Introductory Lectures on $SL(2, Z)$ and modular forms.

W.J. Harvey, King's College London

December 2008 .

1 Chapter 1.

(1.1) We begin with a definition. The *modular group* is the subgroup $SL(2,\mathbb{Z})$ of the matrix group $SL(2,\mathbb{R})$ consisting of matrices with integer entries and determinant 1.

There is an important action of $SL(2,\mathbb{R})$ on the upper half plane $\mathcal{U} = \{z = x + iy \mid y > 0\}$, as *fractional linear* (Mobius) transformations:

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}. \tag{1}$$

It is readily verified that the kernel of this action (the subgroup which acts trivially, fixing all points) is the centre $C_2 \cong \langle \pm I_2 \rangle$. The quotient group $PSL(2,\mathbb{R})$ acts faithfully on $\mathcal{U}$ and also on the boundary $\widehat{\mathbb{R}} = \mathbb{R} \cup \infty \cong S^1$; it is shown in elementary accounts of complex analysis or hyperbolic geometry (see for example [**?**], [**?**]) that the group action on $\widehat{\mathbb{R}}$ is triply transitive. In other words, any ordered triple of points of $\widehat{\mathbb{R}}$ can be mapped to the triple $0.1, \infty$

(1.2) The *hyperbolic metric* ds_h^2 is given by the formula

$$ds_h^2 = y^{-2}(dx^2 + dy^2) \tag{2}$$

on $\mathcal{U}$: this is a Riemannian metric on the upper half plane. Because the action defined in citeeq1 is *transitive* on the points of $\mathcal{U}$ and also on the set of unit tangent directions at each point (two simple exercises for the reader), this space is a symmetric space in the sense of differential geometry.

Invariance of the metric under the (differential of the) action by $SL(2,\mathbb{R})$ is verified by a simple calculation (Exercise 1). Thus the real Mobius transformations form the (direct) *isometry group* of $\mathcal{U}$: it is not hard to show that the action is transitive and so no larger direct isometry group is possible .

Note that the space $\mathcal{U}$ is in fact Poincaré's model of *hyperbolic plane geometry*. Geodesics between any two points are defined by circular arcs (or line

segments) orthogonal to the boundary real line. From this, we are able to use geometric ideas such as polygonal shape, convexity, length and area to illuminate the group activity on $\mathcal{U}$.

(1.3) The upper half plane $\mathcal{U}$ as a parameter space: lattices in the plane and their classification by 'marked shapes'.

The upper half plane serves as parameter space for a range of interesting objects. Gauss used it to classify positive definite binary quadratic forms, and an epic list of classical authors from Jacobi on established the link with complex tori and elliptic functions. We look at this second aspect here.

Complex structures on a torus. The first item we consider is the shapes of complex tori, surfaces of genus 1 with a complex analytic structure.

The classic Weierstrass theory of elliptic (i.e. doubly periodic meromorphic) functions, which will be summarised later on, depends on this standard model for a torus coming from a choice of generating set for the lattice of periods, isomorphic to the (free abelian rank 2) fundamental group of the torus,

$$\pi_1(X) = \mathbb{Z} + \mathbb{Z}.$$

Underlying this, there is the important concept of a *homotopy-marking* for the surface, which underpins the theory of deformations, the rigorous study of varying shapes of torus. Intuitively, this boils down to considering the effect of changing the shape of a fundamental parallelogram tile for this lattice group of plane translations; a stricter method delivers a very precise description, a genuine space of shapes, the first *space of moduli*, a ground-breaking step in algebraic geometry, the precursor of a widespread pattern of description for types of algebraic variety of specified type, a tool with great influence in mathematics generally.

From the *uniformisation theorem* or, alternatively, the Riemann-Roch theorem, one sees directly that a marked complex structure on the torus is tantamount to this choice of two nonzero complex numbers $\{\lambda_1,\ \lambda_2\}$ which are linearly independent over the reals, representing the monodromy of a chosen non-trivial holomorphic 1-form around the two generating loops; this determines a lattice subgroup $\Lambda = \langle m\lambda_1 + n\lambda_2 \mid m, n \in \mathbb{Z}$ of the additive group $\mathbb{C}$ such that the complex torus is isomorphic to the quotient space $\mathbb{C}/\Lambda$. A more topological way to specify a marking begins from a choice of base point and then two simple based loops whose homotopy classes generate $\pi_1(X, x_0)$. This determines (either by lifting paths to the universal covering plane or by integration) two Euclidean line segments, joining 0 to the complex numbers λ_1 and λ_2 respectively, which may be regarded as a *geometric marking* of the torus. The standard picture of the torus $X = X(\lambda_1, \lambda_2)$ is then obtained by identification of opposite sides of the parallelogram with corners at the points $0, \lambda_1, \lambda_2$ and $\lambda_1 + \lambda_2$.

