varepsilon/2 \text{ and } |\beta|_{\varrho_D} \leq d(u, z) + \varepsilon/2.$$

Since the endpoint of $\gamma =$ the initial point of β, we can join the paths and form the combined paths $\gamma * \beta$ whose initial point is z and endpoint is w. Thus

$$d(z,w) \le \int_{\gamma*\beta} \varrho_D |dz| = \int_\gamma \varrho_D |dz| + \int_\beta \varrho_D |dz| \le d(z,u) + d(u,z) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the triangle inequality is true and d is a metric. (Actually, since $i)$ is so far only partially proved, we know only so far that d is a pseudo-metric).

An isometry between metric spaces X and Y is a homeomorphism g which preserves the metric:

(25)
$$d(g(z), g(w)) = d(z,w).$$

Theorem 5.5. *If D is a Möbius disk and $D' = gD$, $g \in M$, then g is a hyperbolic isometry $D \to D'$, that is g satisfies (24).*

Proof. Let $z, w \in D$ and let F be the family of regular paths with initial point z and endpoint w. Then $gF = \{g \circ \gamma : \gamma \in F\}$ is the family of regular paths with $g(z)$ as the initial point and $g(w)$ as the endpoint. Thus these families are in one-to-one correspondence. We have seen that $|\gamma|_{\varrho_D} = |g \circ \gamma|_{\varrho_{D'}}$ and hence the infimums (25) giving $d(z,w)$ and $d(g(z), g(w))$ are the same.

Notation. If the hyperbolic disk D is fixed, we denote

$$|\gamma|_h = |\gamma|_{\varrho_D}$$

for the hyperbolic length.

Geodesics. A hyperbolic geodesic of a Möbius disk D is first of all a subset L of D such that there is a homeomorphism $\sigma : \mathbf{R} \to L$ which is regular so that lengths are defined. Thus if $a = \sigma(t)$ and $b = \sigma(s)$, (we choose the notation so that $t < s$), then the subarc L_{ab} is parametriced by $\gamma_{ts} = \gamma|[t,s]$ and has hyperbolic length

$$|L_{ab}|_h = |\gamma_{ts}|_h.$$

L is a geodesic of the hyperbolic metric if

$$d(a,b) = |L_{ab}|_h.$$

Thus the distance $d(a,b)$ can be realized as the length of the arc L_{ab}. If c is "between" a and b, then L_{ac} is the union of the adjacent arcs L_{ac} and L_{cb} with c as the common point and hence

(26)
$$d(a,b) = |L_{ab}|_h = |L_{ac}|_h + |L_{cb}|_h = d(a,c) + d(c,b).$$

Any two points of a the hyperbolic space D can be joined by a geodesic. We start with

Lemma 5.6. *The positive imaginary axis $I = \{it : t > 0\}$ is a geodesic isometric to $\mathbf{R}$ and*

$$d(it, is) = \left|\log \frac{t}{s}\right|.$$

If γ is a path joining two points $a, b \in I$ contains a point outside γ, then

$$|\gamma_h| > d(a,b).$$

Proof. Let $a = it$ and $b = is$. We can assume that $t < s$. Let $\gamma : [0,1] \to U$ be a regular path such that $\gamma(0) = a$ and $\gamma(1) = b$. Write

$$\gamma(t) = \alpha(t) + i\beta(t)$$

where α and β are real functions. Note that the path $i\beta$, $t \mapsto i\beta(t)$, also joins a and b. Thus $\gamma(t) = \sqrt{\alpha'(t)^2 + \beta'(t)^2}$ and $\varrho_U(\gamma(t)) = \beta(t)$ and hence

$$|\gamma|_h = \int_0^1 \frac{\sqrt{\alpha\tilde{A}^{1\prime}(t)^2 + \beta'(t)^2}}{\beta(t)} dt \geq \int_0^1 \frac{\beta'(t)}{\beta(t)} dt(*) = \log\beta(1) - \log\beta(0) = \log\frac{s}{t}.$$

If γ contains a point outside I, then α is not constant and hence $\alpha'(t) \neq 0$ for some t. Hence $|\gamma'(t)| > \beta'(t)$ and we see that there is a proper inequality in (*) since paths are piecewise continuously differentiable. Thus I is a geodesic and hence $\log(s/t) = d(is, it)$. Thus the map $I \to \mathbf{R}$, $z \mapsto d(z,i)$ is a bijection and an isometry of I onto $\mathbf{R}$.

