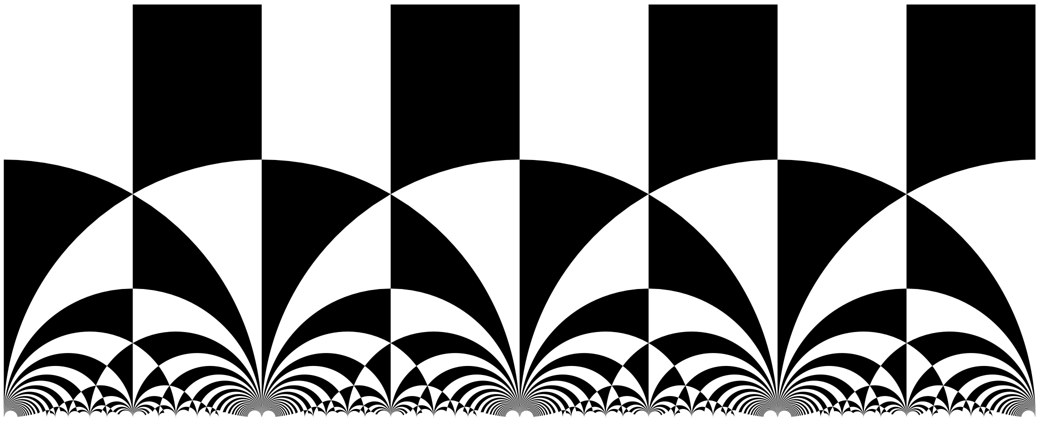
Anders Thorup

Notes on Automorphic Functions



Københavns Universitets Matematiske Institut
1995

Cover figures designed in *Mathematica* by Jakob Dylander and Anders Jensen.
Typeset by $\mathcal{AMS}$-TEX with *MathTime 1.1*

Automorphic Functions
Copenhagen 1994

Möbius transformations

1. Möbius transformations.

(1.1) Setup. The general linear group $\mathrm{GL}_2(\mathbf{C})$ consists of all 2×2 matrices with complex entries and non-zero determinant,

$$\alpha = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ where } ad - bc \neq 0. \tag{1.1.1}$$

The special linear group $\mathrm{SL}_2(\mathbf{C})$ is the subgroup formed by matrices (1.1.1) for which the determinant, $\det \alpha = ad - bc$, is equal to 1. For a matrix α of the form (1.1.1) and a point z of the Riemann Sphere (the extended complex plane $\overline{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$), we define the product,

$$\alpha \cdot z := \frac{az + b}{cz + d}. \tag{1.1.2}$$

The transformation $z \mapsto \alpha \cdot z$ of the Riemann sphere is called the *Möbius transformation* associated to the matrix α. The denominator of the fraction (1.1.2) will play an important role; we define, as a function of the matrix α and the complex number z,

$$J(\alpha, z) := cz + d. \tag{1.1.3}$$

Denote by $\mathbf{C}^2$ the 2-dimensional vector space of columns, and by $(\mathbf{C}^2)^*$ the subset of non-zero columns. Then there is a surjective map $(\mathbf{C}^2)^* \to \overline{\mathbf{C}}$ defined by

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \mapsto z_1/z_2.$$

The non-zero columns that are mapped to a given point z of $\overline{\mathbf{C}}$ are called the *representatives* of z. Clearly, the representatives of a given point form the non-zero columns in a 1-dimensional vector subspace of $\mathbf{C}^2$.

Three matrices deserve a special notation:

$$s := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad t := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad u := \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}.$$

Note that $s = tu$.

(1.2) Example. The translation $z \mapsto z + b$ and the multiplication $z \mapsto az$ for $a \neq 0$ are Möbius transformations, associated to the matrices,

$$\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}.$$

The point ∞ is fixed. On the other hand, if d is given complex number, then d is mapped to ∞ under the Möbius transformation associated to the matrix,

$$\begin{bmatrix} 0 & 1 \\ -1 & d \end{bmatrix}.$$

(1.3) Proposition. (1) *Two matrices α and β define the same Möbius transformation, if and only if they are proportional.*

(2) *The following equations hold for the product defined in* (1.1.2):

$$\alpha \cdot (\beta \cdot z) = (\alpha\beta) \cdot z, \quad 1 \cdot z = z.$$

(3) *Given in $\overline{\mathbf{C}}$ two sets of 3 different points, (u, v, w) and (u', v', w'). Then there is a Möbius transformation $z \mapsto \alpha \cdot z$ under which $(u, v, w) \mapsto (u', v', w')$, and the matrix α is unique up to multiplication by a non-zero scalar. In particular, the Möbius transformation $z \mapsto \alpha \cdot z$ is unique.*

(4) *If a column $\tilde{z}$ represents the point z, then the column $\alpha\tilde{z}$ represents the image point $\alpha \cdot z$.*

Proof. Clearly (4) holds, and (2) is an immediate consequence. In (1), obviously, if the matrices α and β are proportional, then they define the same Möbius transformation. Conversely, it follows from (3) that if two matrices define the same Möbius transformation, then they are proportional.

To prove (3), consider a set of 3 columns $(\tilde{u}, \tilde{v}, \tilde{w})$ representing 3 different points (u, v, w) in $\overline{\mathbf{C}}$. Say that the set of representatives is *balanced*, if

$$\tilde{w} = \tilde{u} + \tilde{v}. \tag{1.3.1}$$

When a representative $\tilde{w}$ of w is given, there is a unique choice of representatives of u and v, such that the resulting set of representatives is balanced. Indeed, the vectors representing a given point form the non-zero vectors in a 1-dimensional vector subspace of $\mathbf{C}^2$. Therefore, since $\mathbf{C}^2$ is a 2-dimensional vector space, the decomposition (1.3.1) is the unique decomposition of the vector $\tilde{w}$ into a sum of two vectors lying in two given different 1-dimensional subspaces of $\mathbf{C}^2$. It follows in particular that a balanced set of representatives is unique up to multiplication by a non-zero scalar.

Choose two balanced sets of representatives, $(\tilde{u}, \tilde{v}, \tilde{w})$ for (u, v, w), and $(\tilde{u'}, \tilde{v'}, \tilde{w'})$ for (u', v', w'). If (u, v, w) is mapped to (u', v', w') under a Möbius transformation $z \mapsto \alpha \cdot z$, then, by (4), $(\alpha\tilde{u}, \alpha\tilde{v}, \alpha\tilde{w})$ is a set of representatives for (u', v', w'), and it is a balanced set, because $(\tilde{u}, \tilde{v}, \tilde{w})$ is balanced. Therefore, by uniqueness of balanced sets, the set $(\alpha\tilde{u}, \alpha\tilde{v}, \alpha\tilde{w})$ is proportional to the set $(\tilde{u'}, \tilde{v'}, \tilde{w'})$. Hence we obtain the matrix relation,

$$\alpha(\tilde{u}, \tilde{v}, \tilde{w}) \sim (\tilde{u'}, \tilde{v'}, \tilde{w'}). \tag{1.3.2}$$

In particular, extracting the equations of the first two columns, we obtain that

$$\alpha \sim (\tilde{u'}, \tilde{v'})(\tilde{u}, \tilde{v})^{-1}. \tag{1.3.3}$$

Conversely, if α is defined by equality in (1.3.3), then equality holds in (1.3.2), since, on both sides, the third column is the sum of the first and the second. Therefore, by (4), the Möbius transformation associated to α maps $(u, v, w) \mapsto (u', v', w')$.

Thus (3) has been proved, and the proof is complete. □

(1.4) Definition. Usually we write αz for the product $\alpha \cdot z$. It follows from Proposition (1.3)(2) that the product $(\alpha, z) \mapsto \alpha z$ defines an action of the group $\mathrm{GL}_2(\mathbf{C})$ on the Riemann sphere $\overline{\mathbf{C}}$. Clearly, the Möbius transformations are analytic automorphisms. Hence the associated representation is a homomorphism of groups,

$$\mathrm{GL}_2(\mathbf{C}) \to \mathrm{Aut}_{\mathrm{an}}(\overline{\mathbf{C}}).$$

It is well known that the homomorphism is surjective, that is, every analytic automorphism of the Riemann sphere is a Möbius transformation. It follows from Proposition (1.3)(1) that the kernel of the homomorphism is the subgroup $\mathbf{C}^*$ of non-zero scalar matrices. For any subgroup G of $\mathrm{GL}_2(\mathbf{C})$, we denote by $\mathrm{P}G$ the quotient of G modulo the subgroup of scalar matrices contained in G, or equivalently, $\mathrm{P}G$ is the image of G in the group of Möbius transformations. The subgroup G is called *inhomogeneous* if it contains no non-trivial scalar matrix, that is, if $G = \mathrm{P}G$.

Note that $\mathrm{PGL}_2(\mathbf{C})$ is the group of all Möbius transformations. The group $\mathrm{PSL}_2(\mathbf{C})$ is the quotient,

$$\mathrm{PSL}_2(\mathbf{C}) = \mathrm{SL}_2(\mathbf{C})/\pm 1.$$

Every matrix in $\mathrm{GL}_2(\mathbf{C})$ is proportional to a matrix in $\mathrm{SL}_2(\mathbf{C})$, as it follows by dividing by a square root of the determinant. Hence $\mathrm{PSL}_2(\mathbf{C}) = \mathrm{PGL}_2(\mathbf{C})$, and every Möbius transformation is associated to a matrix in $\mathrm{SL}_2(\mathbf{C})$. A subgroup G of $\mathrm{SL}_2(\mathbf{C})$ is *homogeneous* if and only if it contains the matrix -1.

(1.5) Example. To determine the Möbius transformation under which $(\infty, 0, i)$ is mapped to $(1, -1, 0)$, consider these two sets of balanced representatives for the two sets of points:

$$\begin{bmatrix} i & 0 & i \\ 0 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 2 \end{bmatrix}.$$

It follows from (1.3)(3) that the Möbius transformation is associated to the matrix,

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}^{-1} \sim \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}.$$

Hence the transformation is the map $z \mapsto (z-i)/(z+i)$. It is called the *Cayley transformation*. To obtain a matrix in $\mathrm{SL}_2(\mathbf{C})$, divide the matrix by the square root of its determinant $2i$, that is, divide the matrix by $1 + i$.

(1.6) Corollary. *Any Möbius transformation preserves the cross ratio,*

$$\mathrm{df}(u, v, w, z) := \frac{u-w}{v-w} \Big/ \frac{u-z}{v-z}\,. \tag{1.6.1}$$

In particular, a Möbius transformation preserves angles, it maps a circle to a circle and the interior of an oriented circle to the interior of the image circle.

Proof. The cross ratio of four different points in $\mathbf{C}$ is the complex number defined by (1.6.1). It is easy to extend the definition to the case when one of the four points is equal to ∞. Clearly, for any set of four columns $(\tilde{u}, \tilde{v}, \tilde{w}, \tilde{z})$ representing (u, v, w, z), we have the equation,

$$\mathrm{df}(u, v, w, z) = \frac{\det(\tilde{u}, \tilde{w})}{\det(\tilde{v}, \tilde{w})} \Big/ \frac{\det(\tilde{u}, \tilde{z})}{\det(\tilde{v}, \tilde{z})}. \tag{1.6.2}$$

The first assertion of the Corollary is the equation $\mathrm{df}(\alpha u, \alpha v, \alpha w, \alpha z) = \mathrm{df}(u, v, w, z)$ for any Möbius transformation $z \mapsto \alpha z$. By (1.3)(4), the first assertion follows from (1.6.2), since $\det(\alpha\tilde{u}, \alpha\tilde{v}) = \det\alpha \det(\tilde{u}, \tilde{v})$.

To prove the remaining assertions, fix a point v in $\mathbf{C}$ and, in $\overline{\mathbf{C}}$, a point $u \neq v$. For any point w in $\overline{\mathbf{C}}$ different from u and v, consider the circle C_w through u, v, and w, oriented by the order uvw. Let t_w be the oriented tangent in the point v to the circle C_w. If z is a fourth point, and t_z is defined similarly from the circle C_z, then the following equation holds for the angle from t_w to t_z:

$$\angle(t_w, t_z) = \arg \mathrm{df}(u, v, w, z). \tag{1.6.3}$$

To prove Equation (1.6.3), consider first the case when $u = \infty$. Then t_w and t_z are the two oriented straight lines from v to w and z, and the left hand side is the angle between them. On the other side, the cross ratio reduces to the ratio $(v-z)/(v-w)$. Hence the argument of the cross ratio is equal to the left hand side. Assume next that $u \neq \infty$, and consider the oriented (straight) line $t = \overrightarrow{uv}$ from u to v. It is the common chord to the two circles C_w and C_z. It follows from elementary plane geometry properties of the circle C_w, that the angle from $\overrightarrow{wv}$ to $\overrightarrow{wu}$ is equal to the angle from t_w to t,

$$\angle(t_w, t) = \angle(\overrightarrow{wv}, \overrightarrow{wu}).$$

Moreover, the angle on the right hand side is equal to the argument of $(u-w)/(v-w)$. Now, the equation (1.6.3) follows from the additivity $\angle(t_w, t) + \angle(t, t_z) = \angle(t_w, t_z)$.

As a consequence of (1.6.3), the cross ratio $\mathrm{df}(u, v, w, z)$ belongs to $\mathbf{R}_+$ if and only if z is on the arc vu of the circle C_w, it belongs to $\mathbf{R}_-$ if and only if z is on the arc uv of the circle, it belongs to the upper half plane if and only if z is an inner point of the circle C_w, and it belongs to the lower half plane if and only if z is exterior to the circle.

Therefore, as a Möbius transformation preserves the cross ratio, it follows that circles and interiors of circles are preserved, and it follows from (1.6.3) that angles between circles are preserved. ▯

(1.7) Remark. In the proof of (1.6) we obtained information on the argument of the cross ratio $\mathrm{df}(u, v, w, z)$ by writing the cross ratio as a quotient of two fractions,

$$\frac{u-w}{v-w}, \qquad \frac{u-z}{v-z}. \tag{1.7.1}$$

Similarly, we can obtain information on the modulus of the cross ratio. Assume for simplicity that the three points u, v, and w are not on a straight line (in particular, they are different from ∞), and consider the (ordinary) triangle defined by them. Denote by W the angle at w and by U_w and V_w the angles at u and v. The two lengths $|u-w|$ and $|v-w|$ are sides of the triangle, and hence the quotient $|u-w|/|v-w|$ is equal to the quotient $\sin V_w/\sin U_w$. Hence, as the sum of the three angles is equal to π, we obtain for the modulus of the first fraction (1.7.1) the equation,

$$\left|\frac{u-w}{v-w}\right| = \frac{\sin V_w}{\sin U_w} = \cos W + \cot U_w \sin W. \tag{1.7.2}$$

Thus the modulus of the cross ratio $\mathrm{df}(u, v, w, z)$ is the quotient of the expression (1.7.2) and the expression obtained similarly replacing w by z.

Fix u and v and an oriented circle C through u and v. Consider the expression (1.7.2) as a function $d_C(w)$ defined for points w different from u and v on the circle C. The angle W is constant, say equal to θ, on the arc from v to u, and on the arc from u to v the angle W is equal to $\pi - \theta$. Consider points on the arc from v to u. Then, as w runs from v to u, the angle U_w increases from 0 to $\pi - \theta$, and consequently, the function $d_C(w)$ decreases from $+\infty$ to 0.

Consider, for 4 different points (u, v, w, z) on C, the cross ratio $\mathrm{df}(u, v, w, z)$. It follows from the proof of Corollary (1.6), that the cross ratio is real and positive if and only if w and z belong to the same of the two arcs determined by u and v. In particular, when w and z belong to the same arc, then the cross ratio is equal to its modulus, and hence equal to $d_C(w)/d_C(z)$. As a consequence, when z is on the arc from v to u containing w, the cross ratio $\mathrm{df}(u, v, w, z)$ increases from 0 to 1 as z runs from v to w, and it increases from 1 to $+\infty$ as z runs from w to u.

It is easy to prove that the latter assertion holds also when the circle is a straight line.

(1.8) Definition. An open disk in $\overline{\mathbf{C}}$ will simply be called a *disk*. Thus a disk is either an open half plane in $\mathbf{C}$, or the interior of a usual circle in $\mathbf{C}$, or the exterior (including ∞) of a usual circle. The boundary $\partial\mathfrak{D}$ of a disk $\mathfrak{D}$ is a circle, always oriented counter clockwise around the disk. If $\mathfrak{D}$ is a disk, we denote by $\mathrm{SL}(\mathfrak{D})$ the stabilizer of $\mathfrak{D}$ in $\mathrm{SL}_2(\mathbf{C})$, that is, the subgroup of $\mathrm{SL}_2(\mathbf{C})$ consisting of matrices α for which $\alpha\mathfrak{D} = \mathfrak{D}$.

Throughout, we denote by $\mathfrak{H}$ the upper half plane: $\Im z > 0$, and by $\mathfrak{E}$ the open unit disk: $|z| < 1$. The boundary (in $\overline{\mathbf{C}}$) of $\mathfrak{H}$ is the extended real line $\overline{\mathbf{R}}$, and the boundary of $\mathfrak{E}$ is the unit circle: $|z| = 1$.

(1.9) Corollary. *Given two triples $(\mathfrak{D}, w, u)$ and $(\mathfrak{D}', w', u')$, each consisting of a disk, a point in the disk, and a point on the boundary of the disk. Then there is a unique Möbius transformation mapping the first triple to the second.*

Proof. The point w is in $\mathfrak{D}$ and the point u is on the boundary $\partial\mathfrak{D}$. Clearly, there is a unique circle C orthogonal to $\partial\mathfrak{D}$, and passing through w and u. The circles C and $\partial\mathfrak{D}$ intersect in two points, one of which is u. Denote by v the second point of intersection. Define C' and v' similarly from the second triple. Clearly, for a Möbius transformation as required, the boundary of $\mathfrak{D}$ is mapped to the boundary of $\mathfrak{D}'$. Hence the circle C is mapped to the circle C', and consequently, v is mapped to v'. Thus, by Proposition (1.3), the Möbius transformation is the unique transformation under which

$$(u, v, w) \mapsto (u', v', w'). \tag{1}$$

Conversely, under the transformation determined by (1), the circle C is mapped to the circle C'. Hence the boundary of $\mathfrak{D}$, which is orthogonal to C, is mapped to the boundary of $\mathfrak{D}'$, and hence $\mathfrak{D}$ is mapped to $\mathfrak{D}'$. Thus the Möbius transformation determined by (1) has the required properties. □

(1.10) Example. Clearly, the Cayley transformation of (1.5) is the unique Möbius transformation mapping $(\mathfrak{H}, i, \infty)$ onto $(\mathfrak{E}, 0, 1)$.

(1.11) Remark. Let α be a matrix of $\mathrm{GL}_2(\mathbf{C})$. The transformation $z \mapsto \alpha\bar{z}$, where $\bar{z}$ denotes the complex conjugate of z, is called the *anti-transformation* associated to α. Note that an anti-transformations is not a Möbius transformation. However, the composition of anti-transformations, associated to matrices α and β, is a Möbius transformation, associated to the product $\alpha\overline{\beta}$.

Clearly, under an anti-transformation, the cross ratio of four points is changed into the complex conjugate. As a consequence, an anti-transformation preserves circles, but angles between circles are reversed. The interior of an oriented circle is mapped to the exterior of the image circle (when the image circle is given the image orientation).

(1.12) Example. Complex conjugation is the anti-transformation associated to the identity matrix. Under complex conjugation, the unit disk $\mathfrak{E}$ is mapped to itself, but the orientation of the boundary is reversed.

(1.13) Exercise. Prove for $z \in \mathfrak{H}$ and $\sigma \in \mathrm{SL}_2(\mathbf{R})$ that $\Im(\sigma z) = |J(\sigma, z)|^{-2}\Im z$.

2. Fixed points.

(2.1) Definition. Let α be a matrix in $\mathrm{GL}_2(\mathbf{C})$, and consider the associated Möbius transformation $z \mapsto \alpha z$. Clearly, a point z of $\overline{\mathbf{C}}$ is a fixed point of the transformation, if and only if the representatives of z in $(\mathbf{C}^2)^*$ are eigenvectors of α. Therefore we get the following classification of transformations:

Case 1. The matrix α has one eigenvalue and the corresponding eigenspace is of dimension 2. In this case, the matrix α is a scalar matrix,

$$\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}.$$

The associated transformation is the identity, and every point in $\overline{\mathbf{C}}$ is fixed.

Case 2. The matrix α has exactly one eigenvalue λ, and the eigenspace is of dimension 1. In this case, the transformation has exactly one fixed point. The transformation (and the matrix) is called *parabolic*. The matrix is similar to a matrix of the form,

$$\begin{bmatrix} \lambda & b \\ 0 & \lambda \end{bmatrix},$$

where $b \neq 0$.

Case 3. The matrix α has two different eigenvalues λ_1 and λ_2 (necessarily both with a one dimensional eigenspace). In this case, the transformation has two fixed points. The matrix α is similar to a matrix for the form,

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Of particular geometric interest is the quotient λ_1/λ_2 of the two eigenvalues (changed to its inverse when the two eigenvalues are interchanged). The matrix α is called *hyperbolic*, if the quotient is real and positive, and *elliptic*, if the quotient is of modulus 1. It is called *loxodromic* if it is neither elliptic nor hyperbolic.

Note that the quotient of the two eigenvalues of a matrix α is equal to 1, if and only the matrix is either a scalar or a parabolic matrix.

(2.2) Example. Obviously, under the action of $\mathrm{SL}_2(\mathbf{C})$ on $\overline{\mathbf{C}}$, the isotropy group of the point ∞ is the subgroup formed by the following matrices:

$$\begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix}. \tag{2.2.1}$$

Clearly, among the matrices in (2.2.1) the parabolic matrices are the following:

$$\pm\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix};$$

their associated transformations are translations $z \mapsto z + b$ for $b \neq 0$.

Consider the matrices in (2.2.1) that fix in addition the point 0. Of these matrices, the hyperbolic matrices are the following:

$$\pm\begin{bmatrix} r & 0 \\ 0 & r^{-1} \end{bmatrix},$$

where $r \in \mathbf{R}_+ - \{1\}$; their associated transformations are multiplications $z \mapsto r^2 z$. The elliptic matrices are the following:

$$\begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix},$$

where $\theta \in \mathbf{R} - \mathbf{Z}\pi$; their associated transformations are rotations $z \mapsto e^{2i\theta} z$.

An invariant closely related to the geometric properties of Möbius transformations is the following, defined for any matrix $\alpha \in \mathrm{GL}_2(\mathbf{C})$:

$$\mathrm{h}(\alpha) := \frac{1}{2}\,\frac{\mathrm{tr}(\alpha^2)}{\det \alpha}\,.$$

By Proposition (1.3)(1), the invariant $\mathrm{h}(\alpha)$ depends on the associated Möbius transformation $z \mapsto \alpha z$ only. Moreover, the invariant is unchanged if α is replaced by a conjugate $\sigma\alpha\sigma^{-1}$. Expressed by the eigenvalues in (2.1), we have that $\mathrm{h}(\alpha) = \frac{1}{2}(\lambda_1/\lambda_2 + \lambda_2/\lambda_1)$. Thus the quotient λ_1/λ_2 and its inverse λ_2/λ_1 are the two roots of the quadratic polynomial,

$$\lambda^2 - 2\,\mathrm{h}(\alpha)\lambda + 1.$$

In particular, if α is not a scalar matrix, then α is parabolic, if and only if $\mathrm{h}(\alpha) = 1$, α is hyperbolic if and only if $\mathrm{h}(\alpha)$ is real and in the interval $1 < h < +\infty$, and α is elliptic, if and only if $\mathrm{h}(\alpha)$ is real and in the interval $-1 \le h < 1$.

(2.3) Lemma. (1) *The stabilizer* $\mathrm{SL}(\mathfrak{H})$ *of the upper half plane* $\mathfrak{H}$ *is the subgroup* $\mathrm{SL}_2(\mathbf{R})$ *consisting of matrices,*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{where } a, b, c, d \in \mathbf{R} \text{ and } ad - bc = 1.$$

The isotropy group $\mathrm{SL}(\mathfrak{H})_\infty$ *of the point* ∞ *on the boundary of* $\mathfrak{H}$ *is the subgroup consisting of matrices,*

$$\begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix} \quad \text{for } a \in \mathbf{R}^*,\ b \in \mathbf{R}. \tag{2.3.1}$$

In particular, the isotropy group $\mathrm{SL}(\mathfrak{H})_\infty$ *is non-compact and non-commutative. The following relation holds,*

$$\begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix}\begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a & b \\ 0 & a^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & a^2 h \\ 0 & 1 \end{bmatrix}. \tag{2.3.2}$$

Moreover, the subset of $\mathrm{SL}(\mathfrak{H})_\infty$ *consisting of matrices that are either parabolic or* ± 1 *is a subgroup, isomorphic to* $\{\pm 1\} \times \mathbf{R}$.

(2) *The stabilizer* $\mathrm{SL}(\mathfrak{E})$ *of the unit disk* $\mathfrak{E}$ *is the subgroup* $\mathrm{SU}_{1,1}(\mathbf{C})$ *consisting of matrices,*

$$\begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix} \text{ for } a, b \in \mathbf{C} \text{ and } |a|^2 - |b|^2 = 1.$$

The isotropy group $\mathrm{SL}(\mathfrak{E})_0$ *of the point* 0 *in* $\mathfrak{E}$ *is the subgroup consisting of matrices,*

$$\begin{bmatrix} a & 0 \\ 0 & \bar{a} \end{bmatrix} \text{ for } a \in \mathbf{C},\ |a| = 1. \tag{2.3.3}$$

In particular, the isotropy group is compact and commutative (and isomorphic to the unit circle $\mathrm{U}_1(\mathbf{C})$*).*

Proof. (1) Let α be a matrix in $\mathrm{SL}(\mathfrak{H})$. The Möbius transformation $z \mapsto \alpha z$ maps the disk $\mathfrak{H}$ onto itself and, consequently, it maps the boundary $\overline{\mathbf{R}}$ onto itself. Therefore, by Section 1, the Möbius transformation is associated to a matrix α' with real entries. Since α' has real entries, the imaginary part of $\alpha' \cdot i$ is equal to a positive scalar times the determinant of α'. It follows that the determinant is positive, and dividing the matrix by a square root of the determinant, we may assume that α' has determinant 1. As α and α' define the same Möbius transformation, it follows that $\alpha = \pm\alpha'$. Hence α belongs to $\mathrm{SL}_2(\mathbf{R})$. Conversely, it is obvious that any matrix in $\mathrm{SL}_2(\mathbf{R})$ belongs to $\mathrm{SL}(\mathfrak{H})$.

The remaining assertions of (1) and the similar assertions of (2) are left as an exercise. □

(2.4) Exercise. (1) Prove that the subgroup $\mathrm{SL}(\mathfrak{H})_i$ of matrices in $\mathrm{SL}(\mathfrak{H})$ having i as fixed point is the subgroup $\mathrm{SO}_2(\mathbf{R})$ of matrices,

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad \text{for } a, b \in \mathbf{R} \text{ and } a^2 + b^2 = 1.$$

Prove that the subgroup $\mathrm{SL}(\mathfrak{H})_{1,-1}$ of matrices in $\mathrm{SL}(\mathfrak{H})$ having 1 and -1 as fixed points is the subgroup $\mathrm{SO}_{1,1}(\mathbf{R})$ of matrices,

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \quad \text{for } a, b \in \mathbf{R} \text{ and } a^2 - b^2 = 1.$$

(2.5) Corollary. *Consider a disk* $\mathfrak{D}$ *in* $\overline{\mathbf{C}}$*. No matrix in* $\mathrm{SL}(\mathfrak{D})$ *is loxodromic. Let* $\sigma \neq \pm 1$ *be a matrix in* $\mathrm{SL}(\mathfrak{D})$*. If* σ *is parabolic, then the fixed point of* σ *belongs to the boundary* $\partial\mathfrak{D}$*. If* σ *is elliptic, then* σ *has one fixed point in* $\mathfrak{D}$ *and the other fixed point belongs to the complement of the closure of* $\mathfrak{D}$*. Finally, if* σ *is hyperbolic, then the two fixed points of* σ *belongs to the boundary of* $\mathfrak{D}$*.*

Proof. After conjugation, we may assume that the disk is the upper half plane $\mathfrak{H}$. Then, by Proposition (2.3), the matrix σ has real entries. Hence, the eigenvalues of σ are either real or a pair of complex conjugate numbers. Clearly, in the first case there are real eigenvectors and, consequently, either σ is parabolic with a fixed point in $\overline{\mathbf{R}}$ or σ is hyperbolic with two fixed points in $\overline{\mathbf{R}}$. In the second case, if a column is an eigenvector corresponding to one eigenvalue, then the conjugate column is an eigenvector corresponding to the complex conjugate eigenvalue. Hence, in the second case, σ has one fixed point in the upper half plane $\mathfrak{H}$ and the complex conjugate fixed point in the lower half plane.

Thus the assertions hold. □

(2.6) Note. By Corollary (1.9), for any two disks $\mathfrak{D}$ and $\mathfrak{D}'$ there is a Möbius transformation $z \mapsto \alpha z$ defining an isomorphism,

$$\alpha : \mathfrak{D}' \to \mathfrak{D},$$

of $\mathfrak{D}'$ onto $\mathfrak{D}$. In addition, for a given pair of points $u \in \mathfrak{D}$ and $u' \in \mathfrak{D}'$ (or $u \in \partial\mathfrak{D}$ and $u' \in \partial\mathfrak{D}'$), we may choose α such that $\alpha u' = u$.

Under the isomorphism α, points $z \in \mathfrak{D}$ correspond to points $z^\alpha := \alpha^{-1} z$ in $\mathfrak{D}'$, functions f defined on $\mathfrak{D}$ correspond to functions $f^\alpha := f\alpha$ on $\mathfrak{D}'$, and automorphisms σ of $\mathfrak{D}$ correspond to automorphisms $\sigma^\alpha := \alpha^{-1}\sigma\alpha$ on $\mathfrak{D}'$. The latter correspondence extends to matrices: matrices σ in $\mathrm{SL}(\mathfrak{D})$ correspond to matrices $\sigma^\alpha := \alpha^{-1}\sigma\alpha$ in $\mathrm{SL}(\mathfrak{D}')$. The correspondence is called *conjugation*. In particular,

$$\mathrm{SL}(\mathfrak{D}') = \mathrm{SL}(\mathfrak{D})^\alpha,$$

is the conjugate subgroup of $\mathrm{SL}(\mathfrak{D})$. Clearly, the property of being a fixed point is preserved under conjugation, and for the isotropy groups we obtain the equation,

$$\mathrm{SL}(\mathfrak{D}')_{u'} = \mathrm{SL}(\mathfrak{D})_u^\alpha,$$

when $u' = u^\alpha$.

Usually, of the two disks in (1.8), we take $(\mathfrak{E}, 0)$ as the model of a disk $\mathfrak{D}$ and a point $u \in \mathfrak{D}$, and we take $(\mathfrak{H}, \infty)$ as the model of a disk $\mathfrak{D}$ and a point $u \in \partial\mathfrak{D}$.

A property of points $u \in \mathfrak{D} \cup \partial\mathfrak{D}$ will often be defined by first defining the property for $0 \in \mathfrak{E}$ and $\infty \in \partial\mathfrak{H}$ and then in general by choosing an isomorphism σ from $(\mathfrak{D}, u)$ to one of the standard models. In these cases, the property will be said to be *defined by conjugation*. In each case, the definition has to be justified by proving that the property is independent of the choice of conjugation σ.

(2.7) Definition. Let $\mathfrak{D}$ be a disk, and let G be a subgroup of $\mathrm{SL}(\mathfrak{D})$. By definition, a *$G$-elliptic point* is a point u in $\mathfrak{D}$ which is fixed under some nontrivial (necessarily elliptic) matrix in G, and a *G-parabolic point* is a point u of the boundary $\partial\mathfrak{D}$ which is fixed under some parabolic matrix in G. A point of $\mathfrak{D}$ which is not G-elliptic may be called a *G-ordinary point*.

As a subgroup of $\mathrm{SL}(\mathfrak{D})$, the group G acts on $\mathfrak{D}$ and on the boundary $\partial\mathfrak{D}$. Clearly, a point u of $\mathfrak{D}$ is G-elliptic if and only if the isotropy group G_u is non-trivial (that is, contains a matrix different from ± 1), and a point u of $\partial\mathfrak{D}$ is G-parabolic, if and only if the isotropy group G_u contains a parabolic matrix.

Note that the properties are preserved under conjugation. If $\mathfrak{D}' = \alpha\mathfrak{D}$, then $G' := \alpha G \alpha^{-1}$ is a subgroup of $\mathrm{SL}(\mathfrak{D}')$, and for points u and $u' := \alpha u$, where u is in $\mathfrak{D}$ or $\partial\mathfrak{D}$, we have that $G'_{u'} = \alpha G_u \alpha^{-1}$. Moreover, the point u is G-parabolic or G-elliptic respectively, if and only if u' is G'-parabolic or G'-elliptic.

3. Non-Euclidean plane geometry.

(3.1) Lemma. *Let C be an ordinary circle in* $\mathbf{C}$*, with its center on the positive real axis and such that C is orthogonal to the unit circle around* 0*. Let $r < r'$ be the two points of intersection of C and the real axis. Let l be a straight line through* 0 *intersecting the circle C in two points z and z' with $\Re z \le \Re z'$. Denote by A the angle between l and the real axis and denote by B the angle between l and the tangent in z to the circle C. Then,*

$$\cos A = \frac{|z|^{-1} + |z|}{r^{-1} + r}, \qquad \sin B = \frac{|z|^{-1} - |z|}{r^{-1} - r}. \tag{3.1.1}$$

Proof. Let z_0 denote the midpoint of the chord zz'. Then $z_0 = (z' + z)/2$. In particular, $r_0 = (r' + r)/2$ is the center of the circle C. Clearly, the straight line from r_0 to z_0 is orthogonal to the line l. Moreover, the angle at r_0 between the lines $r_0 z$ and $r_0 z_0$ is equal to B. Hence we have the equations,

$$\cos A = \frac{|z_0|}{r_0}, \qquad \sin B = \frac{|z_0 - z|}{r_0 - r}. \tag{3.1.2}$$

As is well known, the product of distances, $|z||z'|$, is independent of l. By hypothesis, if l is the tangent to the circle C, then $z = z'$ is on the unit circle, and hence $|z||z'| = 1$. Hence $|z||z'| = 1$ for an arbitrary line l intersecting the circle. In particular, $rr' = 1$. Therefore, the equations (3.1.2) imply the equations (3.1.1) $\square$

(3.2) Setup. Consider for the rest of this section a disk $\mathfrak{D}$. By definition, a *line in* $\mathfrak{D}$ is (the part in $\mathfrak{D}$ of) a circle orthogonal to the boundary of $\mathfrak{D}$. The circle will intersect the boundary of $\mathfrak{D}$ in two points, called the *limit points* of the line. By Corollary (1.6), a Möbius transformation $\alpha\colon \mathfrak{D} \to \mathfrak{D}'$ maps lines of $\mathfrak{D}$ to lines of $\mathfrak{D}'$. As a consequence, assertions about points, lines, and limit points in $\mathfrak{D}$ hold in general, if they hold for one of the two standard disks, $\mathfrak{H}$ and $\mathfrak{E}$. For instance, in the unit disk $\mathfrak{E}$, the lines through the point 0 are the diameters of the unit circle. Hence, for every point $z \neq 0$ in $\mathfrak{E}$ there is a unique line of $\mathfrak{E}$ passing through 0 and z. As a consequence, for any two different points z and w in $\mathfrak{D}$, there is a unique line of $\mathfrak{D}$ passing through z and w. Similarly, for any point w in $\mathfrak{D}$ and any line l not passing through w, there is a unique point z on l such that the line through w and z is orthogonal to l.

(3.3) Definition. Consider the line in $\mathfrak{D}$ passing through two different points w and z of $\mathfrak{D}$. The line has two limit points u, v on the boundary of $\mathfrak{D}$; we may choose the notation such that the open arc wzu of the line is contained in $\mathfrak{D}$. It follows from Remark (1.7) that the cross ratio $\mathrm{df}(u, v, w, z)$ is real and greater than 1. The number,

$$\mathrm{dist}_{\mathfrak{D}}(w, z) := \log \mathrm{df}(u, v, w, z),$$

which is positive, is called the *non-euclidean distance* between w and z. Note that if $w = z$, then the cross ratio on the right hand side is equal to 1 for any pair of different points u and

v on the boundary of $\mathfrak{D}$. Accordingly, as might be expected, the distance is defined to be 0 when $w = z$.

The non-euclidean distance is a metric in $\mathfrak{D}$, that is, the triangle inequality holds for any 3 points w, x, z in $\mathfrak{D}$. This fact, and the fundamental trigonometric formulas of non-euclidean plane geometry will be proved in the following.

Note that any Möbius transformation $\alpha\colon \mathfrak{D} \to \mathfrak{D}'$ is an isometry with respect to the distances of $\mathfrak{D}$ and $\mathfrak{D}'$. In particular, the matrices in $\mathrm{SL}(\mathfrak{D})$ define isometries of the disk $\mathfrak{D}$.

(3.4) Remark. If an oriented line l is given in $\mathfrak{D}$ then there is a signed distance $\mathrm{dist}_l(w, z)$ defined for points w and z on l as follows: Let u and v be the limit points of l, chosen such that the arc from v to u is the part of l in $\mathfrak{D}$. Then $\mathrm{dist}_l(w, z) := \log \mathrm{df}(u, v, w, z)$. Clearly, for any 5 different points in $\overline{\mathbf{C}}$ we have the equation,

$$\mathrm{df}(u, v, w, z) = \mathrm{df}(u, v, w, x)\,\mathrm{df}(u, v, x, z).$$

Hence, for 3 points w, x, z on a line l, we have the additivity for the signed distance,

$$\mathrm{dist}_l(w, z) = \mathrm{dist}_l(w, x) + \mathrm{dist}_l(x, z).$$

In particular, $\mathrm{dist}_l(w, z) = -\,\mathrm{dist}_l(z, w)$.

(3.5) Example. In the unit disk $\mathfrak{E}$, the distance from 0 to a point z in $\mathfrak{E}$ is given by the formula,

$$\mathrm{dist}(0, z) = \log \frac{1 + |z|}{1 - |z|}\,. \tag{3.5.1}$$

Indeed, the two sides of the formula are unchanged under a rotation around 0. Hence we may assume that the point z is real and $0 < z < 1$. Then the line through 0 and z is the real axis, and the limit points are $u = 1$ and $v = -1$. The formula is now obvious.

(3.6) Remark. It follows from the formula (3.5.1) for the unit disk $\mathfrak{E}$ that a *non-euclidean circle* (also called a *geodesic circle*) with center 0, that is, the set of points in $\mathfrak{E}$ of a fixed (non-euclidean) distance to 0, is an ordinary circle. Moreover, the line in $\mathfrak{E}$ from the center 0 to a point on a geodesic circle around 0 is orthogonal to the circle. Similarly, an open geodesic disk around 0, that is, a set of points in $\mathfrak{E}$ whose distance from 0 is strictly less than a given positive number, is an ordinary disk contained in $\mathfrak{E}$, and the system of geodesic disks around 0 form a basis for the system of neighborhoods of 0. As a consequence, for any point w in an arbitrary disk $\mathfrak{D}$, the geodesic circles in $\mathfrak{D}$ around w are (ordinary) circles, and the line in $\mathfrak{D}$ from w to a point on a geodesic circle around w is orthogonal to the circle. Similarly, geodesic disks in $\mathfrak{D}$ are ordinary disks contained in $\mathfrak{D}$, and the system of geodesic disks around w is a basis for the system of neighborhoods of w. In other words, the non-euclidean distance induces a topology in $\mathfrak{D}$ equal to the ordinary topology of $\mathfrak{D}$ as a subspace of $\overline{\mathbf{C}}$.

(3.7) Example. Consider the upper half plane $\mathfrak{H}$, and two points z and w in $\mathfrak{H}$. Then the distance from w to z is given by the formula,

$$\cosh \mathrm{dist}(w, z) = 1 + \frac{|w - z|^2}{2\Im z \Im w}\,. \tag{3.7.1}$$

Indeed, assume first that the two points have different real part. Then the line in $\mathfrak{H}$ from w to z is a circle orthogonal to the real axis. The circle intersects the real axis in two points, labeled u, v in the usual order, see (3.3). Let r be the radius of the circle, and o its center. Let θ_w be the angle vuw and θ_z the angle vuz. Then, clearly,

$$\tan\theta_w = \frac{|v-w|}{|u-w|} \quad\text{and}\quad \tan\theta_z = \frac{|v-z|}{|u-z|}.$$

As the cross ratio $\operatorname{df}(u,v,w,z)$ is positive real, it is equal to $\tan\theta_z/\tan\theta_w$. Therefore, the left hand side of (3.7.1), is equal to the expression,

$$\frac{1}{2}\Big(\frac{\tan\theta_z}{\tan\theta_w} + \frac{\tan\theta_w}{\tan\theta_z}\Big). \tag{3.7.2}$$

On the other side, the angles vow and voz are, respectively, $2\theta_w$ and $2\theta_z$. Hence, the angle woz is equal to $2(\theta_z - \theta_w)$. Therefore,

$$\sin 2\theta_w = \frac{\Im w}{r}, \quad \sin 2\theta_z = \frac{\Im z}{r}, \quad \sin(\theta_z - \theta_w) = \frac{\frac{1}{2}|z-w|}{r}.$$

From these equations it follows that the right hand side of (3.7.1) is equal to the expression,

$$1 + \frac{2\sin^2(\theta_z - \theta_w)}{\sin 2\theta_w \,\sin 2\theta_z}. \tag{3.7.3}$$

By elementary trigonometric formulas, the expressions (3.7.2) and (3.7.3) are equal. Hence (3.7.1) holds.

The equation (3.7.1) is easily seen to hold when z and w have the same real part. Hence the equation holds in general.

(3.8) Exercise. Consider the non-euclidean distance in $\mathfrak{H}$. Prove for two points w and z with the same real part that

$$\operatorname{dist}(w,z) = |\log(\Im z/\Im w)|. \tag{3.8.1}$$

Prove for h positive and real that

$$\operatorname{dist}(z, z+h) = 2\log\bigg(\frac{h}{2\Im z} + \sqrt{\Big(\frac{h}{2\Im z}\Big)^2 + 1}\bigg). \tag{3.8.2}$$

Note in particular that the distance converges to zero for $\Im z \to \infty$.

(3.9) Setup. Consider in $\mathfrak{D}$ a triangle A, B, C, that is, A, B and C are three different points of $\mathfrak{D}$, in general assumed to be not on the same line of $\mathfrak{D}$. Denote by a, b, and c the sides of the triangle, that is, a is the line through B and C, b is the line through A and C, and c is the line through A and B. It is customary to denote by the same symbols also the lengths of the sides of the triangle, that is, a is the distance $\operatorname{dist}_{\mathfrak{D}}(B,C)$ etc. Similarly, A will also denote the (unoriented) angle at A, that is, the angle between the lines b and c, etc.

(3.10) Proposition. *Assume in the setup of* (3.9) *that the angle at C is a right angle. Then the following formulas hold:*

$$\cos A = \frac{\tanh b}{\tanh c}, \qquad \sin A = \frac{\sinh a}{\sinh c}, \qquad \cosh c = \cosh a \cosh b. \tag{3.10.1}$$

Proof. To prove the first formula, we may after conjugation assume that $\mathfrak{D}$ is the unit disk $\mathfrak{E}$ and A is the point 0. Moreover, after a rotation around 0 we may assume that C is a real point r with $0 < r < 1$. Then the line a is an ordinary circle with center on the positive real axis and orthogonal to the unit circle, and B is a point z on the line a. The distances b and c from 0 to C and from 0 to B are, by Example (3.5), given by $\exp b = (1+r)/(1-r)$ and $\exp c = (1+|z|)/(1-|z|)$. Therefore, the first formula of (3.10.1) follows from the first formula of (3.1.1).

Similarly, the second formula, in the symmetric form $\sin B = \sinh b/\sinh c$, follows from the second formula of (3.1.1). Finally, the third formula of (3.10.1) follows from the first two by using the relation $\cos^2 A + \sin^2 A = 1$. ▯

(3.11) Proposition. *In the setup of* (3.9), *we have for an arbitrary triangle ABC in $\mathfrak{D}$ the cosine relation:*

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos C,$$

and the sine relations:

$$\frac{\sin A}{\sinh a} = \frac{\sin B}{\sinh b} = \frac{\sin C}{\sinh c}.$$

Proof. Denote by H the point on the side a such that the line from A to H is orthogonal to a. Then the triangles ACH and ABH have a right angle at H and they have as common side the line h from A to H. Denote by x the side CH of ACH and by y the side BH of ABH.

By the third formula of (3.10.1), applied to ACH and ABH, we obtain that $\cosh h = \cosh b/\cosh x = \cosh c/\cosh y$. Hence,

$$\cosh c = \cosh b\,\frac{\cosh y}{\cosh x}. \tag{1}$$

Assume that the point H on a lies between B and C (the two alternative cases are left to the reader). Then $y = a - x$ by (3.4). Hence, by the addition formula for the hyperbolic cosine, we obtain from (1) that,

$$\cosh c = \frac{\cosh a \cosh x - \sinh a \sinh x}{\cosh x}\cosh b = \cosh a \cosh b - \sinh a \cosh b \tanh x. \tag{2}$$

Moreover, by the first equation of (3.10.1), we have that $\tanh x = \tanh b \cos C$. Hence the equation (2) implies the cosine relation.

To prove the sine relation, note that the second formula of (3.10.1) implies the equations $\sin C \sinh b = \sinh h = \sin B \sinh c$. Hence the second sine relation holds, and by symmetry they all hold. ▯

(3.12) Remark. The triangle inequality for the non-euclidean distance follows from the cosine relation. Indeed, the right hand side of the relation is at most equal to $\cosh a \cosh b + \sinh a \sinh b = \cosh(a+b)$, and consequently $c \le a + b$. Moreover, equality holds if and only if the angle C is equal to π, that is, if and only if C belongs to the line segment from A to B.

(3.13) Setup. In the geometric language, the elliptic transformations of $\mathfrak{D}$ are called *rotations*, the hyperbolic transformations are called *translations*, and the parabolic transformations are called *limit rotations*. A translation has two limit point as fixed points; the line between them is called the *axis*.

We will study these maps in more detail in the following. In addition we will need the *reflections*: Let l be a line in $\mathfrak{D}$. Choose a Möbius transformation $z \mapsto \alpha z$ mapping the unit disk $\mathfrak{E}$ onto the given disk $\mathfrak{D}$ and mapping the real axis of $\mathfrak{E}$ to the given line l of $\mathfrak{D}$. Then the reflection in l is the transformation

$$\rho_l := \alpha(\ )^{\mathrm{c}}\alpha^{-1},$$

where $(\ )^{\mathrm{c}}$ denotes complex conjugation. Note that the resulting transformation of $\mathfrak{D}$ is independent of the choice of α. A different choice of α would be of the form $\alpha\beta$ where β is a matrix in $\mathrm{SL}(\mathfrak{E})$ leaving invariant the real axis. Thus β has real entries, and consequently β commutes with complex conjugation $(\ )^{\mathrm{c}}$.

(3.14) Lemma. (1) *Let $z \mapsto \sigma z$ be a rotation in $\mathfrak{D}$ with fixed point w. Then every line t through w is mapped to a line through w, and the angle from t to σt is a constant, given by the formula,*

$$\cos \angle(t, \sigma t) = \mathrm{h}(\sigma).$$

Moreover, if t and t' are oriented lines through a given point w, then there is a unique rotation around w that maps t to t'. (2) *Let $z \mapsto \sigma z$ be a translation in $\mathfrak{D}$ with axis l. Then every point z in l is mapped to a point in l, and the distance from z to σz is a constant, given by the formula,*

$$\cosh \operatorname{dist}(z, \sigma z) = \mathrm{h}(\sigma).$$

Moreover, if z and z' are given points on a line l, then there is a unique translation with axis l that maps z to z'.

Proof. (1) After conjugation, we may assume that $\mathfrak{D}$ is the unit disk $\mathfrak{E}$ and that $w = 0$. Then σ is a matrix,

$$\sigma = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix},$$

and the associated Möbius transformation is the ordinary rotation $z \mapsto e^{2i\theta}z$. Hence the first and the last assertions of (1) hold, and the angle from t to σt is equal to 2θ. Moreover, $\mathrm{h}(\sigma) = \frac{1}{2}\mathrm{tr}(\sigma^2) = \cos(2\theta)$, and hence the formula of (1) holds.