When are two marked complex torus structures equivalent*?* This means that

the tori are to be conformally homeomorphic, by a mapping which is produced by a conformal (hence complex linear) map in the covering plane between the two given markings. This happens if and only if there is a nonzero complex scaling factor and perhaps a switch of ordering of the numbers, after which we may assume that the marking pair is given by $\lambda_1 = 1,\ \lambda_2 = \tau$ with Im $\tau > 0$. Thus, $\tau \in \mathcal{U}$.

When do two such pairs determine the same quotient Riemann surface? The corresponding normalised lattices $\Lambda_\tau = \langle m + n\tau \mid m, n \in \mathbb{Z}$ and $\Lambda_{\tau'}$ have to coincide , which means that there are integers $a,\ b,\ c,\ d$ such that

$$1 = c\tau + d \tag{3}$$
$$\tau' = a\tau + b, \tag{4}$$

and satisfying $ad - bc = 1$, so that the process can be done in reverse. Thus the two notions of equivalence taken together produce the action given in equation (1) above, of the modular group on the space T_1 of marked complex tori, given by the action of the homogeneous modular group $\Gamma(1) = PSL_2(\mathbb{Z})$ by fractional linear automorphisms:

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}$$

of the hyperbolic plane $\mathcal{U} = \{Im(\tau) > 0\}$.

In this way, we encounter the classic prototype for a discrete group action, as first considered by Klein and by Poincaré, the modular group $\Gamma(1) \cong PSL(2, \mathbb{Z})$ operating on the upper half plane.

4) **Classification of real Möbius group elements** into types or conjugacy classes, and the corresponding mappings of $\mathcal{U}$.

For this, you can either use the classification of real (invertible) matrices by eigenvalues or go by the (closely related) fixed point properties, which motivate the present geometrical approach. Recall first that any real or complex Möbius transformation distinct from the identity has either one or two fixed points in the Riemann sphere $\mathbb{CP}^1$; the real coefficients force restrictions in the geometric types which occur.

DEFINITION. A real Möbius map/isometry of $\mathcal{U}$ is called *elliptic* if it has one fixed point inside $\mathcal{U}$. It is *parabolic* if it has one boundary fixed point, *hyperbolic* if it fixes two boundary points.

Note that elliptic elements have two complex conjugate fixed points , one in each of the upper and lower half planes.

Typical examples of parabolic transformation are real translations $T(z) = z + b$, with $b \neq 0$: each preserves as a set every horizontal line. This is the family of *horocycles* at ∞ . A conjugate map has the same property with respect to the family of circles tangent to the boundary circle at the fixed point.

Hyperbolic transformations are conjugate to real dilations $V_\lambda(z) = \lambda z$, with $\lambda > 1$: Each one fixes a pair of points, and preserves the hyperbolic geodesic joining them, acting on this as a hyperbolic translation from the repelling fixed point (in the examples it is 0) towards the attracting fixed point (at ∞ for V_λ).
Note. The *fixed points* of a Móbius map corresponding to a matrix $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are given by the equation

$$cz^2 + (d-a)z - b = 0.$$

Thus, the classification turns on the value of the discriminant

$$(d-a)^2 + 4bc = (\text{Trace}^2 - 4\text{det})T :$$

if $T(z) \in PSL(2,\mathbb{R})$, then we have

- T is elliptic if and only if $\text{Trace}^2 T < 4$.
- T is parabolic if and only if $\text{Trace}^2 T = 4$.
- T is hyperbolic if and only if $\text{Trace}^2 T > 4$.

Note. An interesting phenomenon associated with a hyperbolic transformation in $SL(2,\mathbb{Z})$ is the so-called *Arno'ld's cat mapping* given by the matrix

$$T = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

This is a hyperbolic element with fixed points at $(1 \pm \sqrt{5})/2$. It generates a semigroup of positive iterates, $T \circ T = T^2, T \circ T^2 = T^3, \ldots$, which distort the unit square (thought of as a tile within the real plane) by stretching in one eigendirection and shrinking in the other, to give a sequence of parallelograms, each of which has area 1 and projects onto the quotient torus, where the result is that one obtains a dissection of the original tile first of all into three triangular pieces; this is extended to a map of the plane $\widetilde{T}$ by periodicity, and this same mapping repeated, generating a more and more stretched out and fragmented pattern as n grows. However, when one takes on the original tile a photographic image, which is produced by an array of $m \times m$ black or white pixellated dots, the effect of this map is really just a permutation of the dots, and some large enough power of it gives the identity mapping on this finite subset although the map itself is greatly different from the identity, so that the photograph reappears, a highly paradoxical effect when viewed among the surrounding chaotic patterns. It is discussed in (for instance) *Anton & Rorres*'s Elementary Linear Algebra with applications (John Wiley & Sons, 2000). We note that this concerns the linear action of the transformation T in the plane projected to the torus and has nothing to say directly about the nature of the corresponding hyperbolic isometry.