Theorem 5.7. *Let D be a Möbius disk which is provided with the hyperbolid metric. Then the geodesics of D are of the form $L = S \cap L$ where S is a Möbius circle intersecting orthogonally ∂D and each such L is isometric to $\mathbf{R}$. Any two distinct points of D are contained in a unique geodesic.*

Proof. If $D' = gD$, $g \in M$, then g is an isometry of the hyperbolic metrics of D and D'. Also, if S is a Möbius circle S is orthogonal to ∂D if and only if gS is orthogonal to ∂gS. Thus it is enough to prove the theorem in the case that $D = U$, the upper half-plane.

We already know that the positive imaginary axis I is a geodesic isometric to $\mathbf{R}$. Hence so is gI for any $g \in M$. Since I is orthogonal to ∂U (at ∞ this needs to be checked using the auxiliary map $1/z$) also gI is orthogonal to ∂U.

So we need to check the following.

1. If S is a Möbius circle orthogonal to ∂U, then $S \cap U = gI$ for some $g \in M(U)$.

2. If $z, w \in U$ are distinct, then there is a one and only one Möbius circle S orthogonal to ∂U such that $z, w \in U$.

3. If $L \subset U$ is a hyperbolic geodesic, then there is a unique Möbius circle S orthogonal to ∂U such that $L \subset U$.

To prove 1, let $\{z_1, z_2\} = S \cap \partial U$. Pick another point $z_3 \in \partial U \setminus \{z_1, z_2\}$. Thus there is a Möbius transformation g such that $g(0) = z_1$, $g(\infty) = z_2$ and $g(1) = z_3$. Thus either $gU = U$ or $gU = L$, the lower halfplane. In the latter case we replace g by $1/g(1/z)$ after which $gU = U$ and g still sends $\{0, \infty\}$ onto $\{z_1, z_2\}$. Thus gI is orthogonal to S and hence $gI = S \cap U$.

Proof of 2. Let $z = x + iy$ and $w = u + iv$. If $x = u$, then z and w are on $L = \{z \in U : \operatorname{Re} z = x = u\}$ which is a Möbius circle intersecting ∂U orthogonally at ∞ and $x = u$ and there is no other such Möbius circle.

If $x \neq u$, then suppose that there is a circle S orthogonal to ∂U and containing z and w. Then the center a of S is a real number and

$$(x-a)^2 + y^2 = (u-a)^2 + v^2.$$

This is equivalent to

$$x^2 - 2xa + y^2 = u^2 - 2ua + v^2$$

and hence

$$a = \frac{x^2 + y^2 - u^2 - v^2}{2(x-u)}.$$

Thus there is just one circle orthogonal to ∂U containing z and w and 2. is proved.

Finally, suppose that L is a hyperbolic geodesic. If $a, b \in L$, let L_{ab} be the closed subarc of L with endpoints a and b and let S_{ab} be the Möbius circle containing a and b and orthogonal to ∂U. It follows from Lemma 5.6 that $L_{ab} \subset S_{ab}$. If c, d are another pair of points of L such that $L_{ab} \subset L_{cd}$, we obviously $S_{cd} = S_{ab}$. Hence we can denote $S = S_{ab}$ independently of $a, b \in L$ and it follows that $L \subset S$. Thus also 3. is true and the theorem is proved.