(2) After conjugation, we may assume that $\mathfrak{D}$ is the upper half plane $\mathfrak{H}$ and that the axis l is the imaginary axis. Then σ is a matrix,

$$\sigma = \begin{bmatrix} r & 0 \\ 0 & r^{-1} \end{bmatrix},$$

and the associated Möbius transformation is the multiplication $z \mapsto r^2 z$. Hence a point iy on the line l is mapped to the point $ir^2 y$ on l, and the distance from iy to $ir^2 y$ is, by (3.8.1), equal to $|\log(r^2 y/y)| = |\log r^2|$. Hence the first and the last assertions of (2) hold. Moreover, $\mathrm{h}(\sigma) = \frac{1}{2}\mathrm{tr}(\sigma^2) = (r^2 + r^{-2})/2$, and hence the formula of (2) holds. □

(3.15) Definition. For a point P in $\mathfrak{D}$, denote by δ_P the *half turn* around P, that is, the non-euclidean rotation with angle π around P. For a line l in $\mathfrak{D}$, denote by ρ_l the reflection in l (Note that ρ_l is not a Möbius transformation, it is an anti-analytic automorphism of $\overline{\mathbf{C}}$). Moreover, for two different points P and Q, denote by τ_{QP} the translation along the line through P and Q that maps P to Q. Finally, for two different oriented lines l and m that intersect at a point P, denote by δ_{ml} the rotation around P that maps l to m.

(3.16) Lemma. (1) *If P and Q are different points of $\mathfrak{D}$, then the we have the equation,*

$$\tau_{QP}^2 = \delta_Q \delta_P. \tag{3.16.1}$$

(2) *If l and m are different oriented lines that intersect at a point P in $\mathfrak{D}$, then we have the equation,*

$$\delta_{ml}^2 = \rho_m \rho_l. \tag{3.16.2}$$

(3) *If m and l are two different lines with no point of intersection in $\mathfrak{D}$ or in $\partial\mathfrak{D}$, then we have the equation,*

$$\tau_{QP}^2 = \rho_m \rho_l, \tag{3.16.3}$$

where P and Q are the unique points of l and m such that the line through P and Q is orthogonal to l and m.

Proof. (1) Let l be the oriented line from P to Q and let u and v be its limit points. Clearly u and v are interchanged by any of the half turns δ_P and δ_Q. Therefore u and v are fixed points for the composition $\delta_Q\delta_P$. Hence the composition is a translation with axis l. Obviously the square τ_{QP}^2 is a translation with axis l. Hence, to prove Equation (3.16.1), it suffices to show that there is one point at which the two sides of the equation takes the same value. Let R be the point on the line l for which the point Q is the midpoint of the line segment PR. Clearly, the point Q is mapped to R by the translation τ_{QP}, and P is mapped to R by the half turn δ_Q. It follows that the point P is mapped to R by any of the two sides of the equation (3.16.1). Therefore, the equation holds.

(2) The composition $\rho_m\rho_l$ of anti-transformations is a Möbius transformation, and it has obviously the point P as fixed point. Hence the composition is a rotation around P. Obviously, under the composition, the line l is mapped to the line $l' = \rho_m l$ through P for which the

angle from l to m is equal to the angle from m to l'. Clearly, the latter line is also the image of l under the composition δ_{ml}^2. Therefore, equation (3.16.2) holds.

(3) As in (2), the composition $\rho_m\rho_l$ is a Möbius transformation. Let n be the line through P and Q, and let u and v be its limit points. As n is orthogonal to the lines l and m, the limit points u and v are interchanged by any of the reflections ρ_m and ρ_l. Hence u and v are fixed points of the composition $\rho_m\rho_l$. Therefore, the composition is a translation with axis n. Moreover, the point P is, by any of the two sides of equation (3.16.3), mapped to the point R on n for which Q is the midpoint of the line segment PR. Therefore, the equation holds. □

(3.17) Corollary. *In the setup of* (3.9)*, the following equations hold:*

$$\tau_{AC}^2\tau_{CB}^2\tau_{BA}^2 = 1, \tag{3.17.1}$$

$$\delta_{ac}^2\delta_{cb}^2\delta_{ba}^2 = 1. \tag{3.17.2}$$

In addition, consider the line h through A orthogonal to a, let H be the point of intersection of h and a, and let a' be the line through A orthogonal to h. Then,

$$\tau_{AH}^2 = \delta_{a'c}^2\delta_{ca}^2. \tag{3.17.3}$$

Proof. Note that, depending on a choice of orientations of the sides a and c, there are two rotations δ_{ac}. If one is by the angle θ, then the other is by the angle $\theta-\pi$. Hence the square δ_{ac}^2 is well defined.

Obviously, the first formula follows from (3.16.1) since the half turns are involutions. Similarly, the second formula follows from (3.16.2), and the third from (3.16.2) and (3.16.3). □

(3.18) Lemma. *Consider two matrices σ and τ in* $\mathrm{SL}(\mathfrak{D})$*. Assume that τ is a translation in $\mathfrak{D}$, and assume that σ is either a rotation around a point on the axis of τ or a translation along an axis orthogonal to the axis of τ. Then,*

$$2\,\mathrm{tr}(\sigma\tau) = \mathrm{tr}(\sigma)\mathrm{tr}(\tau). \tag{3.18.1}$$

Proof. By conjugation we may assume that $\mathfrak{D}$ is the upper half plane $\mathfrak{H}$, that the axis of τ is the imaginary axis $i\overline{\mathbf{R}}$, and that either σ is a rotation around the point i or a translation along the line through i orthogonal to the imaginary axis.

Now τ and σ belong to $\mathrm{SL}_2(\mathbf{R})$. The matrix τ fixes 0 and ∞; hence it is of the form,

$$\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \quad \text{where } ad = 1.$$

The matrix of σ is either a rotation or a translation. Accordingly, by (2.4), it is of one of the following two forms,

$$\begin{bmatrix} c & b \\ -b & c \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} c & b \\ b & c \end{bmatrix},$$

where $c^2+b^2=1$ or $c^2-b^2=1$. Clearly, in both cases the left hand side of (3.18.1) is equal to $2c(a+d)$ and hence equal to the right hand side. □

(3.19) Proposition. *In the setup of* (3.9)*, assume for the triangle ABC that the angle C is a right angle. Then,*

$$\cosh c = \cosh a \cosh b, \qquad \cos B = \cosh b \sin A. \tag{3.19.1}$$

Proof. Clearly, the two equations are equivalent to the equations of (3.10.1). To give an alternative proof, note that the first equation of (3.17) implies the following:

$$\tau_{AB}^2 = \tau_{AC}^2 \tau_{CB}^2.$$

Moreover, as the angle at C is a right angle, Lemma (3.18) applies with $\sigma := \tau_{AC}^2$ and $\tau := \tau_{CB}^2$. As a consequence, we obtain the equation,

$$2\,\mathrm{tr}(\tau_{AB}^2) = \mathrm{tr}(\tau_{AC}^2)\mathrm{tr}(\tau_{CB}^2).$$

Dividing by 4 we obtain an equation for the invariants $\mathrm{h}(\sigma)$ which, by Lemma (3.14)(2) is the first asserted equation (3.19.1).

To prove the second equation, consider the last equation in (3.17). As the angle at C is a right angle, we have that $H = C$. Thus we obtain the equation,

$$\delta_{ca'}^2 \tau_{AC}^2 = \delta_{ca}^2.$$

Again Lemma (3.18) applies, and we obtain the equation,

$$2\,\mathrm{tr}(\delta_{ca}^2) = \mathrm{tr}(\delta_{ca'}^2)\mathrm{tr}(\tau_{AC}^2).$$

Divide by 4 to obtain an equation for the invariants $\mathrm{h}(\sigma)$. By construction of a', the angle from c to a' is equal to $\pi/2 - A$. Hence, by Lemma (3.14)(1) and (2), the equation obtained is the second equation of (3.19.1). ∎

(3.20) Exercise. Consider two lines l and l' of $\mathfrak{D}$, with limit points u, v and u', v'. Assume that the four points u, v, u', v' on the boundary of $\mathfrak{D}$ are different. In addition, assume that v and v' belong to the same of the two arcs in which the boundary $\partial\mathfrak{D}$ is divided by u, u'. Prove that the cross ratio $\mathrm{df}(u, v, u', v')$ is negative, if and only if the lines l and l' intersect. If the cross ratio is negative, then the angle θ between the lines l and l' is determined by the formula,

$$\tan^2(\theta/2) = -\,\mathrm{df}(u, v, u', v').$$

If the cross ratio is positive, then the (non-euclidean) distance θ between the lines is determined by the formula,

$$\tanh^2(\theta/2) = \mathrm{df}(u, v, u', v').$$

4. Proper actions.

(4.1) Proposition. *Let $G := \mathrm{SL}(\mathfrak{D})$ be the stabilizer of a disk $\mathfrak{D}$ in $\overline{\mathbf{C}}$. Then the action of G on $\mathfrak{D}$ is proper, that is, the map $(\alpha, w) \mapsto (\alpha w, w)$ is a proper map,*

$$G \times \mathfrak{D} \to \mathfrak{D} \times \mathfrak{D}. \tag{4.1.1}$$

Proof. Recall that a continuous map between locally compact (Hausdorff) spaces is a *proper map* if the preimage of any compact subset of the target is a compact subset of the source.

The map (4.1.1) is an obvious composition of two maps,

$$G \times \mathfrak{D} \to \mathrm{P}G \times \mathfrak{D} \to \mathfrak{D} \times \mathfrak{D}. \tag{1}$$

Of the two maps, the first is proper, because $G \to \mathrm{P}G$ is the homomorphism with kernel ± 1. Thus is suffices to show that the second map is proper.

Fix a point u on the boundary $\partial\mathfrak{D}$. Then there is a map,

$$\mathrm{P}G \times \mathfrak{D} \to \partial\mathfrak{D} \times \mathfrak{D} \times \mathfrak{D},$$

defined by

$$(\alpha, w) \mapsto (\alpha u, \alpha w, w).$$

It is bijective by Corollary (1.9), and obviously continuous. It is in fact a homeomorphism, that is, the inverse map is continuous. In other words, if we consider for $(u', w', w) \in \partial\mathfrak{D} \times \mathfrak{D} \times \mathfrak{D}$ the unique Möbius transformation $z \mapsto \alpha z$ under which $(\mathfrak{D}, w, u)$ is mapped to $(\mathfrak{D}, w', u')$, then the Möbius transformation depends continuously on (u', w', w). The latter fact is easily proved using the explicit determination of the Möbius transformation given in Section 1.

Clearly, when the source of the second map in (1) is replaced by the homeomorphic space $\partial\mathfrak{D} \times \mathfrak{D} \times \mathfrak{D}$, then the map is simply the projection,

$$\partial\mathfrak{D} \times \mathfrak{D} \times \mathfrak{D} \to \mathfrak{D} \times \mathfrak{D}.$$

Hence the map is proper, because $\partial\mathfrak{D}$ is compact.

Thus the Proposition has been proved. □

(4.2) Corollary. *Let K_1 and K_2 be compact subsets of the disk $\mathfrak{D}$. Then the following subset of $G := \mathrm{SL}(\mathfrak{D})$ is compact:*

$$\{\alpha \mid K_1 \cap \alpha K_2 \neq \emptyset\}.$$

Proof. The product set $K_1 \times K_2$ is a compact subset of $\mathfrak{D} \times \mathfrak{D}$. Hence its preimage under the map in (4.1) is a compact subset of $G \times \mathfrak{D}$. Clearly, the preimage is the following set:

$$\{(\alpha, z) \mid \alpha z \in K_1 \text{ and } z \in K_2\}.$$

Hence the latter set is compact. Consequently, its image in G under the projection $G \times \mathfrak{D} \to G$ is a compact subset of G. The latter image is the set in question. Hence it is compact. □

(4.3) Corollary. *For any point $z \in \mathfrak{D}$, the map $\alpha \mapsto \alpha z$ is a proper map,*

$$G \to \mathfrak{D}.$$

In particular, the isotropy group G_z is a compact subgroup of G.

Proof. The first assertion of the Corollary is the special case of (4.2) obtained by taking K_1 arbitrary compact and $K_2 = \{z\}$. The second assertion is a consequence, since the isotropy group is the preimage of the compact set $\{z\}$ under the map $G \to \mathfrak{D}$. ▯

(4.4) Note. In Section 2 we proved that the isotropy group G_z is in fact conjugated to the compact group $\mathrm{SO}_2(\mathbf{R})$, and in particular, the isotropy group is isomorphic to the unit circle $\mathrm{U}_1(\mathbf{C})$. Note also that we proved in Section 2 that the isotropy group $\mathrm{SL}(\mathfrak{D})_u$ of a point u on the boundary $\partial\mathfrak{D}$ is non-compact. In particular, the action of $\mathrm{SL}(\mathfrak{D})$ on the closure of $\mathfrak{D}$ is not proper.

(4.5) Exercise. Several topological facts were used in the proof of Proposition (4.1). The Riemann sphere $\overline{\mathbf{C}}$ is the one point compactification of $\mathbf{C}$. Hence the disk $\mathfrak{D}$ as an open subset of $\overline{\mathbf{C}}$ is locally compact. The surjection of (1.1), $(\mathbf{C}^2)^* \to \overline{\mathbf{C}}$, is continuous and open. As a consequence, $\overline{\mathbf{C}}$ is equal to the quotient $(\mathbf{C}^2)^*/\mathbf{C}^*$.

The group $\mathrm{SL}_2(\mathbf{C})$ is locally compact: it is a closed subset of the space $\mathbf{C}^4$ of 2×2 matrices. It is a topological group. The group $\mathrm{SL}(\mathfrak{D})$ is a closed subgroup, since it is conjugated to $\mathrm{SL}(\mathfrak{H}) = \mathrm{SL}_2(\mathbf{R})$.

The group $\mathrm{PSL}(\mathfrak{D})$ has two topologies: the quotient topology induced by the surjection $\mathrm{SL}(\mathfrak{D}) \to \mathrm{PSL}(\mathfrak{D})$ and the subset topology induced by the inclusion $\mathrm{PSL}(\mathfrak{D}) \hookrightarrow \mathrm{Aut}_{\mathrm{cont}}(\mathfrak{D})$ (where the automorphism group is given the compact–open topology). The two topologies are equal, and $\mathrm{PSL}(\mathfrak{D})$ is a locally compact topological group.

Prove these topological facts. In addition, prove (or find a reference for) the following general facts: If K is compact and X is locally compact, then the projection $K \times X \to X$ is proper. If K is a compact subgroup of a locally compact group G, then the canonical map $G \to G/K$ is proper.

Discrete subgroups

1. Isotropy groups and their generators.

(1.1) Setup. Fix a disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$. Recall that a subset Γ of a topological space is said to be discrete in the given topological space, if it is a closed subset and the induced topology is discrete. Equivalently, when the topological space is locally compact, the condition is that the intersection of Γ and any compact subset is finite. In particular, a subgroup Γ of $\mathrm{SL}_2(\mathbf{C})$ is discrete, if and only if, for every positive real number R, there is only a finite number of matrices in Γ for which all four entries have modulus at most equal to R.

Recall that a point u is called Γ-parabolic if it is a fixed point of some parabolic matrix in Γ. The set of Γ-parabolic points is denoted $\partial_\Gamma\mathfrak{D}$. It is a subset of the boundary $\partial\mathfrak{D}$. A point is Γ-elliptic if it belongs to $\mathfrak{D}$ and is fixed under some non-trivial (necessarily elliptic) matrix of Γ. A Γ-ordinary point is a point of $\mathfrak{D}$ which is not Γ-elliptic.

(1.2) Lemma. (1) *Assume that Γ is a discrete subgroup of $\mathrm{SL}(\mathfrak{E})$. Then the isotropy group Γ_0 of the point 0 in $\mathfrak{E}$ is a finite cyclic group generated by the matrix,*

$$d_{2\pi/N} := \begin{bmatrix} e^{2\pi i/N} & 0 \\ 0 & e^{-2\pi i/N} \end{bmatrix},$$

where $N = |\Gamma_0|$ is the order of the isotropy group. In particular, the point 0 is Γ-elliptic, if and only if $|\Gamma_0| > 2$. The group Γ is homogeneous, if and only if N is even. The group $\mathrm{P}\Gamma_0$ of associated Möbius transformations is cyclic, generated by the rotation $z \mapsto e^{2\pi i/d}z$, where $d = |\mathrm{P}\Gamma_0|$.

(2) *Assume that Γ is a discrete subgroup of $\mathrm{SL}(\mathfrak{H})$. Assume that the point ∞ on the boundary of $\mathfrak{H}$ is Γ-parabolic. Then the isotropy group Γ_∞ is infinite, and all its matrices different from ± 1 are parabolic. The isotropy group it is either cyclic or dicyclic (that is, isomorphic to $\{\pm 1\} \times \mathbf{Z}$). The group Γ is homogeneous, if and only if the isotropy group Γ_∞ is dicyclic. In the dicyclic case, the isotropy group consists of all matrices of the form $\pm t_h^n$ (for a unique $h > 0$) where t_h is the matrix,*

$$t_h := \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}.$$

In the cyclic case, Γ_∞ consists either of all powers t_h^n [the 'regular' case] or of all powers $(-t_h)^n$ [the 'irregular' case] (for a unique $h > 0$). In all cases, the group $\mathrm{P}\Gamma_\infty$ of associated Möbius transformations is the infinite cyclic group generated by the translation $z \mapsto z + h$.

Proof. Assertion (1) follows immediately from Lemma (Möb.2.3)(2) since Γ_0 is a discrete subgroup of $\mathrm{SL}(\mathfrak{E})_0$, and $\mathrm{SL}(\mathfrak{E})_0$ is isomorphic to the unit circle $U_1(\mathbf{C})$.

To prove Assertion (2), note that the matrices of Γ_∞ are of the form (Möb.2.3.1). Therefore, since the point ∞ is assumed to be Γ-parabolic, there exists a matrix of the form $\pm t_h$ in Γ. Moreover, since Γ is discrete, we can choose in Γ_∞ a matrix $\pm t_h$ with $h > 0$ minimal. Then, again using that Γ is discrete, it follows from the relation (Möb.2.3.2) that no matrix (Möb.2.3.1) with $a \neq \pm 1$ can belong to Γ_∞. Thus all matrices different from ± 1 in Γ_∞ are parabolic. Again, since Γ is discrete and h is minimal, it follows that all matrices of Γ_∞ are, up to multiplication by ± 1, powers of t_h. The remaining assertions of (2) follow easily. □

(1.3) Definition. In (1), the generator $z \mapsto e^{2\pi i/d} z$ of $\mathrm{P}\Gamma_0$ is the Möbius transformation associated to the following matrix:

$$\gamma_0 := \begin{bmatrix} -e^{\pi i/d} & 0 \\ 0 & -e^{-\pi i/d} \end{bmatrix}. \tag{1.3.1}$$

The matrix γ_0 belongs to Γ_0. Indeed, assume first that N is even. Then $-1 \in \Gamma_0$ and $d = N/2$. As $\gamma_0 = -d_{2\pi i/N}$, it follows that $\gamma_0 \in \Gamma_0$. Assume next that N is odd. Then $d = N$. As $\gamma_0 = d_{2\pi i/N}^{(N+1)/2}$, it follows again that $\gamma_0 \in \Gamma_0$.

The matrix γ_0 is called the *canonical generator* at the point 0. It generates the group Γ_0 if N is odd or if $N \equiv 0 \pmod 4$. If $N \equiv 2 \pmod 4$, then γ_0 generates an inhomogeneous subgroup of index 2 in Γ_0.

case, and the matrix $\gamma_\infty := t_h$ in the other cases, will be called the *canonical generator* at the point ∞. The transformation associated to the canonical generator is map $z \mapsto z + h$, and h is the minimal possible step length of the transformations in $\mathrm{P}\Gamma_\infty$. The canonical generator γ_∞ generates a cyclic inhomogeneous subgroup of Γ_∞ that is mapped isomorphically onto $\mathrm{P}\Gamma_\infty$.

(1.4) Definition. For a general disk $\mathfrak{D}$ and a point u in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ we obtain, by conjugation, assertions corresponding to those of Lemma (1.2).

If u is in $\mathfrak{D}$, choose a conjugation $\alpha\colon \mathfrak{D} \to \mathfrak{E}$ such that $\alpha u = 0$. Then Γ is conjugate to the discrete subgroup $\alpha\Gamma\alpha^{-1}$ of $\mathrm{SL}(\mathfrak{E})$, and u is conjugate to 0. Hence the isotropy group Γ_u is conjugate to the isotropy group $(\alpha\Gamma\alpha^{-1})_0$. It follows from (1.2)(1) that Γ_u is a finite cyclic group, and hence its quotient $\mathrm{P}\Gamma_u$ is a finite cyclic group. The quotient $\mathrm{P}\Gamma_u$ is non-trivial if and only if u is Γ-elliptic. In any case, the order $e_u := |\mathrm{P}\Gamma_u|$ is called the *order* of u with respect to Γ. The isotropy groups for the points in the orbit Γu are conjugate. In particular, all points in the orbit Γu have the same order. Moreover, if u is Γ-elliptic, then all points in the orbit Γu are Γ-elliptic. In this case, the orbit is said to be a Γ-*elliptic orbit*.

If u is in $\partial_\Gamma \mathfrak{D}$, choose a conjugation $\alpha\colon \mathfrak{D} \to \mathfrak{H}$ such that $\alpha u = \infty$. It follows similarly that the isotropy group Γ_u is infinite cyclic or dicyclic, and that its quotient $\mathrm{P}\Gamma_u$ is an infinite cyclic group. Moreover, all non-trivial matrices in Γ_u are parabolic. Clearly, all points in the orbit Γu are Γ-parabolic. The orbit is said to be a Γ-*parabolic orbit* or to be a *cusp* for Γ.

For $u \in \mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, define the *canonical generator* of Γ at u as the following matrix $\gamma_u \in \Gamma_u$: If u is in $\mathfrak{D}$, choose a conjugation $\alpha\colon (\mathfrak{D}, u) \to (\mathfrak{E}, 0)$. Then Γ is conjugate to the discrete subgroup $\alpha\Gamma\alpha^{-1}$ of $\mathrm{SL}(\mathfrak{E})$, and u is conjugate to 0. Define γ_u as the conjugate of the canonical generator γ_0 in (1.3). Similarly, if u is in $\partial_\Gamma\mathfrak{D}$, choose a conjugation $\alpha\colon (\mathfrak{D}, u) \to (\mathfrak{H}, \infty)$, and define γ_u as the conjugate of γ_∞.

In both cases, it has to be verified that the matrix obtained is independent of the choice of α. However, two different choices α_1 and α_2 differ by a matrix $\alpha := \alpha_1\alpha_2^{-1}$ that stabilizes the image disk and the image point $\alpha_1 u = \alpha_2 u$. We may assume that α has determinant 1. Thus is has to be proved, in the two cases $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$ and $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$ respectively, that if $\alpha \in \mathrm{SL}(\mathfrak{D})_u$, then, for the canonical generators considered in (1.3), conjugation by α maps the canonical generator of $(\alpha\Gamma\alpha^{-1})_u$ onto the canonical generator of Γ_u.

The latter assertion is obvious for $(\mathfrak{E}, 0)$, since the isotropy $\mathrm{SL}(\mathfrak{E})_0$ is commutative and hence conjugation by α is in fact the identity. For $(\mathfrak{H}, \infty)$, the assertion follows from Equation (Möb.2.3.2) describing conjugation of a canonical generator under a matrix in $\mathrm{SL}(\mathfrak{H})_\infty$.

In the parabolic case, the cusp represented by u is said to be a *regular cusp* if the canonical generator γ_u is conjugate to t_h and to be an *irregular cusp* if γ_u is conjugate to $-t_h$. Note that the width h of t_h is not canonical. A different choice of conjugation changes t_h to t_{a^2h}, see (Möb.2.3.2).

(1.5) Example. The Cayley transformation $z \mapsto (z-i)/(z+i)$ of (Möb.1.5) is associated to the matrix,

$$\alpha := \frac{1}{1+i}\begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}.$$

The Cayley transformation maps $(\mathfrak{H}, i)$ onto $(\mathfrak{E}, 0)$. Consequently, under conjugation by α, the stabilizer group $\mathrm{SL}(\mathfrak{E}) = \mathrm{SU}_{1,1}(\mathbf{C})$ is mapped onto the stabilizer group $\mathrm{SL}(\mathfrak{H}) = \mathrm{SL}_2(\mathbf{R})$, and the isotropy group $\mathrm{SL}(\mathfrak{E})_0 = \mathrm{U}_1(\mathbf{C})$ is mapped onto the isotropy group $\mathrm{SL}(\mathfrak{H})_i = \mathrm{SO}_2(\mathbf{R})$. Under the conjugation, the following matrices correspond:

$$\begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \text{ and } \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Consider the discrete subgroup $\Gamma := \mathrm{SL}_2(\mathbf{Z})$ of $\mathrm{SL}(\mathfrak{H})$. It is homogeneous. Clearly, the point ∞ is Γ-parabolic, and the canonical generator at ∞ is the matrix,

$$t := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{1.5.1}$$

The point $i = e^{2\pi i/4}$ is elliptic: the isotropy group Γ_i is of order 4 and the canonical generator at i is the matrix of order 4,

$$s := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{1.5.2}$$

Indeed, the isotropy group Γ_i consists of the matrices with integer entries in $\mathrm{SO}_2(\mathbf{R})$, and clearly, the latter matrices are the four matrices ± 1, $\pm s$. It follows from the correspondence above that $-s$ is conjugate to the matrix $d_{2\pi/4}$. Hence s is the canonical generator.

The point $\rho = e^{2\pi i/3}$ is elliptic: the isotropy group Γ_ρ is of order 6 and the canonical generator at ρ is the matrix of order 3,

$$u = t^{-1}s = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}. \tag{1.5.3}$$

Indeed, since $\rho^2 + \rho + 1 = 0$, it is easy to describe the isotropy group $\mathrm{SL}(\mathfrak{H})_\rho$. It follows easily that the isotropy group Γ_ρ consists of the following 6 matrices ± 1, $\pm u$, $\pm u^2$. The matrix u is of order 3. Hence Γ_ρ is generated by the two matrices $-u$ and $-u^2$ of order 6. The Möbius transformation associated to $\pm u$ is the rotation by the angle $2\pi/3$ since the half ray from ρ to ∞ is mapped to the half ray from ρ to -1. Hence $-u$ is conjugate to the matrix $d_{2\pi/6}$, and u is the canonical generator.

(1.6) Proposition. *Assume that* Γ *is the discrete subgroup* $\mathrm{SL}_2(\mathbf{Z})$ *of* $\mathrm{SL}(\mathfrak{H})$. *Let F be the closed subset of $\mathfrak{H}$ defined by the inequalities,*

$$F:\ |z| \geq 1, \quad -\frac{1}{2} \leq \Re z \leq \frac{1}{2}\,.$$

Denote by F_0 the subset obtained from F by omitting from the boundary the points z where either $\Re z = 1/2$ or $|z| = 1$, $\Re z > 0$. Then the set F_0 is a complete set of representatives for the set of orbits $\mathfrak{H}/\Gamma$. Moreover, the points i and ρ are the only Γ-elliptic points in F_0. Finally, there is only one cusp, namely the orbit $\Gamma\infty$ consisting of ∞ and the rational numbers.

Proof. In addition to the subsets F and F_0, denote by G the vertical strip of $\mathfrak{H}$ defined by the inequalities: $-\frac{1}{2} \leq \Re z < \frac{1}{2}$. We proceed in a series of steps.

Step 1. For any point z of $\mathfrak{H}$ there is only a finite number of matrices σ in Γ,

$$\sigma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{1}$$

such that

$$\Im(\sigma z) \geq \Im z, \quad \sigma z \in G. \tag{2}$$

Indeed, for a fixed pair of integers (c, d), consider the matrices (1). They correspond to the integer solutions (a, b) of the equation $ad - bc = 1$. Hence, for the fixed pair of prime integers (c, d), there are matrices σ of the form (1), and if σ_0 is any such matrix, then the matrices of the form (1) are exactly the matrices $t^k\sigma_0$ for $k \in \mathbf{Z}$. The matrix t defines the translation $tz = z + 1$ and the vertical strip G is of width 1. Therefore, when (c, d) is fixed, there is only one matrix σ of the form (1) such that $\sigma z \in G$. Now, for any matrix (1) in $\mathrm{SL}_2(\mathbf{R})$, we have that $\Im(\sigma z) = (\Im z)/|cz + d|^2$. Hence the inequality in (2) is equivalent to the following:

$$|cz + d|^2 \leq 1. \tag{3}$$

The latter inequality has only a finite number of integer solutions (c, d). Therefore, the assertion of Step 1 holds.

Step 2. If $|z| < 1$, then $\Im(sz) > \Im z$. This is just a simple observation.

Step 3. Any point z in $\mathfrak{H}$ is Γ-equivalent to a point in F_0. Indeed, for any point z there is a unique point z_1 in G of the form $t^k z$. If $|z_1| < 1$, consider $z_2 := (sz_1)_1$. Define, inductively, $z_{n+1} := (sz_n)_1$ as long as $|z_n| < 1$. By construction, the points z_j belong to G, and they are Γ-equivalent to z. By Step 2 we have that $\Im z_1 < \Im z_2 < \cdots$. Therefore, by Step 1, the sequence $z_1, z_2, \dots$ is finite. It stops with a point z_n in G such that $|z_n| \geq 1$. If $|z_n| > 1$, then z_n belongs to F_0. If $|z_n| = 1$, then either z_n or sz_n belongs to F_0. Hence, in all cases, the point z is Γ-equivalent to a point of F_0.

Step 4. By Step 3, the set F_0 contains a complete set of representatives for $\mathfrak{H}/\Gamma$. Therefore, to show that F_0 is a complete set of representatives and that i and ρ are the only Γ-elliptic points in F_0, it suffices to prove the following assertion: Let $\sigma \neq \pm 1$ be a matrix in Γ, and let w and w' be points of F_0. Then the equation $\sigma w = w'$ implies that

(s) either $w' = w = i$ and $\sigma = \pm s$,
(u) or $w' = w = \rho$ and $\sigma = \pm u, \pm u^2$.

To prove the assertion, we may assume that $\Im w' \geq \Im w$. Then $1 \geq |cw + d|^2$. Hence the first of the following four inequalities holds:

$$1 \geq c^2|w|^2 + cd(w + \overline{w}) + d^2 \geq c^2 + cd(w + \overline{w}) + d^2 \geq c^2 - |cd| + d^2 \geq (|c| - |d|)^2. \quad (4)$$

The second inequality holds because $|w| \geq 1$. The third holds because $|\Re w| \leq \frac{1}{2}$, and the last is trivial. Moreover, the last inequality is strict if $cd \neq 0$.

Since c and d are integers, the two last expressions in (4) are non-negative integers. Moreover c and d are prime. Hence the inequalities leave the three possibilities: $c = 0,\ |d| = 1$, or $|c| = 1,\ d = 0$, or $|c| = |d| = 1$.

Assume first that $c = 0$, $|d| = 1$. Replacing σ by $-\sigma$ we may assume that $d = 1$. Then the matrix σ is of the form,

$$\sigma = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix},$$

and $w' = \sigma w = w + b$. As w' and w belong to F_0 by hypothesis, it follows that $b = 0$, contradicting that $\sigma \neq \pm 1$. Thus the first possibility is excluded.

Assume next that $|c| = 1$, $d = 0$. We may assume that $c = 1$. Then the matrix σ is of the form,

$$\sigma = \begin{bmatrix} a & -1 \\ 1 & 0 \end{bmatrix},$$

and $w' = \sigma w = sw + a$. Clearly, in (4) the second inequality is an equality, and consequently $|w| = 1$. As w' and w belong to F_0 by hypothesis, it follows that either $a = 0$ and $w = w' = i$, or $a = -1$ and $w' = w = \rho$. In the first case, $\sigma = s$ and hence (s) holds, in the second case, $\sigma = u$ and hence (u) holds.

Finally, assume that $|c| = |d| = 1$. We may assume that $d = 1$ and $c = \pm 1$. Then the fourth inequality in (4) is strict. Consequently, the first three inequalities are equalities. Now, the second inequality is strict unless $|w| = 1$ and the third inequality is strict unless

$\Re w = -\frac{1}{2}$ and $cd = 1$. Therefore, it follows that $w = \rho$ and that $c = d = 1$, that is, σ has the form

$$\sigma = \begin{bmatrix} a & a-1 \\ 1 & 1 \end{bmatrix}.$$

Then $w' = \sigma w = (-1)/(w+1) + a = \rho + a$. As w' is in F_0 by hypothesis, it follows that $a = 0$. Thus $w' = w = \rho$ and $\sigma = -u^2$. Hence (u) holds. Thus the proof of Step 4 is completed.

Step 5. To prove the final assertion of the Proposition, note that the point ∞ is Γ-parabolic. Clearly, the orbit $\Gamma\infty$ is equal to $\overline{\mathbf{Q}} := \mathbf{Q} \cup \{\infty\}$. Conversely, let $u \in \mathbf{R}$ be a Γ-parabolic point. Then u is a fixed point of a non-trivial matrix γ in Γ with eigenvalue ± 1. Therefore $J(\gamma, u) = \pm 1$. Hence $u \in \mathbf{Q}$, because $J(\sigma, u)$ has the form $cu + d$, with $c, d \in \mathbf{Z}$ and $c \neq 0$.

Thus all assertion of the Proposition have been proved. □

(1.7) Remark. The group $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ is the modular group. It is generated by the matrices s and t. Indeed, let Γ_1 be the subgroup generated by s and t, and fix a point w in the interior of F_0. Let γ be a matrix of Γ. It follows from the proof of (1.6) applied to $z := \gamma w$ that there is a matrix γ_1 in Γ_1 such that $\gamma_1 \gamma w$ belongs to F_0. As w and $\gamma_1 \gamma w$ both belong to F_0, it follows that $\gamma_1 \gamma = \pm 1$. Hence $\gamma = \pm \gamma_1^{-1}$ belongs to Γ_1.

(1.8) Lemma. *Let Δ be a subgroup of finite index in Γ. Then every Γ-orbit splits into at most the number $|\Gamma{:}\Delta|$ of Δ-orbits. Moreover, every Δ-elliptic point is Γ-elliptic, and a point is Δ-parabolic if and only if it is Γ-parabolic. Finally, the canonical generator of Δ at a point u is up to a sign equal to $\gamma_u^{d_u}$, where d_u is the index, $d_u := |\mathrm{P}\Gamma_u{:}\mathrm{P}\Delta_u|$.*

Proof. Set $d := |\Gamma{:}\Delta|$, and let γ_j be a system of representatives for the right cosets modulo Δ. Consider a point u of $\mathfrak{D} \cup \partial\mathfrak{D}$. Then the orbit Γu is the union, $\Gamma u = \Delta(\gamma_1 u) \cup \cdots \cup \Delta(\gamma_d u)$ of the number d of Δ-orbits. Of these, at most d are different. Hence the first assertion holds.

Clearly, a Δ-elliptic (resp. Δ-parabolic) point u is Γ-elliptic (resp. Γ-parabolic). To prove that the converse holds for a Γ-parabolic point u, note that the isotropy group Δ_u is of finite index (at most d) in Γ_u. Hence Δ_u is infinite, because Γ_u is infinite. Moreover, all nontrivial matrices in Γ_u are parabolic. Therefore Δ_u contains a parabolic matrix.

The final assertion is left to the reader. □

2. Properly discontinuous actions.

(2.1) Setup. Consider, as in Section 1, a disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$.

(2.2) Lemma. *Let K_1 and K_2 be compact subsets of the disk $\mathfrak{D}$. Then there is only a finite number of elements γ in Γ for which the following condition holds:*

$$K_1 \cap \gamma K_2 \neq \emptyset.$$

Proof. The set of elements γ for which the condition holds is the intersection of Γ and the subset of $\mathrm{SL}(\mathfrak{D})$ described in (Möb.4.2). As the latter set is compact, and Γ is discrete, the intersection is finite. □

(2.3) Remark. An action of a group on a locally compact space is said to be *properly discontinuous* if it has the property of Lemma (2.2).

(2.4) Corollary. *For every point u of $\mathfrak{D}$, the stabilizer Γ_u is a finite group and the orbit Γu is discrete in $\mathfrak{D}$.*

Proof. For every compact subset K of $\mathfrak{D}$, the condition $\gamma u \in K$ holds for only a finite number of elements γ in Γ. In particular, the intersection $\Gamma u \cap K$ is finite. Therefore the orbit Γu is discrete in $\mathfrak{D}$. The first assertion of the Corollary was proved in Section 1; alternatively, it follows from (2.2) by taking $K_1 = K_2 := \{u\}$. □

(2.5) Corollary. *For any two points u and u' in $\mathfrak{D}$, there are neighborhoods U of u and U' of u', such that for any matrix γ in Γ, if $\gamma U \cap U' \neq \emptyset$, then $\gamma u = u'$. In particular, every point u of $\mathfrak{D}$ has a neighborhood U such that for any matrix γ in Γ, if $\gamma U \cap U \neq \emptyset$, then $\gamma \in \Gamma_u$.*

Proof. Choose open neighborhoods V of u and V' of u' whose closures relative to $\mathfrak{D}$ are compact. Obviously, the condition,

$$\gamma V \cap V' \neq \emptyset, \tag{1}$$

holds for the matrices γ in Γ for which $\gamma u = u'$. By Proposition (2.2), applied to the closures of V and V', the condition (1) can hold only for a finite number of elements of Γ. In particular, there is only a finite number of elements $\gamma_1, \dots, \gamma_N$ in Γ such that (1) holds and $\gamma_i u \neq u'$. For each $i = 1, \dots, N$, the point $\gamma_i u$ is different from u'. Accordingly, there are open neighborhoods U_i of u and U_i' of u' such that $\gamma_i U_i \cap U_i' = \emptyset$. Now, clearly, the two open sets,

$$U := V \cap U_1 \cap \cdots \cap U_N, \quad U' := V' \cap U_1' \cap \cdots \cap U_N',$$

have the required properties.

When $u' = u$, the asserted special case is obtained by replacing U by $U \cap U'$. □

(2.6) Corollary. *The set of Γ-elliptic points is a discrete subset of $\mathfrak{D}$.*

Proof. Otherwise there would be a point u in $\mathfrak{D}$ which is an accumulation point of the set of Γ-elliptic points. Then every neighborhood U of u contains an infinity of Γ-elliptic points u_i. For each u_i there is an elliptic transformation γ_i in Γ with u_i as fixed point. The transformations γ_i are different, because every elliptic transformation has only one fixed point in $\mathfrak{D}$. In particular, there is an infinite number of elements γ_i in Γ for which $\gamma_i U \cap U \neq \emptyset$. Thus we have obtained a contradiction to the assertion in (2.5). □

(2.7) Lemma. *Consider the group $G = \mathrm{SL}_2(\mathbf{R})$ acting on the upper half plane $\mathfrak{H}$. Let $ll\colon G \to \mathbf{R}$ be the map defined by*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto c\,.$$

Let Δ be a discrete subgroup of G. Assume that the point ∞ is Δ-parabolic. Then the image $ll(\Delta)$ is discrete in $\mathbf{R}$.

Proof. Since the point ∞ is Δ-parabolic, there is in Δ a matrix of the form,

$$\tau = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \text{ where } h > 0.$$

We have to prove, for any given positive real number R, that the intersection $ll(\Delta)$ and $[-R, R]$ is finite.

For $0 < \varepsilon < 1$, denote by K_ε the set of all matrices δ in G for which the following inequalities hold:

$$\varepsilon \le \Im(\delta i) \le 1, \quad 0 \le \Re(\delta i) \le h. \tag{1}$$

The set K_ε is a compact subset of G. Indeed, the inequalities above, for the real and imaginary parts of complex numbers, describe a compact rectangle in $\mathfrak{H}$, and the set K_ε is the preimage of the rectangle under the map $\delta \mapsto \delta i$. As the map $\delta \mapsto \delta i$ is proper by Corollary (Möb.4.3), the preimage is compact.

The intersection $\Delta \cap K_\varepsilon$ is finite, because Δ is discrete in G and K_ε is compact. Therefore, to prove that the intersection $ll(\Delta) \cap [-R, R]$ is finite, it suffices to prove that when ε is chosen sufficiently small, then the intersection $ll(\Delta) \cap [-R, R]$ is contained in the image $ll(\Delta \cap K_\varepsilon)$. We prove that the latter inclusion holds when $\varepsilon \le 1/\big(R^2 + (1 + hR)^2\big)$.

Let c be a value in the intersection $ll(\Delta) \cap [-R, R]$. If $c = 0$, then $c = ll(1)$, and clearly the identity matrix 1 belongs to $\Delta \cap K_\varepsilon$ for any ε. Assume that $c \neq 0$. Then $0 < |c| \le R$ and there exists in Δ a matrix,

$$\delta = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{2}$$

whose lower left entry is the given value c. If δ is replaced by $\delta\tau$, the c is unchanged and the d is replaced by $d + hc$. Therefore, by replacing δ by a product $\delta\tau^n$ for a suitable n, we may assume in (2) that $1 \le d \le 1 + h|c|$. Then,

$$1 \le c^2 + d^2 \le 1/\varepsilon. \tag{3}$$

Indeed, the first inequality holds because $1 \le d$, and the second holds by the choice of ε since $d \le 1 + h|c|$.

As the matrix δ has determinant 1, the imaginary part $\Im(\delta i)$ is equal to $1/(c^2+d^2)$. Hence the first two inequalities of (1) hold by (3). When δ is replaced by $\tau\delta$, then the c and the d are unchanged, but the value δi is replaced by $\delta i + h$. Therefore, by replacing δ by $\tau^m\delta$ for a suitable m, we may assume that the last two inequalities of (1) hold. Hence δ belongs to $\Delta \cap K_\varepsilon$ and, consequently, c belongs to $ll(\Delta \cap K_\varepsilon)$.

Thus the required inclusion has been proved. $\square$

(2.8) Definition. Let u be a point on the boundary ∂D. Then, by definition, a *fundamental neighborhood* of u is the union of the point u and an open disk (strictly) contained in $\mathfrak{D}$ and tangent to $\partial\mathfrak{D}$ at the point u. The boundary of the disk, excluding the point u, is called a *horo cycle*.

Clearly, the topology on $\mathfrak{D}$ is the trace of a topology on the closed disk,

$$\mathfrak{D} \cup \partial\mathfrak{D},$$

in which a neighborhood of a point u of $\partial\mathfrak{D}$ is a subset of $\mathfrak{D} \cup \partial\mathfrak{D}$ containing a fundamental neighborhood of u. By convenience, a fundamental neighborhood of a point u in $\mathfrak{D}$ is a geodesic disk around u, that is, an open ball centered at u with respect to the non-euclidean distance on $\mathfrak{D}$.

Clearly, a Möbius transformation $z \mapsto \alpha z$ maps a fundamental neighborhood U of a point u of $\mathfrak{D} \cup \partial\mathfrak{D}$ onto a fundamental neighborhood of the image point αu in $\alpha\mathfrak{D} \cup \partial\alpha\mathfrak{D}$. In particular, any matrix in $\mathrm{SL}(\mathfrak{D})$ defines a topological automorphism of $\mathfrak{D} \cup \partial\mathfrak{D}$.

Clearly, the subset $\partial_\Gamma\mathfrak{D}$ of Γ-parabolic points is invariant under the action of Γ. Hence Γ acts on the topological space $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$. We denote by $\overline{\mathfrak{D}/\Gamma}$ the corresponding orbit space, with its quotient topology. Note that $\overline{\mathfrak{D}/\Gamma}$ contains the quotient $\mathfrak{D}/\Gamma$ as an open subspace. In addition, it contains one point for each orbit of Γ-parabolic points. A point in $\overline{\mathfrak{D}/\Gamma}$ is called *parabolic*, *elliptic*, or *ordinary* respectively, if it corresponds to an orbit of Γ-parabolic points, an orbit of Γ-elliptic points, or an orbit of Γ-ordinary points.

(2.9) Example. In the topology of the closed disk $\mathfrak{H} \cup \overline{\mathbf{R}}$, the fundamental neighborhoods of the point ∞ on the boundary are the subsets containing ∞ and a half plane $\mathfrak{H}_R : \Im z > R$ for some positive real number R. The horo cycles around ∞ are the horizontal straight lines contained in $\mathfrak{H}$.

(2.10) Lemma. *Let K be a compact subset of $\mathfrak{D}$. Then, for every Γ-parabolic point u, there exists a neighborhood U of u such that for all matrices $\gamma \in \Gamma$,*

$$\gamma U \cap K = \emptyset.$$

Proof. After a conjugation, we may assume that $\mathfrak{D}$ is the upper half plane $\mathfrak{H}$ and that $u = \infty$. From Lemma (2.7), it follows in particular that the exists a positive real number r with the property that if γ is a matrix in Γ and $|ll(\gamma)| < r$, then $ll(\gamma) = 0$. As K is a compact subset of $\mathfrak{H}$, it is contained in a horizontal strip,

$$K \subseteq \{z \in \mathfrak{H} \mid r_1 < \Im z < r_2\} \tag{1}$$

Now choose R such that $R \geq r_2$ and $R \geq 1/(r_1 r^2)$. Let U be the corresponding fundamental neighborhood, $\mathfrak{H}_R : \Im z > R$, of ∞. We have to prove for any point z of U and any matrix γ in Γ,

$$\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

that $\gamma z \notin K$.

Now, since $R \geq r_2$, the half plane U is disjoint from the strip in (1), and in particular disjoint from K. Assume first that $|c| < r$. Then $c = ll(\gamma) = 0$, and hence the transformation $z \mapsto \gamma z$ is of the form $z \mapsto z + h$. Consequently, if $z \in U$, then $\gamma z \in U$, and it follows that $\gamma z \notin K$. Assume next that $|c| \geq r$. Then,

$$\Im(\gamma z) = \frac{1}{|cz + d|^2}\Im z \leq \frac{1}{(|c|\Im z)^2}\Im z \leq \frac{1}{r^2 \Im z}.$$

If $z \in U$, then the right hand side is less than $1/(r^2 R)$, and hence, by the choice of R, the right hand side is less than r_1. Consequently, $\gamma z \notin K$. □

(2.11) Lemma. *Consider an orbit $B = \Gamma w_0$, where w_0 is a point of $\mathfrak{D}$. If U is a fundamental neighborhood of a point u in $\mathfrak{D}$, then U contains only a finite number of points in B. If U is a fundamental neighborhood of a point in $\partial_\Gamma \mathfrak{D}$, then of the horo cycles contained in U, only a finite number contain points of B.*

Proof. The orbit B is discrete by (2.4). Hence the assertion for a point $u \in \mathfrak{D}$ follows, because a fundamental neighborhood U has compact closure.

Assume that u is Γ-parabolic, and let U be any fundamental neighborhood u. After conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then U is a half plane $\mathfrak{H}_R : \Im z > R$, and the horo cycles are the straight horizontal lines. By (2.10) applied with $K = \{w_0\}$, the half plane $\mathfrak{H}_R$ for $R \gg 0$ contains no points of B. Consider the canonical generator γ_∞ of Γ_∞. It is a translation $z \mapsto z + h$. Then every point in $B \cap \mathfrak{H}_r$ is mapped by a suitable power of γ_∞ in the rectangle of $\mathfrak{H}$:

$$0 \leq \Re z \leq h, \; r \leq \Im z \leq R.$$

The rectangle is compact. Hence, by (2.4), it contains only a finite number of points of B. Thus every point of $B \cap U$ is one of the horo cycles through these finitely many points. □

(2.12) Theorem. *For any two points u_1 and u_2 in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, there are neighborhoods U_1 of u_1 and U_2 of u_2, such that for any matrix γ in Γ, if $\gamma U_1 \cap U_2 \neq \emptyset$, then $\gamma u_1 = u_2$. In particular, every point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ has a neighborhood U such that for any matrix γ in Γ, if $\gamma U \cap U \neq \emptyset$, then $\gamma \in \Gamma_u$.*

Proof. If u_1 and u_2 belong to $\mathfrak{D}$, the assertion is the content of Corollary (2.5). If u_1 is Γ-parabolic and $u_2 \in \mathfrak{D}$, the assertion follows from Lemma (2.10) by taking as U_2 any compact neighborhood of u_2.

Assume that u_1 and u_2 are Γ-parabolic. After a conjugation, we may assume that $\mathfrak{D}$ is the upper half plane $\mathfrak{H}$ and that $u_2 = \infty$. Let $\pm\tau_h$ be the canonical generator of Γ_∞. Choose in $\mathfrak{H}$ any straight horizontal line L, and on L any horizontal line segment K of length at least equal to h. Then K is a compact subset of $\mathfrak{H}$, and so Lemma (2.10) applies with $u := u_1$. Accordingly, there is a fundamental neighborhood U_1 of u_1 such that $\gamma U_1 \cap K = \emptyset$ for all γ in Γ. Choose as neighborhood U_2 of $u_2 = \infty$ any fundamental neighborhood $\mathfrak{H}_R : \Im z > R$ lying above the line L, that is, such that R is at least equal to the imaginary part of points on L.