A final pair of exercises to think about:

1. Find all compact subgroups of the Lie group $SL(2,R)$.

[The stabiliser of any point is a compact subgroup, conjugate to the subgroup $PSO(2,\mathbb{R})$ stabilising i. If we let K be any compact subgroup, the set of images of i under all $T \in K$ is a compact subset C of $\mathcal{U}$ (why is this?). Conversely, given any compact set C in the upper half plane, the set of elements $\gamma \in G$ with $\gamma(i) \in C$ is compact. But is it a subgroup? (No!) So how can we pin down the compact subgroups?]

2. Concerning the action of $SL(2,R)$ on $\mathcal{U}$, show that it is *proper*: that is, prove that if K is any compact subset of $\mathcal{U}$, then the set of all g in $SL(2,R)$ such that gK intersects K nontrivially is compact.

[This s the fundamental property which distinguishes actions of a lie group on a space with compact stabilisers.]

2

Week 2: Discontinuity and quotient surfaces.

1) **Fundamental sets for discrete groups.**

DEFINITION. Let Γ be a discrete group acting by isometries on a metric space X. A *fundamental set* for Γ is defined to be a closed set F with two key properties.

(i) the interior of F, F^0, has empty intersection with each translate gF^0, $g \in \Gamma \setminus Id$.

(ii) the union of the $g-$translates of F covers the space.

For the modular group, there is a popular and convenient choice of fundamental set which we construct below.

Note. A group which acts properly and isometrically on a metric space with discrete orbits can be given a fundamental set with a geometric flavour, called a *Dirichlet fundamental set.* Roughly speaking, it is the set of points closer to a designated base point z_0 than to any other point of the $\Gamma-$orbit of z_0. See the exercises for Week 2.

In this special case, we use a more direct approach, following [**?**] . First of all we mention three special elements of Γ: they are

$$T(z) = z + 1; \quad U(z) = \frac{-1}{z}; \quad S = T \circ U,\ S(z) = \frac{z-1}{z}.$$

Next we concentrate attention on a certain *hyperbolic ideal triangle*

$$\mathcal{D} = \{|z| > 1\} \cap \{-\frac{1}{2} \leq \Re(z) \leq \frac{1}{2}\} \cap \mathcal{U}$$

in the upper half plane, with vertices at the points $\rho = e^{i\pi/3}$, $\rho^2 (= \rho - 1)$ and a third vertex at ∞, often called an *ideal vertex*. Edges joining these points are hyperbolic geodesic line segments: the edges to ∞ are vertical half-lines.

It is easily seen that U and S are both elliptic *torsion elements*, of orders 2 and 3 respectively, and their fixed points are i and ρ.

These maps determine side-pairing transformations in $\Gamma(1)$, precise conformal mappings of the triangle $\mathcal{D}$ onto some neighbouring triangle which shares an edge with $\mathcal{D}$, thereby enjoying properties crucial to understanding the whole action of $\Gamma(1)$ on $\mathcal{U}$. In particular, we can express the various transforms of $\mathcal{D}$ by elements of $\Gamma(1)$ in terms of words in these two letters, as we shall prove below.

2) **A fundamental set for $SL(2,\mathbb{Z})$:** Theorem 1 and two corollaries. In fact we will prove that the subgroup Γ_0 of Γ generated by T and U, which of course contains $S = T \circ U$, has $\mathcal{T}$ as fundamental domain.

THEOREM 1. *(a) For each $z \in \mathcal{U}$, there exists $\gamma \in \Gamma_0$ with $\gamma(z) \in \mathcal{D}$.*
(b) If $z, z' \in \mathcal{D}$ with $z' = \gamma(z)$ and $\gamma \in \Gamma_0$, then $z \in \partial\mathcal{D}$.
(c) For all $z \in \mathcal{D}$, the stability subgroup $Stab_\Gamma(z)$ is trivial except for ρ, i and $\rho - 1$.

Proof. (a) We have $\Im\gamma(z) = (\Im z)|cz + d|^{-2}$. Now we claim that the set of all values of this expression for $\Gamma_g \in \Gamma_0$ has a maximum, say at z_0 : this follows from the fact that there are only a finite number of pairs $c, d \in \mathbb{Z}$ with $|cz+d|$ less than a given bound. Furthermore, any point $z \in \mathcal{U}$ has a translate $T^k(z) = z+k$ with real part x in the interval $|x| \leq 1/2$. Hence there is an element γ of Γ' with $z_1 = \gamma(z_0) \in \mathcal{D}$: for if not, then $|z_1| < 1$ and $\Im(-1/z_1) > \Im z_1 = \Im z_0$ contradicting our choice of z_0. This proves (a).