Hyperbolic metric is a metric and not a pseudo-metric. We can now conclude then proof that $d(z, w) > 0$ if $z \neq w$. If $z, w \in I$, this follows from the formula in the lemma and otherwise we can map z, w to I by a Möbius transformation.

Usage of the word geodesic. Our original definition allows that a geodesic need not be maximal. We have seen that such geodesics are contained circles orthogonal to ∂D. Henceforth we mean by geodesics of D sets of the form $S \cap D$ where S is a Möbius circle orthogonal to ∂D. These are maximal geodesics and each of them is isometric in the hyperbolic metric to $\mathbf{R}$.

If D is the upper halfplane, then the hyperbolic lines are parts of vertical lines or of circles with centers on $\mathbf{R}$. If $D = \Delta$, the unit disk, then the hyperbolic lines through 0 are just parts of the euclidean lines through origin and thus they are euclidean open line arcs with endpoints a and $-a$ where $|a| = 1$. It is easy to see from this situation that L is a hyperbolic line and $z \in L$, then there is another hyperbolic line K through z intersecting L orthogonally.

Hyperbolic geometry and trigonometry. Hyperbolic metric of a Möbius disk D gives hyperbolic geometry to D. Its lines are the hyperbolic geodesics which are of the form $S \cap D$ where S and ∂D intersect orthogonally. It satisfies all the axioms of euclidean geometry except the parallel axiom: If L is a hyperbolic line and z is outside L, then there are many (actually infinite) number of hyperbolic lines through z which do not intersect L.

A hyperbolic triangle T consists of 3 hyperbolic line segments a b and c which meet at 3 vertices. It is possible to have formulae similar to euclidean trigonometry for hyperbolic triangles. Let a, b and c be the lengths of the sides of T and let α, β and γ be the angles at vertices of the triangle so that α is opposite to the side whose lengths was denoted by a etc. The hyperbolic cosine rule is

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma$$

where cosh and sinh are the hyperbolic sine and cose. In the limit, when the lengths of the sides approach 0, this gives the euclidean formula

$$c^2 = a^2 + b^2 - 2ab \cos \gamma.$$

We have also the sine rule in the form

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

again this gives in the limit the euclidean sine formula.

A striking difference from the euclidean geometry is the absence of similarity. We cannot multiply the sides of a triangle by a fixed constant and have a new triangle with these multiplied side lengths and same angles. This can be seen for instance from the sine formula.

One can use the hyperbolic metric to calculate areas. We do not go into details but one can show that the area of a triangle T is

$$\pi - (\alpha + \beta + \gamma)$$

when the angles of the triangle are α, β and γ. Thus the sum of the angles of a triangle is always less than π but on the limit, when the area approaches 0, the sum of the angles approaches π.

One can also consider ideal hyperbolic triangles so that one or more of the vertices of the triangle are on the boundary circle. The angle at such a vertex on the boundary is 0. If all the angles are on the boundary, then all the angles are 0 in this case the triangle has the maximal area π.

The fundamental domain. An important application of the hyperbolic metric and its Möbius-invariancy is that by means of it we can find a fairly regular domain D so that D and its transforms gD, $g \in G$, cover D so that distinct gD and hD $(g, h \in G)$ overlap only on the boundary. This gives insight into the geometry of the quotient surface $D/G = \{Gz : z \in D\}$ for a Fuchsian group G of D.