We claim that if $\gamma U_1 \cap U_2 \neq \emptyset$, then $\gamma u_1 = \infty$. Indeed, the image γu_1 is a point on the boundary of $\mathfrak{H}$. Assume that $\gamma u_1 \neq \infty$. Then γu_1 is a point on the real axis, and γU_1 is a (usual) disk in $\mathfrak{H}$ tangent to the real axis. By assumption, the disk contains a point in U_2, that is, a point with imaginary part greater than R. Therefore, the disk γU_1 contains a point on the line L. A suitable power τ_h^n will move the latter point into the set K. Thus $\tau_h^n \gamma U_1 \cap K \neq \emptyset$, contradicting the property of U_1.

Thus the Theorem has been proved. □

(2.13) Corollary. *The quotient $\overline{\mathfrak{D}/\Gamma}$ is a Hausdorff space. Moreover, it is locally compact. In fact, for any point u in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, if U is a sufficiently small fundamental neighborhood of u, then every point different from u in U is ordinary and the image of U in $\overline{\mathfrak{D}/\Gamma}$ is topologically isomorphic to an open disk: $\{q \in \mathbf{C} : |q| < \varepsilon\}$. In particular, the sets of Γ-parabolic points and Γ-elliptic points are discrete in $\overline{\mathfrak{D}/\Gamma}$.*

Proof. It follows from Theorem (2.12) that the topology on the quotient space $\overline{\mathfrak{D}/\Gamma}$ is Hausdorff. Indeed, let u_1 and u_2 represent two given different points of the quotient space. Then u_1 and u_2 are not Γ-equivalent. Therefore, by (2.12) there are open neighborhoods U_1 of u_1 and U_2 of u_2 such that $\gamma U_1 \cap U_2 = \emptyset$ for all γ in Γ. Consequently, the images of U_1 and U_2 in the quotient are disjoint. As the quotient map is an open map, the two images are disjoint neighborhoods of the two given points.

Consider next a given point in the quotient space, represented by a point u. By (2.12), for a sufficiently small fundamental neighborhood U of u, the following condition holds:

$$\gamma U \cap U \neq \emptyset \implies \gamma \in \Gamma_u. \tag{1}$$

Except possibly for u, all the points of U are in $\mathfrak{D}$, and hence not parabolic. Moreover, no point w different from u in U can be elliptic. Indeed, if w in U is elliptic, say $w = \gamma w$ where $\gamma \neq \pm 1$, then it follows from (1) that also u is a fixed point of γ, and then $w = u$ because an elliptic Möbius transformation has its other fixed point in the disc exterior to $\mathfrak{D}$. Hence all points different from u in U are ordinary.

As the quotient map is an open map, the image of U in $\overline{\mathfrak{D}/\Gamma}$ is an open neighborhood of the given point, and topology on the image as a subset is equal to the quotient topology on the image. It follows from (1) that two points in U are equivalent under Γ if and only if they are equivalent under Γ_u. Moreover, since U is a fundamental neighborhood of u, it is stable under Γ_u. Therefore, the image of U in $\overline{\mathfrak{D}/\Gamma}$ is isomorphic to the quotient U/Γ_u.

Assume first that u is Γ-parabolic. After conjugation we may assume that $(\mathfrak{D}, u)$ is $(\mathfrak{H}, \infty)$. Then U, as a fundamental neighborhood of ∞, is a half plane $\Im z > R$. Moreover, the group $\mathrm{P}\Gamma_\infty$ is the infinite cyclic group generated by $z \mapsto z + h$. Consider the exponential,

$$q(z) := e^{2\pi i z/h}.$$

It defines a continuous map q from $\mathfrak{H}$ to the pointed unit disk $\{w : 0 < |w| < 1\}$, and the induced equivalence relation on $\mathfrak{H}$ is precisely the equivalence relation defined by the action of Γ_∞ on $\mathfrak{H}$. Moreover, the map q defines a continuous open map from U onto an open disk $\{w : |w| < \varepsilon\}$. Therefore q induces a topological isomorphism,

$$U/\Gamma_\infty \xrightarrow{\sim} \{w : |w| < \varepsilon\}.$$

Assume next that u is a point in $\mathfrak{D}$. After conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. Then U, as a fundamental neighborhood of 0 in the unit disk, is an ordinary disk: $|z| < \varepsilon$. Moreover, the group $\mathrm{P}\Gamma_0$ is a finite cyclic group of order d, generated by the map $z \mapsto \zeta z$ for some d-th root of unity ζ. Consider the d'th power map,

$$q(z) = z^d.$$

It defines a continuous map q from $\mathfrak{E}$ onto itself, and the induced equivalence relation on $\mathfrak{E}$ is precisely the equivalence relation defined by the action of Γ_0 on $\mathfrak{E}$. Moreover, the map q defines a continuous open map from the disk U onto itself. Therefore q induces a topological isomorphism,

$$U/\Gamma_0 \xrightarrow{\sim} \{w : |w| < \varepsilon\}.$$

Hence we have proved in both cases that the image of U in $\overline{\mathfrak{D}/\Gamma}$ is isomorphic to an open disk as asserted.

Hence all assertions of the Corollary have been proved. ∎

(2.14) Corollary. *If the quotient $\overline{\mathfrak{D}/\Gamma}$ is compact, then the numbers of Γ-parabolic points and Γ-elliptic points in $\overline{\mathfrak{D}/\Gamma}$ are finite. If the quotient $\mathfrak{D}/\Gamma$ is compact, then there are no Γ-parabolic points.*

Proof. Clearly, the first assertion follows from Corollary (2.13). The quotient D/Γ is an open subset of $\overline{\mathfrak{D}/\Gamma}$, and it is dense by (2.13). Therefore, the second assertion holds. ∎

3. Finite normal fundamental domains.

(3.1) Setup. Fix a disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$. A subset F of $\mathfrak{D}$ is called a *fundamental domain* for the action of Γ on $\mathfrak{D}$ if the following three conditions are satisfied:

(1) The transforms γF for γ in Γ cover $\mathfrak{D}$.
(2) The set F is the closure in $\mathfrak{D}$ of its subset U of interior points.
(3) For any $\gamma \neq \pm 1$ in Γ, the transform γU is disjoint from U.

It follows from (2) that a fundamental domain F is a closed subset of $\mathfrak{D}$ and that any neighborhood of a point in the boundary $F - U$ contains points from U and points from the complement of F. It follows from (1) that F contains a system of representatives for the orbit space $\mathfrak{D}/\Gamma$. It follows from (3) that if an orbit has a representative in the interior U, then it has no other representatives in F. However, some orbits might have several representatives belonging to the boundary $F - U$. The orbit space $\mathfrak{D}/\Gamma$ is equal to the quotient of F modulo the identification of Γ-equivalent points on the boundary.

We will show in the next section that any discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ has a fundamental domain. In this section we fix a fundamental domain F, and we consider additional conditions on F. At a minimum we assume the following condition:

(4) The boundary $F - U$ is the union of a finite number of line segments.

(3.2) Definition. The line segments forming the boundary of F will be called the *boundary segments*. Note that a boundary segment can be infinite: one or both of its end points may be limit points, that is, they may belong to the boundary $\partial\mathfrak{D}$. If two of the boundary segments lie on the same line and have points in common, we may replace the two by their union, and conversely, any boundary segment can be divided into two by an interior point. Thus we may assume that if two of the boundary segments meet, then they meet at a common end point. The end points of the boundary segments are called the *vertices* of F. The *finite vertices*, that is, the vertices in $\mathfrak{D}$, belong to F, the *infinite vertices* are limit points of F. An infinite vertex which is isolated among the limit points of F is said to be a *cusp* of F. The domain F is called a *finite domain* if all of its infinite vertices are cusps. Equivalently, F is finite if it has only a finite number of limit points.

(3.3) Observation. Take any point v of F, and a sufficiently small fundamental neighborhood V of v. If v is an interior point of F, then V is contained in U. If v is a boundary point of F, and not a vertex, then there is a boundary segment through v such that $F \cap V$ is one of the two sectors into which V is divided by the line determined by the boundary segment. Finally, if v is a vertex, then $F \cap V$ consists of one or more *angular sectors* of V, each bounded two half rays through v. In the latter case, the sum of the angles of the angular sectors is called the *angle* of F at v. In the two former cases, the angle at v is, respectively, 2π and π.

Take similarly a limit point v of F and a sufficiently small fundamental neighborhood V of v. If v is not a vertex, then (except for v) all point of V are contained in U. If v is a cusp, then $F \cap V$ is a union of finitely many cuspidal sectors each bounded by two half rays through v. Finally, if v is a vertex and not a cusp, then $F \cap V$ consists of a finite number (possibly none) of cuspidal sectors and one or two sectors bounded by a single half ray through v.

It follows from Condition (3) that if a point of F is Γ-elliptic, then it lies on the boundary of F. Moreover, if the order of the point is at least 3, then it is even a vertex of F. Similarly, if a limit point of F is Γ-parabolic, then it is a cusp of F.

(3.4) Definition. A line segment L in $\mathfrak{D}$ is called a *side* of F if it is the intersection of a boundary segment of F and a boundary segment of γF for some $\gamma \neq \pm 1$ in Γ. The points of L are then common boundary points of F and γF, and each end point of L is a vertex of either F or γF. In particular, an end point of a side is Γ-equivalent to a vertex of F. Clearly, the transformation γ is unique. The inverse γ^{-1} is called the *boundary transformation* corresponding to the side L, and it is denoted γ_L. It maps the side L onto a side $\gamma_L L$ of F.

By Condition (3), the intersection $F \cap \gamma F$ is contained in the boundary of F and in the boundary of γF. Each of the two boundaries is a finite union of boundary segments. It follows that the intersection is a union of a finite number of points and a finite number of sides of F.

(3.5) Example. Consider the subset $F = F(1)$ of $\mathfrak{H}$ introduced in Proposition (1.6). It follows from the Proposition that F is a finite fundamental domain for the action of $\mathrm{SL}_2(\mathbf{Z})$ on $\mathfrak{H}$. The boundary of F is the union of 3 line segments, namely the segment from ∞ to ρ, the segment from ρ to $\rho + 1$, and the segment from $\rho + 1$ to ∞. The finite vertices are ρ and $\rho + 1$, and ∞ is the infinite vertex. The angles F at ρ and $\rho + 1$ are equal to $2\pi/6$. Moreover, the vertex ∞ is a cusp of F. Clearly, the three boundary segments are sides of F. The corresponding boundary transformations are $t\colon z \mapsto z + 1$, $s\colon z \mapsto -1/z$, and $t^{-1}\colon z \mapsto z - 1$.

(3.6) Proposition. *For any point w of $\mathfrak{D}$ there is only a finite number of transforms γF such that $w \in \gamma F$. In particular, in F there is only a finite number of points that are Γ-equivalent to w. Moreover, F has only a finite number of sides.*

Proof. Assume that w belongs to the n transforms $\gamma_1 F, \dots, \gamma_n F$, and consider the union,

$$\gamma_1 F \cup \cdots \cup \gamma_n F. \tag{3.6.1}$$

Each transform $\gamma_i F$ is, in a sufficiently small fundamental neighborhood of w, a union of angular sectors, and the angle of $\gamma_i F$ at w is equal to the angle of F at $\gamma_i^{-1} w$. The interiors of different transforms are disjoint. Hence, in a sufficiently small fundamental neighborhood of w, the union (3.6.1) is a union of angular sectors. The angle of the union is at most 2π and it is equal to the sum of the angles of F at the points $\gamma_i^{-1} w$. The latter angles are bounded away from 0, since the angle is at least equal to π except possibly at the finitely many vertices of F. Therefore, there is an upper limit for the number n. Thus there is only a finite number of transforms $\gamma_j F$ containing w.

If $v \in F$ is Γ-equivalent to w, say $w = \gamma v$, then γF is among the $\gamma_j F$, and hence the transformation γ is equal to one of the γ_j. In particular, v is one of the finitely many points $\gamma_j^{-1} w$. Hence, only a finite number of points of F are equivalent to w.

An end point v of a side of F is Γ-equivalent to a vertex of F. If v is a limit point of F, then v is an infinite vertex of F. If v is in F, then v is Γ-equivalent to a finite vertex of F and so, by what was just proved, there is only a finite number of possibilities for v. Hence there is only a finite number of end points of sides. Thus the number of sides of F is finite. □

(3.7) Observation. Consider the union (3.6.1) with n maximal, that is, the $\gamma_j F$ are all the transforms containing w. Then, clearly, the union is a neighborhood of w if and only if the sum of the angles of F at the points $\gamma_j^{-1} w$ is equal to 2π.

(3.8) Note. It follows from Proposition (3.6) that Γ-equivalence on F is finite, that is, every point v of F is Γ-equivalent to only a finite number of points in F. If v is in U, then v is Γ-equivalent only to itself. If v is a relative interior point of a side L of F, then v is equivalent to $\gamma_L v$ and to no other point different from v.

It follows also that F contains only a finite number of Γ-elliptic points v. Indeed, if v is Γ-elliptic of order at least 3, then v is one of the finitely many vertices. So assume that v is Γ-elliptic of order 2, and not a vertex, that is, assume that v is relative interior on a boundary segment B. The non-trivial transformation γ with v as fixed point is a rotation by π around u. Therefore B has a line segment in common with γF. It follows that v belongs to a side of F and that γ is the corresponding boundary transformation. There is only a finite number of sides. In particular, only a finite number of points can be fixed point of a boundary transformation. Hence there is only a finite number of possibilities for v.

(3.9) Proposition. *The following conditions on the domain F are equivalent:*

(5i) *Any compact subset K of $\mathfrak{D}$ meets only a finite number of transforms γF.*

(5ii) *For any point w of $\mathfrak{D}$, the union,*

$$\gamma_1 F \cup \cdots \cup \gamma_n F, \tag{3.9.1}$$

where the union is over the finitely many transforms $\gamma_j F$ containing w, is a neighborhood of w.

(5iii) *The boundary of F is the union of the sides of F.*

Proof. (i) $\Longrightarrow$ (ii): Let W be a fundamental neighborhood of w. The transforms γF cover $\mathfrak{D}$. In particular, they cover W. As the closure of W is compact, it follows from (i) that W is contained in a finite union of transforms $\gamma_j F$. Each transform is a closed subset of $\mathfrak{D}$. Therefore, if we omit from the transforms $\gamma_j F$ those that do no contain w, then the union of the remaining $\gamma_j F$ is still a neighborhood of w. Thus (ii) holds.

(ii) $\Longrightarrow$ (i): For any point w of $\mathfrak{D}$, the union (3.9.1) of the finitely many $\gamma_j F$ contains a fundamental neighborhood W of w. Each transform γF is the closure of its interior. Therefore, if a transform γF meets W, then it is equal to one of the $\gamma_j F$. In particular, W meets only a finite number of transforms γF. Now, let K be a compact subset of $\mathfrak{D}$. Choose for each point w of K a fundamental neighborhood W meeting only a finite number of transforms γF. Since K is compact, it follows that K meets only a finite number of transforms γF. Thus (i) holds.

(ii) $\Longrightarrow$ (iii): Take any point w on the boundary of F. It belongs to some boundary segment B of F. Consider a small fundamental neighborhood W of w. It follows from (ii) that W is contained in the union of F and a finite number of transforms $\gamma_i F$ containing w. In particular, the part of B in W is also part of the boundary of one of the transforms $\gamma_i F$. Therefore, the part of B in W is part of a side of F. Hence w belongs to a side of F. Thus (iii) holds.

(iii) $\Longrightarrow$ (ii): Consider the union (3.9.1), and its part in a sufficiently small fundamental neighborhood W of w. The part is the union of a finite number of angular sectors bounded by a finite number (possibly none) of half rays from w. By (iii), we may assume that each bounding half ray is the part of a side of a transform $\gamma_i F$. However, being part a side of a transform γF, the half ray is also contained in a second transform $\gamma' F$, contradicting the half ray bounded the part in W of the union (3.9.1). Thus there are no bounding half rays. Hence the union contains W. Thus (ii) holds.

Hence the equivalence of the three conditions has been proved. ∎

(3.10) Definition. The fundamental domain F will be called a *normal domain* if the equivalent conditions of Proposition (3.9) hold. Note that we have assumed Condition (4) for F. In general, for an arbitrary fundamental domain, the condition (5i) can be taken as the definition of normality.

If F is a normal domain, then for the transforms $\gamma_j F$ in the union (3.9.1) the sum of the angles of the transforms $\gamma_j F$ at w is equal to 2π. Equivalently, the sum of the angles of F at the points $\gamma_j^{-1} w$ is equal to 2π. The points $\gamma_j^{-1} w$ are exactly the points v of $F \cap \Gamma w$, and each v appears as $\gamma_j^{-1} w$ exactly $|\mathrm{P}\Gamma_w|$ times. Hence, for a normal domain F, the following formula holds:

$$\sum_{v \in F \cap \Gamma w} \mathrm{Angle}_v F = 2\pi / |\mathrm{P}\Gamma_w|. \tag{3.10.1}$$

Conversely, if the equation (3.10.1) holds for all points w, then the condition (5ii) holds, as observed in (3.7).

(3.11) Proposition. *Assume that F is normal fundamental domain for Γ. Then the group $\mathrm{P}\Gamma$ is generated by the boundary transformations γ_L corresponding to the sides of F.*

Proof. Let Δ be the subgroup of generated by the γ_L. It suffices to prove the equation,

$$\mathfrak{D} := \bigcup_{\delta \in \Delta} \delta F. \tag{3.11.1}$$

Indeed, assume that the equation holds. Fix a point u of the interior U of F. Let γ be a matrix in Γ. Then γu belongs to δF for some δ in Δ. As γu is an interior point of γF, it follows that γ, up to ± 1, is equal to δ.

To prove the equation (3.11.1), let $\mathfrak{D}_0$ be the union on the right hand side. Using the properties of the sides L of F, it follows easily that any point of F is an interior point of a finite union of transforms δF for $\delta \in \Delta$. As a consequence, it F meets a transform γF, then $\gamma F = \delta F$ for some δ in Δ.

It follows first that the union $\mathfrak{D}_0$ is an open subset of $\mathfrak{D}$, and next that the complement $\mathfrak{D} - \mathfrak{D}_0$ is a union of transforms $\gamma \mathfrak{D}_0$. Therefore, the complement is open too, and since $\mathfrak{D}$ is connected, it follows that $\mathfrak{D}_0 = \mathfrak{D}$. ∎

(3.12) Proposition. *Let Δ be a subgroup of finite index d in Γ. Assume for simplicity that Δ is homogeneous or that Γ is inhomogeneous. Consider a decomposition into right cosets,*

$$\Gamma = \Delta\gamma_1 \cup \cdots \cup \Delta\gamma_d.$$

Assume that F is a normal fundamental domain for Γ. Then the union of transforms,

$$G := \gamma_1 F \cup \cdots \cup \gamma_d F,$$

is a normal fundamental domain for Δ. Moreover, any side of G is of the form $\gamma_i L$ where L is a side of F, and the corresponding boundary transformation is determined as follows: according to the coset decomposition, we have $\gamma_i \gamma_L^{-1} = \delta\gamma_k$ for a unique k and a unique δ in Δ. Then $\delta = \delta_{i,L}$ is the boundary transformation corresponding to the side $\gamma_i L$.

Proof. The union of the $\gamma_i U$ is an open subset of $\mathfrak{D}$, and G is its closure. The interior V of G contains the union, and therefore G is the closure of V. The conditions (1)–(3) follow easily for G. Moreover, the boundary of G is contained in the union of transforms $\gamma_i L$ where L is a side of F. Of course, the inclusion may be strict: some side $\gamma_i L$ of $\gamma_i L$ may be a side of a different $\gamma_j F$ and hence not a part of the boundary of G. The assertions of the Proposition follow easily. ∎

(3.13) Exercise. Modify Proposition (3.12) to cover the case when Γ is homogeneous and Δ is inhomogeneous.

(3.14) Proposition. *Assume that F is a finite fundamental domain. Then F is normal, if and only if all cusps of F are Γ-parabolic. Moreover, if F is normal, then every Γ-parabolic point is Γ-equivalent to a cusp of F and the quotient $\overline{\mathfrak{D}/\Gamma}$ is compact.*

Proof. By hypothesis, there is only a finite number of limit points of F, and they are all cusps. Assume first that they are all Γ-parabolic. To prove that F is normal, we verify Condition (5i). Let K be a compact subset of $\mathfrak{D}$. Then, by Lemma (2.10), each cusp of F has a fundamental neighborhood U, such that the intersection $\gamma U \cap K$ is empty for all γ. Clearly, if we cut away from F a fundamental neighborhood of each cusp, then what remains of F is a compact subset K_1 of F. By Lemma (2.2), the intersection $\gamma K_1 \cap K$ is non-empty for only finitely many γ. Therefore, the intersection $\gamma F \cap K$ is non-empty for only finitely many γ. Thus Condition (5i) holds.

Assume conversely that F is normal. Let v be a cusp of F. We have to show that there is in Γ a parabolic matrix with v as fixed point. After a suitable conjugation, we may assume that $(\mathfrak{D}, v) = (\mathfrak{H}, \infty)$. Then a fundamental neighborhood of v is a half plane $V : \Im z > R$. When R is sufficiently big, the intersection $F \cap V$ is the union of a finite number of vertical strips, bounded by a finite number of vertical line segments. There is only a finite number of sides of F. So, enlarging R, we may, by (5iii), assume that each vertical bounding line segment is part of a side of F. In particular, the rightmost of the vertical bounding segments is also part of a side of some transform $\gamma_1 F$ different from F. In particular, $v = \gamma_1 v_1$ for some cusp v_1 of F. Repeat the argument to the cusp v of $\gamma_1 F$ and the rightmost bounding vertical segment of $\gamma_1 F$, and continue. As there are only a finite number of cusps of F, it follows that the given cusp $v = \infty$ is a fixed point of some matrix $\gamma \neq \pm 1$ of Γ.

We claim that γ is parabolic. Indeed, otherwise γ would have a second fixed point different from ∞ on the boundary of $\mathfrak{H}$. After a conjugation we may assume that the second fixed point of γ is 0. Hence, $\gamma z = rz$ for some positive real number $r \neq 1$. Replacing γ by γ^{-1} if necessary, we may assume the $r < 1$. Now, F contained a vertical strip, bounded by two vertical lines and (below) by a horizontal straight line. Clearly, when the powers γ^n are applied to the strip, the two bounding vertical lines move towards the imaginary axis and the bounding straight horizontal line moves towards the real axis. Hence, if K is any compact neighborhood of a point on the imaginary axis, then the intersection $K \cap \gamma^n F$ is non-empty for $n \gg 0$. The latter property contradicts Condition (5i). Therefore, γ is parabolic. Thus we have proved conversely that if F is normal, then every cusp of F is Γ-parabolic.

Assume that F is normal. Denote by F^* the union of F and the limit points of F. There is only a finite number of limit points. Therefore F^* is a compact subset of $\mathfrak{D} \cup \partial\mathfrak{D}$. Each cusp is Γ-parabolic, and hence F^* is contained in $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$. Consider the image of F^* in the quotient $\overline{\mathfrak{D}/\Gamma}$. As F^* is compact, the image is compact. Moreover, as F contains a system of representatives for the Γ-orbits in $\mathfrak{D}$, the image contains the open subset $\mathfrak{D}/\Gamma$. The latter subset is dense in $\overline{\mathfrak{D}/\Gamma}$ by (2.13). Hence the image of F^* is all of $\overline{\mathfrak{D}/\Gamma}$. It follows first that every point of $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$ is Γ-equivalent to a point of F^*, and next that the quotient $\overline{\mathfrak{D}/\Gamma}$ is compact. Thus the two remaining assertions of the Proposition have been proved. □

(3.15) Definition. A discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ such that the quotient $\overline{\mathfrak{D}/\Gamma}$ is compact is called a *Fuchsian group of the first kind*. It follows from Corollary (3.14) that if Γ has a finite normal fundamental domain, then Γ is of the first kind. Conversely, we prove in the next section that if Γ is of the first kind, then there exists a finite normal fundamental domain F for Γ.

(3.16) Definition. For the limit points of F there is no notion corresponding to the angle at the points of F. However, for a Γ-parabolic cusp v of F we can define the *width* of F at v as follows: In a fundamental neighborhood V of v, the intersection $F \cap V$ is a finite union of cuspidal sectors. Assume first that $(\mathfrak{D}, v) = (\mathfrak{H}, \infty)$. Then V is a half plane $V : \Im z > R$ in $\mathfrak{H}$, and the cuspidal sectors are vertical strips. The canonical generator at $v = \infty$ is a translation $z \mapsto z + h$. Define the width at ∞ as $1/h$ times the sum of the euclidean widths of the strips (where the euclidean width of a vertical strip is the euclidean distance between its bounding vertical lines). It follows easily from Condition (3) that the width is at most equal to 1.

In general, choose a conjugation $\alpha\colon (\mathfrak{D}, v) \to (\mathfrak{H}, \infty)$, and define the width of F at the Γ-parabolic point v as the width at ∞ of αF with respect to the conjugate subgroup Γ^α. A different choice of α differs from the first choice by a Möbius transformation of the form $z \mapsto rz + b$, where $r > 0$ and $b \in \mathbf{R}$. Hence, for the second choice, the widths of the vertical strips and the number h are both multiplied by r, and consequently, the quotient is unchanged. Hence the width of F at a Γ-parabolic cusp v is well defined.

If F is a finite normal domain, then following formula holds for any Γ-parabolic point w:

$$\sum_{v \in F^* \cap \Gamma w} \mathrm{width}_v F = 1, \tag{3.15.1}$$

where F^* is the union of F and the cusps of F. Indeed, consider the cusps $v_i \in F^*$ that are Γ-equivalent to w, and choose for each a matrix γ_i such that $\gamma_i v_i = w$. Then every transform γF having w is a cusp is of the form $\gamma_w^n \gamma_i F$, where γ_w is the canonical generator. Each transform $\gamma_i F$ is, in a small fundamental neighborhood V of w, a union of cuspidal sectors bounded by vertical line segments. By (5iii), we may assume that each bounding vertical line segment is part of a side of $\gamma i F$. It follows that the transform $\gamma_w^n \gamma_i F$ cover all of V. The formula (3.15.1) is an easy consequence.

Note that, using the canonical generator γ_v at a point v of F, we can normalize the angle at v: the canonical generator γ_v is a rotation around v by the angle $2\pi/|\mathrm{P}\Gamma_u|$. Define the width of F at v to be the angle of F at v divided by the rotation angle of γ_v. Then the formula (3.15.1) hold for all points w of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. In fact, for points $w \in \mathfrak{D}$, (3.15.1) is obtained from (3.10.1) by division by $2\pi/|\mathrm{P}\Gamma_w|$.

4. Canonical fundamental domains.

(4.1) Setup. Fix a disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$. Let b be a point of $\mathfrak{D}$. The corresponding *canonical domain* $F(b)$ is the subset of $\mathfrak{D}$ consisting of all points z satisfying the inequalities,

$$F(b) : \mathrm{dist}(z, b) \leq \mathrm{dist}(z, b') \text{ for all } b' \in \Gamma b. \tag{4.1.1}$$

We will show in this section that if the point b is Γ-ordinary (i.e., not Γ-elliptic), then the set $F(b)$ is a fundamental domain for Γ. In addition, the interior of $F(b)$ is the set $U(b)$ of points z for which the inequalities (4.1.1) are strict for $b' \neq b$. Moreover, we prove that $F(b)$ is a convex polygon with (in general) an infinite number of sides. If Γ is Fuchsian group of the first kind, then $F(b)$ is a finite normal fundamental domain.

For the remainder of the section we fix a Γ-ordinary orbit B, say $B = \Gamma b_0$ where b_0 is not Γ-elliptic. We consider exclusively the canonical domains $F(b)$ for points $b \in B$. Clearly, for any matrix γ in Γ, we have that $\gamma F(b) = F(\gamma b)$. Thus the domains $F(b)$ are the transforms under Γ of the domain $F := F(b_0)$ for any given point b_0 in B.

(4.2) Observation. The inequality (4.1.1), for two different points b, b' of B, defines a (non-euclidean) half plane $F_{b,b'}$ of $\mathfrak{D}$, bounded by the line,

$$L_{b,b'} : \mathrm{dist}(z, b) = \mathrm{dist}(z, b'),$$

of points of equal distance to b and b'. Each half plane $F_{b,b'}$ is closed and convex. In addition, if v is any point of $F_{b,b'}$ then the open line segment from b to v is contained in the interior of $F_{b,b'}$, that is, for points z on the open line segment, the inequality (4.4.1) is strict.

The subset $F(b)$ is the intersection of the closed half planes $F_{b,b'}$ for all $b' \neq b$ in B. Therefore, the subset $F(b)$ is closed and convex in $\mathfrak{D}$. In addition, if w belongs to $F(b)$, then the open line segment from b to w is contained in $U(b)$. The latter property holds also when w is a limit point of $F(b)$, since a point w of $\partial\mathfrak{D}$ is a limit point of $F(b)$ if and only if it is a limit point of each half plane $F_{b,b'}$.

(4.3) Definition. Fix $b_0 \in B$ and set $F := F(b_0)$. Obviously, if $b \neq b_0$, then the intersection $F \cap F(b)$ is contained in the line $L_{b_0,b}$. Moreover, the intersection is closed and convex. Hence the intersection is either empty, or a single point, or a line segment of $\mathfrak{D}$ (possibly with one or both end points on the boundary $\partial\mathfrak{D}$). By definition, a line segment L of $\mathfrak{D}$, which is an intersection $L = F \cap F(b)$ for some $b \neq b_0$ in B, is called a *side* of F. An end point of a side is called a *vertex*. The finite vertices belong to F, the infinite vertices are limit points of F.

The line $L_{b_0,b}$ is orthogonal to the line segment from b_0 to b and intersects the line segment in its midpoint. Hence, when b_0 is fixed, the line $L_{b_0,b}$ determines the point b. As a consequence, different sides of F lie on different lines. It follows easily that a finite vertex of F belongs to exactly two sides of F and is a common end point of the two. In addition, an infinite vertex can belong to at most two sides. A limit point of F which is the common end point of two different sides is called a cusp of F.

We prove below that the boundary of F is the union of the sides of F.

(4.4) Observation. Consider a point u in $\mathfrak{D}$. By Lemma (2.11), any fundamental neighborhood of u contains only a finite number of points from B. It follows that among the points in B there is a finite number of points b_j for which the distance to u is minimal. They are said to be the points of B *nearest* to u. They lie on geodesic circle around u (including the possibility of the circle with radius zero, when u belongs to B). If $u \in B$, there is only a single $b_1 := u$ nearest to u. If $u \notin B$, then, according to an orientation of the geodesic circle, we can index the b_j in a cyclic order $b_1, \dots, b_n$, so that there are no b_j on the open arc from b_i to b_{i+1}. Then, clearly, any fundamental neighborhood V of u is divided into finitely many angular sectors,

$$V_i := \{z \in V \mid \operatorname{dist}(z, b_i) \le \operatorname{dist}(z, b_j) \text{ for all } j\}, \tag{4.4.1}$$

and two different V_i and V_j have a half ray from u in common when $j = i \pm 1$, and only the point u in common otherwise. If there is only one point b_i, then $V_1 = V$, and the angle of V_1 at u is equal to 2π. If there are two points b_i, then V_1 and V_2 have a line segment as common boundary. The angles of V_1 and V_2 at u are equal to π. Finally, when there are more than two points b_i, then each V_i has at u an angle strictly less than π. In any case, the sum of the angles of V_i at u is equal to 2π. It is convenient to define the *width* of V_i as the angle of V_i divided by the rotation angle of the canonical generator γ_u.

Consider next a point u in $\partial_\Gamma\mathfrak{D}$. Then, again by Lemma (2.11), there is a smallest horo cycle around u containing points of B. The points of B on this smallest horo cycle are said to be *nearest* to u. The set of points of B nearest to u is a discrete subset of the horo cycle. Moreover, it is an infinite set, since it is invariant under the canonical generator γ_u. Hence the set of points of B nearest to u can be indexed cyclically b_i for $i \in \mathbf{Z}$ so that there are no b_j on the arc from b_i to b_{i+1}. Then, clearly, any fundamental neighborhood V of u is divided in infinitely many cuspidal sectors V_i by the equations (4.4.1), and two different V_i and V_j have a half ray from u in common when $j = i \pm 1$, and only the point u in common otherwise.

Define the *width* at u of the cuspidal sector V_i as follows: Assume first that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then the canonical generator γ_u is a translation $z \mapsto z + h$, and the points b_i of B nearest to ∞ are on a horizontal straight line of $\mathfrak{H}$. The fundamental neighborhood V is a half plane $\Im z > R$ in $\mathfrak{H}$, and the cuspidal sector V_i is a vertical strip of $\mathfrak{H}$, bounded by the two vertical lines of equal distance to b_{i-1} and b_i and to b_i and b_{i+1}. Define the width of V_i at $u = \infty$ as the euclidean width of the strip divided by h. In general, define the width of a cuspidal sector V_i by using a conjugation as in (3.16).

Note that the width of a cuspidal sector V_i is at most equal to 1, since the set of nearest points is stable under the canonical generator γ_u.

(4.5) Lemma. *Let u be a point of $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$, and γ_u the canonical generator of Γ_u. Consider, in the setup of* (4.4)*, a fundamental neighborhood V of u and the division of V in sectors V_i. Then there is a smallest positive number d such that $\gamma_u V_i = V_{i+d}$ for all i. Moreover, the sum of the widths of V_j for $j = 1, \dots, d$ is equal to* 1.

Proof. If $u \in \mathfrak{D}$, then any geodesic circle around u is invariant under γ_u. In particular, γ_u permutes the points b_i. If $u \in \partial_\Gamma\mathfrak{D}$, then any horo cycle around u is invariant under γ_u. In particular, γ_u permutes the points b_i. It follows, in both cases, that γ_u permutes the sectors V_i.

Assume that $u \in \mathfrak{D}$. Then γ_u is a rotation by the angle $2\pi/|\mathrm{P}\Gamma_u|$ around u. So the number $d \leq n$ defined by $\gamma_u V_i = V_{i+d}$ is independent of i. Clearly the union of the V_j, for $j = 1, \dots, e$ is itself an angular sector, and its angle at u is the rotation angle of γ_u. The assertions of the Lemma follow easily.

Assume that $u \in \partial_\Gamma \mathfrak{D}$. After a conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then γ_u is a translation $z \mapsto z + h$, and the cuspidal sectors are vertical strips. So, the number d defined by $\gamma_u V_i = V_{i+d}$ is independent of i. Clearly the union of the V_j, for $j = 1, \dots, d$, is itself a vertical strip of euclidean width h. The assertions of the Lemma follow easily. □

(4.6) Proposition. *Let u be a point of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. Consider, in the setup of* (4.4)*, a fundamental neighborhood V of u and the division of V in sectors V_i. If V is sufficiently small, then $F(b_i) \cap V = V_i$ and if b is a point of B such that $F(b) \cap V \neq \emptyset$, then b is one of the b_i.*

Proof. The points b_i are the points of B nearest to u. To prove the Proposition, it suffices to show that a sufficiently small fundamental neighborhood V of u has the following special property: for any point w in V, the points of B nearest to w is a subset of the b_i. Indeed, assume that V has the special property. Consider a given i, say $i = 0$. Obviously, the inequalities in (4.1.1) for $b := b_0$ and all $b' \in B$ imply the inequalities in (4.4.1) for all j. Hence, $F(b_0) \cap V \subseteq V_0$. Conversely, assume that w belongs to V_0. By the special property, the points of B nearest to w are among the b_i. By definition of V_0, among the b_i, then point b_0 is nearest to w. Hence b_0 is a point of B nearest to w, that is, $w \in F(b_0)$. Thus $F(b_0) \cap V = V_0$. Moreover, if the intersection $F(b) \cap V$ is non-empty, take a point w in the intersection. Then b is a point of B nearest to w. So, by the special property, b is one of the b_i.

To prove the existence of a fundamental neighborhood V with the special property, assume first that u belongs to $\mathfrak{D}$. Let r be the common (non-euclidean) distance from u to the nearest points b_i. By (2.11), any geodesic disk around u contains only a finite number of points of B. Hence, if $\varepsilon > 0$ is sufficiently small, then the geodesic disk W with radius $r + \varepsilon$ contains of points of B only the b_i. The geodesic disk V around u with radius $\varepsilon/2$ has the special property. Indeed, let w be a point of V. If w belongs to V, then, by the triangle inequality, the distance from w to any b_j is strictly less than $r + \varepsilon/2$, and the distance from w to a point outside W is at least equal to $r + \varepsilon/2$. Therefore, the points of B nearest to w are among the b_i. Thus V has the special property.

Assume next that u is a Γ-parabolic point. After a conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then the canonical generator γ_u is a translation $z \mapsto z + h$, and the smallest horo cycle containing the nearest points b_i is a horizontal (straight) line $\Im z = r$. It follows from (2.11) that for some $r' < r$, the fundamental neighborhood $W : \Im z > r'$ contains of points of B only the b_i. Let ε be the non-euclidean distance between the straight lines $\Im z = r$ and $\Im z = r'$. It is the distance between any two points with the same real part on the two lines; in fact, it is equal to $\log(r/r')$. The distance $\mathrm{dist}(z, z + h)$ converges to zero uniformly as $\Im z \to \infty$. Therefore, there exists an $R > r$ such that if $\Im z > R$ and $0 \leq |k| \leq h/2$, then $\mathrm{dist}(z, z + k) < \varepsilon$. The fundamental neighborhood $V : \Im z > R$ has the special property. Indeed, let w be a point of V. Let s be the non-euclidean distance from w to the line $\Im z = r$ (then $s = \log(\Im u/r)$). Since the set of points b_i is invariant under the translation $z \mapsto z + h$, there is a b_j such that for the difference of real parts, $k := \Re b_j - \Re u$, we have that $0 \le |k| \le h/2$. Now, the distance from w to $w + k$ is strictly less than ε and the distance from $w + k$ to b_j is equal to s. Hence the distance from w to b_j is strictly less than $s + \varepsilon$. On the other hand, the distance from w to the line $\Im z = r'$ is equal to $s + \varepsilon$. Hence the distance from w to any point of B different from the b_i is at least equal to $s + \varepsilon$. Therefore, the points of B nearest to w are among the b_j. Thus V has the special property. □

(4.7) Observation. Fix a point $b_0 \in B$, and set $F := F(b_0)$. The local descriptions in (4.5) and (4.6) apply in particular to points v that are either in F or Γ-parabolic limit points of F. It follows that a sufficiently small fundamental neighborhood V of v decomposes into a finite number of sectors V_i, of which one, say V_0, is equal to $F \cap V$. The width of V_0 at v is called the width of F at v. If v is in F, then the sectors are angular and there is finite number. If v is Γ-parabolic, then the sectors are cuspidal and there is an infinite number.

Consider the following (exhaustive) cases:

(1) There is only one V_i. Then $V = V_0$ is contained in F, and hence v is an inner point of F. Moreover, if $F(b)$ meets V, the $b = b_0$. In particular, no point of V is a side of F or a vertex of F. The width of F at v is equal to 1.

(2) There are two angular sectors V_i, say V_0 and V_1. Then $V_1 = F(b_1) \cap V$, and $V \cap F(b) = \emptyset$ when b is not b_0 or b_1. The common boundary of V_0 and V_1 is the intersection of V and the line C_{b_0,b_1}, and equal to the part in V of the intersection $F(b_0) \cap F(b_1)$. In particular, the latter intersection is a side of F, it contains v, and it is the only side of F containing v. The width of F at v is equal to $1/2$. The interior points of V_0 are interior points of F. In particular, the point v belongs to the closure of the set of interior points of F.

(3) There are more than two angular sectors V_i. Then the common boundaries of V_0 and the two adjacent V_i's are parts of sides of F. In particular, then v is a vertex of F. The width of F at v is strictly less than $1/2$. The interior points of V_0 are interior points of F. In particular, the point v belongs to the closure of the set of interior points of F.

(∞) Assume that v is a Γ-parabolic limit point of F. If V_{-1} and V_1 are the two adjacent V_i, then the intersections $V_0 \cap V_{-1}$ and $V_0 \cap V_1$ are parts of two sides of F. Thus v is the common end point of two different sides; in particular, then v is a cusp of F.

(4.8) Theorem. *Let b_0 be a point of $\mathfrak{D}$ which is not Γ-elliptic. Then the canonical domain $F = F(b_0)$ of* (4.1) *is a fundamental domain for Γ, and its interior is the subset $U(b_0)$. The boundary of F is the union of the sides of F. In addition, every Γ-parabolic limit point of F is a cusp. Finally, every compact subset of $\mathfrak{D}$ meets only a finite number of sides.*

Denote by F^* the union of F and the set of Γ-parabolic limit points of F. Then, for any point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, the intersection $F^* \cap \Gamma u$ is finite, and the following formula holds,

$$\sum_{v \in F^* \cap \Gamma u} \operatorname{width}_v F = 1$$

In particular, every Γ-parabolic point is equivalent to a point in F^*.

Proof. Set $U := U(b_0)$, and consider the conditions (1)–(3) of (3.1). It follows from Proposition (4.6) in particular that any point v of $\mathfrak{D}$ belongs $F(b)$ for some $b = \gamma b_0$, and that it belongs to only a finite number of $F(b)$. As $F(\gamma b_0) = \gamma F(b_0)$, it follows that condition (1) holds. Next, let v be a point in U. Then b_0 is the single point in $B = \Gamma b_0$ nearest to v. It follows from the discussion in (4.7)(1) that a sufficiently small neighborhood of v is contained in F. Hence U is an open subset of $\mathfrak{D}$. If a point of F is not in U, it follows from the discussion that any neighborhood of v contains points of U and points of the complement of F. Hence U is the set of interior points in F and F is the closure of U. Thus condition (2) holds. Finally, the intersection $U \cap \gamma U$ is equal to $U(b_0) \cap U(\gamma b_0)$. It is obvious from the definition in (4.1) that the intersection is empty unless $b_0 = \gamma b_0$. Since b_0 is not Γ-elliptic, the equation $b_0 = \gamma b_0$ implies that $\gamma = \pm 1$. Hence condition (3) holds. Thus F is a fundamental domain for Γ.

The assertions about the boundary of F follow immediately from the discussion in (4.7). By the same discussion, any point u of $\mathfrak{D}$ has a fundamental neighborhood meeting at most two sides of F. It follows that a compact subset of $\mathfrak{D}$ can only meet a finite number of sides of F.

Finally, the formula follows from Lemma (4.5). □

(4.9) Corollary. *The quotient $\overline{\mathfrak{D}/\Gamma}$ is compact, if and only if $F = F(b_0)$ is a finite normal domain.*

Proof. By Proposition (3.14), if the fundamental domain F is a finite normal domain, then the quotient is compact.

Conversely, assume that the quotient is compact. Choose for each point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ a small fundamental neighborhood $V = V_u$ of u having the property of Proposition (4.6). If V has the property for u, and u' is Γ-equivalent to u, say $u' = \gamma u$, then $V' := \gamma V$ has the property for u'. Moreover, V' is independent of the choice of γ, because two different choices differ by a matrix in Γ_u, and the matrices of Γ_u leaves the fundamental neighborhood V invariant. Hence we may assume that $V_{\gamma u} = \gamma V_u$ for all γ in Γ.

For any point v of F^*, the fundamental neighborhood V_v is a union of sectors V_i. Let F_v be the sector contained in F^*.

Consider an arbitrary point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ and the fundamental neighborhood V_u. Then V_u is a union of sectors V_i and V_i is the part in V_u of $F(b_i)$. We have that $b_i = \gamma b_0$ for some γ in Γ, and hence $F(b_i) = \gamma F$. Thus, with $v := \gamma^{-1} u$, we have that $v \in F^*$ and $V_u = \gamma V_v$. It follows that V_i is equal to γF_v. Therefore, in the decomposition of V_u, any of the sectors V_i is a transform of F_v for some point $v \in F^* \cap \Gamma u$.

The intersection $F^* \cap \Gamma u$ is finite by the Theorem. Hence it follows from the preceding argumentation that any point u in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ has a fundamental neighborhood which is a union of transforms of sets F_v for a finite number of points v in F^*. It follows that any point in the quotient has a neighborhood which is a finite union of the images of the F_v. Since the quotient is compact, therefore the quotient is the union of a finite number of images of the F_v. In other words, there is a finite set of points v_k in F^* such that any point in F^* is Γ-equivalent to a point in the union of the F_{v_k}.

Let w be a point which is either on the boundary of F or a limit point of F. Let M denote the open line segment from b_0 to w. Then M is contained in the interior of F. In particular, no points of M can belong to a transform of F different from F. Hence M is contained in the union of the finitely many F_{v_i}. It follows first that w is a boundary point of one of the F_{v_k}, and next that w belongs to one of the F_{v_k}.

Each F_{v_k} meets at most two sides of F and has at most one point on $\partial\mathfrak{D}$. As the boundary of F is the union of the sides, it follows first that the number of sides is finite. Hence the condition (4) of (3.1) holds. Next, it follows that there is only a finite number of limit points for F. Hence F is a finite domain. Moreover, the limit points of F are in F^* and hence Γ-parabolic. Therefore, by Proposition (3.14), F is a finite normal domain.

Thus the Corollary has been proved. □

(4.10) Note. If the canonical domain $F = F(b_0)$ has finite area, then it is a finite normal domain. Indeed, assume that the area of F is finite. For each finite vertex v of F, let α_v denote the angle of F at v (equal to 2π times the width at v). Set $\alpha_v := 0$ for all limit points v of F. Let v_i be a finite number of points that are either vertices or limit points of F. Let F' be the convex hull of the v_i. Then F' is a finite polygon, possibly with some of its vertices on $\partial\mathfrak{D}$. If α'_i is the angle of F' at v_i, then it is well known that the area of F' plus 2π is equal to $\sum(\pi - \alpha'_i)$. Now, since F is convex, we have that $F' \subseteq F$. So, the area of F' is at most the area of F and the angle α'_i is at most α_{v_i}. Therefore, the sum $\sum(\pi - \alpha_{v_i})$ is at most equal to the area of F plus 2π. As a consequence, the following sum over all points v that are either a vertex or a limit point of F is bounded above:

$$\sum(\pi - \alpha_v). \tag{*}$$

In the sum (*), each limit point of F contributes with the term π. Hence there is only a finite number of limit points of F. Divide the finite vertices into Γ-equivalence classes. Each class is finite, and sum of the angles in a given class is equal to $2\pi/d$ where d is the common number of elements of the isotropy group $\mathrm{P}\Gamma_v$ for any vertex v in the class. Let n be the number of vertices in the class. Then the sum of the terms in (*) corresponding to the vertices in the class is equal to $(n - 2/d)\pi$. Each angle is strictly less than π. Therefore, if $d = 1$, then $n \geq 3$, and if $d = 2$ then $n \geq 2$. Hence, $n - 2/d \geq 1$ if $d = 1$ or $d = 2$. If $d > 2$, then $(n - 2/d) \geq 1/3$. Hence, in all cases, the group contributes with at least $\pi/3$ to the sum. It follows that there is only a finite number of groups. Hence the number of finite vertices is finite.

As the number of vertices is finite, so is the number of sides. Moreover, there is only a finite number of limit points of F. So F is a finite domain. Now it follows from Theorem (3.7) that F is a finite normal domain.

(4.11) Exercise. Prove that the considerations in (4.10) imply the following estimate:

$$\mathrm{Area}(F)/\pi + 2 \geq (\text{number of infinite vertices}) + (\text{number of finite vertices})/3.$$

5. The Euler characteristic.

(5.1) Setup. Fix a disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ such that the quotient $X = X(\Gamma) := \overline{\mathfrak{D}/\Gamma}$ is compact.

The quotient is a *surface* by Corollary (2.13): any point in the quotient has an open neighborhood topologically isomorphic to an open disk in the plane. Therefore, as is well known, the *Euler characteristic* $\chi(X)$ is defined. It is the alternating sum of the ranks of the cohomology groups,

$$\chi(X) = \dim H^0(X) - \dim H^1(X) + \dim H^2(X). \tag{5.1.1}$$

Obviously, the quotient X is connected. In addition, the definition of the local isomorphisms in Corollary (2.13) shows that X in a natural way is an oriented surface. Therefore, as is well known, the surface X is (topologically) a sphere with a certain number of handles attached. The number of handles is called the *genus* of X and is denoted $g = g(X) = g(\Gamma)$. In addition, the cohomology groups $H^0(X)$ and $H^2(X)$ are 1-dimensional, and $H^1(X)$ is of dimension $2g$. Hence the genus and the characteristic are related by the formula,

$$\chi = 2 - 2g. \tag{5.1.2}$$

Moreover, it is well known that the characteristic can be determined from any a triangulation of X by the formula,

$$\chi(X) = \#(\text{vertices}) - \#(\text{edges}) + \#(\text{faces}). \tag{5.1.3}$$

Fix a finite normal fundamental domain F for Γ, cf. Proposition (3.14) and Corollary (4.9). The boundary of F is the union of finitely many sides. We will take as definition of a vertex here a point which is an end point of a side. Recall that for any side L of F there is a unique boundary transformation γ_L in Γ such that $\gamma_L L$ is a side of F. We will say that L is an *elliptic side* if $\gamma_L L = L$. Equivalently, a side L is elliptic, if it contains an elliptic point v which is not a vertex. Then, necessarily, the point v is elliptic of order 2 and γ_L is the rotation by the angle π around v.