(b) Let $z, g(z) \in \mathcal{D}$, with $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Without loss of generality, we may assume that $\Im g(z) \geq \Im z$. Therefore $|cz + d| \leq 1$ from the earlier calculation; hence it follows that $|c| \leq 1$, since $|z| \geq 1$ and c, d are integers so $|c| \geq 2$ is impossible. Now examining in turn the three cases $c = 0, \pm 1$ completes the argument: for instance, if $c = 0$ then $d = \pm 1$ so g is a translation and z lies on a vertical edge.

(c) This follows from the case analysis in (b). $\diamondsuit$

COROLLARY 1. *$\Gamma_0 = \Gamma(1)$, that is, the modular group is generated by the two elements S and T; equally, it is generated by the pair T and U, or by the two torsion elements S and U.*

COROLLARY 2. *The projection mapping $\pi : \mathcal{U} \to X = \mathcal{U}/\Gamma(1)$ is surjective when restricted to $\mathcal{D}$: $\pi : \mathcal{D} \to X$. by pasting of edges with the maps T and U.*

Consider now the topology of the resulting quotient surface: the act of past-

ing vertical edges gives an infinite cylinder, with lower end the section of the unit circle. The lower end is then pasted shut using U, to give a topological surface homeomorphic to the plane. [How would you justify this statement?] We will consider later the process of compactifying this surface.

[Note: we also need to discuss further the role of ∞, images of corner points of the triangle and the Riemann surface structure there.]

A closer analysis of this combinatorial structure on $\mathcal{U}$ shows that the group $\Gamma(1)$ has presentation $\langle U, S \mid |U^2 = S^3 = \mathrm{Id}\rangle$.

3) **Automorphic forms: the definition and interpretation.**

These are defined to be, in the first instance, functions on the upper half plane which satisfy a certain functional equation with respect to the action of our discrete group Γ:

$$f(\gamma(z)) = (cz+d)^k f(z) \text{ for all } \gamma = \frac{az+b}{cz+d} \in \Gamma. \tag{5}$$

Later there will be adjectives (like holomorphic, meromorphic ,...) attached to this. Notice the specific multiplier which occurs: $(cz+d)^{-2} = \gamma'(z)$, implying that there is an interpretation of this formula in the case $k = 2$ as a *differential form* on the quotient surface $X = \mathcal{U}/\Gamma$, a section of the cotangent bundle. For if we had a function F satisfying the equation with $k = 0$, which would say that $F(\gamma(z)) = F(z)$ for all $\gamma \in \Gamma$, that would project to give a genuine function on X, while differentiation gives that F' satisfies with $k = 2$. For arbitrary powers of k we say that such a function on X defines an *automorphic form of weight* $k/2$ for Γ.

Poincaré and Klein studied these forms for arbitrary (finitely generated) discrete groups $\Gamma \subset PSL(2, \mathbb{R})$ in the late 19th Century, laying the ground for a complete understanding of the basic structure of the fields of meromorphic functions and forms on the Riemann surfaces $X = \mathcal{U}/-$. This has developed today into a vast area involving the associated representation theory and an elaborate conjectural picture of many aspects of mathematics (including number theory, arithmetic geometry, conformal field theory, etc.) within a categorical framework loosely called the Langlands Program.

3

Week 3. Automorphic forms as differentials on the quotient Riemann surface.

The finiteness condition. Orbifold cone points and punctures.

THEOREM 3. *Finite area implies finite type RS quotient.*

Gauss-Bonnet Theorem.

Area of $\mathcal{D}$ for the modular group.

Cusps and horocycle cusp-neighbourhoods. The $q-$expansion. The compact quotient $\mathbb{C} \cup \{\infty\}$ is isomorphic to the Riemann sphere.

Analogue of cusp-nbd for elliptic points: there exists some open set $V \subset \mathcal{U}$, a nbd of a reference point ζ in $\mathcal{U}$ with nontrivial Γ-stabiliser $Stab_\Gamma(\zeta) = G = \langle \gamma | \gamma^m = 1 \rangle$, such that V is G-stable and projection $p : V \to V/G$ is m-to-1 onto a nbd of the image point Q.

Examples of modular forms: Eisenstein series. Convergence properties and modularity.

[**?**] L.V. Ahlfors , *Complex Analysis* (3rd Ed.) John Wiley & Sons.

[**?**] E.G. Rees, *Lectures on Geometry*, Springer, Universitext Series.

[**?**] J-P. Serre, *A Course in Arithmetic*, Springer, Universitext Series.