Before we start to construct such a fundamental domain, we need an auxiliary result. We use the following notation. We consider a Möbius disk D and the hyperbolic metric on it. We know that two distinct points $z, w \in D$ are on a well-defined hyperbolic line L_{zw} of D. Thus there is a well-defined subarc of L_{zw} with endpoints z and w; we denote it by J_{zw}. The midpoint of J_{zw} is the well-defined point a such that $d(a, z) = d(z, w) = d(z, w)/2$. There is a hyperbolic line through the midpoint z intersecting L_{zw} orthogonally. We call it the (hyperbolic) orthogonal bisector of J_{zw} and denote B_{zw}

Lemma 5.8. *Let $z, w \in D$, then*

$$(*) \qquad B_{zw} = \{\zeta \in D : d(\zeta, z) = d(\zeta, w)\}$$

is the orthogonal bisector of the hyperbolic line segment J_{zw}*.*

Proof. Let K be the orthogonal bisector of J_{zw}. Thus J_{zw} and K intersect orthogonally at $a \in J_{zw}$. We can apply a Möbius transformations which maps D to U and K to the positive imaginary axis I (we refer to the reasoning in the preceding theorem). The maps $\sigma(x+iy) = -x+iy$, the reflection of U on I, obviously a hyperbolic isometry since $|\sigma'(z)| = 1$ at all points and since it preserves the imaginary part of a complex number:

$$\varrho_U(\sigma(x+iy))|\sigma'(x+iy)| = \frac{1}{y} = \varrho_U(x+iy).$$

Thus $|\gamma|_h = |\sigma \circ \gamma|h$ and it follows that σ preserves the hyperbolic metric. The hyperbolic line L_{zw} containing z and w is orthogonal to K and hence $\sigma(L_{zw}) = L_{zw}$ and it follows that $\sigma(J_{za}) = J_{wa}$ and so $\sigma(z) = w$ and $\sigma(w) = \sigma(z)$. Obviously $\sigma K = K$.

Thus if $\zeta \in K$, then

$$d(z,\zeta) = d(\sigma(z), \sigma(\zeta)) = d(w,\zeta)$$

and so we need to prove only that if $\zeta \in U \setminus K$, then $d(\zeta, z) \neq d(\zeta, w)$. Suppose that this is not the case. Then one of the hyperbolic line segments $J_{z\zeta}$ and $J_{w\zeta}$ intersects K, say $J_{w\zeta} \cap K = \{b\}$. Thus, as we have seen,

$$d(w,b) = d(z,b)$$

and hence

$$d(z,\zeta) = d(w,\zeta) = d(w,b) + d(b,\zeta) = d(z,b) + d(b,\zeta).$$

Thus in the triangle with vertices z, b and ζ, then length of one side is the same as the sum of the lenghts of two other sides. This contradicts Lemma 5.7 which says that if a path γ connects two points z and b of a hyperbolic line L_0, then $|\gamma|_h > d(z,b)$ if γ goes outside L_0.

The hyperbolic line B_{zw} divides the Möbius disk D into two components, one of them contains z and the other w. We denote the one containing z by H_{zw} (here the order is important H_{wz} is the other component. The function $f(\zeta) = d(\zeta, w) - d(\zeta, z)$ vanishes on B_{zw} but is non-zero elsewhere. Obviously $f(z) > 0$ and hence we can conclude that

$$H_{zw} = \{\zeta \in D : d(\zeta, z) < d(\zeta, w), (*)\},$$

that is, H_{zw} contains $\zeta \in D$ such that ζ is closer to z than w.

We obtain the fundamental domain F of a Fuchsian group of D so that we fix a point $s \in D$ such that $g(a) \neq z$ for $g \in G \setminus \{\mathrm{id}\}$. Since the set of elliptic fixpoints is discrete (Lemma 4.3), we can find such points. Let

$$H_g = \overline{H_{ag(a)}} = \{z \in D : d(z,a) \leq d(z, g(a))$$

where $g \in G \setminus \{\mathrm{id}\}$. The fundamental domain of G with center a is

$$F_a = \bigcap_{g \in G'} H_g = \{z \in D : d(z,a) \leq d(z, g(a)) \text{ for } g \in G\} = \{z \in D : d(z,a) \leq d(g(z), a) \text{ for } g \in G\}$$

where $G' = G \setminus \{\mathrm{id}\}$. In the last equality we have used the fact that the hyperbolic metric is Möbius-invariant and that $G = \{g^{-1} : g \in G\}$. The last expression shows that we we pick to F_a the point (or points) from each orbit Gz which is closest to a. This characterzation implies the first property of F_a:

1°. $gF_a = F_{g(a)}$ for $g \in G$.