(5.2) Definition. The *Euler characteristic* of the domain F is defined from any triangulation of F as the number,

$$\chi(F) = \#(\text{inner vertices}) - \#(\text{inner edges}) + \#(\text{faces}). \tag{5.2.1}$$

The Euler characteristic is in fact a topological invariant of the interior U of F. If U is connected, then the characteristic is equal to 1 minus the number of holes in U. At any rate, it is a consequence of (the proof of) the next formula that the characteristic of F is independent of the choice of triangulation.

(5.3) Proposition.

$$\mathrm{Area}(F) = -2\pi\chi(F) + \#(\text{sides of } F)\pi - \sum \mathrm{Angle}_v F, \tag{5.3.1}$$

where the sum is over the vertices of F and the angle of F at a cusp is defined as 0.

Proof. Consider a triangulation of F. Each face is a triangle. An edge of a triangle is either an inner edge or part of the boundary of F. Clearly, if we divide a side of F into two by a point on the side, then the number of sides is increased by 1 and the sum of the angles of is increased by π; hence the right side of (5.3.1) is unchanged. Therefore, we may assume that the sides of F are exactly those edges of the triangulation that are contained in the boundary.

If we sum the areas of the faces of the triangulation, we get the area of F, that is the left side of (5.3.1). On the other hand, each face is a triangle and its area is equal to π minus the sum of the three angle of the triangle. Hence the sum of the areas of the faces is equal to π times the number of faces minus the sum of the angles of the faces. In the latter sum, each inner vertex contributes with 2π and each vertex on the boundary of F contributes with the angle of F at the vertex. Hence we have obtained the equation,

$$\mathrm{Area}(F) = \#(\text{faces})\pi - \#(\text{inner vertices})\, 2\pi - \sum \mathrm{Angle}_v F.$$

The equation (5.3.1) is a consequence, since obviously, the equation,

$$3\#(\text{faces}) = 2\#(\text{inner edges}) + \#(\text{sides}),$$

holds for a triangulation whose outer edges are the sides of F. □

(5.4) Proposition. *Take as vertices of F the end points of the sides. Then the Euler characteristics of $X = \overline{\mathfrak{D}/\Gamma}$ and the fundamental domain F are related by the formula,*

$$\chi(X) = \chi(F) - \frac{\#(\text{non-elliptic sides of } F)}{2} + \#\Big(\frac{\text{vertices of } F}{\Gamma}\Big), \tag{5.4.1}$$

where the last fraction denotes the set of Γ-equivalence classes of vertices of F.

Proof. Consider an elliptic side L of F. It does not contribute to the right hand side of (5.4.1). The boundary transformation γ_L is a rotation by π around an elliptic point v of order 2 on L. The point v divides L into two line segments L' and L'' with v as common end point. The two segments L' and L'' are interchanged by γ_L. Hence, considering the sides of F, we may replace the side L by the two sides L' and L'' and add the point v as a vertex. The point v is Γ-equivalent to no other point of F. Therefore, the replacement does not change the right hand side of (5.4.1). Thus, to prove (5.4.1), we may assume that no side of F is elliptic.

Consider a side L and its transform $L' = \gamma_L L$. A given point v interior on L divides L into two line segments L_1 and L_2 with v as common end point. Then L' is divided into the two line segments $L_1' := \gamma_L L_1$ and $L_2' = \gamma_L L_2$ with $v' := \gamma_L v$ as common end point. Hence, considering the sides of F, we may replace the two sides L and L' by the four sides

L_1, L_1', L_2, L_2' and add the two points v and v' as a vertices. The points v and v' form a single Γ-equivalence class of vertices. Therefore, the replacement does not change the right hand side of (5.4.1). Thus, to prove (5.4.1), we can divide any side L of F into a finite number of sides as long as we divide similarly the side γL. In particular, just dividing some of the sides into two sides if necessary, we may assume that no two different points on a side are Γ-equivalent.

Consider a triangulation $\mathcal{T}$ of F. As noted above, we may assume that any edge of $\mathcal{T}$ which is contained in the boundary of F is a side of F. Then every triangle of $\mathcal{T}$ is mapped homeomorphically onto its image in the quotient X, and the images form a triangulation of X. Clearly, in the image triangulation, the faces correspond to the faces of $\mathcal{T}$, the edges correspond to inner edges of $\mathcal{T}$ and pairs $L, \gamma_L L$ of sides of F, and the vertices correspond to inner vertices of $\mathcal{T}$ and Γ-equivalence classes of vertices of F. Thus the required equation (5.4.1) follows from (5.1.3) and the definition of $\chi(F)$. ▯

(5.5) Corollary. *The following formula holds:*

$$\frac{\mathrm{Area}(F)}{2\pi} = -\chi(X) + \sum_{w \bmod \Gamma} \Bigl(1 - \frac{1}{|\mathrm{P}\Gamma_w|}\Bigr), \tag{5.5.1}$$

where the sum is over all orbits that are Γ-elliptic or Γ-parabolic, and $|\mathrm{P}\Gamma_w|$ is the common order of the isotropy groups of the points in an orbit. In the sum, at the Γ-parabolic orbits where the isotropy group is infinite, the fraction $1/|\mathrm{P}\Gamma_w|$ is counted as 0.

Proof. We apply the two formulas (5.2.1) and (5.4.1). It is clear, and noted in the proof of (5.4), that to apply (5.4.1) we may assume that no side of F is Γ-elliptic. Hence we obtain the equation,

$$\frac{\mathrm{Area}(F)}{2\pi} = -\chi(X) + \#\Bigl(\frac{\text{vertices of } F}{\Gamma}\Bigr) - \sum \mathrm{Angle}_v F\,/2\pi.$$

In the sum, group the terms according to their Γ-equivalence class. Consider a Γ-equivalence class of vertices represented by a point w, that is, a class consisting of all vertices Γ-equivalent to w. If w is in $\mathfrak{D}$, then, by (3.10), the corresponding contribution to the sum is equal to $1/|\mathrm{P}\Gamma_w|$. If w is in $\partial_\Gamma\mathfrak{D}$, corresponding to an equivalence class of cusps of F, then all the angles are zero, and the corresponding contribution in the sum is zero, and hence by convention equal to $1/|\mathrm{P}\Gamma_w|$. Hence, we obtain the equation,

$$\frac{\mathrm{Area}(F)}{2\pi} = -\chi(X) + \sum_{w \bmod \Gamma} \Bigl(1 - \frac{1}{|\mathrm{P}\Gamma_w|}\Bigr),$$

where the sum is over all orbits containing a vertex of F. The asserted equation (5.5.1) is a consequence since, by assumption, every elliptic point is Γ-equivalent to a vertex. ▯

(5.6) Note. It follows from (5.5) that the area of the fundamental domain F is always a rational multiple of 2π, that is, the quotient $\mathrm{Area}(F)/2\pi$ is a rational number. The quotient is denoted $\mu = \mu(\Gamma)$. In addition, it is customary to denote by $\nu_e(\Gamma)$ the number of Γ-elliptic orbits of order e. In addition, the number of Γ-parabolic orbits is denoted $\nu_\infty(\Gamma)$. Hence, using (5.1.2), the following equation is just a rewriting of (5.5.1),

$$\mu = 2g - 2 + \nu_\infty + \sum_{e\geq 2} \nu_e\Big(1 - \frac{1}{e}\Big). \tag{5.6.1}$$

(5.7) Observation. If Δ is a subgroup of finite index in Γ, then

$$\mu(\Delta) = |\mathrm{P}\Gamma{:}\mathrm{P}\Delta|\,\mu(\Gamma). \tag{5.7.1}$$

Indeed, if the transformations γ_i, for $i = 1, \dots, d$, represent the right cosets of $\mathrm{P}\Gamma$ modulo $\mathrm{P}\Delta$, then, as observed in (3.12), the union $G := \bigcup \gamma_i F$ is a normal fundamental domain for Δ. Obviously, the area of G is d times the area of F. As a consequence, (5.7.1) holds.

(5.8) Example. The group $\Gamma(1) := \mathrm{SL}_2(\mathbf{Z})$ is a discrete subgroup of $\mathrm{SL}(\mathfrak{H})$. A fundamental domain for $\Gamma(1)$ was determined in (1.6). The domain was a triangle with one cusp and two finite Γ-equivalent vertices. Obviously for a triangle, and more generally for any polygon with a simply connected interior, the Euler characteristic is equal to 1. The side connecting the two finite vertices is elliptic, and there are two equivalence classes of vertices. Hence, by the formula (5.4.1), the Euler characteristic of $\Gamma(1)$ is $1 - 1 + 2 = 2$. Thus the genus is equal to 0, confirming that the quotient $\overline{\mathfrak{H}/\Gamma(1)}$ obviously is a sphere.

The angles of the fundamental domain at the two finite vertices are equal to $2\pi/6$. Hence the area of the domain is equal to $\pi/3$. In other words, $\mu = \frac{1}{6}$. There is one parabolic orbit, one elliptic orbit of order 2 and one of order 3 and no other elliptic orbits. Hence $\nu_\infty = \nu_2 = \nu_3 = 1$, and $\nu_e = 0$ for $e > 3$.

Consider a subgroup Γ of finite index in $\Gamma(1)$. Set $d := |\mathrm{P}\Gamma(1){:}\mathrm{P}\Gamma|$. Then $\mu(\Gamma) = d/6$ by (5.7.1). Moreover, since every Γ-elliptic point is $\Gamma(1)$-elliptic, we have $\nu_e = 0$ for $e > 3$. Hence, with $\nu_e = \nu_e(\Gamma)$, equation (5.6.1) is equivalent to the following,

$$g(\Gamma) = 1 + \frac{d}{12} - \frac{\nu_2}{4} - \frac{\nu_3}{3} - \frac{\nu_\infty}{2}\,. \tag{5.8.1}$$

(5.9) Note. The area of the fundamental domain F can be obtained by integration. Assume first the disk is the upper half plane $\mathfrak{H}$. Then,

$$\mathrm{Area}(F) = \int_{\partial F} \frac{dz}{\Im z} = \int_{\partial F} \frac{dx}{y} = \int_F \frac{dx\,dy}{y^2}. \tag{5.9.1}$$

The two middle path integrals are over the boundary of F, orientated counter clockwise around F. The first path integral in (5.9.1) should be taken with care. If a side L of F has an infinite vertex as end point, then the integral $\int_L dz/y$ is not convergent. When F has infinite vertices, the path integral along ∂F is defined as follows: consider the path integrals obtained by integration along the boundaries of the subdomains obtained from F by cutting away a small fundamental neighborhood of each infinite vertex. The limit of these path integrals, as the fundamental neighborhoods shrink around the infinite vertices, is then the path integral along ∂F.

In the first integral of (5.9.1), the integrand is a sum of two forms,

$$\frac{dz}{y} = \frac{dx}{y} + i\frac{dy}{y}.$$

The second form is a total differential in $\mathfrak{H}$, $y^{-1}dy = d(\log y)$. Hence its integral over a closed path is equal to zero. Therefore, the second equality of (5.9.1) holds.

Consider the integral $\int_L dx/y$ over any oriented line segment L in $\mathfrak{H}$. Say L is the segment from u to v, where u and v are allowed to be limit points. Then the path integral $\int_L dx/y$ is convergent, and given by the formula,

$$\int_L \frac{dx}{y} = \theta_{L,v} - \theta_{L,u}, \tag{5.9.2}$$

where, for any point w on L, the angle $\theta_{L,w}$ is the oriented angle from L to the (vertical) line from w to ∞. The formula follows easily from the definition of the path integral, noting that L is part of a circle orthogonal to the real axis (the trivial case when L is part of a vertical straight line has to be treated separately). From the formula (5.9.2) it follows that if F is a triangle, then the path integral $\int_{\partial F} dx/y$ is equal to π minus the sum of the interior angles. Thus the equation $\int_{\partial F} dx/y = \mathrm{Area}(F)$ holds when F is a triangle. Therefore, using a triangulation of F, the equation holds in general. As $d(y^{-1}dx) = -y^{-2}\,dy{\wedge}dx = y^{-2}\,dx{\wedge}dy$, the last equality in (5.9.1) follows from Stoke's theorem. Thus the equations of (5.9.1) have been proved.

The first equation of (5.9.1) implies the following:

$$\mathrm{Area}(F) = \sum i \int_L \frac{J(\gamma_L, z)'}{J(\gamma_L, z)}\,dz. \tag{5.9.3}$$

where the sum is over the sides L of F, and γ_L is the boundary transformation. Indeed, if $\gamma \in \mathrm{SL}(\mathfrak{H})$ then $J(\gamma, z) = cz + d$ with real numbers c and d. It follows that $2iJ(\gamma, z)' = 2ic = (J(\gamma, z) - \overline{J(\gamma, z)})/\Im z$. Moreover, $\Im(\gamma z) = |J(\gamma, z)|^{-2}\Im z$ and $d(\gamma z) = J(\gamma, z)^{-2}dz$. Hence we obtain the equations,

$$2i\int_L \frac{J(\gamma, z)'dz}{J(\gamma, z)} = \int_L \Big(1 - \frac{\overline{J(\gamma, z)}}{J(\gamma, z)}\Big)\frac{dz}{\Im z} = \int_L \Big(\frac{dz}{\Im z} - \frac{d(\gamma z)}{\Im(\gamma z)}\Big) = \int_L \frac{dz}{\Im z} - \int_{\gamma L} \frac{dz}{\Im z}.$$

Take $\gamma := \gamma_L$. Then $\gamma_L L$ is a side L' of F. However, the orientation on L' as a side of F is the reverse of the orientation on L' as an image $\gamma_L L$. Hence the difference on the right hand side of the equations is the sum $\int_L dz/\Im z + \int_{L'} dz/\Im z$. Consequently, (5.9.3) follows by summation over all the sides L of F. In fact, it follows that

$$\frac{\mathrm{Area}(F)}{2} = \sum{}' i \int_L \frac{J(\gamma_L, z)'}{J(\gamma_L, z)}\, dz. \tag{5.9.4}$$

where the sum is over unordered pairs $\{L, L'\}$ of sides with $L' = \gamma_L L$.

It is not hard to see, using a Möbius transformation to transform the integrals, that (5.9.3) and (5.9.4) holds for any disk $\mathfrak{D}$.

Modular groups

1. Finite projective lines.

(1.1) Setup. Recall that the group $\mathrm{SL}(\mathfrak{H})$ is the group $\mathrm{SL}_2(\mathbf{R})$. The subgroup $\mathrm{SL}_2(\mathbf{Z})$ is a discrete subgroup. It is called the *modular group*.

Fix a natural number N. We are going to consider various subgroups of $\mathrm{SL}_2(\mathbf{Z})$ related to subgroups of $\mathrm{SL}_2(\mathbf{Z}/N)$. Note that, by the Chinese Remainder Theorem, if $N = \prod p^\nu$ is the prime factorization of N, then

$$\mathrm{GL}_2(\mathbf{Z}/N) = \prod \mathrm{GL}_2(\mathbf{Z}/p^\nu), \quad \mathrm{SL}_2(\mathbf{Z}/N) = \prod \mathrm{SL}_2(\mathbf{Z}/p^\nu).$$

Clearly, the definition of the action of $\mathrm{GL}_2(\mathbf{C})$ on the Riemann sphere $\overline{\mathbf{C}}$ extends to an arbitrary field. More generally, if R is any commutative ring, we define $\mathrm{PGL}_n(R)$ as the quotient of $\mathrm{GL}_n(R)$ modulo the scalar matrices:

$$\mathrm{PGL}_n(R) := \mathrm{GL}_n(R)/R^*.$$

By definition, $\mathrm{PSL}_n(R)$ is the image of $\mathrm{SL}_n(R)$ in $\mathrm{PGL}_n(R)$, that is,

$$\mathrm{PSL}_n(R) := \mathrm{SL}_n(R)/\mu_n(R),$$

where $\mu_n(R)$ is the subgroup of R^* consisting of n'th roots of unity. Note in particular, for $n = 2$, that the group $\mathrm{PSL}_2(R)$ is in general a non-trivial quotient of $\mathrm{SL}_2(R)/\pm 1$, since the equation $\alpha^2 = 1$ in R may have non-trivial solutions.

Assume that R is a local ring with maximal ideal $\mathfrak{m}$. Let $\overline{R}$ be the disjoint union, $\overline{R} := R \cup \frac{1}{\mathfrak{m}}$, where $\frac{1}{\mathfrak{m}}$ is a set consisting of one symbol for each element in $\mathfrak{m}$. Denote by R^2 the R-module of columns, and by $(R^2)^*$ the subset of columns for which at least one entry is a unit. Then there is a surjective map $(R^2)^* \to \overline{R}$,

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \mapsto z_1/z_2. \tag{1.1.1}$$

The quotient $z_1/z_2 \in \overline{R}$ is defined as $z_1 z_2^{-1}$ if $z_2 \in R^*$ and as the symbol $1/(z_1^{-1}z_2)$ if $z_2 \in \mathfrak{m}$. Obviously, the group $\mathrm{GL}_2(R)$ acts on the set $(R^2)^*$ and the scalar matrices permute the elements of the fibers of the map (1.1.1). Hence the action descends to an action of $\mathrm{PGL}_2(R)$ on $\overline{R}$. As in the case of Möbius transformations, it is a faithful representation,

$$\mathrm{PGL}_2(R) \hookrightarrow \mathrm{Aut}(\overline{R}). \tag{1.1.2}$$

The set $\overline{R}$ for a local ring R is the *projective line* $I\!P^1(R)$. The action of $\mathrm{PGL}_2(R)$ on the projective line generalizes to an action of $\mathrm{PGL}_{n+1}(R)$ on projective n-space $I\!P^n(R) := (R^{n+1})^*/R^*$.

(1.2) Note. Consider a field F. Then it follows, as in the proof of (Möb.1.3)(3), that the action of $\mathrm{PGL}_2(F)$ on $\overline{F}$ is triply transitive. However, the action of the subgroup $\mathrm{PSL}_2(F)$ is in general only doubly transitive: if (u, v) and (u', v') are two pairs of different points of $\overline{F}$ then the matrices α that map the first pair to the second are the matrices of the following form (with an obvious notion for the representatives of the four points):

$$\alpha = [\tilde{u'}\ \tilde{v'}]\begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}[\tilde{u}\ \tilde{v}]^{-1},$$

where $\lambda, \mu \in F^*$. Clearly, with suitable choices of λ and μ, the determinant of α is equal to 1. For instance, if F is a finite field with q elements, then there are $(q-1)(q-1)$ matrices α, and $q-1$ of these have determinant 1. So, if q is odd, there are $(q-1)/2$ automorphisms in $\mathrm{PSL}_2(F)$ mapping the first pair to the second.

(1.3) Proposition. *Consider in* $\mathrm{SL}_2(\mathbf{Z})$ *the three matrices,*

$$s := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad t := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad u := \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}.$$

Then $s^2 = -1$, $u^3 = 1$, *and* $s = tu$. *In particular,* s *is of order* 4 *and* u *is of order* 3. *Moreover, the group* $\mathrm{SL}_2(\mathbf{Z})$ *is generated by the two matrices* s *and* t, *and hence also by* s *and* u.

Proof. The three asserted equations follow by a simple computation. In addition,

$$t^b = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}, \quad st^{-1}s^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{1.3.1}$$

Denote by Γ the subgroup generated by s and t. It follows from (1.3.1) that the elementary row and column operations on matrices in $\mathrm{GL}_2(\mathbf{Z})$ can be achieved by multiplication from the left and right by matrices in Γ. By the Euclidean algorithm, every matrix in $\mathrm{GL}_2(\mathbf{Z})$ can be changed into a diagonal matrix by the elementary operations. In particular, any matrix σ in $\mathrm{SL}_2(\mathbf{Z})$ can, by multiplications from the right and left by matrices in Γ, be changed into ± 1. As $-1 = s^2$, it follows that σ belongs to Γ. □

(1.4) Sublemma. *The group* $\mathrm{SL}_2(\mathbf{Z}/N)$ *is generated by the following matrices:*

$$s := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad t := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad d_\mu := \begin{bmatrix} \mu & 0 \\ 0 & \mu^{-1} \end{bmatrix} \textit{for } \mu \in (\mathbf{Z}/N)^*. \tag{1.4.1}$$

Proof. Note first that if the integers a, c, N are relatively prime, then there is a number k such that $a + kc$ is relatively prime to N. Indeed, it suffices to take $k := \prod p$, where the product is over all primes p such that p is a divisor of N and not a divisor of a.

Consider an arbitrary matrix σ in $\mathrm{SL}_2(\mathbf{Z}/N)$:

$$\sigma = \begin{bmatrix} \bar{a} & \bar{b} \\ \bar{c} & \bar{d} \end{bmatrix},$$

where $ad - bc \equiv 1 \pmod N$. We have to show that we can obtain the identity matrix by multiplying σ a finite number of times by the matrices in (1.4.1) and their inverses.

Clearly, the integers a, c, N are relatively prime. Hence there exists a number k such that $a+kc$ is prime to N. Therefore, replacing σ by $t^k\sigma$, we may assume that $\bar{a} \in (\mathbf{Z}/N)^*$. Next, replacing σ by $d_{\bar{a}}^{-1}\sigma$, we may assume that $a = 1$. Furthermore, since

$$\sigma t^{-b} = \begin{bmatrix} 1 & \bar{b} \\ \bar{c} & \bar{d} \end{bmatrix} \begin{bmatrix} 1 & -\bar{b} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \bar{c} & 1 \end{bmatrix},$$

we may assume that $a = d = 1$ and $b = 0$. Finally,

$$s\sigma s^{-1} t^c = s\begin{bmatrix} 1 & 0 \\ \bar{c} & 1 \end{bmatrix} s^{-1} t^c = \begin{bmatrix} 1 & -\bar{c} \\ 0 & 1 \end{bmatrix} t^c = 1,$$

and hence the identity matrix has been obtained, as required. □

(1.5) Proposition. *Reduction modulo N is a surjective homomorphism of groups,*

$$\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N).$$

As a consequence, the group $\mathrm{SL}_2(\mathbf{Z}/N)$ is generated by the first two matrices s and t of (1.4.1).

Proof. To prove that reduction modulo N is surjective, it suffices to lift the generators of (1.4.1). Obviously s and t lift. For a diagonal matrix d_μ, we have that $\mu = \overline{m}$, where m is prime to N. Then m is prime to N^2. Hence there are numbers x and y such that $xm - yN^2 = 1$. Clearly, the following matrix of $\mathrm{SL}_2(\mathbf{Z})$ is a lift of d_μ:

$$\begin{bmatrix} m & yN \\ N & x \end{bmatrix}.$$

Hence the reduction is surjective. It follows that $\mathrm{SL}_2(\mathbf{Z}/N)$ is generated by s and t, since the lifts of s and t generate $\mathrm{SL}_2(\mathbf{Z})$ by (1.3). □

(1.6) Corollary. *The group $\mathrm{GL}_2(\mathbf{Z}/N)$ is generated by the matrices,*

$$s = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad d_{\mu,1} := \begin{bmatrix} \mu & 0 \\ 0 & 1 \end{bmatrix} \textit{for } \mu \in (\mathbf{Z}/N)^*. \tag{1.6.1}$$

Proof. Clearly, the assertion follows from the last assertion of Proposition (1.5). □

(1.7) Corollary. *Let p be a prime number. Then, for $\nu \geq 1$, the surjective ring homomorphism $\mathbf{Z}/p^\nu \to \mathbf{Z}/p$ induces a surjective group homomorphism,*

$$\mathrm{GL}_2(\mathbf{Z}/p^\nu) \to \mathrm{GL}_2(\mathbf{Z}/p), \tag{1.7.1}$$

and the kernel of the induced homomorphism consists of all matrices in $\mathrm{Mat}_2(\mathbf{Z}/p^\nu)$ *of the form,*

$$1+\alpha, \tag{1.7.2}$$

where all entries in α *belong to the kernel of* $\mathbf{Z}/p^\nu \to \mathbf{Z}/p$.

Proof. Clearly, the generators s and t of (1.6.1) for $N := p$ lift, and the generator $d_{1,\mu}$ lifts, because $(\mathbf{Z}/p^\nu)^* \to (\mathbf{Z}/p)^*$ is surjective. Hence the homomorphism (1.7.1) is surjective. Obviously, any matrix in the kernel has the form (1.7.2). Conversely, for a matrix $\sigma = 1+\alpha$ of the form (1.7.2), the determinant is congruent to 1 modulo p; hence σ belongs to $\mathrm{GL}_2(\mathbf{Z}/p^\nu)$, and obviously σ belongs to the kernel of (1.7.1). ∎

(1.8) Proposition. *The orders of the groups* $\mathrm{GL}_2(\mathbf{Z}/N)$ *and* $\mathrm{SL}_2(\mathbf{Z}/N)$ *are given by the following formulas,*

$$|\,\mathrm{GL}_2(\mathbf{Z}/N)| = N^4 \textstyle\prod_{p|N}(1-\frac{1}{p})(1-\frac{1}{p^2}), \quad |\,\mathrm{SL}_2(\mathbf{Z}/N)| = N^3 \prod_{p|N}(1-\frac{1}{p^2}).$$

As a consequence,

$$|\,\mathrm{SL}_2(\mathbf{Z}/N)/\pm 1| = \begin{cases} 6 & \text{for } N=2,\\ \frac{1}{2}N^3\prod_{p|N}(1-\frac{1}{p^2}) & \text{for } N>2. \end{cases}$$

Proof. By the Chinese Remainder Theorem, it suffices to prove the formulas when N is a prime power, $N = p^\nu$. The quotient $\mathbf{Z}/p$ is the field $\mathbf{F}_p$ with p elements. Hence the order of the group $\mathrm{GL}_2(\mathbf{Z}/p)$ is equal to the number of bases of the 2-dimensional vector space $\mathbf{F}_p^2$, that is, the order is equal to $(p^2-1)(p^2-p)$. By Lagrange, the kernel of the homomorphism $\mathbf{Z}/p^\nu \to \mathbf{Z}/p$ is of order $p^{\nu-1}$. Hence, there are $(p^{\nu-1})^4$ matrices of the form (1.7.2). Consequently, by Corollary (1.7),

$$|\,\mathrm{GL}_2(\mathbf{Z}/p^\nu)| = (p^{\nu-1})^4(p^2-1)(p^2-p) = (p^\nu)^4(1-p^{-2})(1-p^{-1}).$$

Thus the first asserted formula holds. The second formula follows from the first, because $\mathrm{SL}_2(\mathbf{Z}/p^\nu)$ is the kernel of the surjective homomorphism,

$$\det\colon\ \mathrm{GL}(\mathbf{Z}/p^\nu) \to (\mathbf{Z}/p^\nu)^*,$$

and $(\mathbf{Z}/p^\nu)^*$ has order $p^\nu - p^{\nu-1} = p^\nu(1-p^{-1})$.

Clearly, the final formula is a consequence, since $1=-1$ in $\mathbf{Z}/N$ only when $N=2$ (or $N=1$). ∎

(1.9) Remark. The order of $\mathrm{P\,SL}_2(\mathbf{Z}/N)$ is obtained from the order of $\mathrm{SL}_2(\mathbf{Z}/N)$ by dividing by the order of the group $\mu_2(\mathbf{Z}/N)$ defined in (1.1). Clearly, the latter order is a multiplicative function of N. Obviously, for an odd prime power p^ν, the order of $\mu_2(\mathbf{Z}/p^\nu)$ is equal to 2. For a power of 2, the order of $\mu_2(\mathbf{Z}/2^\nu)$ is equal to 1 for $\nu=1$, equal to 2 for $\nu=2$ and equal to 4 for $\nu \geq 3$.

(1.10) Lemma. (1) *The group* $\mathrm{SL}_2(\mathbf{Z}/1)$ *is the trivial group of order* 1.

(2) *The group* $\mathrm{PSL}_2(\mathbf{Z}/2) = \mathrm{SL}(\mathbf{Z}/2)$ *is the dihedral group* $D_3 = S_3$ *of order* 6.

(3) *The group* $\mathrm{PSL}_2(\mathbf{Z}/3) = \mathrm{SL}(\mathbf{Z}/3)/\pm 1$ *is the tetrahedral group* A_4 *of order* 12.

(4) *The group* $\mathrm{PSL}_2(\mathbf{Z}/4) = \mathrm{SL}(\mathbf{Z}/4)/\pm 1$ *is the hexahedral group* S_4 *of order* 24.

(5) *The group* $\mathrm{PSL}_2(\mathbf{Z}/5) = \mathrm{SL}(\mathbf{Z}/5)/\pm 1$ *is the dodecahedral group* A_5 *of order* 60.

(6) *The group* $\mathrm{PSL}_2(\mathbf{Z}/6) = \mathrm{SL}(\mathbf{Z}/6)/\pm 1$ *is isomorphic to the group* $S_3 \times A_4$ *of order* 72.

Proof. The assertion (1) is trivial. By the Chinese Remainder Theorem, the assertion (6) follows from (2) and (3). To prove the remaining assertions, consider for $N = 2, 3, 4, 5$ the ring $\mathbf{Z}/N$. Then $\mathbf{Z}/N$ is a field for $N = 2, 3, 5$, and a local ring for $N = 4$. In each of the four cases, the order of the group $\mathrm{PSL}_2(\mathbf{Z}/N)$ is given by the formula of (1.8). To identify the group with the asserted permutation group, consider the projective line $\overline{\mathbf{Z}/N}$ and the representation of (1.1.2),

$$\mathrm{PSL}_2(\mathbf{Z}/N) \hookrightarrow \mathrm{Aut}(\overline{\mathbf{Z}/N}). \tag{1.10.1}$$

$N = 2$: The group $\mathrm{PSL}_2(\mathbf{Z}/2)$ has order 6. The projective line $\overline{\mathbf{Z}/2}$ has 3 elements: $\infty, 0, 1$. Hence the right hand side of (1.10.1) is the symmetric group S_3. Thus the two sides have the same order, and hence the inclusion is an isomorphism.

$N = 3$: The group $\mathrm{PSL}_2(\mathbf{Z}/3)$ has order 12. The projective line $\overline{\mathbf{Z}/3}$ has 4 elements: $\infty, 0, 1, -1$. Hence the right hand side of (1.10.1) is the symmetric group S_4. As the left hand side is of order 12, it is equal to the unique subgroup A_4 of index 2 in S_4.

$N = 4$: The group $\mathrm{PSL}_2(\mathbf{Z}/4)$ has order 24. The projective line $X := \overline{\mathbf{Z}/4}$ has 6 elements: $\infty, \frac{1}{2}, 0, 2, 1, -1$, and there is an obvious map $\overline{\mathbf{Z}/4} \to \overline{\mathbf{Z}/2}$. Given any point x in X there is a unique second point x' having the same image as x in $\overline{\mathbf{Z}/2}$. Consider subsets Z of X consisting of 3 elements lying over the 3 elements $\infty, 0, 1$ of the projective line $\overline{\mathbf{Z}/2}$. Obviously, if $Z = \{x, y, z\}$ is such a subset, then so is the complement $Z' = \{x', y', z'\}$. Hence, there are 4 elements in the set $\mathcal{D}$ of all decompositions of X into two such subsets,

$$X = \{x, y, z\} \cup \{x', y', z'\}.$$

Clearly, the group $\mathrm{PSL}_2(\mathbf{Z}/4)$ acts on the set $\mathcal{D}$. Consequently, we obtain a representation,

$$\mathrm{PSL}_2(\mathbf{Z}/4) \to \mathrm{Aut}(\mathcal{D}) = S_4. \tag{1.10.2}$$

The representation (1.10.2) is injective. Indeed, assume that σ in $\mathrm{SL}_2(\mathbf{Z}/4)$ acts as the identity on $\mathcal{D}$. Take a decomposition in $\mathcal{D}$, say $\{a, b, c\} \cup \{a', b', c'\}$. Then $\{a, b', c'\} \cup \{a', b, c\}$ is a second decomposition. As the first decomposition is fixed under σ, we have either $\sigma\{a, b, c\} = \{a, b, c\}$ or $\sigma\{a, b, c\} = \{a', b', c'\}$. Assume that $\sigma\{a, b, c\} = \{a', b', c'\}$. Then, since the second decomposition is fixed under σ, it follows first that $\sigma\{a, b', c'\} = \{a', b, c\}$ and next that $\sigma a = a'$. Similarly, $\sigma b = b'$ and $\sigma c = c'$, and hence $\sigma x = x'$ for all x. However, it is easily verified that the permutation $x \mapsto x'$ can not be obtained from a matrix with determinant 1. So the case $\sigma\{a,b,c\} = \{a',b',c'\}$ is excluded. Therefore, $\sigma\{a,b,c\} = \{a,b,c\}$. As the second decomposition is fixed under σ, it follows that $\sigma a = a$ and $\sigma a' = a'$. Similar equations are obtained for b and c, and hence σ is the identity in $\mathrm{PSL}_2(\mathbf{Z}/4)$.

As the representation (1.10.2) is injective and the two groups have the same number of elements, the representation is an isomorphism.

$N = 5$: The group $G = \mathrm{PSL}_2(\mathbf{Z}/5)$ has order 60. The projective line $X := \overline{\mathbf{Z}/5}$ has 6 elements: $\infty, 0, 1, 2, -2, -1$. The action of G on X is doubly transitively. In fact, as observed at the end of (1.2), if (u, v) and (u', v') are any two pairs of different points of X, then there are two elements of G mapping the first pair to the second. Take a subset $\{x_1, x_2\}$ with two elements. There are two elements of G for which x_1 and x_2 are fixed points. One element is the identity, and the second is (as seen by reducing to the case $x_1 = \infty$ and $x_2 = 0$) a double transposition $(y_1, y_2)(z_1, z_2)$. Similarly, there are two elements of G that interchanges x_1 and x_2, namely the two double transpositions $(x_1, x_2)(y_1, y_2)$ and $(x_1, x_2)(z_1, z_2)$. Therefore, the subset $\{x_1, x_2\}$ is part of a unique decomposition of X into three parts,

$$X = \{x_1, x_2\} \cup \{y_1, y_2\} \cup \{z_1, z_2\}, \tag{1}$$

with the special property that any element in G that stabilizes one part will stabilize all three parts. There are 15 subsets with two elements. Hence, there are 5 elements in the set $\mathcal{D}$ of all decompositions (1) with the special property. Clearly, the group G acts on the set $\mathcal{D}$. Consequently, we obtain a representation,

$$\mathrm{PSL}_2(\mathbf{Z}/5) = G \to \mathrm{Aut}(\mathcal{D}) = S_5. \tag{1.10.3}$$

The representation is injective. Indeed, assume that α is a nontrivial element of G that acts as the identity on $\mathcal{D}$. Take a point x_1 such that $\alpha x_1 \neq x_1$. Consider first the decomposition (1) with $x_2 := \alpha x_1$. Since the decomposition is invariant, it follows that $\alpha\{x_1, x_2\} = \{x_1, x_2\}$. Hence α has two fixed points, namely either y_1 and y_2, or z_1 and z_2. Consider next the decomposition (1) with a fixed point of α as x_2. Then, again since the decomposition is invariant, it follows that $\alpha\{x_1, x_2\} = \{x_1, x_2\}$, which is a contradiction since $\alpha x_1 \neq x_1$ and $\alpha x_2 = x_2$. Hence the representation is injective.

As the representation (1.10.3) is injective, necessarily its image is the unique subgroup A_5 of index 2 in S_5.

Hence all the isomorphisms of the Lemma have be established. $\square$

(1.11) Exercise. (1) The projective line $X := \overline{\mathbf{Z}/3}$ has 4 elements: ∞, 0, 1, -1. Prove that, as permutations of X, we have $s = (\infty, 0)(1, -1)$ and $t = (0, 1, -1)$. The tetrahedral group T is the automorphism group of a tetrahedron. The elements of T permute the 4 faces of the tetrahedron, and so T is a subgroup of S_4. Prove that the permutations s and t are 'tetrahedral', that is, under a suitable labeling of the four faces, s and t can be realized as rotations of the tetrahedron. Conclude that $\mathrm{PSL}_2(\mathbf{Z}/3) = T$.

(2) The projective line $X := \overline{\mathbf{Z}/4}$ has 6 elements: ∞, $\frac{1}{2}$, 0, 2, 1, -1. Prove that, as permutations of X, we have $s = (\infty, 0)(2, \frac{1}{2})(1, -1)$ and $t = (0, 1, 2, -1)$. The hexahedral

group H is the automorphism group of a hexahedron. The elements of H permute the 6 faces of the hexahedron, and so H is a subgroup of S_6. Prove that the permutations s and t are 'hexahedral', and conclude that $\mathrm{PSL}_2(\mathbf{Z}/4) = H$.

(3) The projective line $X := \overline{\mathbf{Z}/5}$ has 6 elements: ∞, 0, 1, 2, -2, -1. Prove that, as permutations of X, we have $s = (\infty, 0)(1, -1)$ and $t = (0, 1, 2, -2, -1)$. The dodecahedral group O is the automorphism group of a dodecahedron. The elements of O permute the 6 pairs of opposite faces of the dodecahedron, and so O is a subgroup of S_6. Prove that the permutations s and t are 'dodecahedral', and conclude that $\mathrm{PSL}_2(\mathbf{Z}/5) = O$.

2. Small modular groups.

(2.1) Example. Consider the group $G := \mathrm{SL}_2(\mathbf{Z}/2)$. By (1.10)(2), G it is the symmetric group S_3 of order 6. Hence G has a unique (normal) subgroup $G^{(2)}$ of index 2 (and order 3), namely the alternating group A_3 generated by u. The quotient $G/G^{(2)}$ is the cyclic group C_2, and so the canonical map to the quotient may be viewed as a surjective character,

$$\chi_2 \colon G \to C_2.$$

In addition, G has three subgroups of index 3 (and order 2), generated by the three involutions,

$$s = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad sts^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

(Note that $1 = -1$ in $\mathbf{Z}/2$.) The latter three subgroups are denoted respectively G_θ, G_0, and G^0.

(2.2) Example. Consider the group $G = \mathrm{SL}_2(\mathbf{Z}/3)$. It is of order 24 by (1.8). By (1.10)(3), its quotient modulo ± 1 is the alternating group A_4 of order 12. The alternating group A_4 contains Klein's 'Vier'-group, the kernel of the homomorphism $S_4 \to S_3$, and the image of A_4 under the homomorphism is the alternating group $A_3 = C_3$. Hence, by composition, we obtain a surjective character,

$$\chi_3 \colon G \to \mathrm{PSL}_2(\mathbf{Z}/3) = A_4 \to A_3 = C_3.$$

The kernel is denoted $G^{(3)}$. It is of index 3 (and order 8).

(2.3) Example. Consider the group $G := \mathrm{SL}_2(\mathbf{Z}/4)$. It is of order 48 by (1.8). We will describe some of its subgroups.

(i) The group G acts on the set $X := ((\mathbf{Z}/4)^2)^*$ of columns where at least one coordinate is a unit. There are 12 columns in X, and they are the representatives of the 6 points of the projective line $\overline{\mathbf{Z}/4}$. Hence they also represent the three points ∞, 0, 1 on the projective line $\overline{\mathbf{Z}/2}$. Define a 3-*orbit* as a subset $A = \{a, a', a''\}$ with 3 elements of X such that $a + a' + a'' = 0$. For instance, the following matrix is a 3-orbit:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}. \tag{1}$$

In $\mathbf{Z}/4$, the sum of two units is always in the maximal ideal. Hence the sum of any three units is a unit. It follows easily for any 3-orbit $A = \{a, a', a''\}$ that its three columns represent the three points ∞, 0, 1 in $\overline{\mathbf{Z}/2}$; consequently, there is a unique ordering of the 3 elements of A such that a, a', a'' represent respectively ∞, 0, 1. Clearly, the set of all 3-orbits has $4^2 = 16$ elements. The group G acts on the set of 3-orbits. The stabilizer of the 3-orbit (1) consists of the matrices in $G = \mathrm{SL}_2(\mathbf{Z}/4)$ whose columns are two of the three columns in (1). Hence, the stabilizer is of order 3. As a consequence, the group G acts transitively on the set of 3-orbits.

Now, for any given 3-orbit A there is a unique decomposition of X into four 3-orbits,

$$X = A \cup B \cup C \cup D, \tag{2}$$

with the special property that if E is any 3-orbit in the decomposition, then $-E$ is *not* in the decomposition. Indeed, since G acts transitively on the set of 3-orbits, it suffices to prove the assertion when $A = \{a, a', a''\}$ is the 3-orbit (1). Then the unique decomposition is the following:

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \cup \begin{bmatrix} -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \cup \begin{bmatrix} -1 & 0 & 1 \\ 2 & -1 & -1 \end{bmatrix} \cup \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}. \tag{3}$$

To get the decomposition, take as A the 3-orbit (1). For $B = \{b, b', b''\}$, take $b := -a$. Then $b' \neq a'$ since $a' \in A$, and $b' \neq -a'$ by the special property. Hence the first coordinate of b' is equal to 2. Therefore, the first coordinate of b'' is equal to -1. As $b'' \neq a''$, it follows that the second coordinate of b'' is equal to 1. Therefore, the second coordinate of b' is equal to -1. Now it is easy to fill in the remaining 6 columns.

Let Y be the set of all decompositions with the special property. It follows that the set Y has 4 elements. The group G acts on the set Y. By the special property, if $\mathcal{D}$ is a decomposition in Y, then the composition $-\mathcal{D}$ is different from $\mathcal{D}$. Hence the matrix $-1 = s^2$ acts non-trivially on Y. Consider the matrix u of order 3. There are 4 decompositions $\mathcal{D}$ in the set Y, so at least one decomposition $\mathcal{D}$ is invariant under u. But then also $-\mathcal{D}$ is invariant under u, and hence also the two remaining decompositions in the set Y are invariant under u. Thus u acts trivially on Y. Since G is generated by u and s, it follows that the image of the representation,

$$G \to \mathrm{Aut}(Y) = S_4,$$

is a cyclic group of order 4. Thus the representation may be viewed as a surjective character,

$$\chi_4 \colon G \to C_4,$$

where the image is generated by the image of s (or by the image of t, since $s = tu$). The kernel of χ_4 is a normal subgroup $G^{(4)}$ of order 12 in G. As $G^{(4)}$ contains all order-3 elements of G, it is generated by the order-3 elements of G. In particular, $G^{(4)}$ is the unique normal subgroup of index 4 in G.

(ii) Reduction modulo 2 is a surjective homomorphism $G \to \mathrm{SL}_2(\mathbf{Z}/2)$. Hence the kernel of the reduction is a normal subgroup $G(2)$ of order 8. It is easily described: it consists of the matrices having two equal units in the diagonal and two non-units off the diagonal. In particular, all its elements are of order 2. Thus $G(2)$ it is commutative and elementary abelian (isomorphic to $C_2 \times C_2 \times C_2$).

(iii) Consider the intersection,

$$V := G^{(4)} \cap G(2).$$

It is a normal subgroup of G. It can be described as the kernel of the restriction $G^{(4)} \to \mathrm{SL}_2(\mathbf{Z}/2) = S_3$. As $G^{(4)}$ is generated by the order-3 elements of G, the image of the restriction $G^{(4)} \to S_3$ is the subgroup A_3 of order 3 in S_3. Hence V is of order 4 and, being a subgroup of $G(2)$, V is isomorphic to $C_2 \times C_2$. Since the decomposition (3) is invariant under V, it follows that V consists of the following four matrices:

$$V: \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 2 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}. \tag{4}$$

(iv) Denote by G_θ the subgroup of G consisting of matrices whose reduction modulo 2 is equal to one of the two matrices,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

In other words, with respect to the reduction map $\mathrm{SL}_2(\mathbf{Z}/4) \to \mathrm{SL}_2(\mathbf{Z}/2)$, the subgroup G_θ is the preimage of the subgroup of the target denoted G_θ in Example (2.1). Alternatively, G_θ can be described as the subgroup of G stabilizing the following set of 4 columns in $((\mathbf{Z}/4)^2)^*$:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}. \tag{5}$$

Clearly, G_θ is a subgroup of index 3 (and order 16), it contains the subgroup $G(2)$ and in addition the matrix s. It follows that restriction defines a surjective character,

$$\chi_4 : G_\theta \to C_4.$$

The kernel $G_\theta \cap G^{(4)}$ is of order 4, and hence equal to V.

(v) The subgroup G_θ has a second character defined as follows: When the four columns of (5) are multiplied by 2, they become equal, namely equal to the column n whose two coordinates are equal to 2. Now, there are four subsets $\{x, x'\}$ with two elements of X such that $x + x' = n$, namely the following four:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix}. \tag{6}$$

Clearly, the column n is invariant under the matrices of G_θ. Therefore, the group G_θ acts on the set Z consisting of the four subsets (6). Hence we obtain a representation,

$$G_\theta \to \mathrm{Aut}(Z) = S_4.$$

Obviously, the matrix $s^2 = -1$ acts nontrivially on Z. Moreover, the set W consisting of following four matrices is easily seen to act trivially on Z:

$$W: \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}. \tag{7}$$

Since G_θ is of order 16, it follows that the representation is in fact a surjective character,

$$\chi_\theta : G_\theta \to C_4,$$

and that its kernel is equal to W.

(2.4) Remark. Let $\Gamma = \mathrm{SL}_2(\mathbf{Z})$ be the modular group. Its center is the subgroup ± 1. Consider the commutator quotient $\overline{\Gamma} := \Gamma/\Gamma'$. As Γ is generated by the matrices s and u, it follows that $\overline{\Gamma}$ is generated by the images $\bar{s}$ and $\bar{u}$. Moreover, since $s^4 = u^3 = 1$, the order of $\bar{s}$ is a divisor of 4 and the order of $\bar{u}$ of order a divisor of 3. Therefore, being commutative, the quotient is cyclic and its order is a divisor of 12.

Now, by Example (2.1), the quotient $\mathrm{SL}_2(\mathbf{Z}/2)$ of Γ has a surjective character onto C_2. Hence, by composition, we obtain a surjective character

$$\chi_2 \colon \Gamma \to \mathrm{SL}_2(\mathbf{Z}/2) \to C_2.$$

The kernel is denoted $\Gamma^{(2)}$. Similarly, from Example (2.2), we obtain a surjective character,

$$\chi_3 \colon \Gamma \to \mathrm{SL}_2(\mathbf{Z}/3) \to C_3,$$

whose kernel is denoted $\Gamma^{(3)}$.

It follows from the Chinese Remainder Theorem that the intersection $\Gamma^{(6)} := \Gamma^{(2)} \cap \Gamma^{(3)}$ is in fact the kernel of the surjective character,

$$\chi_6 \colon \Gamma \to \mathrm{SL}_2(\mathbf{Z}/6) \to C_2 \times C_3 = C_6.$$

As a consequence, the order of $\bar{u}$ is equal to 3 and the order of $\bar{s}$ is equal to 2 or 4. Clearly, as $s^2 = -1$, the order of $\bar{s}$ is equal to 4 if and only if $-1 \notin \Gamma'$. In fact, it follows from the more complicated example (2.3) that there is a surjective character $\chi_4 \colon \Gamma \to C_4$. Hence the order of $\bar{s}$ is equal to 4 and the commutator quotient $\overline{\Gamma}$ is cyclic of order 12.

Therefore, in an obvious notation, the commutator subgroup Γ' is equal to $\Gamma^{(12)}$, where $\Gamma^{(12)} = \Gamma^{(4)} \cap \Gamma^{(3)}$ is the kernel of a surjective character

$$\chi_{12} \colon \Gamma \to C_4 \times C_3 = C_{12}.$$

3. Congruence subgroups.

(3.1) Definition. By Proposition (1.5), reduction modulo N is a surjective homomorphism of groups,

$$\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/N).$$

The kernel of the reduction map is denoted $\Gamma(N)$. Thus $\Gamma(1)$ is the full modular group $\mathrm{SL}_2(\mathbf{Z})$, and $\Gamma(N)$ is a normal subgroup. A subgroup of $\Gamma(1)$ containing $\Gamma(N)$ is called a *level-N (congruence) subgroup*. As the reduction map is surjective, the level-N subgroups correspond bijectively to the subgroups of $\mathrm{SL}_2(\mathbf{Z}/N)$.

Important series of level-N subgroups are defined by the following congruence conditions modulo N on the usual four entries a, b, c, d of a matrix in $\mathrm{SL}_2(\mathbf{Z})$:

$$\begin{aligned}
\tilde{\Gamma}(N): &\quad a \equiv d \equiv \pm 1, \quad b \equiv c \equiv 0,\\
\Gamma_0(N): &\quad c \equiv 0,\\
\Gamma_{01}(N): &\quad c \equiv 0,\ d \equiv 1,\\
\Gamma_0^0(N): &\quad c \equiv 0,\ b \equiv 0.
\end{aligned}$$

The four subgroups correspond respectively to the following subgroups of $\mathrm{SL}_2(\mathbf{Z}/N)$: the subgroup ± 1, the subgroup of upper triangular matrices, the unipotent subgroup (upper triangular matrices with 1 in the diagonal), and the subgroup of diagonal matrices. Note that the subscripts on the middle two subgroups refer to the second row (c, d) modulo N of the matrix (in the literature, Γ_{01} is sometimes denoted Γ_1). Similar congruence subgroups $\Gamma^0(N)$ and $\Gamma^{10}(N)$ are defined using the first row (a, b).