We now keep a fixed and denote $F = F_a$ for short. We examine its properties.

A set $A \subset D$ is hyperbolically convex if, whenever $z, w \in A$, the hyperbolic line segment J_{zw} with endpoints z and w is in A. The next property of F_a is:

2°. *F is hyperbolically convex.*

To see 2°, we note that each closed half-plane is hyperbolically convex and hence their intersection has the same property.

3°. *If $z \in D$, then there are $g_1, \ldots, g_n$, $n \geq 1$, such that $z \in g_i F$ and such that*

$$d(z, g_1(a)) = \ldots = d(z, g_n(a))$$

and z has a neighborhood U such that

$$d(z, g_i(a)) < d(z, g(a))$$

for $g \in G \setminus \{g_1, \ldots, g_n\}$.

To see 3°, we note that since Ga is a discrete subset of D, then

$$m = \inf_{g \in G} d(z, g(a))$$

is attained by some $g \in G$. Since $d(z, w) \to \infty$ as w tends toward the boundary (see for instance Ex. 4 in the last exercise set), there can be only a finite number of $g \in G$ for which $m = d(z, g(a))$. So there are such g_i as claimed. The same reasoning shows that there is $M > m$ such that $d(z, g(a)) \geq M$ if $g \in G \setminus \{g_1, \ldots, g_n\}$ and hence there is such a neighborhood U of a as claimed.

4°. *If $g, h \in G$, then $gF \cap hF_a$ is either empty, a point, or hyperbolic line, ray or segment.*

This follows since $gF \cap hF$ is a convex closed subset of the bisector $B_{g(a)h(a)}$ which is a hyperbolic line. It follows that only the possibilities mentioned can occur.

A *side* of gF is a set of form $gF \cap hF$ where $h \in G \setminus \{g\}$ and h contains more than one point. In the following ∂A denotes *boundary in D*.

5°. *∂gF is the union of sides of gF. If $s = gF \cap hF$, $h \in G \setminus \{g\}$ is a side of gF, and if $z \in s$ is not endpoint of s, then*

$$d(z, g(a)) = d(z, h(a)) < d(z, f(a)) \tag{$**$}$$

if $f \in G \setminus \{h, g\}$. Hence s is a side only of gF and hF.

Let $z \in F$. Let $g_1, \dots, g_n$ and the neighborhood U of z be as in 3° where $g = g_1$. If $n = 1$, then $U \subset gF$ and hence $z \in \mathrm{int}F$. Suppose that $n = 2$. Then we see that $d(z, g(a)) = d(z, h(a))$, $h \in G \setminus \{g\}$. If $d(z, g(a) < d(z, f(a))$ for $f \in G \setminus \{g, h\}$, then this is true in a neighborhood U of z and $U \cap B_{g_1(a)g_2(a)} \subset gF \cap hF$. So $gF \cap hF$ is a side of gF. Obviously, s is a side only of gF and hF.

If $n > 2$, then all the bisectors $B_k = B_{g(a)g_k(a)}$ have the common point z. It is geometrically evident that the bisectors are distinct (we bypass the verification of this simple fact). It is again geometrically evident that there are two bisectors B_p and B_q such that $B_q \cap U \subset gF$ and $B_p \cap U \subset gF$. Hence $gF \cap g_pF$ and $gF \cap g_qF$ are sides of gF whose endpoint z is. We can see from this that except for endpoints of s, (**) is valid.

6°. *The sets gF. $g \in G$, are distinct and cover D and if $z \in \mathrm{int}\, gD$ for at most one g.*

This follows since $g(a)$, $g \in G$, are distinct and obviously $g(a)$ is an interior point of gD. By 5°, $z \in \mathrm{int}\, gD$ is not a point of hD, $h \neq g$.