The subgroups $\Gamma(N)$ for $N > 2$ are inhomogeneous, that is, they do not contain the matrix -1. In general, if Γ is a subgroup of $\mathrm{SL}_2(\mathbf{C})$, we denote by $\tilde{\Gamma}$ the *homogenized group* $\Gamma \cup (-\Gamma)$. Then $\tilde{\Gamma}$ is a homogeneous group, and the two groups Γ and $\tilde{\Gamma}$ have the same image in $\mathrm{PSL}_2(\mathbf{C})$, that is, $\mathrm{P}\Gamma = \mathrm{P}\tilde{\Gamma}$. If Γ is inhomogeneous, then Γ is of index 2 in $\tilde{\Gamma}$, and $\Gamma \xrightarrow{\sim} \mathrm{P}\Gamma$. Clearly, the congruence group $\tilde{\Gamma}(N)$ is the homogenized group of $\Gamma(N)$. Note that, for $N = 2$, we have that $\Gamma(2) = \tilde{\Gamma}(2)$. The groups $\Gamma_{01}(N)$ and $\Gamma^{10}(N)$ are inhomogeneous, and they define homogeneous groups $\tilde{\Gamma}_{01}(N)$ and $\tilde{\Gamma}^{10}(N)$.

(3.2) Observation. Congruence subgroups have finite index in $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$, because the quotient $\Gamma(1)/\Gamma(N)$ is the finite group $G_N := \mathrm{SL}_2(\mathbf{Z}/N)$. In fact, the order of G_N is determined in Proposition (1.8), and it follows that

(1) the index of $\Gamma(N)$ is equal to $N^3 \prod_{p|N}(1 - \frac{1}{p^2})$.

It is easy to determine the orders of the various subgroups of G_N that define the special level-N subgroups of (3.1). It follows that

(2) the index of $\tilde{\Gamma}(N)$ is equal to $\frac{1}{2}N^3 \prod_{p|N}(1 - \frac{1}{p^2})$ when $N > 2$.
(3) the index of $\Gamma_0(N)$ is equal to $N \prod_{p|N}(1 + \frac{1}{p})$.
(4) the index of $\Gamma_{01}(N)$ is equal to $N^2 \prod_{p|N}(1 - \frac{1}{p^2})$.
(5) the index of $\Gamma_0^0(N)$ is equal to $N^2 \prod_{p|N}(1 + \frac{1}{p})$.

Indeed, the group $\tilde{\Gamma}(N)$ is the preimage of the subgroup ± 1 of G_N, and so the index in (2) is obtained by dividing the order of G_N by 2. Similarly, in G_N the number of upper triangular matrices is equal to the product of the orders of $\mathbf{Z}/N$ and $(\mathbf{Z}/N)^*$. Hence the index in (3) is obtained by dividing the order of G_N by $N\varphi(N) = N^2 \prod_{p|N}(1 - \frac{1}{p})$. The indices in (4) and (5) are obtained similarly.

(3.3) Example. Level-2 subgroups correspond to the subgroups of $G = \mathrm{SL}_2(\mathbf{Z}/2)$ considered in Example (2.1). The preimage, under the reduction modulo 2, of the subgroup $G^{(2)}$ is a normal subgroup $\Gamma^{(2)}$ of index 2 in $\Gamma(1)$. It is equal to the kernel of a surjective character,

$$\chi_2 \colon \Gamma(1) \to C_2.$$

The preimages of the subgroups G_0 and G^0 are the subgroups $\Gamma_{01}(2) = \Gamma_0(2)$ and $\Gamma^{01}(2) = \Gamma^0(2)$. In addition, the preimage of the subgroup G_θ is a level-2 subgroup Γ_θ, called the *θ-group*. The latter three subgroups are non-normal subgroups of index 3.

(3.4) Example. The subgroup $\Gamma(3)$ is of index 24, and the homogeneous group $\tilde{\Gamma}(3)$ is of index 12. The homogeneous level-3 subgroups correspond to the subgroups of the group $\mathrm{PSL}_2(\mathbf{Z}/3)$ considered in (1.10)(3). In particular, the normal subgroup $\Gamma^{(3)}$ introduced in Remark (2.4) is a level-3 subgroup of index 3, equal to the kernel of a surjective character,

$$\chi_3 \colon \Gamma(1) \to C_3.$$

(3.5) Example. The subgroup $\Gamma(4)$ is of index 48. Various level-4 subgroups correspond to the subgroups of $G = \mathrm{SL}_2(\mathbf{Z}/4)$ considered in Example (2.3). The preimage, under the reduction modulo 4, of the subgroup $G^{(4)}$ is a normal subgroup $\Gamma^{(4)}$ of index 4, equal to the kernel of a surjective character,

$$\chi_4 \colon \Gamma(1) \to C_4.$$

The preimage of $G(2)$ is the level-2 subgroup $\Gamma(2)$ of index 6. The intersection $\Gamma^{(4)} \cap \Gamma(2)$ is the preimage of V; it is a normal subgroup of index 12 in $\Gamma(1)$, and it is denoted Γ_V. Note that $\Gamma^{(4)}$ and Γ_V are non-homogeneous. Their homogenized groups are respectively $\Gamma^{(2)}$ and $\Gamma(2)$.

The preimage of the subgroup G_θ is the θ-group Γ_θ of index 3. It contains the group $\Gamma(2)$, and in particular the normal subgroup Γ_V. Hence restriction defines a surjective character,

$$\chi_4 \colon \Gamma_\theta \to C_4.$$

In addition, there is a surjective character,

$$\chi_\theta \colon \Gamma_\theta \to C_4,$$

whose kernel is the preimage Γ_W of W. Note that Γ_W is of index 12 in $\Gamma(1)$, but not normal.

(3.6) Exercise. Prove that the conjugate group $\Gamma_\theta^u = u^{-1}\Gamma_\theta u$ is equal to $\Gamma_0(2)$. [Hint: reduce $s^u = u^{-1}su$ modulo 2.]

(3.7) Exercise. The group $\Gamma(2)$ is the homogenized group of Γ_V. The character χ_4 on $\Gamma(1)$ restricts to a surjective character,

$$\chi_4 \colon \Gamma(2) \to C_2.$$

By example (2.3)(iii), the kernel of the restricted character is equal to Γ_V. Prove that the restricted character on $\Gamma(2)$ is determined by the formula,

$$\chi_4 \begin{bmatrix} a & b \\ c & d \end{bmatrix} = (-1)^{(a+b+c-1)/2}.$$

[Hint: Clearly, for a matrix in $\Gamma(2)$, we have modulo 2 that $a \equiv= 1$ and $c \equiv d \equiv 0$, and hence $a+b+c \equiv 1$. Therefore, modulo 4 we have that $a+b+c \equiv \pm 1$. Now show from the description of V in (2.3) that $a+b+c \equiv 1$ holds for the matrices in Γ_V.]

(3.8) Exercise. Clearly, for a matrix in Γ_θ we have for the numbers c and d that exactly one is odd. Prove that the character $\chi_\theta \colon \Gamma_\theta \to C_4$ is given by the formula,

$$\chi_\theta \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{cases} 1 & \text{for } d \equiv 1 \pmod 4, \\ i & \text{for } c \equiv 1 \pmod 4, \\ -1 & \text{for } d \equiv -1 \pmod 4, \\ -i & \text{for } c \equiv -1 \pmod 4, \end{cases}$$

where $i := \chi_\theta(s)$. [Hint: inspect the cosets of G_θ / W.]

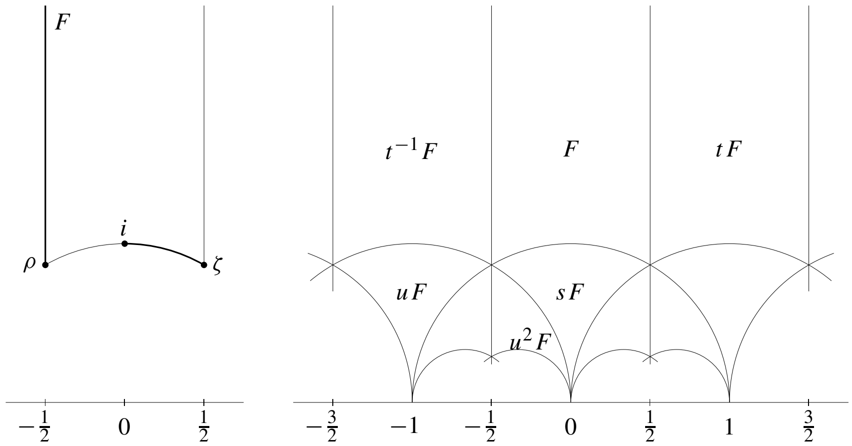
4. Some modular fundamental domains.

(4.0) Setup. In this section we describe fundamental domains for some congruence subgroups of the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ acting on the upper half plane $\mathfrak{H}$. The congruence subgroups are of finite index in $\Gamma(1)$. Recall that if F is a fundamental domain for a discrete group Γ and Δ is a subgroup of finite index in Γ with a given system γ_i of representatives for the right cosets of Δ in Γ, then the union G of the transforms $\gamma_i F$ is a fundamental domain for Δ. Moreover, if γ_L is the systems of boundary transformations corresponding to the side L of F, then a system of boundary transformations of G are obtained as follows: Consider a side of G, say $\gamma_j L$ where L is a side of F. Then $\gamma_L L$ is a side of F. According to the coset representation, we have

$$\gamma_j \gamma_L^{-1} = \delta_{j,L}^{-1} \gamma_k$$

with a unique $\delta_{j,L} \in \Delta$. Then $\delta_{j,L}\gamma_j L = \gamma_k \gamma_L L$ is a side of G, and so $\delta_{j,L}$ is the boundary transformation corresponding to the side $\gamma_j L$.

(4.1) Example. Take as F the fundamental domain F for the full modular group $\Gamma(1)$ described in Proposition (Discr.1.6). Each transform γF for γ in $\Gamma(1)$ is a again a fundamental domain, and the transforms cover $\mathfrak{H}$.



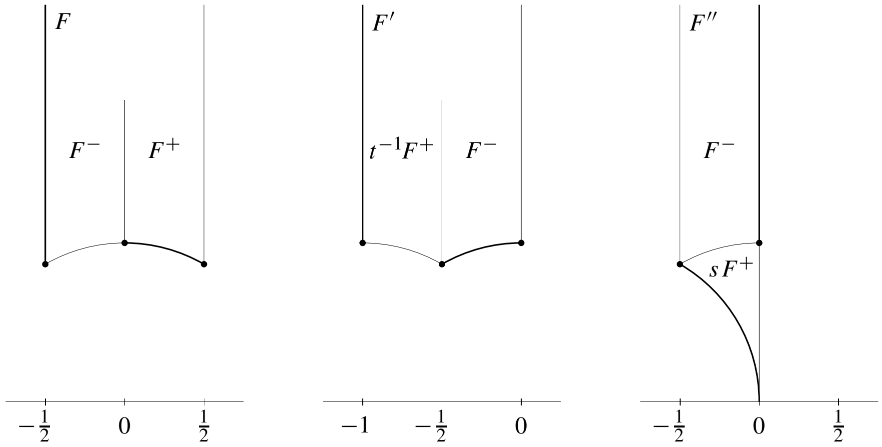
The fundamental domain F for $\Gamma(1)$, and some of its transforms.

It follows from the description that under the action of $\Gamma(1)$ there are exactly two elliptic orbits, namely one of order 2 represented by the point i and one of order 3 represented by the point ρ. In addition, there is only one parabolic orbit, represented by the point ∞.

Moreover, it follows from the description of the equivalence on the boundary of F that the orbit space $\mathfrak{H}/\Gamma$ (topologically) is a 2-sphere minus a point.

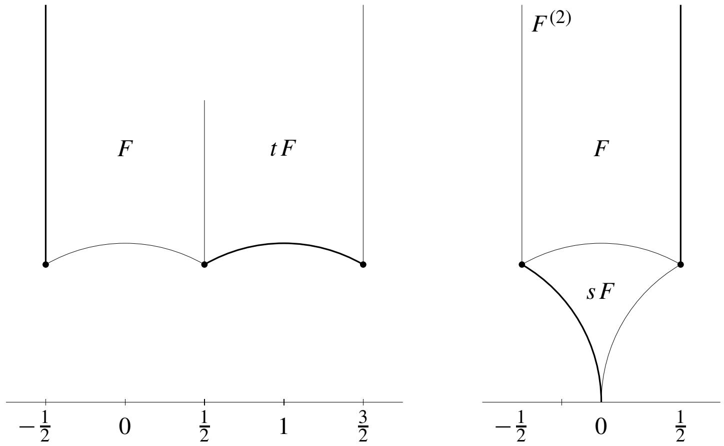
(4.2) Example. Occasionally, to obtain different fundamental domains, it is convenient to cut a domain in subdomains, and then to apply different γ's to the different pieces. For instance, F can be cut into two domains, $F = F^- \cup F^+$, defined respectively by the inequalities: $\Re z \geq 0$ and $\Re z \leq 0$. Then $F' := t^{-1}F^+ \cup F^-$ is a different fundamental domain for $\Gamma(1)$

and so is $F'' := sF^+ \cup F^-$. For the two latter domains, the boundary transformations are t, u and s, u.



The fundamental domains F, F' and F'' for $\Gamma(1)$.

(4.3) Example. The group $\Gamma^{(2)}$ is a normal level-2 subgroup of index 2 in $\Gamma(1)$, equal to the kernel of the character $\chi_2 \colon \Gamma(1) \to C_2$. Representatives for the cosets are $1, t$. Hence the union $F \cup tF$ is a fundamental domain for $\Gamma^{(2)}$. Similarly, the union $F^{(2)} := F \cup sF$ is a fundamental domain.



Fundamental domains for $\Gamma^{(2)}$.

Consider the first domain $F \cup tF$. It has three finite vertices ρ, ζ and $\rho+2$, and one infinite vertex ∞. Its boundary transformations are t^2 and $v := st^{-1}$. The angles at the vertices ρ and $\rho+1$ are equal to $2\pi/6$, the angle at ζ is equal to $2\pi/3$ and the width at ∞ is equal to 1. Hence, there are two elliptic orbits of order 3, represented by ρ and ζ, and one parabolic orbit represented by ∞. Clearly, the boundary transformations are t^2 and st^{-1}. Topologically, the orbit space $\mathfrak{H}/\Gamma^{(2)}$ is a 2-sphere minus a point.

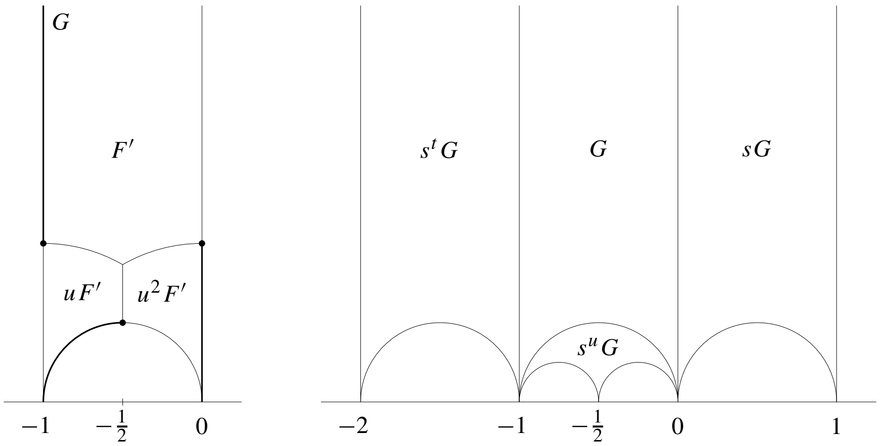
Similarly, the second domain $F^{(2)}$ has 2 finite vertices ρ and ζ, and two infinite vertices ∞ and 0. The angles at ρ and ζ are equal to $2\pi/3$ and the widths at 0 and ∞ are equal to $\frac{1}{2}$.

The boundary transformations are defined by the two matrices

$$u = t^{-1}s = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \quad v = st^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}. \tag{4.3.1}$$

As a consequence, modulo ± 1, the group $\Gamma^{(2)}$ is generated by u and v.

(4.4) Example. The group $\Gamma^{(3)}$ is a normal level-3 subgroup of index 3 in $\Gamma(1)$, equal to the kernel of the character $\chi_3 \colon \Gamma(1) \to C_3$. Representatives for the cosets are the three matrices 1, u, and u^2. From the fundamental domain F' of (4.1), we obtain the fundamental domain $G := F' \cup uF' \cup u^2F'$ for $\Gamma^{(3)}$.



The fundamental domain G for $\Gamma^{(3)}$, and some of its transforms.

The domain G has no finite vertices, and 3 infinite vertices -1, 0, and ∞. The boundary transformations correspond to the matrices,

$$s = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad s^t = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}, \quad s^u = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}. \tag{4.4.1}$$

It follows that there are 3 elliptic orbits, all of order 2, represented by the points i, $i-1$, and $-\frac{1}{2} + \frac{i}{2}$. In addition, the three infinite vertices are equivalent, so there is only one parabolic orbit. The orbit space $\mathfrak{H}/\Gamma^{(3)}$ is a 2-sphere minus a point.

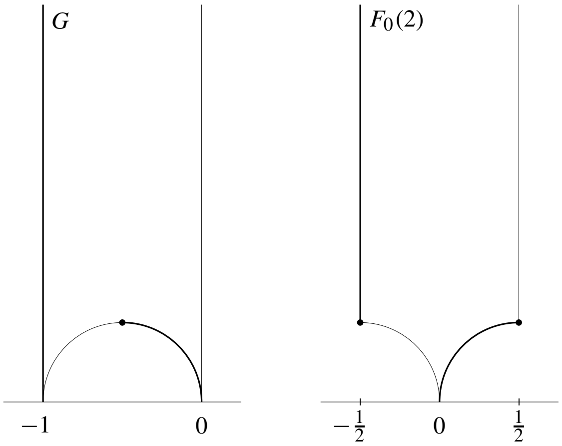
As a consequence, the group $\Gamma^{(3)}$ is generated by the 3 matrices of (4.4.1).

(4.5) Example. The congruence subgroup $\Gamma_0(2)$ is a level-2 subgroup of index 3 in $\Gamma(1)$. The three matrices $1, u, u^2$ represents the right cosets modulo $\Gamma_0(2)$. So, the domain G of Example (4.4) is also a fundamental domain for $\Gamma_0(2)$, but, of course, the boundary transformations are different. For G as a fundamental domain for $\Gamma_0(2)$, the boundary transformations are associated to the matrices,

$$t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad s^u = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}. \tag{4.5.1}$$

There is one elliptic orbit, of order 2, represented by the point $-\frac{1}{2}+\frac{i}{2}$ on the boundary. It is fixed under s^u. Of the infinite vertices, -1 and 0 are equivalent. So there are 2 cusps, and $\mathfrak{H}/\Gamma_0(2)$ is a 2-sphere minus 2 points.

As $t \in \Gamma_0(2)$, we can, as in Example (4.2), obtain a different fundamental domain $F_0(2)$ by cutting G by a vertical line and translating the left piece by t.



The fundamental domains G and $F_0(2)$ for $\Gamma_0(2)$.

The domain $F_0(2)$ has two finite vertices, $\pm\frac{1}{2}+\frac{i}{2}$, both of angle $2\pi/4$ and representing (of course) the same elliptic point. In addition, $F_0(2)$ has two infinite vertices 0 and ∞. The boundary transformations of $F_0(2)$ correspond to the matrices,

$$t=\begin{bmatrix}1&1\\0&1\end{bmatrix},\quad s^u t^{-1}=\begin{bmatrix}-1&0\\2&-1\end{bmatrix} \tag{4.5.2}$$

As a consequence, the two matrices in (4.5.1) (or, up to ± 1 the two matrices in (4.5.2)) generate the group $\Gamma_0(2)$.

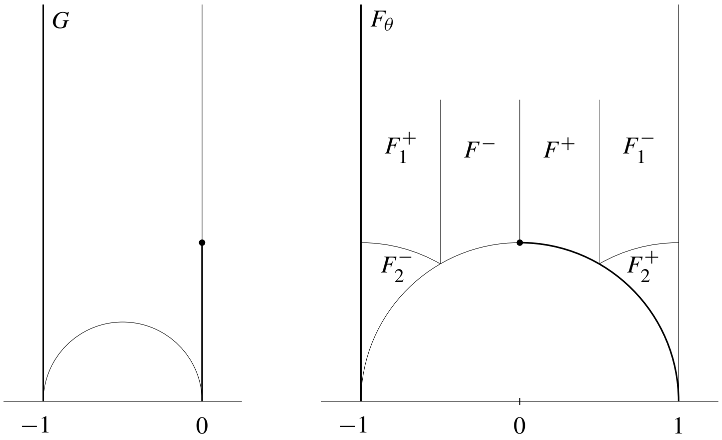
(4.6) Example. The θ-group Γ_θ is a level-2 subgroup of index 3, and as in the previous two examples, the matrices $1, u, u^2$ represent the right cosets. Hence, again, the domain G is also a fundamental domain for Γ_θ.

From the decomposition $F = F^+ \cup F^-$ of Example (4.2), a different fundamental Γ_θ can be obtained as follows: The three matrices $1, t^{-1}, ts$ form a different set of representatives for the cosets modulo Γ_θ, and so do the three matrices $1, t, t^{-1}s$. Apply the first set of matrices to the subdomain F^+ of F and the second set to the subdomain F^-. The result is a fundamental domain F_θ for Γ_θ which is the union of the following 6 pieces: F^+, $F_1^+ := t^{-1}F^+$, $F_2^+ := tsF^+$, F^-, $F_1^- := tF^-$, and $F_2^- := t^{-1}sF^-$. The pieces fit together, and they form a domain like G with three infinite vertices and no finite vertices.

For G as a fundamental domain for Γ_θ, the boundary transformations correspond to the matrices,

$$s=\begin{bmatrix}0&-1\\1&0\end{bmatrix},\quad st^2=\begin{bmatrix}0&-1\\1&2\end{bmatrix}. \tag{4.6.1}$$

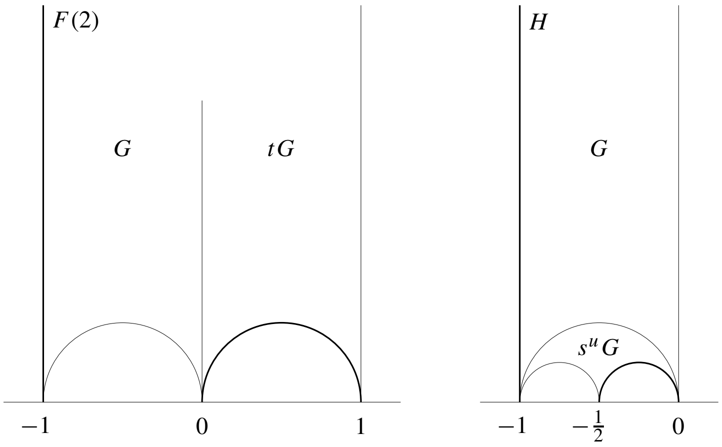
There is one elliptic orbit, of order 2, represented by the point i. Of the infinite vertices, ∞ and 0 are equivalent. The orbit space $\mathfrak{H}/\Gamma_\theta$ is a 2-sphere minus 2 points.



The fundamental domains G and F_θ for Γ_θ.

For the fundamental domain F_θ, the boundary transformations are obviously s and t^2. In particular, the θ-group Γ_θ is generated by the two matrices s and t^2.

(4.7) Example. The group $\Gamma(2)$ is a normal level-2 subgroup of index 6, equal to the kernel of $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/2) = S_3$. Coset representatives are the six matrices $1, u, u^2, t, tu, tu^2$. The group $\Gamma(2)$ is a subgroup of $\Gamma_0(2)$ and 1, t represent the cosets in $\Gamma_0(2)$. Hence, with G is the fundamental domain of (4.5), the union $F(2) := G \cup tG$ is a fundamental domain for $\Gamma(2)$. Clearly, the matrix s^u of (4.3.1) is in $\Gamma_0(2)$ and not in $\Gamma(2)$. Hence, the union $H := G \cup s^u G$ is a second fundamental domain for $\Gamma(2)$.



The fundamental domains $F(2)$ and H for $\Gamma(2)$.

The domain $F(2)$ has no finite vertices; its infinite vertices are $-1, 0, 1, \infty$. The boundary transformations are defined by the two matrices,

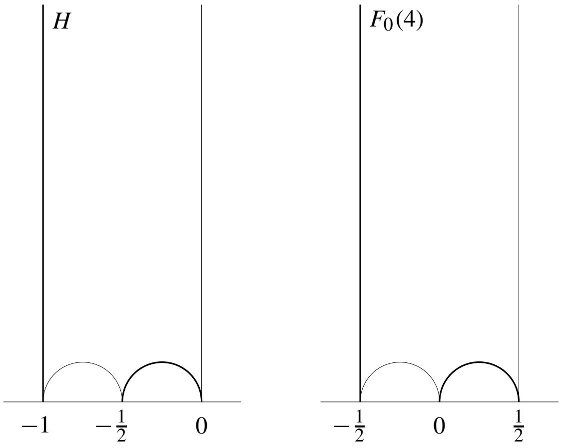
$$t^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad st^{-2}s = \begin{bmatrix} -1 & 0 \\ 2 & -1 \end{bmatrix}. \tag{4.7.1}$$

Thus, up to ± 1, the group $\Gamma(2)$ is generated by the matrices (4.7.1). Clearly, there are no elliptic points. Of the infinite vertices, -1 and 1 are equivalent. Hence, the orbit space $\mathfrak{H}/\Gamma(2)$ is a sphere minus 3 points.

The group $\Gamma(2)$ is the homogenization of the inhomogeneous level-4 subgroup Γ_V defined in Example (3.4). Hence the two groups have the same fundamental domain. As a consequence, Γ_V is generated by the two matrices,

$$-t^{-2} = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \quad st^{-2}s = \begin{bmatrix} -1 & 0 \\ 2 & -1 \end{bmatrix}. \tag{4.7.2}$$

(4.8) Example. The group $\Gamma_0(4)$ is a level-4 subgroup of index 6. It is a subgroup of $\Gamma_0(2)$, of index 2. Clearly, the matrix s^u of (4.3.1) belongs to $\Gamma_0(2)$ and not to $\Gamma_0(4)$, and hence it represents the non-trivial coset of $\Gamma_0(4)$ in $\Gamma_0(2)$. Hence the fundamental domain $H := G \cup s^u G$ for $\Gamma(2)$ described in Example (4.7) is also a fundamental domain for $\Gamma_0(4)$. Since t is in $\Gamma_0(4)$, the latter domain can be cut in two by a vertical line, and we can obtain a second fundamental domain $F_0(4)$ by translating the left piece by t.



The fundamental domains H and $F_0(4)$ for $\Gamma_0(4)$.

The domain H has 4 infinite vertices, ∞, -1, $-\frac{1}{2}$, and 0. Its boundary transformations are t and $s^u t (s^u)^{-1}$. For the domain $F_0(4)$, the boundary transformations correspond to the matrices,

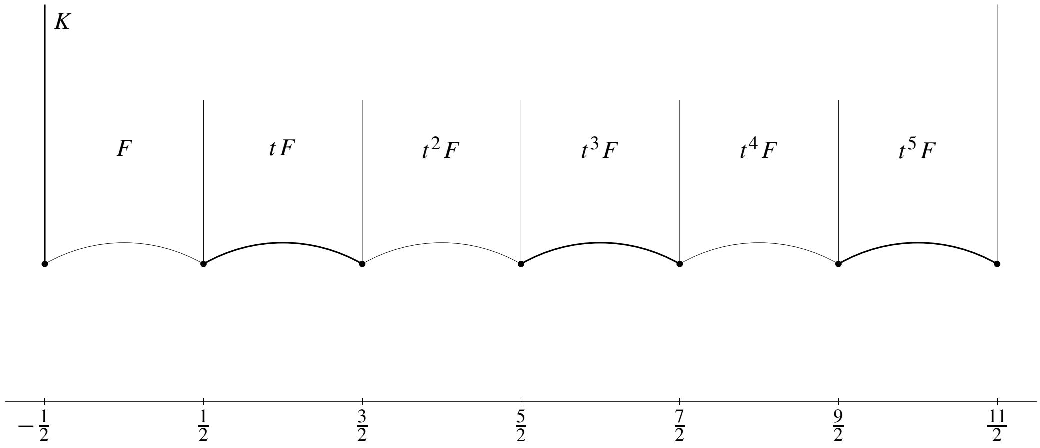
$$t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad s^u t^{-1} (s^u)^{-1} t^{-1} = \begin{bmatrix} -1 & 0 \\ 4 & -1 \end{bmatrix}. \tag{4.8.1}$$

In particular, $\Gamma_0(4)$ modulo ± 1 is generated by the two matrices (4.8.1).

The orbit space $\mathfrak{H}/\Gamma_0(4)$ is a sphere minus 3 points.

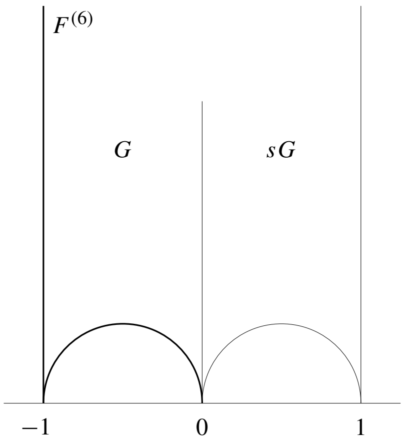
(4.9) Example. The group $\Gamma^{(6)}$ is a normal level-6 subgroup. It is the kernel of a surjective character $\chi_6\colon \mathrm{SL}_2(\mathbf{Z}) \to C_6$. The cyclic group C_6 is generated by the image of t. Thus the powers t^i for $i = 0, \dots, 5$ are representatives for the cosets, and a fundamental domain K for $\Gamma^{(6)}$ is obtained as the union of the translates $t^i F$ for $i = 0, \dots, 5$.

The domain K has ∞ as its only infinite vertex. The finite vertices are the 7 points with imaginary part $\frac{1}{2}\sqrt{3}$ and real parts $\frac{1}{2}j$ for $j = -1, 1, 3, 5, 7, 9, 11$. To get the boundary transformations, note that modulo $\Gamma^{(6)}$ we have that $s \equiv t^3$. It follows easily that the boundary transformations are t^6, st^{-3}, t^4st^{-1}, and t^2st^{-5}. Clearly, of the seven finite vertices, the three with real parts $\frac{1}{2}$, $\frac{5}{2}$ and $\frac{9}{2}$ are equivalent and the remaining four are equivalent. The angles in both classes add up to 2π. Hence, there are no elliptic points for $\Gamma^{(6)}$. It does require some feeling for cutting and pasting to see from the domain K that the orbit space $\mathfrak{H}/\Gamma^{(6)}$ is a torus minus one point.



The fundamental domain K for $\Gamma^{(6)}$.

A more manageable fundamental domain for $\Gamma^{(6)}$ is the domain $F^{(6)}$:



The fundamental domain $F^{(6)}$ for $\Gamma^{(6)}$.

It is obtained as follows: The level-6 group $\Gamma^{(6)}$ is contained in the group $\Gamma^{(}3)$ of Example (4.4), and G was a fundamental domain for the latter group. The boundary transformation s of G is in $\Gamma^{(3)}$ and not in $\Gamma^{(6)}$. Hence the two cosets of $\Gamma^{(6)}$ in $\Gamma^{(3)}$ are represented by the two matrices 1 and t. Hence a different fundamental domain for $\Gamma^{(6)}$ is obtained as the union $F^{(6)} := G \cup sG$.

As $sG = tG$, the union $F^{(6)} = G \cup sG$ is equal to the fundamental domain $F(2)$ considered in Example (4.7), but the boundary transformations are different. The domain has 4 infinite vertices, -1, 0, 1, and ∞, and it has no finite vertices.

The boundary transformations correspond to the two matrices,

$$tst^2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad t^2st = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}. \tag{4.9.1}$$

It follows (again) for $\Gamma^{(6)}$ that there are no elliptic points and exactly one cusp, and the orbit space is a torus minus a point. Modulo ± 1, the group $\Gamma^{(6)}$ is generated by the two matrices (4.9.1).

Automorphic forms

1. Automorphic factors.

(1.1) Definition. Fix an action of a group G on a set X. As usual, when a second set Y is given, the group G acts on the right on the set Y^X of all functions $f\colon X \to Y$ by the definition,

$$f\sigma := f\sigma_X, \tag{1.1.0}$$

where σ_X is the automorphism $x \mapsto \sigma x$ of X.

Let H be a second group. Then an H-valued *automorphic factor* for the action of G on X is a map $j\colon G \times X \to H$ satisfying for all σ, τ in G and all x in X the *automorphy equation*,

$$j(\sigma\tau, x) = j(\sigma, \tau x)j(\tau, x). \tag{1.1.1}$$

Alternatively, the map j may be viewed as a map $\sigma \mapsto j_\sigma$ from G to the group H^X of all maps $X \to H$, and the automorphy equation is the following:

$$j_{\sigma\tau} = (j_\sigma \tau_X)j_\tau. \tag{1.1.2}$$

Assume that H acts on the left on the set Y. Then, for every H-valued automorphic factor j, a right action of G on the set Y^X of all functions $f\colon X \to Y$ is obtained by the definition,

$$f \cdot_j \sigma = j_\sigma^{-1}(f\sigma_X), \tag{1.1.3}$$

or, with arguments,

$$(f \cdot_j \sigma)(x) = j(\sigma, x)^{-1} f(\sigma x). \tag{1.1.4}$$

The action of G on Y^X is called the *automorphic action* defined by the factor j. Note that a function $f\colon X \to Y$ is G-invariant with respect to the automorphic action if and only if, for all x in X and all σ in G,

$$f(\sigma x) = j(\sigma, x)f(x). \tag{1.1.5}$$

(1.2) Note. Clearly, the constant map $j(\sigma, x) = 1$ is an automorphic factor. The corresponding automorphic action on Y^X is given by (1.1.0). More generally, automorphic factors $j(\sigma, x)$ that are independent of x correspond to homomorphisms of groups $\chi\colon G \to H$. Functions $f\colon X \to Y$ that are invariant under the automorphic action corresponding to χ are often called *semi invariant*. They are characterized by the equation,

$$f(\sigma x) = \chi(\sigma)f(x).$$

(1.3) Note. Assume that the group G acts on the right on a group N, by group automorphisms of N. In other words, assume given, for σ in G, a group automorphism of N denoted $n \mapsto n^\sigma$, such that $n^{\sigma\tau} = (n^\sigma)^\tau$. Then the *semi-direct product* is the product set $G \times N$ with the group composition given by the equation,

$$(\sigma, n) \cdot (\tau, m) := (\sigma\tau, n^\tau m). \tag{1.3.1}$$

Obviously, the two maps $\sigma \mapsto (\sigma, 1)$ and $n \mapsto (1, n)$ identify G and N as subgroups of product, and the pair (σ, n) is the product $\sigma \cdot n$. Under the identification, the composition (1.3.1) is essentially given by the commutation rule,

$$n \cdot \sigma = \sigma \cdot n^\sigma,$$

or equivalently, by $n^\sigma = \sigma^{-1} \cdot n \cdot \sigma$. The projection $p : (\sigma, n) \mapsto \sigma$ is a surjective homomorphism of groups $p\colon G \times N \to G$, and its kernel is equal to N. The inclusion $\sigma \mapsto (\sigma, 1)$ is a *section* of p, that is, it is a homomorphism which composed with p is the identity of G. Clearly, the general sections of p are the maps of the form $\sigma \mapsto (\sigma, \nu_\sigma)$, where $\sigma \mapsto \nu_\sigma$ is a 1-*cocycle*, that is, a map $G \to N$ satisfying the following condition,

$$\nu_{\sigma\tau} = (\nu_\sigma)^\tau \nu_\tau.$$

Consider, in the setup of (1.1), the semi-direct product $G \times H^X$. Then, by (1.1.2), the automorphic factors $G \to H^X$ are precisely the 1-cocycles. Equivalently, a map $\sigma \mapsto j_\sigma$ is an automorphic factor if and only if the map $\sigma \mapsto (\sigma, j_\sigma)$ is a homomorphism of groups $G \to G \times H^X$. As a consequence, an automorphic factor j_σ is uniquely determined by its values on a system of generators σ_j for G.

(1.4) Note. Let $j(\sigma, x)$ be an H-valued automorphic factor. Clearly, it follows from the automorphy equation first, that $j(1, x) = 1$, and next that

$$j(\sigma^{-1}, x) = j(\sigma, \sigma^{-1}x)^{-1}. \tag{1.4.1}$$

Assume that H acts on Y, and consider the corresponding automorphic action of G on Y^X. As usual, a right action of G is changed into a left action by composing with the anti-involution $\sigma \mapsto \sigma^{-1}$ of G. It follows from (1.4.1) that the corresponding left action of G on Y^X is given by the equation,

$$\sigma \cdot_j f := (j_\sigma f)\sigma_X^{-1}. \tag{1.4.2}$$

(1.5) Note. In the setup of (1.1), assume that Y is a homogeneous space over H, that is, for any two elements y and y' of Y there is a unique element h in H such that $y' = hy$. Clearly, if some function $f\colon X \to Y$ is G-invariant with respect to the automorphic action defined by j, then j is unique, and given by the equation,

$$j(\sigma, x) = f(\sigma x) f(x)^{-1}. \tag{1.5.1}$$

Conversely, when $f\colon X \to Y$ is any function, then the equation (1.5.1) defines an automorphic factor j.

(1.6) Lemma. *Let α and β be matrices of* $\mathrm{GL}_2(\mathbf{C})$. *Let z be a point of* $\mathbf{C}$, *and assume that the two points βz and $\alpha\beta z$ are different from ∞. Then the following equation holds:*

$$J(\alpha\beta, z) = J(\alpha, \beta z) J(\beta, z). \tag{1.6.1}$$

Moreover, if z is a fixed point of β, then the representatives of z are eigenvectors of β corresponding to the eigenvalue $J(\beta, z)$.

Proof. The points βz and $\alpha\beta z$ are represented by the columns

$$\beta\begin{bmatrix} z \\ 1 \end{bmatrix} = J(\beta, z)\begin{bmatrix} \beta z \\ 1 \end{bmatrix}, \quad \alpha\beta\begin{bmatrix} z \\ 1 \end{bmatrix} = J(\alpha\beta, z)\begin{bmatrix} \alpha\beta z \\ 1 \end{bmatrix}. \tag{1.6.2}$$

It follows that the second equation is obtained by multiplying the first by α. Therefore (1.6.1) holds. Clearly, the last assertion of the Lemma is a consequence of the first equation of (1.6.2). □

(1.7) Observation. Consider a *finite disk* $\mathfrak{D}$, that is, a disk not containing the point ∞. Thus $\mathfrak{D}$ is either the interior of a usual circle in $\mathbf{C}$ or the points of an open half plane in $\mathbf{C}$. In the former case, the point ∞ does not belong to the boundary of $\mathfrak{D}$, in the latter case it does.

It follows from (1.6.1) that, for any subgroup Γ of $\mathrm{SL}(\mathfrak{D})$, the function $J(\gamma, z)$ is a $\mathbf{C}^*$-valued automorphic factor for the action of Γ on $\mathfrak{D}$. More generally, for any integer k, the function $J(\gamma, z)^k$ is an automorphic factor.

Note that the absolute value $|J(\gamma, z)|^k$, for any real number k, is an automorphic factor. Our principal interest is, for certain subgroups Γ, automorphic factors j such that $j(\gamma, z)$ is, at least up to a complex sign, a determination of the function $J(\gamma, z)^k$ where the exponent k is an arbitrary real number.

(1.8) Remark. Fix a real number k. Let w be a non-zero complex number. Recall that a complex number u is called a *value* of w^k, if $|u| = |w|^k$ and if, for some argument θ of w, the real number $k\theta$ is an argument for u. If k is integral, there is only one value of w^k, if k is rational there is a finite number of values, and if k is irrational there is an infinite number of values.

Let $g\colon X \to \mathbf{C}^*$ be a continuous map. Recall that a map $h\colon X \to \mathbf{C}^*$ is called a *determination* of g^k, if h is continuous and, for every x in X, $h(x)$ is a value of $g(x)^k$. If h_0 is a determination of g^k then, for every integer n, the following product is a determination of g^k:

$$e^{2\pi i n k} h_0(x); \tag{*}$$

moreover, if X is connected, then all other determinations are of the form (*).

Determinations of g^k exist if X is simply connected, or if the image of $g(X)$ is contained in a simply connected subset of $\mathbf{C}^*$. In the latter case there exists a continuous argument on $g(X)$, that is, a real valued continuous function Θ defined on $g(X)$ such that $\Theta(w)$ is an argument for w for all w in $g(X)$, and then the following function is a determination,

$$h(x) := |g(x)|^k e^{i\Theta(g(x))k},$$

In most of our applications, the map g will be holomorphic, and the image of g will avoid a half ray starting at the point 0. In this case, the existence of a continuous argument is obvious, and any determination of g^k is again a holomorphic function. For example, if the image $g(X)$ avoids the negative real axis, we may always consider the *principal determination* of g^k, defined using as argument of $g(x)$ the principal argument θ, where $-\pi < \theta < \pi$.

For instance, let $\mathfrak{D}$ be a finite disk and let α be a matrix in $\mathrm{GL}_2(\mathbf{C})$ such that $\alpha z \neq \infty$ for all z in $\mathfrak{D}$. Then the function $J(\alpha, z)$ is (finite) and non-zero everywhere in $\mathfrak{D}$. It has the form $Cz + D$. If $C = 0$, then the function $J(\alpha, z)$ is the non-zero constant D. If $C \neq 0$, then $J(\alpha, z)$ is a Möbius transformation, and in particular, the image of $\mathfrak{D}$ is finite disk not containing the point 0. Therefore, in both cases, there are determinations of the function $J(\alpha, z)^k$. In the first case, any determination is constant, and hence defined on all of $\mathbf{C}$. In the second case, the function $J(\alpha, z)$ has a zero v outside $\mathfrak{D}$. Let V be the open subset of $\mathbf{C}$ obtained by cutting away a half ray starting at v and having no other points in common with the boundary of $\mathfrak{D}$. Then there is a determination of $J(\alpha, z)^k$ defined on all of V. In particular, there is a determination $J(\alpha, z)^k$ defined on $\mathfrak{D}$, and it extends continuously to the points of $\partial\mathfrak{D}$ except possibly to the two points ∞ and $v = \alpha^{-1}\infty$ if they belong to $\partial\mathfrak{D}$.

(1.9) Definition. Let $\mathfrak{D}$ be a finite disk, and let Γ be a subgroup of $\mathrm{SL}(\mathfrak{D})$. A *factor of weight* k on Γ is a function $j(\gamma, z)$ defined for $\gamma \in \Gamma$ and $z \in \mathfrak{D}$ and satisfying the following three conditions:

(1) The function j is a $\mathbf{C}^*$-valued automorphic factor for the action of Γ on $\mathfrak{D}$.
(2) For fixed γ in Γ, the function $j(\gamma, z)$ is holomorphic.
(3) The absolute value $|j|$ is equal to $|J|^k$, that is, for all γ and z,

$$|j(\gamma, z)| = |J(\gamma, z)|^k.$$

The group $\mathbf{C}^*$ acts on itself, it acts on $\mathbf{C}$, and it acts on $\overline{\mathbf{C}}$. In particular, from a given factor j on Γ we obtain a corresponding automorphic action of Γ on the numeric functions on $\mathfrak{D}$. The action is given by the formula,

$$(f \cdot_j \gamma)(z) = \frac{1}{j(\gamma, z)} f(\gamma z). \tag{1.9.1}$$

Numeric functions on $\mathfrak{D}$ that are invariant under the action are called (Γ, j)-*invariant*, or j-*invariant*. They are characterized by the equations, for $\gamma \in \Gamma$ and $z \in \mathfrak{D}$,

$$f(\gamma z) = j(\gamma, z) f(z). \tag{1.9.2}$$

Obviously, the action (1.9.1) is linear. Hence, the j-invariant functions form a vector space over $\mathbf{C}$.

(1.10) Lemma. *Let $\mathfrak{D}$ be a finite disk, and Γ a subgroup of $\mathrm{SL}(\mathfrak{D})$. Assume that j is a factor of weight k on Γ. Fix a matrix γ in Γ, and consider a determination $J(\gamma, z)^k$ on $\mathfrak{D}$. Then there is a complex sign ε (i.e., $|\varepsilon| = 1$), and an equation,*

$$j(\gamma, z) = \varepsilon J(z, \gamma)^k \quad \textit{for all } z \in \mathfrak{D}, \tag{1.10.1}$$

In particular, the function $j(\gamma, z)$ is completely determined by any of its values $j(\gamma, z_0)$ (and the given weight k).

Proof. By Condition (1.9)(3), the quotient $j(\gamma, z)/J(\gamma, z)^k$ is of modulus 1, and holomorphic by Condition (2). Therefore, the quotient is a constant function, and hence equation (1.10.1) holds. □

(1.11) Observation. (1) The constant automorphic factor $j(\gamma, z) = 1$ is a factor of weight 0, and the corresponding invariant functions are simply Γ-invariant functions: $f(\gamma z) = f(z)$ for all γ in Γ.

More generally, any factor $j(\gamma, z)$ of weight 0 is of constant modulus 1, and hence constant. Therefore, the factors of weight 0 on Γ are precisely the unitary characters $\chi : \Gamma \to \mathbf{C}^*$.

(2) Clearly, the product $j_1 j_2$ of factors of weights k_1 and k_2 is a factor of weight $k_1 + k_2$. If f_1 is j_1-invariant and f_2 is j_2-invariant, then the product function $f_1 f_2$ is $j_1 j_2$-invariant.

Similarly, the quotient j_1/j_2 is a factor of weight $k_1 - k_2$. As a consequence, if $j_0(\gamma, z)$ is given factor of weight k, then the factors of weight k are precisely the functions,

$$j(\gamma, z) = \chi(\gamma) j_0(\gamma, z),$$

where $\chi : \Gamma \to \mathbf{C}^*$ is a unitary character. In particular, if the commutator subgroup Γ' is of finite index in Γ, then the number of factors of weight k is either zero or equal to the index $|\Gamma{:}\Gamma'|$.

(3) If k is an integer, then the function $J(\gamma, z)^k$ is a factor of weight k. In particular, it follows from (2) that the factors of integral weight k on Γ are the functions,

$$\chi(\gamma) J(\gamma, z)^k, \tag{1.11.1}$$

where $\chi : \Gamma \to \mathbf{C}^*$ is a unitary character. For an integer k, the action (1.9.1) of Γ on numeric functions corresponding to the factor $J(\gamma, z)^k$ is denoted $f \cdot_k \gamma$, that is,

$$(f \cdot_k \gamma)(z) = \frac{1}{J(\gamma, z)^k} f(\gamma z).$$

Numeric functions on $\mathfrak{D}$ that are invariant with respect to the factor $J(\gamma, z)^k$ are said to be Γ-*invariant of weight* k.

(4) As noted in (1.4), it follows from the automorphy equations that $j(1, z) = 1$. Assume that Γ is homogeneous, that is, $-1 \in \Gamma$. Then it follows similarly that $j(-1, z)^2 = 1$. Consequently, the function $j(-1, z)$ is the constant 1 or -1. If a function f is j-invariant, and f is non-zero, say $f(z_0) \neq 0$, then it follows from the equation $f(z_0) = f((-1)z_0) = j(-1, z_0) f(z_0)$ that $j(-1, z_0) = 1$. Hence $j(-1, z)$ is the constant 1. Therefore, when Γ is homogeneous, we are mostly interested in *homogeneous factors*, that is, factors j for which $j(-1, z) = 1$.

Note that the factor $J(\gamma, z)^k$ for an integer k is only homogeneous when k is even. The homogeneous factors of odd (integral) weight k are the functions (1.11.1) where $\chi : \Gamma \to \mathbf{C}^*$

is an odd unitary character (that is, $\chi(-1) = -1$). In particular, if -1 belongs to the commutator subgroup Γ', then there are no homogeneous factors of odd integral weight on Γ.

(5) Clearly, if Γ is inhomogeneous, and $\tilde{\Gamma}$ denotes the homogenized group $\tilde{\Gamma} = \Gamma \cup (-\Gamma)$, then every factor j on Γ extends to a homogeneous factor j on $\tilde{\Gamma}$ by defining $j(-\gamma, z) := j(\gamma, z)$ for γ in Γ.

(6) The function $|J(\gamma, z)|^k$ is an automorphic factor on Γ, but it is not holomorphic, and hence it is not a factor in the sense of Definition (1.9). Note that the condition (1.9)(3) is that the automorphic factor $|j(\gamma, z)|$ is equal to the automorphic factor $|J(\gamma, z)|^k$. Clearly, for an automorphic factor $j(\gamma, z)$ to satisfy condition (1.9)(3), it suffices that the equations,

$$|j(\gamma, z)| = |J(\gamma, z)|^k, \tag{1.11.2}$$

hold for a system of matrices γ generating the group Γ.

(7) As noted in (1.5), if $f : \mathfrak{D} \to \mathbf{C}^*$ is any function, then the equation,

$$j(\gamma, z) := f(\gamma z)/f(z), \tag{1.11.3}$$

defines an automorphic factor on any subgroup of $\mathrm{SL}(\mathfrak{D})$. If f is holomorphic, then j is holomorphic in z. In fact, if f is meromorphic and the right hand side of (1.11.3) has no zeros or poles, then (1.11.3) defines a holomorphic automorphic factor j. But of course, for j to be a factor in the sense of Definition (1.9) it is required that the equations,

$$|f(\gamma z)/f(z)| = |J(\gamma, z)|^k, \tag{1.11.4}$$

hold for all matrices γ in the group Γ. As observed in (6), it suffices that the equations (1.11.4) hold for a system of generators for Γ.

2. Examples I.

(2.1) Example. Let k be an integer. Let Ω be a lattice in $\mathbf{C}$, say $\Omega = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ where the complex numbers ω_1 and ω_2 are linearly independent over $\mathbf{R}$. Consider the following sum,

$$E_k(\Omega) = E_k\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix} = \sum_{\omega\neq 0} \frac{1}{\omega^k}, \tag{2.1.1}$$

where the sum is over all $\omega \neq 0$ in Ω. The sum is absolutely convergent for $k \geq 3$. For typographical reasons, the sum will also be denoted $E(\omega_1, \omega_2)$. In particular, for $z \in \mathfrak{H}$, we define the *Eisenstein series*,

$$E_k(z) = E_k(z, 1) = \sum{}' \frac{1}{(nz+m)^k}, \tag{2.1.2}$$

where the sum is over pairs of integers $(m, n) \neq (0, 0)$. The function $E_k(z)$, for $k \geq 3$, is holomorphic in $\mathfrak{H}$.

Clearly, if λ is a non-zero complex number, then $E_k(\lambda\Omega) = \lambda^{-k} E_k(\Omega)$. In addition, if γ is a matrix in $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ then the two pairs (ω_1, ω_2) and $(\omega_1, \omega_2)\gamma^{\mathrm{tr}}$ generate the same lattice. As a consequence, $E(\omega_1, \omega_2) = E(\omega_1, \omega_2)\gamma^{\mathrm{tr}}$. In particular, we obtain the equations,

$$E_k(\gamma z) = E_k\begin{bmatrix}\gamma z\\ 1\end{bmatrix} = E_k \frac{1}{J(\gamma, z)}\gamma\begin{bmatrix} z\\ 1\end{bmatrix} = J(\gamma, z)^k E_k(z).$$

In other words, for an integer $k \geq 3$, the function $E_k(z)$ is $\Gamma(1)$-invariant of weight k.

Note that the function $E_k(z)$ is equal to zero when k is odd. We will see later that the function in non-zero when k is even.

(2.2) Example. The Eisenstein series (2.1.2) is closely related to the following series, defined for an integer $k \geq 3$,

$$G_k(z) := \frac{1}{2} \sum_{(m,n)=1} \frac{1}{(mz+n)^k}, \tag{2.2.1}$$

where the sum is over all pairs of relatively prime integers (m, n). Indeed, if we group in (2.1.2) the terms corresponding to the greatest common divisor d of (m, n), we obtain the equation,

$$E_k(z) = 2\zeta(k) G_k(z), \tag{2.2.2}$$

where $\zeta(k) = \sum_{d\geq 1} d^{-k}$. Hence, the function G_k is $\Gamma(1)$-invariant of weight k. It vanishes when k is odd. If k is even, then the value $\zeta(k)$ is a well known rational multiple of π^k. More precisely, then $2\zeta(k) = -(2\pi i)^k B_k/k!$ where B_k is the k'th Bernoulli number.

(2.3) Example. Dedekind's *η-function* is the function in the upper half plane $\mathfrak{H}$ defined by the product,

$$\eta(z) = e^{2\pi i z/24} \prod_{n\geq 1} (1 - e^{2\pi i n z}). \tag{2.3.1}$$

The sum $\sum_n e^{2\pi i n z}$ converges normally in $\mathfrak{H}$; hence, so does the product. Consequently, the η-function is holomorphic and everywhere non-zero in $\mathfrak{H}$.

The η-function satisfies the following functional equations,

$$\eta(z+1) = e^{2\pi i/24}\eta(z), \quad \eta(-1/z) = \sqrt{z/i}\,\eta(z), \tag{2.3.2}$$

where $\sqrt{z/i}$ is the principal determination. The first equation is obvious, but the second is far from trivial. It follows from the functional equations that

$$|\eta(z+1)| = |\eta(z)|, \quad |\eta(-1/z)| = |z|^{\frac{1}{2}}|\eta(z)|.$$

The modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ is generated by the two matrices t and s, and $tz = z+1$ and $sz = -1/z$. Moreover, $J(t,z) = 1$ and $J(s,z) = z$. Hence, the equation,

$$|\eta(\gamma z)/\eta(z)| = |J(\gamma,z)|^{\frac{1}{2}},$$

holds for the two generators s and t. It follows that the equation holds for all matrices in $\Gamma(1)$. Therefore, as observed (1.11)(7), a factor j_η of weight $\frac{1}{2}$ on $\Gamma(1)$ is defined by the equation,

$$j_\eta(\gamma,z) = \eta(\gamma z)/\eta(z). \tag{2.3.3}$$

Of course, the *η-factor* j_η is fully determined by the automorphy equations from the special values,

$$j_\eta(t,z) = e^{2\pi i/24}, \quad j_\eta(s,z) = \sqrt{\frac{z}{i}},$$

but a priori, it is far from obvious that these two values define an automorphic factor on $\Gamma(1)$.

By construction, the function η is j_η-invariant. As the weight of j_η is $\frac{1}{2}$, the even powers j_η^{2k} are of integral weight k, and hence of the form in (1.11.1). In particular, the square j_η^2 is a factor of weight 1 on $\Gamma(1)$, of the form,

$$j_\eta^2(\gamma,z) = \chi(\gamma)J(\gamma,z),$$

where $\chi\colon \Gamma(1) \to \mathbf{C}^*$ is a unitary character. For $\gamma := t$ we obtain the equation $e^{2\pi i/12} = \chi(t)$. For $\gamma = s$ we obtain the equation $z/i = \chi(s)z$, that is, $\chi(s) = e^{-2\pi i/4}$. As t and s generate $\Gamma(1)$, it follows that χ maps into the cyclic subgroup C_{12} of $\mathbf{C}^*$. A priori, the character group of $\Gamma(1)$ is cyclic of order a divisor of 12, since $\Gamma(1)$ is generated by the matrices s and u of orders 4 and 3. Hence, using the functional equation of the η-function, we recover the result of (Mdlar.2.4) that the character group is cyclic of order 12.

As a consequence, the power $j_\eta(\gamma,z)^{2k}$, where k is an integer divisible by 12, is equal to $J(\gamma,z)^k$. In particular, the *discriminant* $\Delta(z)$, defined by the equation,

$$\Delta(z) := \eta(z)^{24},$$

is a $\Gamma(1)$-invariant function of weight 12 on $\mathfrak{H}$. Like $\eta(z)$, the discriminant is everywhere non-zero.

(2.4) Example. As the η-function is everywhere non-zero in $\mathfrak{H}$, there is, for an arbitrary real number k, a determination of the function $\eta(z)^{2k}$. In fact, a canonical determination of η^{2k} is selected by chosing for the first factor in the product (2.3.1) the determination $e^{2\pi ikz/12}$ and for the n'th factor in the product the principal determination $(1-e^{2\pi inz})^{2k}$ (given by the binomial expansion). It follows that a factor of weight k on $\Gamma(1)$ is defined by the equation,

$$j_\eta^{2k}(\gamma, z) := \eta(\gamma z)^{2k}/\eta(z)^{2k}.$$

It follows in particular that for any given real number k there are exactly 12 factors of weight k on $\Gamma(1)$.

(2.5) Example. The function G_4 of Example (2.2) is $\Gamma(1)$-invariant of weight 4. Hence the cube G_4^3 is of weight 12. The discriminant Δ is of weight 12, and it is everywhere non-zero. Therefore, the following function,

$$j(z) := G_4(z)^3/\Delta(z), \tag{2.5.1}$$

is of weight 0, that is, the function $j(z)$ is a $\Gamma(1)$-invariant function in the usual sense: $j(\gamma z) = j(z)$ for all matrices γ in $\Gamma(1)$. The function j is called *Klein's j-invariant*. We prove later the following equation,

$$12^3\Delta(z) = G_4(z)^3 - G_6(z)^2. \tag{2.5.2}$$

It allows an expression of $j(z)$ in terms of G_4 and G_6. In terms of the Eisenstein series, it is customary to write,

$$g_2(z) := 60E_4(z), \quad g_3(z) = 140E_6(z).$$

Then, using the values of the Bernoulli numbers, $B_4 = -1/30$ and $B_6 = 1/42$, it is easy to derive the equation,

$$j(z) = \frac{12^3\, g_2(z)^3}{g_2(z)^3 - 27g_3(z)^2}. \tag{2.5.3}$$

We show later that Klein's invariant defines an isomorphism,

$$\overline{\mathfrak{H}/\Gamma(1)} \xrightarrow{\sim} \overline{\mathbf{C}},$$

from the orbit space to the Riemann sphere.

(2.6) Example. The *θ-function* is the function defined in the upper half plane $\mathfrak{H}$ as the sum over all integers n,

$$\theta(z) = \sum e^{\pi in^2z}. \tag{2.6.1}$$

The θ-function is holomorphic in $\mathfrak{H}$ and it satisfies the two functional equations,

$$\theta(z+2) = \theta(z), \qquad \theta(-1/z) = \sqrt{z/i}\,\theta(z), \tag{2.6.2}$$

where $\sqrt{z/i}$ is the principal determination. As in Example (2.3), the first equation follows immediately from the definition, but the second is far from trivial. In fact, the second equation is equivalent to the functional equation for Riemann's ζ-function. It follows from the two equations that $|\theta(z+2)| = |\theta(z)|$ and $|\theta(-1/z)| = |z|^{1/2}|\theta(z)|$.

The Möbius transformations $z \mapsto z+2$ and $z \mapsto -1/z$ are associated to the matrices t^2 and s. Moreover, $J(t^2, z) = 1$ and $J(s, z) = z$. Therefore, the two equations imply that the θ-function satisfies the equation

$$|\theta(\gamma z)/\theta(z)| = |J(\gamma, z)|^{\frac{1}{2}},$$

for the two matrices $\gamma = t^2$ and $\gamma = s$. The latter two matrices generate the θ-group Γ_θ by (Mdlar.4.6). It follows that the θ-function determines a unique factor j_θ of weight $\frac{1}{2}$ on the group Γ_θ, given by the equation,

$$j_\theta(\gamma, z) = \theta(\gamma z)/\theta(z). \tag{2.6.3}$$

The θ-*factor* j_θ is of course fully determined by its values on the two generators,

$$j_\theta(t^2, z) = 1, \quad j_\theta(s, z) = \sqrt{z/i},$$

but a priori, it is far from obvious that these two values define an automorphic factor on Γ_θ.

(2.7) Remark. The sum (2.6.1) defining the θ-function can be split into two parts, $\theta = \theta_{\mathrm{ev}} + \theta_{\mathrm{odd}}$, where the sums are respectively over the even and odd integers. Obviously, $\theta_{\mathrm{ev}}(z) = \theta(4z)$, and hence $\theta_{\mathrm{odd}}(z) = \theta(z) - \theta(4z)$. Moreover, $\theta_{\mathrm{ev}}(z+1) = \theta_{\mathrm{ev}}(z)$ and $\theta_{\mathrm{odd}}(z+1) = -\theta_{\mathrm{odd}}(z)$. Hence we obtain the functional equation,

$$\theta(tz) = 2\theta(4z) - \theta(z). \tag{2.7.1}$$

From the functional equation (2.7.1) for $\theta(tz)$ and the functional equation (2.6.2) for $\theta(sz)$ we can obtain a functional equation for $\theta(\gamma z)$ for any matrix γ in $\Gamma(1)$. Consider for instance the matrix $u = t^{-1}s$ defining the Möbius transformation $uz = -1 - 1/z$. From (2.7.1) and (2.6.2) we obtain the equations,

$$\theta(-1-1/z) = \theta(-1/z+1) = 2\theta(-4/z) - \theta(-1/z) = 2\sqrt{z/4i}\,\theta(z/4) - \sqrt{z/i}\theta(z).$$

Hence,

$$\theta(uz) = \sqrt{z/i}\Big(\theta(z/4) - \theta(z)\Big). \tag{2.7.2}$$

It follows easily from (2.7.1) and (2.7.2) that the two functions, $\theta(tz)$ and $\theta(uz)$, are given by the sums,

$$\theta(z+1) = \sum_n (-1)^n e^{\pi i n^2 z}, \quad \theta(-1-1/z) = \sqrt{\frac{z}{i}} \sum_{n \text{ odd}} e^{\pi i n^2 z/4}. \tag{2.7.3}$$

3. The signs of a factor.

(3.1) Setup. Fix a finite disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$. In addition, fix a real number k and a factor $j(\gamma, z)$ of weight k on Γ.

(3.2) Lemma. *Let γ be a matrix in Γ, and let u be a fixed point of γ in the closure of $\mathfrak{D}$. Then the following limit exists:*

$$j(\gamma, u) := \lim_{z \to u} j(\gamma, z), \tag{3.2.1}$$

where the limit is taken over points z of $\mathfrak{D}$. If γ is elliptic or equal to ± 1, then $j(\gamma, u)$ is a d'th root of unity where d is the order of γ. If γ is parabolic, then $|j(\gamma, u)| = 1$.

Proof. Consider a determination of $J(\gamma, z)^k$ on $\mathfrak{D}$. By Lemma (1.10), there is an equation,

$$j(\gamma, z) = \varepsilon J(\gamma, z)^k,$$

where $|\varepsilon| = 1$. As remarked in (1.8), it follows that the function $j(\gamma, z)$ extends continuously to the points of $\partial\mathfrak{D}$, except possibly to the points ∞ and $\gamma^{-1}\infty$. Since u is a fixed point, it is only exceptional when it is equal to ∞. In the exceptional case, the function $J(\gamma, z)$ is constant. Therefore $j(\gamma, z)$ is constant and hence it extends trivially to all of $\overline{\mathbf{C}}$. Thus, in all cases, the limit (3.2.1) is simply the value at u of the extended function $j(\gamma, z)$.

The remaining assertions hold trivially if $\gamma = \pm 1$. Assume next that γ is elliptic. Then the fixed point u belongs to $\mathfrak{D}$. Clearly, since u is a fixed point, it follows from the automorphy equation that $j(\gamma, u)^d = j(\gamma^d, u) = 1$. Hence $j(\gamma, u)$ is a d'th root of unity.

Assume next that γ is parabolic. Then the fixed point u belongs to the boundary $\partial\mathfrak{D}$. If $u \neq \infty$, then by (1.6), the number $J(\gamma, u)$ is an eigenvalue of γ, and hence equal to ± 1 since γ is parabolic. Consequently, if $u \neq \infty$ then $j(\gamma, u) = \varepsilon J(\gamma, u)^k$ is of modulus 1. Clearly, if the fixed point u is equal to ∞, then the function $J(\gamma, z)$ is constant, and, since γ is parabolic, it is the constant ± 1. Hence, $j(\gamma, z)$ is a constant of modulus 1. In particular, the limit $j(\gamma, u)$ is of modulus equal to 1.

Thus the assertions have been proved in all cases. □

(3.3) Definition. Recall that for any point u in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ we have defined the canonical generator γ_u of Γ at u. It belongs to the isotropy group Γ_u. It follows from Lemma (3.2) that the following number,

$$\omega_u = \omega_u(j, \Gamma) := j(\gamma_u, u),$$

is of modulus 1. It is called the *sign of the factor j at the point u*. The unique real number $\kappa_u = \kappa_u(j)$ determined by the conditions,

$$\omega_u = e^{2\pi i \kappa_u}, \quad 0 \le \kappa_u < 1,$$

is called the *parameter of the factor j at the point u*.

At a Γ-ordinary point u, the canonical generator is equal to 1. Hence $\omega_u = 1$ and $\kappa_u = 0$. Assume that u is a Γ-elliptic point of $\mathfrak{D}$. Then, by (3.2), the sign is a d_u'th root of unity, where d_u is the order of γ_u. Hence the parameter κ_u is of the form $\kappa_u = a/d_u$ where a is an integer and $0 \le a < d_u$. In particular, we obtain for any point u of $\mathfrak{D}$ the inequalities,

$$0 \le \kappa_u \le 1 - 1/d_u. \tag{3.3.1}$$

Clearly, if $-1 \in \Gamma$ and the factor j is homogeneous, then the sign ω_u is an e_u'th root of unity, where $e_u = |\mathrm{P}\Gamma_u|$ is the order of the Γ-elliptic point u, and in the inequalities we can replace d_u by e_u.

(3.4) Proposition. *Assume that Γ is the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ acting on the upper half plane $\mathfrak{H}$. Then the sign of the factor j at the point ∞ is of the form,*

$$\omega_\infty = \zeta^{-1} e^{2\pi i k/12}, \tag{3.4.1}$$

where ζ is a 12th root of unity. Moreover, the factor j is completely determined by the root ζ and the given weight k. At the elliptic points i and ρ, the signs of the factor j are the following:

$$\omega_i = \zeta^3, \qquad \omega_\rho = \zeta^4.$$

Finally, $j(-1, z) = \zeta^6$ for all z.

Proof. The point ∞ is parabolic, and the matrix t is the canonical generator γ_∞. As $J(t, z) = 1$, it follows that $j(t, z)$ is constant, and hence we obtain the equation,

$$j(t, z) = \omega_\infty.$$

The point i is elliptic, and the matrix s of order 4 is the canonical generator γ_i. As $J(s, z) = z$, the function $j(s, z)$ is, by Lemma (1.10), of the form εz^k where $|\varepsilon| = 1$ and z^k is the principal determination in $\mathfrak{H}$. Evaluation at i yields $\omega_i = \varepsilon i^k$. Hence,

$$j(s, z) = \omega_i z^k / i^k,$$

and the sign ω_i is a 4th root of unity. Now $s = tu$. Therefore, from the automorphy equation, we obtain that

$$j(u, z) = j(s, z)/j(t, uz) = \omega_\infty^{-1} \omega_i z^k / i^k.$$

The point ρ is elliptic, and the matrix u of order 3 is the canonical generator γ_ρ. Hence the sign $\omega_\rho = j(u, \rho)$ is a 3rd root of unity. Take $z = \rho$ in the expression for $j(u, z)$ to obtain the equation,

$$\omega_\infty = \omega_i / \omega_\rho \cdot \rho^k / i^k.$$

Clearly, $\rho^k / i^k = e^{2\pi i k/12}$ and $\zeta := \omega_\rho / \omega_i$ is a 12th root of unity. Hence ω_∞ has the form (3.4.1). Moreover, the assertions about ω_i and ω_ρ follow from the definition of ζ. Furthermore, j is uniquely determined, because Γ is generated by s and t, and $j(t, z)$ and $j(s, z)$ were determined above by the signs at ∞ and i. Finally, $j(-1, z)$ is constant; the constant is equal to $j(-1, i) = j(s^2, i) = j(s, i)^2 = \omega_i^2 = \zeta^6$. $\square$

(3.5) Note. In (3.4), there are twelve possible values for ζ, and hence at most 12 possible factors of a given weight k. As noted in example (2.4), there are in fact 12 factors of any given real weight k.

Clearly, for an integral weight k, the factor $J(\gamma, z)^k$ corresponds to $\zeta = e^{2\pi i k/12}$.

(3.6) Example. The factor j_η associated with Dedekind's η-function is a factor of weight $k = \frac{1}{2}$ on the modular group $\Gamma(1)$. As $j_\eta(t, z) = e^{2\pi i/24}$, we have, in the notation of (3.4), that $\zeta = 1$. Hence we obtain for j_η the following signs,

$$\omega_\infty = e^{2\pi i/24}, \quad \omega_i = 1, \quad \omega_\rho = 1.$$

The corresponding parameters are $1/24$, 0, and 0.

The square j_η^2 is one of the possible 12 factors of weight 1 on $\Gamma(1)$. In fact, by (2.3),

$$j_\eta^2(\gamma, z) = \chi_{12}(\gamma) J(\gamma, z),$$

where $\chi_{12} \colon \Gamma(1) \to C_{12}$ is the unique character for which $\chi_{12}(t) = e^{2\pi i/12}$. As $\omega_\infty = e^{2\pi i/12}$, we have $\zeta = 1$ in (2.4.1), and we obtain for j_η^2 the following signs:

$$\omega_\infty = e^{2\pi i/12}, \quad \omega_i = 1, \quad \omega_\rho = 1.$$

(3.7) Example. The θ-group $\Gamma_\theta \subset \mathrm{SL}_2(\mathbf{Z})$, acting on the upper half plane $\mathfrak{H}$, has one elliptic orbit represented by the point i, and two cusps represented by the points ∞ and -1. At the three points, the canonical generators are the matrices,

$$\gamma_i = s = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \gamma_\infty = t^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \gamma_{-1} = utu^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}.$$

Indeed, the first two equations are obvious. To find the canonical generator at -1, apply conjugation by u. Under the conjugation, the point ∞ corresponds to $u(\infty) = -1$ and Γ_θ corresponds to the conjugate group $\Gamma_\theta^u = u^{-1}\Gamma_\theta u$. Modulo 2, we have that $u^{-1}su \equiv t$, and it follows that the conjugate group Γ_θ^u is equal to $\Gamma_0(2)$, cf. Exercise (Mdlar.3.6). Clearly, for the conjugate group the canonical generator at ∞ is the matrix t. Thus $\gamma_{-1} = utu^{-1}$.

Consider the θ-factor j_θ on Γ_θ. It is of weight $\frac{1}{2}$ and determined by the equations of (2.6),

$$j_\theta(\gamma_\infty, z) = 1, \quad j_\theta(\gamma_i, z) = \sqrt{z/i}.$$

Clearly, $\gamma_{-1} = -\gamma_\infty^{-1}\gamma_i$. Therefore, from the automorphy equation, we obtain that

$$j_\theta(\gamma_{-1}, z) = j_\theta(\gamma_\infty^{-1}, \gamma_i z) j(\gamma_i, z) = 1 \cdot \sqrt{z/i} = \sqrt{z/i}.$$

As $\sqrt{z/i} \to e^{2\pi i/8}$ for $z \to -1$, we obtain for j_θ the following signs,

$$\omega_\infty = 1, \quad \omega_{-1} = e^{2\pi i/8}, \quad \omega_i = 1.$$

The corresponding parameters are 0, $1/8$, and 0.

(3.8) Lemma. *Let L be a linear function, $L(z) = Cz + D$, where $C \neq 0$. Assume that $L(z) \neq 0$ everywhere on $\mathfrak{D}$, and consider a determination of $L(z)^k$ on $\mathfrak{D}$. Let γ be a matrix in* $\mathrm{SL}(\mathfrak{D})$ *and let u be a fixed point of γ belonging to the closure of $\mathfrak{D}$. If u belongs to the boundary of $\mathfrak{D}$, assume moreover that γ is parabolic. Then,*

$$\lim_{z\to u} L(\gamma z)^k/L(z)^k = 1, \tag{3.8.1}$$

where the limit is taken over points z of $\mathfrak{D}$.

Proof. The function L has one zero $v = -D/C$ in $\mathbf{C}$. Let V be the open subset of $\mathbf{C}$ obtained by cutting away from $\mathbf{C}$ a half ray starting at v and having no other points in common with the boundary of $\mathfrak{D}$. Then $\mathfrak{D}$ is contained in V, and clearly there is a unique extension of $L(z)^k$ to a determination defined on all of V. Therefore the assertion (3.8.1) is elementary when u belongs to V: in the fraction, both the denominator and the numerator converges to the non-zero value $L(u)^k$.

It remains to consider two further possible cases: $u = \infty$ and $u = v$. Clearly, to prove (3.8.1), it suffices to prove that the quotient,

$$L(\gamma z)/L(z), \tag{3.8.2}$$

converges to 1 for $z \to u$. The map L is a Möbius transformation. Therefore, after a conjugation, replacing $\mathfrak{D}$ by $L(\mathfrak{D})$, Γ by $L\Gamma L^{-1}$, and γ by $L\gamma L^{-1}$, we may assume that L is the identity, $L(z) = z$. Thus $\mathfrak{D}$ is a finite disk not containing the point 0, and the two possible cases are $u = \infty$ and $u = 0$.

Assume first that $u = \infty$. Then $\mathfrak{D}$ is a half plane and, since γ is parabolic, the associated transformation is of the form $\gamma z = z + b$. Then the fraction (3.8.2) is equal to $1 + b/z$, and it converges to 1 for $z \to \infty$. Assume next that $u = 0$. Then $\mathfrak{D}$ is a disk containing 0 on its boundary. Since γ is parabolic with 0 as fixed point, the associated transformation is of the form $\gamma z = z/(cz + 1)$. Then the fraction in (3.8.2) is equal to $1/(cz + 1)$, and it converges to 1 for $z \to 0$.

Thus the assertion has been proved in all cases. $\square$

(3.9) Exercise. In the setup of (3.8), assume that the fixed point u is on the boundary of $\mathfrak{D}$, but assume that γ is hyperbolic. Prove the limit in (3.8.1) exists, and find its value.

(3.10) Definition. Let α be a matrix of $\mathrm{GL}_2(\mathbf{C})$ mapping a finite disk $\mathfrak{D}'$ onto the finite disk $\mathfrak{D}$. Let $\Gamma' := \Gamma^\alpha = \alpha^{-1}\Gamma\alpha$ be the conjugate subgroup. It is a discrete subgroup of $\mathrm{SL}(\mathfrak{D}')$. Consider a determination $J(\alpha, z')^k$ on $\mathfrak{D}'$. Define the *conjugate* j^α of j as the function on $\Gamma' \times \mathfrak{D}'$ given by the equation,

$$j^\alpha(\gamma', z') := \frac{J(\alpha, z')^k}{J(\alpha, \gamma' z')^k} j(\alpha\gamma'\alpha^{-1}, \alpha z'). \tag{3.10.1}$$

Clearly, the fraction on the right is independent of the choice of determination.

Note that the conjugate factor is not obtained simply by a transport of structure. The latter transport would yield the function $j(\alpha\gamma'\alpha^{-1}, \alpha z')$.

(3.11) Lemma. *In the setup of* (3.10)*, the conjugate* $j' := j^\alpha$ *is a factor of weight* k *on the conjugate group* $\Gamma' := \Gamma^\alpha$*. Moreover, if* u' *and* $u = \alpha u'$ *are corresponding points of* $\mathfrak{D}' \cup \partial_{\Gamma'}\mathfrak{D}'$ *and* $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$*, then the signs are equal:*

$$\omega_{u'}(j') = \omega_u(j). \tag{3.11.1}$$

If the matrix α *belongs to* Γ*, then* $\Gamma' = \Gamma$ *and* $j' = j$*. In particular, the signs of* j *at two* Γ*-equivalent points are equal.*

Proof. In (3.10.1), the function $j(\alpha\gamma'\alpha^{-1}, \alpha z')$ is an automorphic factor for the action of Γ' on $\mathfrak{D}'$, because it is obtained by a simple transport of structure. Moreover, as noted in (1.5), the fraction in (3.10.1) is an automorphic factor. Therefore, the function j^α is an automorphic factor on Γ'. Clearly, the function $j^\alpha(\gamma', z')$ is holomorphic in z'. Finally, the condition (1.9)(3) for j^α follows by applying (1.6) to $\alpha\gamma' = \gamma\alpha$, where $\gamma := \alpha\gamma'\alpha^{-1}$. Hence the conjugate factor j^α is a factor of weight k.

Assume that $u = \alpha u'$. Then, if γ_u is the canonical generator of Γ_u, the conjugate matrix $\alpha^{-1}\gamma_u\alpha$ is the canonical generator $\gamma_{u'}$ of $\Gamma'_{u'}$. Therefore, the equality (3.11.1) follows from Lemma (3.8).

Assume that α belongs to Γ. Then $\Gamma' = \Gamma$. It follows from Lemma (1.11) that in the definition of j^α in (3.10.1), we may replace the fraction $J(\alpha, z)^k / J(\alpha, \gamma z)^k$ by the fraction $j(\alpha, z)/j(\alpha, \gamma z)$. Then the equation $j^\alpha = j$ follows by applying the automorphy equation to $\alpha\gamma' = \gamma\alpha$, where $\gamma =: \alpha\gamma'\alpha^{-1}$.

Clearly, the last assertion of the Lemma is a consequence of (3.11.1). ∎

(3.12) Note. (1) For an integral weight k, the factor $J(\gamma, z)^k$ is invariant under conjugation, as it follows from (1.3).

(2) In the setup of (3.10), let β be a matrix mapping a finite disk $\mathfrak{D}''$ onto $\mathfrak{D}'$. Then $j^{\alpha\beta} = (j^\alpha)^\beta$, as it follows by an easy computation.

(3) It follows from the last part of Lemma (3.11) that the signs ω_u of j are completely determined by the signs at one point in each parabolic or elliptic orbit.

(4) Clearly, if j_1 and j_2 are factors of weights k_1 and k_2 on Γ, then $j_1 j_2$ is a factor of weight $k_1 + k_2$, and for the signs we have the equation $\omega_u(j_1 j_2) = \omega_u(j_1)\omega_u(j_2)$.

(5) The signs $\omega_u(\Gamma, j)$ depend on the group Γ, that is, they change if j is restricted to a subgroup Δ of Γ. Assume that Δ is of finite index in Γ. Then $\partial_\Delta\mathfrak{D} = \partial_\Gamma\mathfrak{D}$. Consider a point u in $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$. Denote by d the index $d := |\mathrm{P}\Gamma_u{:}\mathrm{P}\Delta_u|$. Assume for simplicity that Δ is homogeneous and that the factor j is homogeneous. If γ_u is the canonical generator of Γ_u, then γ_u^d is the canonical generator of Δ_u. Therefore we obtain the equation,

$$\omega_u(\Delta, j) = \omega_u(\Gamma, j)^d.$$

(3.13) Proposition. *For* $j(\gamma, z) = J(\gamma, z)^k$ *where* k *is an integer, the sign is equal to* 1 *at a regular cusp, it is equal to* $(-1)^k$ *at an irregular cusp, and at a point* u *of* $\mathfrak{D}$ *it is equal to* $e^{\pi i k(1-1/e_u)}$ *where* $e_u = |\mathrm{P}\Gamma_u|$*. In particular, for* $k = 2$*, the parameter is equal to* 0 *at any cusp, and equal to* $1 - 1/e_u$ *at a point* $u \in \mathfrak{D}$*.*

Proof. For a cusp u we may, by conjugation, assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Hence the canonical generator γ_u is equal to t_h is the regular case and equal to $-t_h$ in the irregular case. Therefore, in the irregular case, we have that $J(\gamma_u, z)$ is the constant -1; it follows that sign at u of J^k is equal to $(-1)^k$. In particular, the parameter is equal to 0 when k is even, and equal to $\frac{1}{2}$ when k is odd. The assertions in the regular case are proved similarly.

For a point u of $\mathfrak{D}$, we may, by conjugation, assume that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. Then the canonical generator γ_u is the matrix of (Discr.1.3.1) with $d := e_u$. Thus $J(\gamma_u, z)$ is the constant $-e^{-\pi i/e_u} = e^{\pi i(1-1/e_u)}$. It follows that the sign of J^k at u is equal to $e^{\pi i k(1-1/e_u)}$. In particular, the parameter is the fractional part of $\frac{k}{2}(1 - 1/e_u)$. Thus, for $k = 2$, the parameter is equal to $1 - 1/e_u$. ▯

(3.14) Example. Assume that j is a factor on $\Gamma(1)$, determined as in Proposition (3.4) by its weight k and a 12th root of unity ζ. Restrict j to the subgroup Γ_θ. At the point i, the canonical generators are equal and so we obtain the equation $\omega_i(\Gamma_\theta, j) = \omega_i(\Gamma(1), j)$. At the point ∞, the canonical generator of Γ_θ is equal to t^2, and hence $\omega_\infty(\Gamma_\theta, j) = \omega_\infty(\Gamma(1), j)^2$. Finally, at the point -1, the two groups have the same canonical generator, and hence $\omega_{-1}(\Gamma_\theta, j) = \omega_{-1}(\Gamma(1), j) = \omega_\infty(\Gamma(1), j)$. Hence, by the results of Proposition (3.4), we obtain for the restriction of j to Γ_θ the following signs:

$$\omega_\infty = \zeta^{-2} e^{2\pi i k/6}, \quad \omega_{-1} = \zeta^{-1} e^{2\pi i k/12}, \quad \omega_i = \zeta^3.$$

In particular, the factor j_θ of Example (3.7) is not the restriction of a factor on $\Gamma(1)$.

(3.15) Exercise. On the θ-group Γ_θ, the two factors j_η and j_θ differ by a unitary character $\chi\colon \Gamma_\theta \to \mathbf{C}^*$. Identify the character.

(3.16) Definition. Assume that the matrix α in $\mathrm{SL}_2(\mathbf{C})$ maps the finite disk $\mathfrak{D}$ onto the finite disk $\mathfrak{D}'$. Consider a determination $J(\alpha, z')^k$ on $\mathfrak{D}'$ and a complex sign ε. For any numeric function f on $\mathfrak{D}$, we define the *weight-k conjugate* to f as the function,

$$f^\alpha(z) := \frac{\varepsilon}{J(\alpha, z')^k} f(\alpha z'). \tag{3.16.1}$$

The definition of a weight-k conjugate function does not depend on the given subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ and in particular, it is independent of the factor j. However, it should be noted that contrary to the definition of the conjugate factor j^α, the definition of the conjugate function is ambiguous: f^α depends on the choice of determination $J(\alpha, z')^k$. As different determinations differ by a complex sign there is, strictly speaking, only a well defined class of conjugate functions that differ by a complex sign.

Clearly, a function f is (Γ, j)-invariant if and only if the weight-k conjugate function f^α is $(\Gamma^\alpha, j^\alpha)$-invariant. Note also that if α belongs to Γ (in which case $\mathfrak{D}' = \mathfrak{D}$), then $f \cdot_j \alpha$ is a weight-k conjugate to f.

4. Automorphic forms.

(4.1) Definition. A numeric function f (with values in $\overline{\mathbf{C}}$) defined on the upper half plane $\mathfrak{H}$ will be called *exponentially bounded* at ∞, if there exists a real number C such that

$$f(z) = \mathrm{O}(|e^{iz}|^C) \text{ for } \Im z \to \infty, \tag{4.1.1}$$

uniformly on any vertical strip in $\mathfrak{H}$. The positive function $|e^{iz}|^C$ is equal to $e^{-C\Im z}$ and the condition on C implied by the O-notation is equivalent to the following: For every vertical strip in $\mathfrak{H}$, there are positive real numbers R and M so that the inequality,

$$|f(z)| \le Me^{-C\Im z},$$

holds when z is in the strip and $\Im z \ge R$.

In the most important of the cases to be considered, the function f will in fact be periodic, with a real period. Clearly, in this case, the condition holds for C if and only if (4.1.1) holds uniformly on all of $\mathfrak{H}$.

The supremum of the numbers C for which the condition holds, will be called the *order* of f at ∞ and it will be denoted $\operatorname{ord}_\infty^{\mathfrak{H}} f$. The number $\operatorname{ord}_\infty^{\mathfrak{H}} f$ is equal to $+\infty$ if the condition holds for all C, and equal to $-\infty$ if it holds for no C.

Clearly, if the order is positive, then $f(z) \to 0$ for $\Im z \to \infty$, uniformly in any vertical strip. Conversely, if f is bounded in every vertical strip for $\Im z \gg 0$, then the order of f at ∞ is non-negative.

(4.2) Note. The order defined in (4.1) is in some sense analogous to the usual order at a finite point: Let f be a numeric function defined in an open neighborhood of a point u of $\mathbf{C}$. Consider real numbers C such that

$$f(z) = \mathrm{O}(|z-u|^C) \text{ for } z \to u.$$

In other words, assume for some M and R that $|f(z)| \le M|z-u|^C$ for $0 < |z-u| < R$. Clearly, if C exists, then $f(z)$ has finite values in a pointed neighborhood of u. The supremum of the possible numbers C is the order $\operatorname{ord}_u f$. Assume that f is holomorphic near u. Then the order is $-\infty$ if u is an essential singularity of f. If f is meromorphic, then the order is the usual order and, in particular, the order is an integer if f is not the zero function.

(4.3) Lemma. *The order at ∞ of functions on $\mathfrak{H}$ has the following properties:*

(1) $\operatorname{ord}_\infty^{\mathfrak{H}}(e^{iCz}) = C$.
(2) $\operatorname{ord}_\infty^{\mathfrak{H}}(f+g) \ge \inf\{\operatorname{ord}_\infty^{\mathfrak{H}} f\,,\ \operatorname{ord}_\infty^{\mathfrak{H}} g\}$.
(3) $\operatorname{ord}_\infty^{\mathfrak{H}}(fg) \ge \operatorname{ord}_\infty^{\mathfrak{H}} f + \operatorname{ord}_\infty^{\mathfrak{H}} g$, *with equality if* $\operatorname{ord}_\infty^{\mathfrak{H}}(1/f) = -\operatorname{ord}_\infty^{\mathfrak{H}}(f)$.
(4) *If $z \mapsto \alpha z$ is any Möbius transformation and k is a real number, then* $\operatorname{ord}_\infty^{\mathfrak{H}} |\alpha z|^k = 0$.
(5) $\operatorname{ord}_\infty^{\mathfrak{H}} f(rz+b) = r \operatorname{ord}_\infty^{\mathfrak{H}} f(z)$ *for real numbers $r > 0$ and b.*

Proof. In (4), set $f(z) := |\alpha z|^k$. There are two cases to consider: $\alpha\infty \neq \infty$ and $\alpha z = \infty$. In the first case, αz has a finite limit as $z \to \infty$. In particular, (4.1.1) holds for $C = 0$ uniformly on all of $\mathfrak{H}$. Thus the order is non-negative, and obviously, it is not positive. Hence the order is equal to 0. In the second case, $\alpha z = cz + d$ with complex numbers $c \neq 0$ and d. Clearly, in any vertical strip, we have $|z| = \mathrm{O}(\Im z)$, and hence $f(z) = \mathrm{O}(|\Im z|^k)$. Hence (4.1.1) holds for any $C < 0$. Thus the order is non-negative, and obviously, it is not positive. Hence the order is equal to 0.

The remaining assertions are easy consequences of the definition. □

(4.4) Definition. Let f be a numeric function defined on a finite disk $\mathfrak{D}$. Let u be a point on the boundary of $\mathfrak{D}$. Choose a Möbius transformation α mapping $(\mathfrak{H}, \infty)$ onto $(\mathfrak{D}, u)$. Then f will be said to be *exponentially bounded* at the point u, if the conjugate function $f\alpha$ on $\mathfrak{H}$ is exponentially bounded at ∞.

A second choice of α would differ from the first by a Möbius transformation of the form $z \mapsto rz + b$. Hence it follows from the result (4.3)(5) that the definition of exponentially boundedness is independent of the choice of α.

(4.5) Definition. In the setup of (4.4), from the same result (4.3)(5) quoted at the end, it follows that we cannot define the order of f at u simply as the order of $f\alpha$ at ∞.

However, if a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ is given, and u is a Γ-parabolic point, we can normalize the order and obtain the Γ*-order* of f at u as follows: The conjugate group Γ^α is a discrete subgroup of $\mathrm{SL}(\mathfrak{H})$, and the conjugate of the canonical generator γ_u is canonical generator of Γ^α_∞. Moreover, the latter canonical generator defines a Möbius transformation of the form $z \mapsto z + h$, where $h > 0$. Set,

$$\mathrm{ord}^\Gamma_u f := \frac{h}{2\pi} \mathrm{ord}^{\mathfrak{H}}_\infty(f\alpha). \tag{4.5.1}$$

It is a consequence of the proof of the following lemma, that the Γ-order at u is well defined. Obviously, it is different from $-\infty$ if and only if f is exponentially bounded at u.

If u is a point of $\mathfrak{D}$, we define the Γ-order of f at u by the formula,

$$\mathrm{ord}^\Gamma_u f := \frac{1}{|\mathrm{P}\Gamma_u|} \mathrm{ord}_u f, \tag{4.5.2}$$

where the order on the right hand side is usual order at u of the function f, see (4.2). Note that the order of the group $\mathrm{P}\Gamma_u$ is equal to the order of the canonical generator $z \mapsto \gamma_u(z)$. If f is meromorphic, then the order on the right side of (4.5.2) is the usual order of f. In particular, then it is an integer (unless $f = 0$), and the Γ-order of f at a point u of $\mathfrak{D}$ is a rational number. On the contrary, the Γ-order at a cusp can be an arbitrary real number.

(4.6) Lemma. *Consider a finite disk $\mathfrak{D}$, a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$ and a point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. Let $z \mapsto \alpha z$ be a Möbius transformation mapping a finite disk $\mathfrak{D}'$ onto $\mathfrak{D}$. Let $u' := \alpha^{-1} u$ be the conjugate point and $\Gamma' := \Gamma^\alpha$ be the conjugate subgroup. Consider a function f on $\mathfrak{D}$ and, for a given real number k, a weight-k conjugate $f' = f^\alpha$ of f,*

$$f'(z') = \frac{\varepsilon}{J(\alpha, z')^k} f(\alpha z'). \tag{4.6.1}$$

Then,

$$\operatorname{ord}_{u'}^{\Gamma'} f' = \operatorname{ord}_u^{\Gamma} f. \tag{4.6.2}$$

Proof. Assume first that u belongs to $\mathfrak{D}$. The Möbius transformation $z \mapsto \alpha z$ is an analytic isomorphism. Hence, on the right side of (4.6.1), the order of $f'(\alpha z')$ at u' is equal to the order of f at u. Moreover, the fraction is holomorphic and everywhere non-zero on $\mathfrak{D}'$. In particular, its order at u' is equal to 0. Hence, the order of f' at u' is equal to the order of f at u. Moreover, the stabilizer groups $\Gamma'_{u'}$ and Γ_u are conjugate, and hence of the same order. Therefore, it follows from the definition (4.4) that (4.6.2) holds.

Assume next that u is Γ-parabolic. Choose a Möbius transformation $z \mapsto \beta z$ mapping $(\mathfrak{H}, \infty)$ onto $(\mathfrak{D}, u)$. Then the conjugate subgroup $\Delta = \Gamma^\alpha$ is a subgroup of $\mathrm{SL}(\mathfrak{H})$, and the conjugate δ of the canonical generator of Γ_u defines a Möbius transformation $z \mapsto z + h$ with $h > 0$. By definition, the right hand side of (4.6.2) is equal to $(h/2\pi) \operatorname{ord}_\infty^{\mathfrak{H}} f(\beta z)$.

Choose and define similarly β', Δ', δ', and h'. By (4.6.1),

$$f'(\beta' z) = \frac{\varepsilon}{J(\alpha, \beta' z)^k} f(\alpha\beta' z). \tag{4.6.3}$$

Now, the two matrices $\alpha\beta'$ and β differ by the matrix $\sigma := \beta^{-1}\alpha\beta'$. The latter matrix belongs to $\mathrm{SL}(\mathfrak{H})$, and it fixes the point ∞. Hence, $\sigma z = rz + b$. Therefore, on the right side of (4.6.3) we have $f(\alpha\beta' z) = f\beta(rz + b)$. Moreover, in the fraction, $J(\alpha, \beta' z)$ is a Möbius transformation or a constant function. Therefore, by Lemma (4.4), (4) and (5),

$$\operatorname{ord}_\infty^{\mathfrak{H}}(f'\beta') = r \operatorname{ord}_\infty^{\mathfrak{H}}(f\beta). \tag{4.6.4}$$

Clearly, for the two subgroups Δ and Δ' we have that $\Delta' = \sigma^{-1}\Delta\sigma$. Hence their canonical generators δ' and δ are conjugate: $\delta' = \sigma^{-1}\delta\sigma$. Therefore, $h' = h/r$. Now (4.6.2) follows from (4.6.4). ∎

(4.7) Lemma. *Let Δ be a subgroup of finite index in Γ. Then, for any meromorphic function f on $\mathfrak{D}$ and any point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$,*

$$\operatorname{ord}_u^{\Delta} f = |\mathrm{P}\Gamma_u{:}\mathrm{P}\Delta_u| \operatorname{ord}_u^{\Gamma} f.$$

Proof. Set $d := |\mathrm{P}\Gamma_u{:}\mathrm{P}\Delta_u|$. If u is a point of $\mathfrak{D}$, the assertion follows from the definition and Lagrange: $|\mathrm{P}\Gamma_u| = d|\mathrm{P}\Delta_u|$. Assume that u is Γ-parabolic. Then $\mathrm{P}\Gamma_u$ is an infinite cyclic subgroup. Therefore, if γ_u defines the canonical generator of $\mathrm{P}\Gamma_u$, then γ_u^d defines the canonical generator of $\mathrm{P}\Delta_u$. Hence, in the notation of Definition (4.5), if h is the h corresponding to Γ, then the h corresponding to Δ is equal to dh. Thus the assertion holds for a Γ-parabolic point. ∎

(4.8) Setup. Fix a finite disk $\mathfrak{D}$ and a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$. Let j be a factor of real weight k on Γ.

A numeric function f defined on $\mathfrak{D}$ is called a *j-automorphic form*, or a (Γ, j)-*automorphic form*, if the following conditions hold:

(1) f is j-invariant: $f(\gamma z) = j(\gamma, z)f(z)$ for $\gamma \in \Gamma$,
(2) f is meromorphic in $\mathfrak{D}$,
(3) f is exponentially bounded at all Γ-parabolic points of $\mathfrak{D}$.

If the group Γ is homogeneous, we will often assume that the factor j is homogeneous, since only the zero function is j-invariant if $j(-1, z) = -1$, see (1.11)(4).

Note that to verify condition (3) for a given j-invariant function f, it suffices to verify the condition for one point in each Γ-parabolic orbit. Indeed, assume that f is exponentially bounded at some point u on the boundary of $\mathfrak{D}$. Assume that $u = \gamma u'$ for a matrix γ in Γ. It follows immediately from the definition in (4.4) that the function $f\gamma$ is exponentially bounded at the point u'. Since f is j-invariant, $f(\gamma z) = j(\gamma, z)f(z)$. Therefore, it follows from Lemma (4.3)(4) that f is exponentially bounded at u'.

The condition (3) for a given meromorphic function f is usually expressed by saying that f is meromorphic at the cusps, and the automorphic forms defined above are sometimes said to be meromorphic automorphic forms. An automorphic form f is said to be an *integral form*, if f is holomorphic in $\mathfrak{D}$ and the Γ-orders of f at all cusps are non-negative. If, in addition, the Γ-orders at all cusps are positive, then f is called a *cusp form*. The spaces of j-automorphic forms that are meromorphic, or integral, or cusp forms, are denoted respectively,

$$\mathcal{M}(\Gamma, j), \quad \mathcal{G}(\Gamma, j), \quad \mathcal{S}(\Gamma, j).$$

Clearly, the spaces are vector spaces over $\mathbf{C}$.

If k is an integer, then $J(\gamma, z)^k$ is a factor of weight k on Γ. More generally, then the factors of weight k on Γ are exactly the function $j(\gamma, z) = \chi(\gamma)J(\gamma, z)^k$ for a unitary character $\chi\colon \Gamma \to \mathbf{C}^*$. Accordingly, the corresponding spaces are denoted $\mathcal{M}_k(\Gamma, \chi)$, $\mathcal{G}_k(\Gamma, \chi)$, and $\mathcal{S}_k(\Gamma, \chi)$. Omission of χ indicates the trivial character $\chi = 1$. The forms of $\mathcal{M}_k(\Gamma)$ are called Γ-automorphic forms of weight k.

When the weight k is equal to 0, a j-automorphic form is also called a *j-automorphic function*. In particular, a Γ-*automorphic function* is a Γ-invariant function satisfying conditions (2) and (3). The space of Γ-automorphic functions is denoted $\mathcal{M}(\Gamma)$.