7°. *If s is a side of F, then there is another side s' and $g \in G \setminus \{\mathrm{id}\}$ such that $g(s) = s'$. In addition, $s \neq s'$ except if g is elliptic of order two such that the fixpoint of g is on s.*

To prove 6°, let $s = F \cap gF$. Thus $s' = g^{-1} \cap gF$ is also a side of F and $s = gs'$. If $s = s'$, then $g^2 = \mathrm{id}$ since otherwise s would be a side in 3 sets of the form hF, $h \in G$. Thus $g^2 = \mathrm{id}$ and hence g is elliptic of order two and one easily sees that the fixpoint is on s.

If g is as in 7°, we say that g identifies s with s' so that the point $z \in s$ is identified with $g(z)$. We can form the quotient $D/G = \{Gz : z \in S\}$ as F/R where the relation xRy means that there is $g \in G$ such $x = g(y)$. The equivalence classes are $Gz \cap F$ but now the relation is much easier to handle: If $\tilde{x}$ is the equivalence class of x, then $\tilde{x} = \{x\}$ for interior points of F and $\tilde{x}$ contains two points if $x \in s$ for a side of F unless x is an endpoint or elliptic fixpoint in which case $\tilde{x}$ is still finite.

An example. For simplicity we present the following example of a fundamental domain. It is not essential that G is Fuchsian or that the fundamental domain is obtained by means of the hyperbolic metric as described above. What is essential is kind of polygonal structure so that sides are identified in pairs. If G is the group of translations $T_{n+im}(z) = z + n + mi$, $n, m \in \mathbf{Z}$, then a fundamental domain would be

$$D = I \times I$$

where $I = [0, 1]$. D has sides $s_1 = \{0\} \times I$, $s_2 = \{1\} \times I$, $s_3 = I \times \{0\}$ and $s_4 = I \times \{1\}$. To obtain the quotient, s_1 would be identified with s_2 and s_3 with s_4. The first identification gives a cylinder and the second identification makes a torus out of the cylinder.

Fuchsian group as the universal cover group. We mention here some connections of Fuchsian groups to the surface topology and Riemann surface. We refer to some topological facts. If you do not know these just skip this section.

Suppose that G is a Fuchsian group of D. Since D is G-invariant, we can define the quotient space

$$D/G = \{Gx : x \in D\}.$$

As $\Omega(G)/G$, also D/G is a surface in the quotient topology. Let $p(x) = Gx$ be the canonical projection $D \to D/G$. If G does not contain elliptic elements, then p is a covering projection. We have now the extra information that D is simply connected which may not be true in the general case. Thus in this case G is isomorphic to the fundamental group of the quotient surface, usually denoted $\pi(D/G)$.

Suppose that S is a compact orientable surface. Then it is known that S is homeomorphic to a quotient D/G, G Fuchsian group of D, except if S the 2-sphere S^2 or torus $S^1 \times S^1$. One can classify orientable compact surfaces by a certain topological feature called the genus of the surface. Genus of S^2 is 0 and that of torus is 1. If the genus > 2, then S is homeomorphic to the quotient of a Fuchsian group.

In the theory of Riemann surfaces, one is interested of the conformal type of the surface. Two surfaces may be homeomorphic but do not have the same conformal type. i.e there is not a conformal homeomorphism between them. It is known that if the genus of a Riemann surface S is at least 2, then S is conformally equivalent to some quotient D/G; recall that D/G had a natural conformal structure.

These facts are the reason why Fuchsian groups are important in the theory of surfaces and Riemann surfaces. It is often simpler to study the group whose quotient is the surface than the surface itself.

References

[1] L. V. Ahlfors and G. Weill, *A uniqueness theorem for Beltrami equations*, Proc. Amer. Math. Soc. **13** (1962), 975–978.