(4.9) Observation. (1) The product $j_1 j_2$ of factors of weights k_1 and k_2 on Γ is a factor of weight $k_1 + k_2$. It follows that the product $f_1 f_2$ of forms $f_i \in \mathcal{M}(\Gamma, j_i)$ is a form in $\mathcal{M}(\Gamma, j_1 j_2)$. Similarly, the quotient f_1/f_2, when $f_2 \neq 0$, is a form in $\mathcal{M}(\Gamma, j_1/j_2)$.

It follows in particular that the Γ-automorphic functions form a field $\mathcal{M}(\Gamma)$. Moreover, if $\mathcal{M}(\Gamma, j) \neq (0)$, then $\mathcal{M}(\Gamma, j)$ is a one-dimensional vector space over $\mathcal{M}(\Gamma)$. In other words, if f is a non-zero form in $\mathcal{M}(\Gamma, j)$, then the map $\varphi \mapsto \varphi f$ is an isomorphism,

$$\mathcal{M}(\Gamma) \xrightarrow{\sim} \mathcal{M}(\Gamma, j).$$

(2) Let Δ be a subgroup of finite index in Γ. Obviously, a (Γ, j)-invariant function is (Δ, j)-invariant. The two groups Δ and Γ have the same set of parabolic points. Hence the condition (4.8)(3) for a function f holds for Γ if and only if it holds for Δ. Consequently,

$$\mathcal{M}(\Gamma, j) \subseteq \mathcal{M}(\Delta, j).$$

Similarly, it follows from Lemma (4.7) that $\mathcal{G}(\Gamma, j) \subseteq \mathcal{G}(\Delta, j)$ and $\mathcal{S}(\Gamma, j) \subseteq \mathcal{S}(\Delta, j)$.

(3) Consider a finite disk $\mathfrak{D}$ and a Möbius transformation $\alpha : \mathfrak{D}' \to \mathfrak{D}$ defined by a matrix $\alpha \in \mathrm{SL}_2(\mathbf{C})$. Then the conjugate factor j^α is a factor of weight k on the conjugate group $\Gamma^\alpha = \alpha^{-1}\Gamma\alpha$. Choose a determination of $J(\alpha, z')^k$ on $\mathfrak{D}'$. Then for every function f on $\mathfrak{D}$, the weight-k conjugate function f^α is a function on $\mathfrak{D}'$. Clearly, conjugation $f \mapsto f^\alpha$ defines an isomorphism $\mathcal{M}(\Gamma, j) \xrightarrow{\sim} \mathcal{M}(\Gamma^\alpha, j^\alpha)$, and under conjugation, integral forms correspond to integral forms and cusp forms correspond to cusp forms.

(4.10) Lemma. *Let f be a (Γ, j)-automorphic form on $\mathfrak{D}$. If u and v are Γ-equivalent points of $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$, then*

$$\mathrm{ord}_u^\Gamma f = \mathrm{ord}_v^\Gamma f. \tag{4.10.1}$$

Proof. The equation follows from Lemma (4.6). Indeed, let $v = \gamma u$ for some matrix γ in Γ. Then (4.10.1) holds if the function f on the left hand side is replaced by any weight-k conjugate f^γ of f. However, the function $f \cdot_j \gamma$ is weight-k conjugate, and it is equal to f. □

5. Fourier expansions.

(5.1). Keep the setup of (3.1). Assume for the moment that $\mathfrak{D}$ is the upper half plane $\mathfrak{H}$ and that the point ∞ on the boundary of $\mathfrak{H}$ is Γ-parabolic. Then the Möbius transformation associated to the canonical generator γ_∞ is of the form,

$$\gamma_\infty z = z + h, \text{ for some } h > 0.$$

The sign, $\omega_\infty(j)$, of j at infinity is of modulus 1, and the parameter $\kappa = \kappa_\infty(j)$ is defined be the conditions,

$$\omega_\infty(j) = e^{2\pi i\kappa}, \quad 0 \le \kappa < 1.$$

Since $J(\gamma_\infty, z) = \pm 1$, it follows that $j(\gamma_\infty, z)$ is constant,

$$j(\gamma_\infty, z) = e^{2\pi i\kappa}.$$

Now, let f be a non-zero j-automorphic form on $\mathfrak{H}$. Then, in particular, $f(\gamma_\infty z) = j(\gamma_\infty, z)f(z)$, that is,

$$f(z+h) = e^{2\pi i\kappa} f(z). \tag{5.1.1}$$

Let $F(z)$ be the function in $\mathfrak{H}$ defined by the equation,

$$F(z) := e^{-2\pi i\kappa z/h} f(z). \tag{5.1.2}$$

It follows from (5.1.1) that F is periodic:

$$F(z+h) = F(z).$$

In addition, the function $F(z)$ is meromorphic because $f(z)$ is. Consequently, there exists a unique meromorphic function $g(q)$ on the pointed unit disk: $0 < |q| < 1$ such that

$$F(z) = g(e^{2\pi i z/h}). \tag{5.1.3}$$

The function $F(z)$ is exponentially bounded because $f(z)$ is. In particular, in a strip of width h, it follows that $F(z)$ has no poles for $\Im z > R$. Hence, being periodic, it follows that $F(z)$ is holomorphic in the domain $\Im z > R$. Therefore, the function g is holomorphic in a pointed disk: $0 < |q| < \varepsilon$, and the possible singularity of g at 0 is isolated. It follows that g has a convergent Laurent expansion,

$$g(q) = \sum_{-\infty<n<\infty} a_n q^n \text{ for } 0 < |q| < \varepsilon. \tag{5.1.4}$$

The possible singularity 0 of g is at most a pole, that is, the Laurent expansion (5.1.4) has only a finite number of negative terms. Indeed, since $F(z)$ is exponentially bounded, there is a real number C, say of the form $-2\pi N/h$ for some integer N, so that $F(z)/e^{iCz}$ is bounded for $\Im z > R$. It follows that $q^N g(q)$ is bounded near 0. Hence, in the Laurent expansion, $a_n = 0$ for $n < N$. Thus the function g is meromorphic in the whole unit disk: $|q| < 1$, and the order, $\operatorname{ord}_0 g$, of g at 0 is the least integer n such that $a_n \neq 0$.

Let N be the order of g at 0. Then $g(q) = q^N \tilde{g}(q)$, where $\tilde{g}$ is holomorphic and non-zero at 0. Hence $F(z) = e^{2\pi i N z/h}\tilde{g}(e^{2\pi i z/h})$. Obviously, then $\operatorname{ord}_\infty^{\mathfrak{H}} F(z) = 2\pi N/h$. Combined with the definition of F, it follows that

$$\operatorname{ord}_\infty^\Gamma f = \kappa + N.$$

To summarize, we have proved the following: There is a Laurent series,

$$g(q) = \sum_{n \geq N} a_n q^n, \tag{5.1.5}$$

convergent in a pointed disk $0 < |q| < \varepsilon$, so that for $\Im z \gg 0$ (more precisely, for $\Im z > R$ where $R = (-\log \varepsilon)h/2\pi$) we have the following expansion of f,

$$f(z) = e^{2\pi i \kappa z/h} \sum_{n \geq N} a_n e^{2\pi i n z/h} \tag{5.1.6}$$

Clearly, the Γ-order of f at ∞ is related to the order of g at 0 by the formula,

$$\operatorname{ord}_\infty^\Gamma f = \kappa + \operatorname{ord}_0 g \tag{5.1.7}$$

(5.2) Definition. The series (5.1.6), defined in the setup of (5.1), is called the *Fourier expansion* of the form f, and the coefficients a_n are called the *Fourier coefficients*. The function $q = q(z) = e^{2\pi i z/h}$ is called the *local parameter* at ∞. As a function on $\mathfrak{H}$ it has, for any real number l the obvious determination $q^l = e^{2\pi i l z/h}$. Accordingly, we may write the Fourier expansion in the form,

$$q^\kappa \sum_{n \geq N} a_n q^n. \tag{5.2.1}$$

Note that (5.2.1) does not make sense as a function of q for $0 < |q| < \varepsilon$ unless κ is an integer.

Now assume again that $\mathfrak{D}$ is an arbitrary finite disk, and let f be a non-zero j-automorphic form on $\mathfrak{D}$. Let u be a Γ-parabolic point. Consider the sign $\omega_u = \omega_u(j)$ and the parameter $\kappa_u = \kappa_u(j)$. Choose a Möbius transformation $z \mapsto \beta z$ mapping $(\mathfrak{H}, \infty)$ onto $(\mathfrak{D}, u)$, and choose a determination of $J(\beta, z)^k$ on $\mathfrak{H}$. Then the weight-k conjugate,

$$f^\beta(z) = \frac{1}{J(\beta, z)^k} f(\beta z),$$

is a j^β-automorphic form on $\mathfrak{H}$. Hence the preceding discussion applies to f^β. Under conjugation, the sign and the parameter are unchanged by (3.11), and the order is unchanged by (4.8). The series (5.2.1) for f^β, or the series (5.1.6), is called the *Fourier series* of f at the point u. It follows from (5.1.7) that,

$$\operatorname{ord}_u^\Gamma f = \kappa_u + \operatorname{ord}_0 g.$$

In particular, the Γ-order $\operatorname{ord}_u^\Gamma f$ of a j-automorphic form f is congruent to the parameter κ_u modulo $\mathbf{Z}$.

(5.3). Assume for the moment that $\mathfrak{D}$ is the unit disk $\mathfrak{E}$, and consider the point 0 in $\mathfrak{E}$. The Möbius transformation associated to the canonical generator γ_0 is of the form,

$$\gamma_0 z = e^{2\pi i/e} z, \text{ where } e = |\mathrm{P}\Gamma_0|.$$

[Warning: there are two e's: one for Euler, and one for elliptic.] The sign, $\omega_0(j)$, of j at 0 is an e'th root of unity. The parameter $\kappa = \kappa_0(j)$ is defined be the conditions,

$$\omega = e^{2\pi i\kappa}, \quad 0 \le \kappa < 1,$$

and hence $\kappa = a/e$ for some integer a with $0 \le a < e$. The matrix γ_0 is a diagonal matrix, and hence $J(\gamma_0, z)$ is constant. It follows that $j(\gamma_0, z)$ is constant,

$$j(\gamma_0, z) = e^{2\pi i\kappa}.$$

Now, let f be a non-zero j-automorphic form on $\mathfrak{E}$. Then, in particular, $f(\gamma_0 z) = j(\gamma_0, z) f(z)$, that is,

$$f(e^{2\pi i/e} z) = e^{2\pi i\kappa} f(z). \tag{5.3.1}$$

Let $F(z)$ be the function in $\mathfrak{E}$ defined by the equation,

$$F(z) := z^{-e\kappa} f(z); \tag{5.3.2}$$

note that $e\kappa$ is an integer, since $\kappa = a/e$. It follows from (5.3.1) that $F(z)$ satisfies the equation:

$$F(e^{2\pi i/e} z) = F(z).$$

In addition, the function $F(z)$ is meromorphic because $f(z)$ is. Consequently, there exists a unique meromorphic function $g(w)$ on the pointed unit disk $0 < |w| < 1$ such that, for $z \ne 0$,

$$F(z) = g(z^e). \tag{5.3.3}$$

The possible singularity for g at 0 is a pole. Indeed, since $F(z)$ is meromorphic at 0, there is an integer, say of the form eN, so that $z^{eN} F(z)$ is bounded near 0. Hence, it follows from (5.3.2) that $w^N g(w)$ is bounded near 0. Therefore, the function g has a convergent Laurent expansion in a small pointed disk,

$$g(w) = \sum_{n \ge N} a_n w^n \text{ for } 0 < |w| < \varepsilon, \tag{5.3.4}$$

Assume that N is the order of g at 0, i.e., $a_N \ne 0$. Then it follows from (5.3.3) $F(z)$ is of order eN at 0, and then from (5.3.1) that f is of order $e\kappa + eN$. Hence, for the Γ-order, we obtain the equation,

$$\operatorname{ord}_0^\Gamma f = \kappa + N.$$

To summarize, we have proved the following: The Laurent series for f near 0 is of the form $f(z) = z^{e\kappa} g(z^e)$, that is,

$$f(z) = z^{e\kappa} \sum_{n \geq N} a_n z^{en}, \tag{5.3.5}$$

where $e := |\mathrm{P}\Gamma_0|$. The Γ-order of f at 0 is related to the order of g at 0 by the formula,

$$\operatorname{ord}_0^\Gamma f = \kappa + \operatorname{ord}_0 g \tag{5.3.6}$$

(5.4) Definition. Assume again that $\mathfrak{D}$ is an arbitrary finite disk, and let f be a non-zero j-automorphic form on $\mathfrak{D}$. Let u be a point of $\mathfrak{D}$. Consider the sign $\omega_u = \omega_u(j)$ and the parameter $\kappa_u = \kappa_u(j)$. Choose a Möbius transformation $z \mapsto \beta z$ mapping the unit disk $(\mathfrak{E}, 0)$ onto $(\mathfrak{D}, u)$, and consider as in (5.2) a weight-k conjugate f^β. Then f^β is a j^β-automorphic form on $\mathfrak{E}$. Hence the preceding discussion applies to f^β. The Laurent series (5.3.5) for f^β, or the expansion (5.3.4), is called the *normalized Laurent* series for f at the point u. It follows from (5.3.5) that,

$$\operatorname{ord}_u^\Gamma f = \kappa_u + \operatorname{ord}_0 g.$$

In particular, the Γ-order $\operatorname{ord}_u^\Gamma f$ of a j-automorphic form f is congruent to the parameter κ_u modulo $\mathbf{Z}$.

(5.5) Note. Let f be a (Γ, j)-automorphic form on $\mathfrak{D}$, and let u be a point of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. Then the Fourier expansion of f at u defined in (5.2) when u is Γ-parabolic and the normalized Laurent expansion defined in (5.4) when u belongs to $\mathfrak{D}$ depend on the choice of conjugation β. (In addition, they depend on the choice of a determination of $J(\beta, z)^k$, but the latter ambiguity is only up to a complex sign.) To normalize, we may assume that $\det \beta = 1$.

Consider the case when u is Γ-parabolic. Then a second choice β' is of the form $\beta\sigma$, where $\sigma z = rz + b$ $(r > 0)$. Clearly then, a weight-k conjugate $f^{\beta'}$ can be obtained as a weight-k conjugate $(f^\beta)^\sigma$. Hence, to simplify, we may assume that $\mathfrak{D} = \mathfrak{H}$ and that $\beta = 1$. Then $\beta' = \sigma$ and the canonical generator of Γ^σ has $h' = h/r$. We want to compare the Fourier expansion (5.1.6) of f with the Fourier expansion of the weight-k conjugate f^σ. The matrix σ is a diagonal matrix. Hence $J(\sigma, z)$ is constant, and equal to $\pm 1/\sqrt{r}$. Hence, up to a complex sign, the determination $J(\sigma, z)^k$ is constant and equal to $r^{-k/2}$. Therefore, a weight-k conjugate f^σ is given by

$$f^\sigma(z) = r^{k/2} f(rz + b).$$

Accordingly, we obtain the Fourier expansion of f^σ essentially by substitution of $rz + b$ for z in (5.1.6). If we let $\lambda := r^{k/2} e^{2\pi i\kappa b/h}$ and $\varepsilon := e^{2\pi i b/h}$, then the result is the following:

$$f^\sigma(z) = \lambda e^{2\pi i\kappa z/h'} \sum_{n \geq N} (\varepsilon^n a_n) e^{2\pi i n z/h}.$$

In particular, the Fourier coefficients of the new series are given by

$$a'_n = \lambda \varepsilon^n a_n.$$

Here $|\lambda| = r^{k/2}$ and $|\varepsilon| = 1$.

(5.6) Example. The Eisenstein series $E_k(z)$, for an integer $k \geq 4$, and the normalized series,

$$G_k(z) := \frac{1}{2} \sum_{(m,n)=1} \frac{1}{(mz+n)^k},$$

were considered in Example (2.1). They are $\Gamma(1)$-invariant functions of weight k. Clearly, they are holomorphic. The Fourier expansions of the functions are determined in (App.2.5). The series vanish when k is odd. For $k \geq 4$ even, the expansion is the following, for $q = e^{2\pi i z}$,

$$G_k(z) = 1 + \frac{-2k}{B_k} \sum_{n \geq 1} \sigma_{k-1}(n) q^n, \tag{5.6.1}$$

where B_k is the k'th Bernoulli number. In particular, the order at ∞ of the functions are equal to 0. As ∞ represents the only cusp of $\Gamma(1)$, it follows that the functions $E_k(z)$ and $G_k(z)$ are integral automorphic forms.

For $k = 4$ and $k = 6$, the expansions are the following,

$$G_4(z) = 1 + 240 \sum_{r \geq 1} \sigma_3(r) q^r,$$
$$G_6(z) = 1 - 504 \sum_{r \geq 1} \sigma_5(r) q^r.$$

The functions $G_4(z)^3$ and $G_6(z)^2$ are integral automorphic forms of weight 12. Clearly, in their Fourier expansion, the constant term is equal to 1. Hence, for the difference, $G_4(z)^3 - G_6(z)^2$, there is no constant term in the expansion, that is, the difference is a cusp form of weight 12 for $\Gamma(1)$. It is easy to see that the coefficient to q is equal to $1728 = 12^3$. So the difference has an expansion of the form,

$$G_4(z)^3 - G_6(z)^2 = 12^3 q + \cdots. \tag{5.6.2}$$

(5.7) Example. Dedekind's η-function,

$$\eta(z) = e^{2\pi i z/24} \prod_{n \geq 1} (1 - e^{2\pi i n z}),$$

was considered in Example (2.3). It is holomorphic in $\mathfrak{H}$ and everywhere non-zero. The factor j_η is of weight $\frac{1}{2}$ on $\Gamma(1)$ and, by its construction, the function η is j_η-invariant. Obviously, the order, $\operatorname{ord}_\infty^{\mathfrak{H}} \eta$, at ∞ is equal to $2\pi/24$. As ∞ represents the only cusp for $\Gamma(1)$, it follows that η is a j_η-automorphic form. The canonical generator at ∞ is the matrix t, and $tz = z+1$. Hence $h = 1$, and the $\Gamma(1)$-order of η at ∞ is equal to $1/24$. As the order is positive, it follows that η is a cusp form. At the points of $\mathfrak{H}$, the function η is non-zero, that is, the order is equal to 0.

The Fourier expansion of η is obtained by developing the product $\prod_{n\geq 1}(1-q^n)$ into a power series. With its first terms it becomes,

$$\eta(z) = q^{1/24}(1 - q - q^2 + q^5 + q^7 - \cdots). \tag{5.7.1}$$

In fact, the coefficients in the expansion are given by Euler's well known pentagonal number formula,

$$\prod_{n\geq 1}(1-q^n) = \sum_m (-1)^m q^{(3m^2-m)/2}.$$

It follows from the expansion of η that the discriminant $\Delta(z)$ defined in Example (2.3) as the power $\Delta(z) = \eta(z)^{24}$ has a Fourier expansion of the form,

$$\Delta(z) = q - 24q^2 + \cdots . \tag{5.7.2}$$

We prove later that the difference (5.6.2) is in fact equal to 12^3 times the discriminant (5.7.2).

(5.8) Example. The θ-function,

$$\theta(z) = \sum_n e^{\pi i n^2 z},$$

was considered in Example (2.6). It is holomorphic in $\mathfrak{H}$. The factor j_θ is of weight $\frac{1}{2}$ on the group Γ_θ, and by its construction, the function θ is j_θ-invariant. There are two cusps for Γ_θ, represented by the points ∞ and -1. Clearly, the θ-function converges uniformly to 1 as $\Im z \to \infty$. In particular, the θ-function is exponentially bounded at ∞ and the order at ∞ is equal to 0. Essentially, the sum defining θ is the Fourier expansion at ∞. In fact, at ∞ the canonical generator of Γ_θ is t^2, and hence $h = 2$. Thus the local parameter at ∞ is $q = e^{2\pi i z/2} = e^{\pi i z}$, and the Fourier expansion at ∞ is the series,

$$1 + 2q + 2q^2 + 2q^4 + 2q^9 + \cdots . \tag{5.8.1}$$

For the second cusp -1, take the conjugation by the matrix u. Then $u\infty = -1$. A weight-$\frac{1}{2}$ conjugate θ^u is then given by the equation,

$$\theta^u(z) = \frac{\varepsilon}{\sqrt{J(u,z)}}\theta(uz) = \frac{\varepsilon}{\sqrt{z}}\theta(-1-1/z).$$

for any complex sign ε. In particular, the function $\sqrt{i/z}\,\theta(-1-1/z)$ is a conjugate. It follows from (2.7.3) that the latter conjugate has the expansion,

$$\theta^u(z) = \sum_{n \text{ odd}} e^{\pi i n^2 z/4}. \tag{5.8.2}$$

In particular, the conjugate is exponentially bounded at ∞, and hence θ is exponentially bounded at -1. Therefore, θ is a j_θ-automorphic form. To get the Fourier expansion of θ at -1, note that under conjugation by u, the canonical generator of Γ_θ at -1 corresponds to t.

Hence $h = 1$, and the local parameter at ∞ is $q = e^{2\pi i z}$. Now, when n is odd, we have that $n^2 \equiv 1$ modulo 8. The n'th term in (5.8.2) is equal to $e^{2\pi i z(n^2-1)/8} e^{2\pi i z/8}$. Thus (5.8.2) is the Fourier expansion of θ at -1, and it takes the form,

$$\theta^u(z) = e^{2\pi i z/8} \sum_{n \text{ odd}} e^{2\pi i z(n^2-1)/8} = q^{1/8}(2 + 2q + 2q^3 + 2q^6 + 2q^{10} + \cdots).$$

In particular,

$$\operatorname{ord}_{-1}^{\Gamma_\theta} \theta = 1/8.$$

As the orders at the two cusps are nonnegative, θ is an integral form.

6. The Main Theorems.

(6.1). Keep the setup of (3.1). In other words, $\mathfrak{D}$ is a finite disk, Γ is a discrete subgroup of $\mathrm{SL}(\mathfrak{D})$, and j is a factor of real weight k for Γ. In this section, we assume throughout that Γ is a Fuchsian group of the first kind. In addition, we assume that the factor j is homogeneous, that is, if $-1 \in \Gamma$, then $j(-1, z) = 1$.

The group Γ acts on $\mathfrak{D}$ and on the closure $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. Since Γ is of the first kind, the orbit space $X = \overline{\mathfrak{D}/\Gamma}$ is a compact surface. Moreover, there exists a finite fundamental domain F for Γ. Accordingly, there are two fundamental invariants associated with Γ. The first is the nonnegative integer $g = g(\Gamma)$ defined as the genus of X. The second is the positive rational number $\mu = \mu(\Gamma)$ defined as the area of F divided by 2π. The two invariants are related by the formula, see (Discr.5.5),

$$\mu(\Gamma) = 2g(\Gamma) - 2 + \sum_{u \bmod \Gamma} \Big(1 - \frac{1}{e_u}\Big), \tag{6.1.1}$$

where $e_u = |\mathrm{P}\Gamma_u|$. In the sum (6.1.1), and in the sums below, the summation is over the orbits in X, or equivalently, over one point u in each Γ-orbit in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. For a Γ-parabolic point u, the fraction $1/e_u$ is interpreted as zero. So, in the sum (6.1.1), each of the finitely many cusps contributes with the term 1, and for the orbits in $\mathfrak{D}$, only the finitely many Γ-elliptic orbits contribute with non-zero terms.

In this section we state the main theorems on automorphic forms. The proofs are postponed to a separate section.

(6.2) Theorem A. *The field $\mathcal{M}(\Gamma)$ of Γ-automorphic functions on $\mathfrak{D}$ is finitely generated (as a field) over* $\mathbf{C}$, *and of transcendence degree* 1.

The quotient $X = X(\Gamma)$ is a compact Riemann surface, and by construction, $\mathcal{M}(\Gamma)$ is the field of meromorphic functions on X. The assertion of Theorem A is a well known consequence.

(6.3) Theorem B. *The space $\mathcal{M}(\Gamma, j)$ of j-automorphic forms on $\mathfrak{D}$ is non-zero. Moreover, if f is any non-zero j-automorphic form, then*

$$\sum_{u \bmod \Gamma} \mathrm{ord}_u^\Gamma f = \frac{k\mu(\Gamma)}{2}. \tag{6.3.1}$$

The existence of a non-zero j-automorphic form will follow from the construction of Poincaré Series. The proof of (6.3.1) will be given later in this chapter.

(6.4) Corollary. *The space $\mathcal{G}(\Gamma, j)$ of integral j-automorphic forms is of finite dimension over* $\mathbf{C}$. *If $k < 0$, then the dimension is equal to zero. If $k > 0$ then the dimension is at most equal to $k\mu/2 + 1$. If $k = 0$, in which case $j(\gamma, z) = \chi(\gamma)$ with a unitary character χ on Γ, then the dimension is zero if $\chi \neq 1$. Finally, the functions of $\mathcal{G}(\Gamma)$ are exactly the constant functions.*

Proof. Consider a nonzero function f in $\mathcal{G}(\Gamma, j)$. Then the orders $\mathrm{ord}_u^\Gamma f$ are non-negative, and the sum in (6.3.1) is equal to $k\mu/2$. Clearly, therefore, no such f can exist if $k < 0$.

Assume $k \geq 0$. Then $\mathrm{ord}_u^\Gamma f \leq k\mu/2$. Let N be the largest integer less than or equal to $k\mu/2$. To prove that the dimension is finite we may assume that the disk is the unit disk $\mathfrak{E}$. Since the order of f at zero is nonnegative, the normalized Laurent series around 0 is of the form,

$$f(z) = z^{e\kappa} \sum_{n\geq 0} a_n w^n \quad \text{for } w = z^e, \tag{6.4.1}$$

where $\kappa = \kappa_0(j)$ and $e = e_0$. The Γ-order of f at zero is equal to $\kappa + n$ where n is least index such that $a_n \neq 0$; on the other side, the Γ-order is at most equal to $k\mu/2$. It follows that one of the $N + 1$ coefficients a_n for $0 \leq n \leq k\mu/2$ is nonzero. The $N + 1$ coefficients may be viewed as a $\mathbf{C}$-linear map $\mathcal{G}(\Gamma, j) \to \mathbf{C}^{N+1}$, and we have just seen that the map is injective. Hence the dimension of $\mathcal{G}(\Gamma, j)$ is finite and at most equal to $N + 1$.

Assume that $k = 0$. Let f be a function in $\mathcal{G}(\Gamma, \chi)$. Then f is holomorphic in $\mathfrak{D}$, and defines a continuous function on $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$. Moreover, since χ is unitary, it follows from the equation $f(\gamma z) = \chi(\gamma)f(z)$ that the function $|f|$ is Γ-invariant. As the quotient $\overline{\mathfrak{D}/\Gamma}$ is compact, it follows that $|f|$ attains its maximum value at a point u of $\mathfrak{D} \cup \partial_\Gamma\mathfrak{D}$. It follows from the maximum principle that f is constant. Indeed, if u is in $\mathfrak{D}$, the maximum principle applies directly. If u is in $\partial_\Gamma\mathfrak{D}$, we may, after conjugation, assume that $\mathfrak{D} = \mathfrak{H}$ and $u = \infty$. If $\kappa_u = 0$ then, in a neighborhood of u, $f = g(q)$ where $q = e^{2\pi i z/h}$ and g is holomorphic at the origin; hence the maximum principle applies to g. If $\kappa_u > 0$, then f vanishes at u; by the choice of u, therefore $f = 0$.

Hence every function in $\mathcal{G}(\Gamma, \chi)$ is constant. Clearly, if $\chi \neq 1$, then no non-zero constant is χ-invariant.

Thus all assertions of the Corollary have been proved. □

(6.5) Note. Assume that $k > 0$. It follows from (6.1.1) that the number $k\mu/2$ is at most equal to the following number,

$$k(g-1) + \frac{k}{2}\#(\text{parabolic or elliptic } \Gamma\text{-orbits}). \tag{6.5.1}$$

Hence, if N is the integer part of the number (6.5.1), then the dimension of $\mathcal{G}(\Gamma, j)$ is at most $N+1$. Clearly, the argument given in the proof of (6.4) applies equally well to a Γ-parabolic point u, replacing the Laurent expansion by the Fourier expansion. In particular, if for two functions f and g in $\mathcal{G}(\Gamma, j)$ it is known that their first $N + 1$ Fourier coefficients are equal, then the two functions are equal.

(6.6) Special case. *Assume that the disk is the upper half plane $\mathfrak{H}$ and that Γ is a subgroup of finite index in the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$. Denote by d the homogeneous index, $d = |\Gamma(1) : \tilde{\Gamma}|$. Then, for any non-zero j-automorphic form on $\mathfrak{H}$,*

$$\sum_{u \bmod \Gamma} \mathrm{ord}_u^\Gamma f = \frac{kd}{12}\,. \tag{6.6.1}$$

Moreover, if $k > 0$, then the dimension of $\mathcal{G}(\Gamma, j)$ is at most equal to $[kd/12] + 1$.

Proof. In the general setup of (6.1), let Δ be a subgroup of finite index in Γ. Set $d := |\mathrm{P}\Gamma{:}\mathrm{P}\Delta|$. Then, by Proposition (Discr.3.12), a fundamental domain G for Δ can be obtained as the union of d transforms of the fundamental domain F for Γ. Hence the area of G is d times the area of F. Therefore, $\mu(\Delta) = d\,\mu(\Gamma)$.

A fundamental domain for the modular group $\Gamma(1)$ was described in (Mdlar.4.1). Clearly, its area is equal to $2\pi/6$. Consequently, $\mu(\Gamma(1)) = \frac{1}{6}$. Therefore, the assertion is a special case Theorem B and its Corollary. ∎

(6.7) Theorem C. *The number $\delta = \delta(\Gamma, j)$, defined by the following expression, is an integer:*

$$\delta := 1 - g + \frac{k\mu}{2} - \sum_{u \bmod \Gamma} \kappa_u(j), \tag{6.7.1}$$

where $\kappa_u(j)$ is the parameter of j at u. If $\delta < 1 - g$, then $\mathcal{G}(\Gamma, j) = (0)$. Assume that $\delta \geq 1 - g$. Then the following inequalities hold:

$$\delta \leq \dim \mathcal{G}(\Gamma, j) \leq \delta + g.$$

Moreover, if $\delta \geq g$, then

$$\dim \mathcal{G}(\Gamma, j) = \delta. \tag{6.7.2}$$

The number δ is an integer. Indeed, let f be a non-zero j-automorphic form (Theorem A). By Theorem B, the number $k\mu/2$ is the sum of the Γ-orders of f. As observed in Section 5, the Γ-order $\mathrm{ord}_u^\Gamma f$ is congruent modulo $\mathbf{Z}$ to the parameter $\kappa_u(j)$. Thus the difference $\mathrm{ord}_u^\Gamma f - \kappa_u(j)$ is an integer. Hence δ is an integer.

If f is an integral non-zero j-automorphic form, then $\mathrm{ord}_u^\Gamma f \geq 0$. As the order is congruent to κ_u and $0 \leq \kappa_u < 1$, it follows that $\mathrm{ord}_u^\Gamma f \geq \kappa_u(j)$. Hence $\delta \geq 1 - g$. Therefore, the assertion that $\mathcal{G}(\Gamma, j) = (0)$ when $\delta < 1 - g$ is a consequence of Theorem B.

The remaining assertions of Theorem C are consequences of Riemann's part of the Riemann–Roch Theorem. Note that the assertions give full information on the dimension of $\mathcal{G}(\Gamma, j)$ when $g = 0$. If $g > 0$, the Theorem gives only an estimate of the dimension when $1 - g \leq \delta \leq g - 1$.

(6.8) Note. The following alternative expressions for δ are easily obtained from the formula (6.1.1):

$$\begin{aligned}\delta &= g - 1 + \left(\tfrac{k}{2} - 1\right)\mu + \sum\left(1 - \tfrac{1}{e_u} - \kappa_u\right)\\ &= (k-1)(g-1) + \sum\left(\tfrac{k}{2}\left(1 - \tfrac{1}{e_u}\right) - \kappa_u\right).\end{aligned}$$

(6.9) Corollary. *The dimension of $\mathcal{G}(\Gamma, j)$ is equal to δ and at least equal to g under any of the following conditions:* (1) $k > 2$, *or* (2) $k = 2$ *and there are Γ-parabolic points, or* (3) $k = 2$ *and there are Γ-elliptic points u at which $\kappa_u(j) < 1 - 1/e_u$.*

Proof. Consider the first expression for δ in (6.8). In the sum, the term corresponding to a Γ-parabolic point u is equal to $1 - \kappa_u$, and hence positive. The term corresponding to a Γ-elliptic point is non-negative by (3.3), and positive if $\kappa_u < 1 - 1/e_u$. It follows easily the $\delta > g - 1$ under any of the conditions given for the first assertion. Therefore, since δ is an integer, the first assertion of the Corollary follows from Theorem C. ∎

(6.10) Definition. A Γ-parabolic point u is said to represent a *j-regular cusp*, if the sign $\omega_u(j)$ is equal to 1, or, equivalently, if $\kappa_u(j) = 0$. Obviously, j-regularity is a property of the cusp represented by u.

It follows from (3.13) that a cusp is J-regular if and only if it is regular. Moreover, if k is an even integer, then all cusps are J^k-regular. If k is an odd integer, then a cusp is J^k-regular if and only if it is a regular cusp for Γ.

Denote by $\delta' = \delta'(\Gamma, j)$ the following number,

$$\delta' = \delta - \#(j\text{-regular cusps}).$$

(6.11) Theorem D. *Consider the space $\mathcal{S}(\Gamma, j)$ of j-automorphic cusp forms on $\mathfrak{D}$. If $\delta' < 1 - g$, then $\mathcal{S}(\Gamma, j) = (0)$. Assume that $\delta' \geq 1 - g$. Then the following inequalities hold:*

$$\delta' \leq \dim \mathcal{S}(\Gamma, j) \leq \delta' + g.$$

Moreover, if $\delta' \geq g$, then

$$\dim \mathcal{S}(\Gamma, j) = \delta'. \tag{6.11.1}$$

(6.12) Corollary. *The dimension of $\mathcal{S}(\Gamma, j)$ is equal to δ' and at least equal to g under any of the following conditions:* (1) $k > 2$, *or* (2) $k = 2$ *and there are Γ-parabolic points that are not j-regular, or* (3) $k = 2$ *and there are Γ-elliptic points u at which $\kappa_u(j) < 1 - 1/e_u$.*

Proof. Consider the first expression for δ in (6.8). Clearly, in the sum, each regular cusp u contributes with the term 1. Therefore, if we omit from the sum the terms corresponding to the regular cusps, the resulting expression is equal to δ'. The remaining part of the proof is now identical to the proof of Corollary (6.9). ∎

(6.13) Note. An integral j-automorphic form f has non-negative Γ-order at every point u of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. In particular, f has a well defined value at points u of $\mathfrak{D}$. For a point $u \in \partial_\Gamma \mathfrak{D}$, choose a conjugation $\alpha\colon (\mathfrak{H}, \infty) \to (\mathfrak{D}, u)$. Then the weight-$k$ conjugate function f^α has a value at ∞ which, by an abuse of language, will be referred to as the value of f at u. The value is zero if and only if the order of f at u is positive. The order is at least equal to the parameter κ_u. In particular, the value at a point u representing a j-irregular cusp is necessarily zero. It follows that f is a cusp form if and only if the value is equal to zero at every j-regular cusp. In other words, if Z_{reg} is a finite set of Γ-parabolic points representing the subset of j-regular cusps, then the evaluation map,

$$\mathcal{G}(\Gamma, j) \to \mathbf{C}^{Z_{\mathrm{reg}}}, \tag{6.10.1}$$

has as kernel the space $\mathcal{S}(\Gamma, j)$ of cusp forms. In particular, the codimension of $\mathcal{S}(\Gamma, j)$ in $\mathcal{G}(\Gamma, j)$ is at most equal to the number of j-regular cusps, and equality holds if and only if the evaluation map is surjective. Clearly, the evaluation map is surjective, if and only if, for every j-regular cusp u for Γ, there exists an integral j-automorphic form that has a non-zero value at u and vanishes at all other cusps.

Since $\delta - \delta'$ is equal to the number of j-regular cusps, it follows that if both equalities (6.7.2) and (6.11.1) hold, then the evaluation map is surjective. In particular, therefore the evaluation map is surjective if $\delta' \geq g$, and especially, it is always surjective when $k > 2$. It should be emphasized that the evaluation map is not surjective in general.

(6.14). Consider in particular the factor $j = J^k$. Thus k is assumed to be an integer; if k is odd, the group Γ is assumed to be inhomogeneous. The parameters at cusps u were considered in (6.10): If k is even, then all cusps are J^k-regular (and $\kappa_u = 0$). If k is odd, then the J^k-regular cusps are exactly the regular cusps for Γ. At the irregular cusps, the parameter is equal to $\frac{1}{2}$.

Consider the parameter at a point u of $\mathfrak{D}$. It follows from (3.13) that $\kappa_u(J^k)$ is equal to $\bigl\{\frac{k}{2}(1 - 1/e_u)\bigr\}$, where $\{x\}$ denotes the fractional part, $\{x\} := x - [x]$.

An expression for $\delta_k = \delta(\Gamma, J^k)$ is obtained from the second expression for δ in (6.8). If k is even,

$$\delta_k = (k-1)(g-1) + \sum \bigl[\tfrac{k}{2}\bigl(1 - \tfrac{1}{e_u}\bigr)\bigr]. \tag{6.14.1}$$

Note that each cusp of Γ contributes with the integer $k/2$ to the sum. If k is odd,

$$\delta_k = (k-1)(g-1) + \sum \bigl[\tfrac{k}{2}\bigl(1 - \tfrac{1}{e_u}\bigr)\bigr] + \tfrac{1}{2}\,\#(\text{regular cusps}). \tag{6.14.2}$$

Note that each cusp contributes with the integer $\bigl[\frac{k}{2}\bigr] = \frac{k}{2} - \frac{1}{2}$ to the sum.

The numbers for $k = 1$ and $k = 2$ are the following:

$$\delta_1 = \frac{1}{2}\,\#(\text{regular cusps}), \quad \delta_2 = g - 1 + \#(\text{cusps}). \tag{6.14.3}$$

The dimensions of $\mathcal{G}_k(\Gamma)$ and $\mathcal{S}_k(\Gamma)$ are determined by the preceding results for $k \leq 0$ and $k \geq 3$. In addition, if there are cusps for Γ, then the results imply that $\dim \mathcal{G}_2(\Gamma) = \delta_2$.

(6.15) Theorem E. *The dimension of $\mathcal{S}_2(\Gamma)$ is equal to g. In addition, if Γ is inhomogeneous, then the codimension of $\mathcal{S}_1(\Gamma)$ in $\mathcal{G}_1(\Gamma)$ is equal to half the number of regular cusps.*

The proof of Theorem E uses the Roch part of the Riemann–Roch Theorem.

Note that the dimension of $\mathcal{G}_2(\Gamma)$ is determined in all cases. If there are cusps, then the dimension is equal to the number δ_2 of (6.14.3). If there are no cusps, then of course $\mathcal{G}_2(\Gamma) = \mathcal{S}_2(\Gamma)$, and the dimension is equal to g.

As noted in (6.13), the codimension of $\mathcal{S}_k(\Gamma)$ in $\mathcal{G}_k(\Gamma)$ is equal to the number of J^k-regular cusps if $k \geq 3$. It follows from Theorem E that the assertion does not hold for $k = 2$ if there are cusps, and it does not hold for $k = 1$ if there are regular cusps.

Theorem E determines the codimension of $\mathcal{S}_1(\Gamma)$ in $\mathcal{G}_1(\Gamma)$, but no general method is known to determine the actual dimension.

(6.16) Theorem F. *Assume that the disk is the upper half plane $\mathfrak{H}$. Let f be a non-zero j-automorphic cusp form. Then the function $f(z)(\Im z)^{k/2}$ is bounded in $\mathfrak{H}$. Consider the Fourier expansion of f around a cusp u,*

$$f^\alpha = e^{2\pi i\kappa z/h} \sum_{n=0}^{\infty} a_n e^{2\pi i n z/h} \tag{6.16.1}$$

where $\kappa = \mathrm{ord}_u^\Gamma f$ is positive. Then, for the coefficients, we have the estimate,

$$a_n = \mathrm{O}(n^{k/2}). \tag{6.16.2}$$

Proof. Note that the weight k is non-negative, since $\mathcal{S}(\Gamma, j) \neq (0)$. Consider the function $\rho(z) = |f(z)|(\Im z)^{k/2}$. As $\Im(\gamma z) = |J(\gamma, z)|^{-2}\Im(z)$ and f is j-invariant, it follows that $\rho(z)$ is Γ-invariant. Obviously, ρ is continuous in $\mathfrak{D}$. We claim that $\rho(z)$ converges to zero when z approaches a cusp u. Clearly, if α is a matrix of $\mathrm{SL}(\mathfrak{H})$ such that $\alpha\infty = u$, it suffices to prove that $\rho(\alpha z) \to 0$ uniformly as $\Im z \to \infty$. Clearly,

$$\rho(\alpha z) = |f(z)|\Im(\alpha z)^{k/2} = \frac{|f(z)|}{|J(\alpha, z)|^k}(\Im z)^{k/2}. \tag{6.16.3}$$

The fraction on the right is the absolute value of a weight-k conjugate f^α. It follows from the Fourier expansion (6.16.1) that $f^\alpha = \mathrm{O}(|e^{2\pi\kappa z/h}|)$. Thus $\rho(\alpha z) \to 0$ for $\Im z \to \infty$.

It follows that ρ extends to a continuous function on the quotient $\overline{\mathfrak{D}/\Gamma}$. As the quotient is compact, therefore ρ is bounded on $\mathfrak{H}$.

Assume that $\rho(z) \leq K$ for all z in $\mathfrak{H}$. Let $g(q)$ be the function defined in the unit disk: $|q| < 1$ by $f^\alpha(z) = e^{2\pi i\kappa z/h} g(e^{2\pi i z/h})$. Then g is holomorphic in the unit disk, since f in holomorphic in $\mathfrak{H}$. The Fourier coefficients a_n are the coefficients in the power series expansion of g. Hence they can be obtained by integration,

$$a_n = \frac{1}{2\pi i} \int \frac{g(q)}{q^n} \frac{dq}{q},$$

where the path integral can be taken over any circle around 0 in the unit disk. Equivalently,

$$a_n = \frac{1}{h} \int \frac{g(e^{2\pi i z/h})dz}{e^{2\pi i n z/h}} = \frac{1}{h} \int \frac{f^\alpha(z)dz}{e^{2\pi i(n+\kappa)z/h}},$$

where the path integral can be taken over any (euclidean) vertical line segment of length h in $\mathfrak{H}$. Integrate from $i\varepsilon$ to $i\varepsilon + h$. On the path we have, by (6.16.3), the estimate,

$$|f^{\alpha}(z)| = \rho(\alpha z)(\Im z)^{-k/2} \leq K\varepsilon^{-k/2},$$

and the equality $|e^{2\pi i(n+\kappa)z/h}| = e^{-2\pi(\kappa+n)\varepsilon/h}$. Therefore, we obtain the estimate for the coefficient,

$$|a_n| \leq K\varepsilon^{-k/2}e^{2\pi(\kappa+n)\varepsilon/h}.$$

Fix ε and apply the estimate with $\varepsilon := \varepsilon/n$ for $n \geq 1$. Since $(\kappa + n)/n \leq 2$, it follows that

$$|a_n| \leq K\varepsilon^{-k/2}e^{4\pi\varepsilon/h}\, n^{k/2}.$$

Therefore (6.16.2) holds. □

7. Examples II.

(7.1) Example. For the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ acting on $\mathfrak{H}$, there are two elliptic orbits represented by i of order 2 and ρ of order 3, and one cusp represented by ∞. The genus is $g = 0$; the area is $\pi/3$, and so $\mu = \frac{1}{6}$, confirming that

$$\tfrac{1}{6} = 0 - 2 + (1 - \tfrac{1}{2}) + (1 - \tfrac{1}{3}) + 1.$$

Let k be a non-negative even integer, and consider the spaces $\mathcal{G}_k := \mathcal{G}(\Gamma(1), J^k)$ and $\mathcal{S}_k := \mathcal{S}(\Gamma(1), J^k)$. Apply the Main Theorems to the factor J^k. The number $\delta_k = \delta(\Gamma(1), J^k)$ is given by (6.14.1):

$$\delta_k = -(k-1) + \left[\tfrac{k}{2}\left(1 - \tfrac{1}{2}\right)\right] + \left[\tfrac{k}{2}\left(1 - \tfrac{1}{3}\right)\right] + \tfrac{k}{2} = \begin{cases} 1 + \left[\frac{k}{12}\right] & \text{when } k \not\equiv 2 \pmod{12}, \\ \left[\frac{k}{12}\right] & \text{when } k \equiv 2 \pmod{12}. \end{cases}$$

Note that $\delta_k \geq 0 = g$ since k is non-negative. Hence it follows from Theorem C that

$$\dim \mathcal{G}_k = \delta_k.$$

Clearly, if $k \geq 4$, then $\delta_k - 1 \geq 0 = g$. Therefore, by Theorem D, if $k \geq 4$, then

$$\dim \mathcal{S}_k = \delta_k - 1.$$

Obviously, $\mathcal{S}_0 = 0$, and by Theorem E,

$$\dim \mathcal{S}_2 = 0.$$

For small values of k, the dimensions are the following:

k	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28
$\dim \mathcal{G}_k$	1	0	1	1	1	1	2	1	2	2	2	2	3	2	3
$\dim \mathcal{S}_k$	0	0	0	0	0	0	1	0	1	1	1	1	2	1	2

The Eisenstein series $E_k(z)$ and modified series $G_k(z)$, for $k \geq 4$, are functions in $\mathcal{G}_k$. They are not cusp forms, since G_k has the value 1 at ∞. Clearly, the function $G_k(z)$ generates the 1-dimensional space $\mathcal{G}_k$ for $k = 4, 6, 8, 10, 14$. In particular, comparing the values at ∞, we obtain the equations $G_8(z) = G_4(z)^2$, $G_{10}(z) = G_4(z)G_6(z)$, and $G_{14}(z) = G_4(z)^2 G_6(z)$.

The first weight k at which there is a non-trivial cusp form is $k = 12$. The discriminant $\Delta(z)$, defined by $\Delta(z) = \eta(z)^{24}$, is a cusp form of weight 12, with the Fourier expansion $\Delta(z) = q - 24q + \cdots$. Hence it generates the 1-dimensional space $\mathcal{S}_{12}$. On the other hand, the difference $G_4^3 - G_6^2$ is a cusp form; it is easily seen that the coefficient to q in its Fourier expansion is equal to 1728. Therefore, since the space $\mathcal{S}_{12}$ is 1-dimensional, the following equation is a consequence:

$$\Delta(z) = 12^{-3}\left(G_4^3 - G_6^2\right). \tag{7.1.1}$$

The Fourier expansion of $\Delta(z)$ is of the following form,

$$\Delta(z) = \sum_{n \geq 1} \tau(n) q^n, \quad \text{with } \tau(1) = 1. \tag{7.1.2}$$

The function $\tau(n)$ is *Ramanujan's τ-function*. From the equation $\Delta(z) = \eta(z)^{24}$, it follows immediately that $\tau(n) \in \mathbf{Z}$. It is not hard to see directly the right hand side of (7.1.1) has a Fourier expansion with integral coefficients.

(7.2) Example. The Fourier expansions of the functions $G_k(z)$ for $k = 4, 6, 8, \dots$ are determined in (App.2.6): If we write $\Sigma_k := \sum_{r\geq 1} \sigma_k(r) q^k$, then for even $k \geq 4$,

$$G(z) = 1 + \alpha_k \Sigma_{k-1}, \tag{7.2.1}$$

where $\alpha_k = -2k/B_k$. In terms of the integers A_k of (App.1),

$$\alpha_k = \frac{(-1)^{k/2} 2^{k+1}(2^k - 1)}{A_k}.$$

In particular, as $A_4 = 2$ and $A_6 = 2^4$, it follows that $\alpha_4 = 2^4 \cdot 3 \cdot 5 = 240$ and $\alpha_6 = -2^3 \cdot 3^2 \cdot 7 = -504$. The number α_k is the coefficient of q in the Fourier expansion (7.2.1). In particular, from the equations $G_8 = G_4^2$, $G_{10} = G_4 G_6$, and $G_{14} = G_8 G_6$ observed in (7.1), it follows that $\alpha_8 = 2\alpha_4 = 2^5 \cdot 3 \cdot 5$, $\alpha_{10} = \alpha_4 + \alpha_6 = -2^3 \cdot 3 \cdot 11$, and $\alpha_{14} = -2^3 \cdot 3$. Of course, the values of α_8 and α_{10} confirm the values $A_8 = 2^4 \cdot 17$ and $A_{10} = 2^8 \cdot 31$ in (App.1.4). Since $A_{12} = 2^9 \cdot 691$, it follows that

$$\alpha_{12} = \frac{2^4 \cdot 3^2 \cdot 5 \cdot 7 \cdot 13}{691}. \tag{7.2.2}$$

Relations among the series $G_k(z)$ imply relations among their Fourier coefficients. For instance, from the relation $G_8 = G_4^2$, it follows that $\alpha_8 \Sigma_7 = (\alpha_4 \Sigma_3)^2 + 2\alpha_4 \Sigma_3$. Hence,

$$\sigma_7(r) = \sigma_3(r) + 2^3 \cdot 3 \cdot 5 \sum_{t=1}^{r-1} \sigma_3(t)\sigma_3(r-t).$$

For the discriminant $\Delta(z)$, say given by (7.1.1), it follows that

$$(12)^3 \Delta(z) = G_4^3 - G_6^2 = \alpha_4^3 \Sigma_3^3 + 3\alpha_4^2 \Sigma_3^2 + 3\alpha_4 \Sigma_3 - \alpha_6^2 \Sigma_5^2 - 2\alpha_6 \Sigma_5.$$

As a consequence,

$$\Delta(z) = \frac{5\Sigma_3 + 7\Sigma_5}{12} + 8000\Sigma_3^3 + 100\Sigma_3^2 - 147\Sigma_5^2.$$

It is easily seen that the fraction has integer coefficients. Hence the coefficients $\tau(n)$ of $\Delta(z)$ are integers. It follows easily from the equation that $\tau(1) = 1$, $\tau(2) = -24$, and $\tau(3) = 252$. It can be proved that the function $\tau(n)$ is multiplicative. Moreover, for a prime p the following equation holds:

$$\tau(p^{t+1}) = \tau(p)\tau(p^t) - p^{11}\tau(p^{t-1}).$$

Thus $\tau(n)$ is completely determined by its values on primes p. Since $\Delta(z)$ is a cusp form of weight 12, Theorem F implies the estimate $\tau(n) = \mathrm{O}(n^{12/2})$. In fact, Ramanujan's conjecture,

$$|\tau(p)| < 2p^{11/2} \quad \text{for all primes } p,$$

was proved by P. Deligne in 1974. Far more trivial is Ramanujan's congruence,

$$\tau(n) \equiv \sigma_{11}(n) \pmod{691}.$$

It can be proved as follows. The form $G_{12} - G_4^3$ is a cusp form of weight 12, and hence equal to a constant times Δ. The constant can be determined by comparing the coefficients to q. It follows that $G_{12} - G_4^3 = (\alpha_{12} - 3\alpha_4)\Delta$. By (7.2.2) the number $a := 691\alpha_{12}$ is an integer, and prime to 691. Hence, by multiplying by 691, we obtain the equations of forms with integer coefficients:

$$(a - 691 \cdot 3 \cdot \alpha_4)\Delta = 691G_{12} - 691G_4^3 = a\Sigma_{11} - 691(G_4^3 - 1).$$

Modulo 691, it follows that $a\tau(n) \equiv a\sigma_{11}(n)$. Hence Ramanujan's congruence holds.

(7.3) Example. Consider again the modular group $\Gamma(1)$. Let f be a non-zero function in $\mathcal{M}_k = \mathcal{M}_k(\Gamma(1))$. By the Special Case (6.6), we have the equation $\sum_u \operatorname{ord}_u^{\Gamma(1)} f = k/12$. The $\Gamma(1)$-order at ∞, which we will denote by $\operatorname{ord}_\infty f$, is an integer, since the parameter at ∞ is equal to zero. The $\Gamma(1)$-order at i is equal to $\frac{1}{2}\operatorname{ord}_i f$ and the $\Gamma(1)$-order at ρ is equal to $\frac{1}{3}\operatorname{ord}_\rho f$. At points of $\mathfrak{H}$ that are not $\Gamma(1)$-equivalent to i or ρ, the $\Gamma(1)$-order is simply the usual order. Hence the equation may be written as follows:

$$\sum_u \operatorname{ord}_u f + \tfrac{1}{2}\operatorname{ord}_i f + \tfrac{1}{3}\operatorname{ord}_\rho f + \operatorname{ord}_\infty f = \tfrac{k}{12}, \tag{7.3.1}$$

where the sum is over a system of representatives for the $\Gamma(1)$-ordinary orbits. If f is an integral form, then the orders are non-negative.

For example, consider a non-zero form f in $\mathcal{S}_{12}$. The right hand side of (7.3.1) is equal to 1. On the left, the order $\operatorname{ord}_\infty f$ is an integer, and positive since f is a cusp form. Hence it follows from (7.3.1) that $\operatorname{ord}_\infty f = 1$ and that f is of order 0 at all points of $\mathfrak{H}$. In other words, f has no zeros in $\mathfrak{H}$. Of course we know, by example (7.1), that f is a multiple of $\Delta(z)$, and it is obvious from $\Delta(z) = \eta(z)^{24}$ that $\Delta(z)$ has no zeros.

Consider the Eisenstein series G_k for $k \geq 4$. The functional equations, for $\gamma = s$ and $\gamma = u$, are the following:

$$G_k(sz) = z^k G_k(z), \quad G_k(uz) = z^k G_k(z).$$

The point i is a fixed point of s. Hence, by the first equation, if $k \not\equiv 0 \pmod 4$, then $G_k(i) = 0$. Similarly, if $k \not\equiv 0 \pmod 6$, then $G_k(\rho) = 0$. In particular,

$$G_4(\rho) = 0, \qquad G_6(i) = 0. \tag{7.3.2}$$

It follows from (7.3.1) that the points in the orbit represented by ρ are the only zeros of G_4 and, moreover, that these points are simple zeros. Similarly, the function G_6 has simple zeros at the points in the orbit represented by i, and no other zeros.

(7.4) Example. Klein's j-invariant is the function defined in $\mathfrak{H}$ by

$$j(z) := G_4(z)^3/\Delta(z).$$

It is of weight 0. Since $\Delta(z)$ has no zeros in $\mathfrak{H}$, the function $j(z)$ is holomorphic in $\mathfrak{H}$. It has a simple pole at ∞ (since $\Delta(z)$ has a simple zero), with the Fourier expansion,

$$j(z) = q^{-1} + 744 + 196884q + \cdots .$$

The coefficients are integers. Indeed, it follows from (7.1.2) that $\Delta(z)^{-1}$ has an expansion $\Delta(z)^{-1} = q^{-1} + \cdots$ with integral coefficients. As the Fourier coefficients of G_4 are integral, therefore, so are the coefficients of $j(z)$.

The two values $j(\rho) = 0$ and $j(i) = 1728$ are immediate from (7.1.1) and (7.3.2). Consider, for $\lambda \in \mathbf{C}$, the function $f(z) = j(z) - \lambda$ and the equation (7.3.1). The right hand side is 0, and the order at infinity is -1. It follows that exactly one of the remaining orders is non-zero. Therefore, the function $j(z)$ has the value λ exactly at the points of one single orbit. In other words, Klein's j-invariant defines a bijection,

$$\overline{\mathfrak{H}/\Gamma(1)} \xrightarrow{\sim} \overline{\mathbf{C}}. \tag{7.4.1}$$

In fact, it follows that $j(z)$ has ρ as a triple zero, it takes the value 1728 with multiplicity 2 at the point i, and it takes any other value with multiplicity 1 at the points of the corresponding orbit. It is easy to deduce that (7.4.1) is an analytic isomorphism.

(7.5) Note. The result of the previous example has as corollary the following theorem of Picard: *Any holomorphic function $f(z)$ in $\mathbf{C}$ which avoids at least two values is constant.*

Indeed, we may assume that $f(z)$ avoids the two values $w_1 := j(\rho) = 0$ and $w_2 := j(i) = 1728$. Let X by the open subset of $\mathfrak{H}$ obtained by subtracting the two orbits containing i and ρ, and let Y be the open subset of $\mathbf{C}$ obtained by subtracting to two points w_1 and w_2. Then, by (7.4), the j-invariant is an analytic isomorphism of $X/\Gamma(1)$ onto Y. Since $\Gamma(1)$ acts properly discontinuous on X, and without fixed points, it follows that j defines a covering $j\colon X \to Y$. By assumption, f maps $\mathbf{C}$ into Y. Therefore, since $\mathbf{C}$ is simply connected, f lifts to a holomorphic map $\tilde{f}\colon \mathbf{C} \to X$. Consider the function $e^{i\tilde{f}(z)}$. It is defined on $\mathbf{C}$, and bounded, because $\tilde{f}(z)$ takes values in the upper half plane; hence it is constant. It follows that $\tilde{f}(z)$ is constant. Therefore, $f(z)$ is constant.

(7.6) Example. The field of $\Gamma(1)$-automorphic functions is generated by Klein's invariant,

$$\mathcal{M}(\Gamma(1)) = \mathbf{C}(j). \tag{7.6.1}$$

Indeed, the function $j(z)$ has a simple pole at ∞ and no other poles. If $u \in \mathfrak{H}$, then the function $j(z) - j(u)$ has a zero at u, and the $\Gamma(1)$-order of the zero is equal to 1. It follows that the function $(j(z) - j(u))^{-m}$ has a pole of order m (i.e., a zero of $\Gamma(1)$-order $-m$) at the points of the orbit represented by u, and no other poles. Let f be a non-zero $\Gamma(1)$-automorphic function. It follows that by subtracting from f a linear combination of functions of the form,

$$j^n, \quad (j - j(u))^{-m},$$

we can obtain a difference without poles. Thus the difference is a constant function. Therefore f belongs to $\mathbf{C}(j)$.

(7.7) Example. Consider again the modular group $\Gamma(1)$. Let k be an arbitrary positive real number. By Proposition (3.4), a factor of weight k on $\Gamma(1)$ is completely determined by the sign at ∞ which is of the form,

$$\omega_\infty = e^{2\pi i(k-a)/12}, \tag{7.7.1}$$

where a is an integer, $0 \le a < 12$. To get a homogeneous factor, assume that a is even. Denote by $j_{k,a}$ the corresponding factor, and set $\mathcal{M}_k(a) := \mathcal{M}(\Gamma(1), j_{k,a})$ etc. By (3.3), the parameters at ∞, i, and ρ of the factor $j_{k,a}$ are, respectively,

$$\kappa_\infty = \{(k-a)/12\}, \quad \kappa_i = \{a/4\}, \quad \kappa_\rho = \{a/3\}.$$

Hence the number $\delta_{k,a} = \delta(\Gamma(1), j_{k,a})$ is given by the equation,

$$\delta_k(a) = 1 + k/12 - \{(k-a)/12\} - \{a/4\} - \{a/3\}.$$

When k is an even integer, the factor J^k is obtained by taking $a \equiv k \pmod{12}$, and the expression above reduces to that of Example (7.1).

Take $k = \frac{1}{2}$ and $a = 0$. The corresponding factor of weight $\frac{1}{2}$ is then the η-factor, see Example (3.6). Clearly, $\delta_{1/2}(0) = 1$. Hence $\mathcal{G}(\Gamma(1), j_\eta) = \mathcal{S}(\Gamma(1), j_\eta)$ is a 1-dimensional space generated by the η-function.

(7.8) Example. Consider the θ-group Γ_θ. The genus is $g = 0$, and $\mu = \frac{1}{2}$ since the index of Γ_θ in $\Gamma(1)$ is equal to 3. For the θ-group, there is one elliptic orbit, represented by i of order 2 with canonical generator $\gamma_i = s$. In addition, there are two cusps: one is represented by ∞, where the canonical generator is t^2, and one is represented by -1 where the canonical generator is conjugate to t. The θ-function belongs to $\mathcal{G}(\Gamma_\theta, j_\theta)$. The factor j_θ has weight $k = \frac{1}{2}$, and the parameters at the three orbits are respectively,

$$\kappa_i = 0, \quad \kappa_\infty = 0, \quad \kappa_{-1} = \tfrac{1}{8}.$$

Consequently, $\delta(\Gamma_\theta, j_\theta) = 1 + \frac{1}{2}\frac{1}{2}\frac{1}{2} - \frac{1}{8} = 1$. Therefore, the space $\mathcal{G}(\Gamma_\theta, j_\theta)$ is 1-dimensional, and the θ-function is a generator. The θ-function is not a cusp form since it has the value 1 at ∞. Hence there are no non-trivial j_θ-automorphic cusp forms, confirming that $\delta' = 0$ since the cusp ∞ is j_θ-regular and the cusp -1 is not.

The θ-function has a zero of order $1/8$ at the cusp -1. Therefore, from the equation $\sum_u \operatorname{ord}_u^{\Gamma_\theta} f = 1/8$, it follows that $\theta(z)$ has no zeros in $\mathfrak{H}$.

Since $\theta(z)$ has no zeros in $\mathfrak{H}$ and the value 1 at ∞, there is, for any real number l, a unique determination of $\theta(z)^l$ with the property that the limit of $\theta(z)^l$ for $\Im z \to \infty$ is equal to 1. Obviously, $\theta(z)^l$ is a j_θ^l-automorphic form, where j_θ^l is the factor of weight $l/2$ defined by $j_\theta^l(\gamma, z) = \theta(\gamma z)^l/\theta(z)^l$. The factor j_θ^l has the following signs:

$$\omega_i(j_\theta^l) = 1, \quad \omega_\infty(j_\theta^l) = 1, \quad \omega_{-1}(j_\theta^l) = e^{2\pi i l/8}.$$

(The first two equations are obvious, the third can be seen from (2.7.3); of course it is obvious when l is an integer.) For convenience, take $l = 2k$ where k is positive real. Then j_θ^{2k} is a factor of weight k on Γ_θ, and the corresponding parameters are respectively 0, 0, and $\{k/4\}$. It follows that

$$\dim \mathcal{G}(\Gamma_\theta, j_\theta^{2k}) = 1 + \tfrac{k}{4} - \{\tfrac{k}{4}\} = 1 + [k/4].$$

The cusp ∞ is always regular, the cusp -1 is regular if and only if $k \in 4\mathbf{Z}$. Hence,

$$\dim \mathcal{S}(\Gamma_\theta, j_\theta^{2k}) = \begin{cases} [k/4] - 1 & \text{if } k \in 4\mathbf{Z}, \\ [k/4] & \text{otherwise.} \end{cases}$$

Note that, if $k \in 4\mathbf{Z}$, then $j_\theta^{2k} = J^k$.

(7.9) Exercise. Prove the results of Example (7.1), using only the Special Case (6.6) of Theorem B. [Hint: Estimate the dimension $\dim \mathcal{G}_k$ for $k = 2, \dots, 12$. Use Equation (7.3.1) to show that the dimension is 0 for $k = 2$ and use the Eisenstein series to conclude that the dimension is 1 for $k = 4, 6, 8, 10$. Use the Eisenstein series and $\Delta(z)$, say defined by (7.1.1), to conclude that the dimension is 2 for $k = 12$. Investigate zeros and poles of $\Delta(z)$, and conclude that there is, for $k \geq 0$, an exact sequence,

$$0 \longrightarrow \mathcal{G}_k \xrightarrow{\Delta} \mathcal{G}_{k+12} \longrightarrow \mathbf{C} \longrightarrow 0.$$

Now deduce the results of Example (7.1).]

8. The Proofs.

(8.1). Keep the setup of (6.1). In the calculations below we will need the formula,

$$\sum_{u\in U}\operatorname{ord}_u f = \frac{1}{2\pi i}\int_{\partial U}\frac{f'(z)\,dz}{f(z)}\,. \tag{8.1.1}$$

In the formula, U is an open subset of $\mathbf{C}$ with compact closure, f is a non-zero meromorphic function defined in an open domain containing the closure of U. The integral is the path integral over the oriented boundary of U, assumed to be sufficiently regular. It is assumed that the function f has no zeros or poles on the boundary ∂U. The formula is a well known consequence of Cauchy's residue formula.

The path integral on the right hand side of (8.1.1) can be formed over any path C that lies in the domain of f and avoids the zeros and poles of f. We define

$$I_C(f) := \frac{1}{2\pi i}\int_C \frac{f'(z)\,dz}{f(z)}\,. \tag{8.1.2}$$

The notation extends linearly to a *chain* C, defined as a formal integral linear combination of oriented paths. Note the following two equations,

$$I_C(f_1 f_2) = I_C(f_1) + I_C(f_2), \qquad I_{\phi C}(f) = I_C(f\circ\phi). \tag{8.1.3}$$

The first equation follows because the logarithmic derivative of the product function is the sum of the two logarithmic derivatives. In the second equation, the map ϕ is a holomorphic map from some domain into the domain of f. The equation follows from the definition of the path integral, noting that $d(\phi(\alpha(t)) = \phi'(\alpha(t))d\alpha(t)$ when ϕ is holomorphic and α is a $\mathcal{C}^\infty$-map of the real variable t.

(8.2) *Proof of Theorem B.* We have to prove, for a non-zero j-automorphic form f, the following formula:

$$\sum_{u \bmod \Gamma}\operatorname{ord}_u^\Gamma f = \frac{k\mu}{2}\,. \tag{8.2.1}$$

The main observation used in the proof is the following: Let C be a path in $\mathfrak{D}$ avoiding the zeros and poles of f and let γ be a matrix of Γ. Then the following equation holds:

$$I_{\gamma C}(f) = I_C(f) + kI_C(J_\gamma). \tag{8.2.2}$$

Indeed, $f\circ\gamma = j_\gamma f$ since f is j-invariant. Hence, by (8.1.3), the left side is equal to $I_C(f)+I_C(j_\gamma)$. Moreover, $j(\gamma,z) = \varepsilon J(\gamma,z)^k$ for some constant ε and some determination $J(\gamma,z)^k$. Hence the logarithmic derivative of $j(\gamma,z)$ is equal to k times the logarithmic derivative of $J(\gamma,z)$. Thus $I_C(j_\gamma) = kI_C(J_\gamma)$.

Let F be a finite normal fundamental domain for Γ, and denote by U the interior of F. The domain F has a finite number of sides and vertices. Consider an infinite vertex u of F. By (Discr.3.14), u is a cusp for Γ. Therefore, since f is exponentially bounded at the cusps, it follows that f is holomorphic and non-zero in some fundamental neighborhood of u. When a small fundamental neighborhood of each cusp of F is cut away from F, there remains is compact set. As a consequence, the function f has only a finite number of zeros or poles in F (and in particular on the boundary of F).

A side of F can be divided into two by adding to the vertices a point of a side and the image of the point under the boundary transformation corresponding to the side. Hence we may assume that all zeros or poles of f on the boundary of F are finite vertices of F. In addition, we may assume that no side is mapped onto itself under the boundary transformation.

Let F_0 be the subdomain obtained by cutting away from F a small fundamental neighborhood of each vertex. Choose the neighborhoods such that if V is the chosen neighborhood of a vertex u and if γu belongs to F, then γV is the chosen fundamental neighborhood of γu. There is only a finite number of zeros and poles of f in U. So, when the fundamental neighborhoods are chosen sufficiently small, then all zeros and poles in U belong to the interior U_0 of F_0. The points of U are Γ-ordinary, and hence the Γ-order of f at a point u of U_0 is the ordinary order. Therefore, by (8.1.1),

$$\sum_{u\in U} \operatorname{ord}_u^\Gamma f = I_{\partial U_0}(f). \tag{8.2.3}$$

We will prove the formula (8.2.1) by evaluating carefully the integrals on the right side over the various components of ∂U_0.

The boundary ∂U_0 has three types of components: there are line segments left from the sides of F when the neighborhoods of the two end points are cut away, there are arcs consisting of the parts in F of the geodesic circles bounding the fundamental neighborhoods of the finite vertices, and there are segments consisting of the parts in F of the horo circles bounding the fundamental neighborhoods of the infinite vertices of F. The contributions to the path integral in (8.2.3) coming from the components are grouped as follows: Denote by $I_{\text{sides}}(f)$ the sum of the contributions coming from the sides. Choose in each Γ-equivalence class of vertices of F one vertex u, and denote by $I_u(f)$ the sum of the corresponding contributions from the class; it is the sum of the path integrals along the parts in F of the boundaries of all fundamental neighborhoods of the vertices Γ-equivalent to u. Accordingly,

$$I_{\partial U_0}(f) = I_{\text{sides}}(f) + \sum_u I_u(f), \tag{8.2.4}$$

where the sum is over the chosen system of representatives of the Γ-equivalence classes of vertices of F. Now, by (8.2.3) and (8.2.4), to prove Theorem B, it suffices to prove that, as the neighborhoods shrink around the vertices,

$$I_{\text{sides}}(f) \to \frac{k\mu}{2}, \qquad I_u(f) \to -\operatorname{ord}_u^\Gamma f\,. \tag{8.2.5}$$

Consider first the contribution coming from the sides of F. For each side L of F there is a component $C = C_L$ of ∂U_0 lying on L and a corresponding component C' lying on the side $\gamma_L L$, where γ_L is the boundary transformation corresponding to L. The orientation of C' is the reverse of the orientation of $\gamma_L C$. Hence, by (8.2.2), the two sides L and $L' = \gamma_L L$ contribute to $I_{\text{sides}}(f)$ with the sum,

$$I_C(f) + I_{C'}(f) = I_C(f) - I_{\gamma_L C}(f) = -kI_C(J_{\gamma_L}).$$

Therefore,

$$I_{\text{sides}}(f) = -k \sum\nolimits' I_{C_L}(J_{\gamma_L}),$$

where the sum is over unordered pairs $\{L, L'\}$ of sides of F with $L' = \gamma_L L$. Clearly, the sum on the right hand side converges, when the fundamental neighborhoods shrink around the vertices, to the following sum

$$\sum\nolimits' I_L(J_{\gamma_L}) = \frac{1}{2\pi i} \sum\nolimits' \int_L \frac{J(\gamma_L, z)'}{J(\gamma_L, z)} dz\,.$$

The latter sum is, by (Discr.5.9.4), equal to $-\mu/2$. Therefore, $I_{\text{sides}}(f)$ converges to $k\mu/2$, and the first relation in (8.2.5) has been proved.

Consider next the contribution $I_u(f)$ coming from a the vertices Γ-equivalent to u. Choose (finitely many) matrices γ_i in Γ so that the vertices Γ-equivalent to u are the points $u_i = \gamma_i u$. The fundamental neighborhoods of the points u_i are of the images $\gamma_i V$ of a fundamental neighborhood V of u. Hence the components around the u_i are of the form $\gamma_i D_i$ where the D_i are segments of the boundary of V. The contribution $I_u(f)$ is then the sum, $\sum_i I_{\gamma_i D_i}(f)$. Let D be the union of the D_i. Then, by (8.2.2),

$$I_u(f) = I_D(f) + k \sum I_{D_i}(J_{\gamma_i}). \tag{8.2.6}$$

The terms in the sum on the right side converge to zero when the neighborhoods V shrinks around u. Indeed, the assertion is obvious if u is a finite vertex, because then the integrand is bounded and the length of the integration path goes to 0. If u is an infinite vertex, we may assume that the disk is the upper half plane and that $u = \infty$; in this case the assertion is easily verified.

Therefore, to prove the second relation in (8.2.5), it suffices to prove the following equation,

$$I_D(f) = -\operatorname{ord}_u^{\Gamma}(f). \tag{8.2.7}$$

To prove (8.2.7), assume first that u is an infinite vertex. We may, after conjugation, assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then the canonical generator γ_u is a translation $z \mapsto z + h$. The fundamental neighborhood V is a half plane: $\operatorname{Im} z > R$, and the components D_i are horizontal straight line segments on the boundary: $\Im z = R$; they are oriented from the right to the left. As observed in (Discr.3.16), the union D is a system of representatives for the action of the canonical generator γ_u on the line $\Im z = R$. Set $q(z) := e^{2\pi i z/h}$. It follows from the local analysis in (5.1) that there is an equation,

$$f(z) = e^{2\pi i \kappa_u z/h} g(q), \tag{8.2.9}$$

where $\kappa_u = \kappa_u(j)$ is the parameter and g is meromorphic in the unit disk: $|q| < 1$, say of order N at the origin. By the choice of fundamental neighborhood V, the function f has no zeros or poles in the closed half plane $\Im z \geq R$. The half plane is mapped by q onto a pointed disk: $0 < |q| \leq \varepsilon$. Hence g has no zeros or poles in the closed disk: $|q| \leq \varepsilon$ except possibly at the origin. The image qD is the full boundary: $|q| = \varepsilon$ of the disc, clockwise oriented. Therefore, by (8.1.1) and (8.1.3), $I_D(g \circ q) = -N$. The logarithmic derivative of the factor $e^{2\pi i \kappa z/h}$ is equal to $2\pi i \kappa/h$. As D is the union of horizontal line segments, oriented from right to left, of lengths adding up to h, it follows the $I_D(e^{2\pi i \kappa_u z/h}) = -\kappa_u$. Therefore, by (8.2.9),

$$I_D(f) = -\kappa_u - N = -\operatorname{ord}_u^\Gamma f,$$

and thus (8.2.7) holds.

Assume next that u is a finite vertex. We may, after conjugation, assume that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. Then the canonical generator γ_u is a rotation $z \mapsto e^{2\pi i/e_u} z$. The fundamental neighborhood V is a disk: $|z| < \varepsilon$, and the components D_i are arcs on the boundary: $|z| = \varepsilon$; they are clockwise oriented. As observed in (Discr.3.16), the union D is a system of representatives for the action of the canonical generator γ_u on the circle. Set $w(z) := z^{e_u}$. It follows from the local analysis in (5.3) that there is an equation

$$f(z) = z^{\kappa_u e_u} g(w), \tag{8.2.10}$$

where $\kappa_u = \kappa_u(j)$ is the parameter and g is meromorphic in the unit disk: $|w| < 1$, say of order N at the origin. By the choice of fundamental neighborhood V, the function f has no zeros or poles in the pointed disk: $0 < |z| \leq \varepsilon$. The pointed disk is mapped by w onto a pointed disk with radius ε^{e_u}. Hence g has no zeros or poles in the image disk: $|w| \leq \varepsilon^{e_u}$ except possibly at the origin. The image wD is the full boundary of the image circle $|w| = \varepsilon^{e_u}$, clockwise oriented. Therefore, by (8.1.1) and (8.1.3), $I_D(g \circ w) = -N$. The logarithmic derivative of the factor $z^{\kappa_u e_u}$ is equal to $\kappa_u e_u z^{-1}$. As D is the union of arcs, clockwise oriented, of angles adding up to $2\pi/e_u$, it follows that $I_D(z^{\kappa_u e_u}) = -\kappa_u$. Therefore, by (8.2.10),

$$I_D(f) = -\kappa_u - N = -\operatorname{ord}_u^\Gamma f,$$

and thus (8.2.7) holds.

Hence (8.2.7) holds in both cases, and the proof of Theorem B is complete. □

(8.3). The proofs of Theorems A, C, D, E assume familiarity with the theory of Riemann surfaces. Let X be a compact (connected) Riemann surface. It is well known that the only global holomorphic functions on X are the constants. Denote by $\mathcal{M} = \mathcal{M}(X)$ the field of meromorphic functions on X. If $\varphi \neq 0$ is a meromorphic function on X, then φ has finite order $\operatorname{ord}_u \varphi$ at every point u of X. The order is zero except for a finite number of points u, and the sum of the orders is equal to 0:

$$\sum_{u \in X} \operatorname{ord}_u \varphi = 0. \tag{8.3.1}$$

A *divisor* D on X is a finite formal sum, $D = \sum n_u.u$, of points u of X. The coefficients $\operatorname{ord}_u(D) := n_u$ are integers, and equal to 0 except for a finite number of points u. The *degree* of the divisor is the sum, $\deg D = \sum \operatorname{ord}_u(D)$, of the coefficients. To every non-zero meromorphic function φ there is associated a *principal divisor*, $\operatorname{div}\varphi$, defined by $\operatorname{ord}_u(\operatorname{div}\varphi) = \operatorname{ord}_u(\varphi)$, that is, by

$$\operatorname{div}\varphi = \sum (\operatorname{ord}_u \varphi).u.$$

It follows from (8.3.1) that the degree of a principal divisor is equal to 0,

$$\deg(\operatorname{div}\varphi) = 0. \tag{8.3.2}$$

Two divisors D and D' are called *linearly equivalent*, if the difference $D - D'$ is a principal divisor. Thus linearly equivalent divisors have the same degree.

Associate with a given divisor D the following vector space over $\mathbf{C}$ of rational functions on X:

$$H^0(D) := \{\varphi \mid \operatorname{div}\varphi + D \geq 0\},$$

where the inequality $\operatorname{div}\varphi + D \geq 0$ for divisors means the corresponding inequalities $\operatorname{ord}_u \varphi + \operatorname{ord}_u D \geq 0$ for all the coefficients. Note that the inequality $\operatorname{ord}_u \varphi + n \geq 0$ for $n < 0$ requires φ to have a zero at u of order at least $-n$, and for $n \geq 0$ allows φ to have a pole of order at most n at u. In particular, a function φ in $H^0(D)$ is holomorphic except possibly at the finitely many points u where $\operatorname{ord}_u(D) > 0$. If D and D' are linearly equivalent, say $D - D' = \operatorname{div}\psi$, then, clearly, the multiplication, $\varphi \mapsto \psi\varphi$, defines a $\mathbf{C}$-linear isomorphism $H^0(D) \xrightarrow{\sim} H^0(D')$.

By definition, if φ is a non-zero function in $H^0(D)$, then $\operatorname{ord}_u \varphi + \operatorname{ord}_u(D) \geq 0$. By taking the sum over u, it follows that $\deg(\operatorname{div}\varphi) + \deg D \geq 0$. Hence, by (8.3.2), $\deg D \geq 0$. Therefore,

$$H^0(D) = 0 \quad \text{if } \deg D < 0. \tag{8.3.3}$$

Let ω be a non-zero meromorphic differential form. Locally, around a point u of X, we have that $\omega = f\,dz$, where z is a local parameter at u. By definition, the order $\operatorname{ord}_u \omega$ is the order of f at u. Associate with ω the divisor,

$$\operatorname{div}\omega = \sum (\operatorname{ord}_u \omega).u.$$

Any meromorphic differential form ω' is of the form $\omega' = \varphi\omega$ with a meromorphic function φ. It follows that $\operatorname{div}\omega' = \operatorname{div}\varphi + \operatorname{div}\omega$. Hence the divisors of differential forms form one single class of divisors modulo principal divisors, called the canonical class. Any divisor in the canonical class, that is, any divisor of a differential form,

$$K = \operatorname{div}\omega,$$

is called a *canonical divisor*. Obviously, $\operatorname{div}(\varphi\omega) = (\operatorname{div}\varphi) + (\operatorname{div}\omega)$. It follows that the map $\varphi \mapsto \varphi\omega$ induces an isomorphism from $H^0(K)$ to the space of holomorphic differential forms.

The Riemann–Roch Theorem. *The vector spaces* $H^0(D)$ *are of finite dimension, and they vanish if* $\deg D < 0$*. If* $\deg D \geq 0$*, then the dimension of* $H^0(D)$ *is a most equal to* $\deg D + 1$*. Moreover, there exists a number* $g = g(X)$ *such that for any canonical divisor* K *and any divisor* D*,*

$$\dim H^0(D) = \deg D + 1 - g + \dim H^0(K - D). \tag{8.3.4}$$

The number $g = g(X)$ is called the *genus* of the Riemann surface X. It can be shown to be equal to the topological genus of X as a surface.

Corollary. *The genus* $g = g(X)$ *is equal to the dimension of the space of holomorphic differential forms, and the degree of a canonical divisor is equal to* $2g - 2$*,*

$$g = \dim H^0(K), \quad \deg K = 2g - 2. \tag{8.3.5}$$

Moreover, if $\deg D \geq 2g - 1$*, then* $\dim H^0(D) = \deg D + 1 - g$*.*

Proof. The first equality follows from (8.3.4) by taking $D := 0$. Next, the equality $\deg K = 2g-2$ follows by taking $D := K$ in (8.3.4). Finally, if $\deg D \geq 2g-1$, then $\deg(K-D) < 0$. Hence the last assertion of the Corollary follows from (8.3.3). ▯

It follows from the Theorem and the Corollary that if $\deg D \geq 0$, then

$$\deg D + 1 - g \leq \dim H^0(D) \leq \deg D + 1,$$

and the first inequality is an equality if $\deg D \geq 2g - 1$. The latter result is referred to as Riemann's part of the Theorem.

(8.4). To apply the theory of Riemann surfaces to automorphic forms, note that the quotient $X = \overline{\mathfrak{D}/\Gamma}$ is, by the construction given in the proof of Corollary (Discr.2.13), a connected Riemann surface. To get a local parameter around a point of X, choose a representative u in $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$. Assume first that representative u is in $\mathfrak{D}$. After a suitable conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. Then the map,

$$w = z^{e_u},$$

where the canonical generator at 0 is $z \mapsto e^{2\pi i e_u} z$, identifies, for a small fundamental neighborhood U of u, the open subset U/Γ_u of X with a small neighborhood of 0 in $\mathbf{C}$; thus w is a local parameter around the given point of X. Assume next the u is in $\partial_\Gamma \mathfrak{D}$. After a suitable conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then the map,

$$q = e^{2\pi i z/h},$$

where the canonical generator at ∞ is $z \mapsto z + h$, identifies, for a small fundamental neighborhood U of u, the open subset U/Γ_u of X with a small neighborhood of 0 in $\mathbf{C}$; thus q is a local parameter around the given point of X.

Since Γ is assumed to be a Fuchsian group of the first kind, the Riemann surface X is even compact.

By construction, the field $\mathcal{M}(\Gamma)$ of Γ-automorphic functions is the field of meromorphic functions on X. Moreover, the Γ-order at u of a non-zero automorphic function φ is equal to the order of φ as a meromorphic function on X.

(8.5) *Proof of Theorem A.* As is well known, a compact Riemann surface is algebraic and, as a consequence, its field of meromorphic functions is finitely generated and of transcendence degree 1 over $\mathbf{C}$. Therefore, since $X = X(\Gamma)$ is compact, Theorem A holds. □

(8.6). Let $X = X(\Gamma)$ be the Riemann surface of (8.4). Denote by C_{cusp} the *cuspidal divisor* of X, defined as the sum,

$$C_{\text{cusp}} := \sum_{u \text{ a cusp}} 1.u\,.$$

Associate with a non-zero j-automorphic form f the following divisor on X:

$$D_f := \sum_u \big[\operatorname{ord}_u^\Gamma f\big].u,$$

where $[x]$ is the integral part. Recall that the Γ-order is congruent to the parameter $\kappa_u(j)$; whence $\big[\operatorname{ord}_u^\Gamma f\big] = \operatorname{ord}_u f - \kappa_u(j)$. Therefore, by Theorem B, we obtain the equation,

$$\deg D_f = \frac{k\mu}{2} - \sum_u \kappa_u(j). \tag{8.6.1}$$

As a consequence,

$$\deg D_f + 1 - g = \delta(j). \tag{8.6.2}$$

Fix the non-zero j-automorphic form f. By (4.9)(1), the multiplication $\varphi \mapsto \varphi f$ is an isomorphism $\mathcal{M} \to \mathcal{M}(\Gamma, j)$. Under the multiplication, the product φf is an integral form, if and only if, for all u, $\operatorname{ord}_u^\Gamma(\varphi f) \geq 0$, that is, if and only if the following inequality holds:

$$\operatorname{ord}_u \varphi + \operatorname{ord}_u^\Gamma f \geq 0. \tag{8.6.3}$$

The order $\operatorname{ord}_u \varphi$ is an integer. Hence the inequality (8.6.3) is equivalent to the following inequality:

$$\operatorname{ord}_u \varphi + \big[\operatorname{ord}_u^\Gamma f\big] \geq 0. \tag{8.6.4}$$

It follows that φf is in $\mathcal{G}(\Gamma, j)$ if and only if φ is in $H^0(D_f)$. In other words, multiplication by the fixed form f induces an isomorphism,

$$H^0(D_f) \xrightarrow{\sim} \mathcal{G}(\Gamma, j). \tag{8.6.5}$$

In particular, $\dim H^0(D_f) = \dim \mathcal{G}(\Gamma, j)$.

Clearly, if φf is an integral form, then φf is a cusp form, if and only if the inequality (8.6.3) is strict for all cusps u. If a cusp u is j-irregular, then the parameter $\kappa_u(j)$ is non-zero, and hence the order $\operatorname{ord}_u^\Gamma f$ is not an integer; hence, the inequality in (8.6.3) is strict, if and only if the inequality (8.6.4) holds. If a cusp u is j-regular, then the inequality in (8.6.3) is strict if and only if the inequality (8.6.4) is strict, that is, if and only if,

$$\operatorname{ord}_u \varphi + \big[\operatorname{ord}_u^\Gamma f\big] - 1 \geq 0. \tag{8.6.6}$$

Hence φf is a cusp form if and only if (8.6.6) holds at all j-regular cusps u and (8.6.4) holds at all other orbits. Let $C_{j\text{-reg}}$ denote the j-regular part of the cuspidal divisor C_{cusp}, that is, $C_{j\text{-reg}} = \sum 1.u$ where the sum is over the j-regular cusps u. It follows that φf is cusp form if and only if φ belongs to $H^0(D_f - C_{j\text{-reg}})$. In other words, multiplication by the fixed form f defines an isomorphism,

$$H^0(D_f - C_{j\text{-reg}}) \xrightarrow{\sim} \mathcal{S}(\Gamma, j). \tag{8.6.7}$$

In particular, $\dim H^0(D_f - C_{j\text{-reg}}) = \dim \mathcal{S}(\Gamma, j)$.

Obviously, the degree of $C_{j\text{-reg}}$ is equal to the number of j-regular cusps. Hence, by (8.6.2),

$$\deg(D_f - C_{j\text{-reg}}) + 1 - g = \delta'(j). \tag{8.6.8}$$

(8.7) *Proofs of Theorems C and D.* The assertions follow from Riemann's part of the Riemann–Roch theorem, given the expressions (8.6.2) and (8.6.8) for $\delta(j)$ and $\delta'(j)$, and the isomorphisms (8.6.5) and (8.6.7). □

(8.8). To use the full strength of the Riemann-Roch Theorem, we have to identify a canonical divisor K on X. Let f be a non-zero Γ-automorphic form of weight 2, that is, $f \in \mathcal{M}(\Gamma, J^2)$. Clearly, the differential form $f\,dz$ on $\mathfrak{D}$ is Γ-invariant. Hence it descends to a meromorphic differential form on $\mathfrak{D}/\Gamma$, and in fact, as the following calculation shows, to a meromorphic differential form on X.

Consider first a point of $\mathfrak{D}/\Gamma$, represented by a point u of $\mathfrak{D}$. After conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. Then $w = z^{e_u}$ is a local parameter at u. There is a normalized Laurent expansion around u,

$$f(z) = z^{\kappa_u e_u} g(w),$$

and the parameter $\kappa_u = \kappa_u(J^2)$ is equal to $1 - 1/e_u$. Thus $f(z) = (z/w)g(w)$; as $dw = e_u z^{e_u - 1} dz = e_u (w/z) dz$, it follows that $e_u f(z) dz = g(w) dw$. Hence the order at u of $f(z)dz$ as a differential form on X is equal to the order of $g(w)$ at 0. On the other hand, the Γ-order $\mathrm{ord}_u^\Gamma f$ is equal to κ_u plus the order of $g(w)$. Therefore,

$$\mathrm{ord}_u^\Gamma f = \mathrm{ord}_u(f\,dz) + (1 - \tfrac{1}{e_u}). \tag{8.8.1}$$

Consider next a cusp represented by a point u of $\partial_\Gamma \mathfrak{D}$. After conjugation, we may assume that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$. Then $q = e^{2\pi i z/h}$ is a local parameter at u. There is a normalized Fourier expansion around u, and it has the form,

$$f(z) = g(q),$$

since the parameter $\kappa_u = \kappa_u(J^2)$ is equal to 0. As $dq = (2\pi i/h) q\,dz$, it follows that $(2\pi i/h) f(z) dz = q^{-1} g(q) dq$. Hence the order at u of $f(z)dz$ as a differential form on X is equal to the order of $g(q)$ minus 1. On the other hand, the Γ-order of f is equal to the order of $g(q)$. Therefore, at a cusp u,

$$\mathrm{ord}_u^\Gamma f = \mathrm{ord}_u(f\,dz) + 1. \tag{8.8.2}$$

If $1/e_u$ is interpreted as 0 at a cusp u, then (8.8.2) is simply (8.8.1). The divisor $K_f := \operatorname{div}(f\,dz)$ is a canonical divisor on X. From the two equations (8.8.1) and (8.8.2), it follows that

$$D_f = K_f + C_{\text{cusp}}. \tag{8.8.3}$$

Take the degree in (8.8.3) and apply (8.6.1) with $j := J^2$ to obtain the equation,

$$\mu = 2g - 2 + \sum\Bigl(1 - \frac{1}{e_u}\Bigr). \tag{8.8.4}$$

The equation is identical to (6.1.1). However, (6.1.1) was obtained with the topological genus and (8.8.4) was obtained with the holomorphic genus of X. Hence the two genera are equal, that is, the topological genus is equal to the dimension of the vector space of holomorphic differential forms.

In addition, since all cusps are J^2-regular, the isomorphism $H^0(K) \xrightarrow{\sim} \mathcal{S}_2(\Gamma)$ follows from (8.8.3) and (8.6.7). Hence the space $\mathcal{S}_2(\Gamma)$ of cusp forms of weight 2 is isomorphic to the space of holomorphic differential forms on X; in particular, $\dim \mathcal{S}_2(\Gamma) = g$.

(8.9) Theorem H. *Let $\check{j}$ be the factor defined by $\check{j} = J^2/j$. Then a cusp u is j-regular if and only if it is $\check{j}$-regular, and $\delta(j) + \delta(\check{j})$ is equal to the number of j-regular cusps. Moreover, the following equation holds,*

$$\dim \mathcal{G}(\Gamma, j) = \delta(j) + \dim \mathcal{S}(\Gamma, \check{j}). \tag{8.9.1}$$

Proof. The factor $\check{j}$ is of weight $2 - k$ and $j\check{j} = J^2$. Fix two non-zero automorphic forms, $f \in \mathcal{M}(\Gamma, j)$ and $\check{f} \in \mathcal{M}(\Gamma, \check{j})$. Then the product $f\check{f}$ belongs to $\mathcal{M}(\Gamma, J^2)$. Hence the divisor $K := \operatorname{div}(f\check{f}\,dz)$ is a canonical divisor. Now, for any point u,

$$\operatorname{ord}_u^\Gamma(f\check{f}) = \operatorname{ord}_u^\Gamma f + \operatorname{ord}_u^\Gamma \check{f}. \tag{8.9.2}$$

The fractional parts of the three orders in (8.9.2) are the parameters, $\kappa_u(J^2)$, $\kappa_u(j)$, and $\kappa_u(\check{j})$. Hence either $\kappa_u(J^2) = \kappa_u(j) + \kappa_u(\check{j})$ or $\kappa_u(J^2) + 1 = \kappa_u(j) + \kappa_u(\check{j})$. Assume that u is in $\mathfrak{D}$. Then $\kappa_u(J^2) = 1 - 1/e_u$. Moreover, any parameter at u is, by (3.3.1), at most equal to $1 - 1/e_u$. Therefore, $\kappa_u(J^2) = \kappa_u(j) + \kappa_u(\check{j})$. Assume next that u is a cusp. Then $\kappa_u(J^2) = 0$. If u is j-regular, then $\kappa_u(j) = 0$; it follows that also $\kappa_u(\check{j}) = 0$ and that $\kappa_u(J^2) = \kappa_u(j) + \kappa_u(\check{j})$. If u is j-irregular, then it follows that u is also $\check{j}$-irregular and that $\kappa_u(j) + \kappa_u(\check{j}) = 1$. Hence $\kappa_u(J^2) + 1 = \kappa_u(j) + \kappa_u(\check{j})$.

Now take integral parts of the orders in (8.9.2), and consider the corresponding divisors. On the left we obtain, by (8.8.3), the divisor $K + C_{\text{cusp}}$. On the right we obtain, by the discussion above, the divisor $D_f + D_{\check{f}} + C_{j\text{-irreg}}$. As a consequence, we obtain the equation,

$$K + C_{j\text{-reg}} = D_f + D_{\check{f}}. \tag{8.9.3}$$

It was seen in the discussion above that a cusp u is j-regular if and only if it is $\check{j}$-regular. Take degrees in (8.9.3) and use (8.3.5) and (8.6.2) to see that $\delta(j) + \delta(\check{j})$ is equal to the number of j-regular cusps. Finally, apply (8.6.5) to (f, j) and (8.6.7) to $(\check{f}, \check{j})$. By the Riemann-Roch Theorem and (8.9.2), the final assertion of the Mail Theorem is a consequence. ∎

(8.10) Note. The factor $\check{j}$ introduced in the Theorem may be called the *dual factor*. Since $\delta(j)+\delta(\check{j})$ is equal to the number of j-regular cusps, it follows that

$$\delta(\check{j}) = -\delta'(j). \tag{8.10.1}$$

The Main Theorem implies the Theorems C, D, and E. To obtain Theorem E, note first that J^2 is the dual to the trivial factor $j = 1$. The constant function $f = 1$ is Γ-automorphic of weight 0, and the corresponding divisor is $D_1 = 0$. Hence, $\delta(1) = 1 - g$ and since $\dim H^0(0) = 1$, the equation $\dim \mathcal{S}_2(\Gamma) = g$ follows from (8.9.1).

Note next that the factor J is self-dual. Hence $\delta(J)$ is equal to half the number of regular cusps of Γ, cf. (6.14.3). Therefore, the last part of Theorem E is a consequence of (8.9.1). Note that, more generally, any factor of weight 1 is of the form $\chi(\gamma)J(\gamma,z)$, and the dual is the factor $\overline{\chi(\gamma)}J(\gamma,z)$. In particular, if the character χ is quadratic, then the factor $\chi(\gamma)J(\gamma,z)$ is self dual.

Poincaré Series and Eisenstein Series

1. Poincaré series.

(1.1). Fix a finite disk $\mathfrak{D}$, a discrete subgroup Γ of $\mathrm{SL}(\mathfrak{D})$, and a homogeneous factor j on Γ of real weight k. Assume that Γ is a Fuchsian group of the first kind.

(1.2) Definition. Recall that the factor j defines an right action of Γ on meromorphic functions ϕ on $\mathfrak{D}$, determined by $(\phi \cdot_j \gamma)(z) = j(\gamma, z)^{-1}\phi(z)$. Clearly, the series,

$$\sum_{\gamma\in\Gamma} \phi \cdot_j \gamma,$$

is formally Γ-invariant. In general, of course, the series can not be expected to be convergent.

Assume more generally that Λ is a given subgroup of Γ and that ϕ is (Λ, j)-invariant, that is, $\phi(\gamma z) = j(\gamma, z)\phi(z)$ for all $\gamma \in \Lambda$. Consider the series,

$$G(\Gamma, j, \Lambda, \phi) = \sum_{\gamma\in\Lambda\backslash\Gamma} \phi \cdot_j \gamma, \tag{1.2.1}$$

where the sum is over a system of representatives for the right cosets of Λ in Γ. Since ϕ is (Λ, j)-invariant, the term in the sum corresponding to γ depends only on the coset containing γ. Again, the series is formally Γ-invariant. The series (1.2.1) is the general *Poincaré series*.

The main questions associated with Poincaré series are the following:

(1) Is the series normally convergent in $\mathfrak{D}$? If it is, then the sum $G(z)$ is meromorphic in $\mathfrak{D}$, and (Γ, j)-invariant. Moreover, if the given invariant function ϕ is holomorphic in $\mathfrak{D}$, then $G(z)$ is holomorphic in $\mathfrak{D}$.

(2) Is $G(z)$ exponentially bounded at the cusps of Γ? If it is, then $G(z)$ is a (Γ, j)-automorphic form. If the given function ϕ is holomorphic, is $G(z)$ then in integral form, or even a cusp form?

(3) Finally, is $G(z)$ not the zero function?

(1.3). Consider an isomorphism $\alpha\colon \mathfrak{D}' \to \mathfrak{D}$ from a finite disk $\mathfrak{D}'$ onto $\mathfrak{D}$, defined by a matrix α in $\mathrm{SL}_2(\mathbf{C})$. Recall that α defines a conjugate factor j^α on the conjugate group Γ^α. Moreover, for every meromorphic function ϕ on $\mathfrak{D}$, there is a weight-k conjugate function ϕ^α on $\mathfrak{D}'$ defined by $\phi^\alpha(z') = \varepsilon J(\alpha, z')^k\phi(\alpha z')$ with a fixed complex sign ε and a fixed determination of $J(\alpha, z)^k$. Clearly, if ϕ is (Λ, j)-invariant, then the conjugate function ϕ^α is $(\Lambda^\alpha, j^\alpha)$-invariant, and

$$G(\Gamma, j, \Lambda, \phi)^\alpha = G(\Gamma^\alpha, j^\alpha, \Lambda^\alpha, \phi^\alpha). \tag{1.3.1}$$

(1.4) Definition. Let u be a point of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$ and take the isotropy group Γ_u as Λ. Natural (Γ_u, j)-invariant functions are obtained as follows.

Assume first that $(\mathfrak{D}, u) = (\mathfrak{E}, 0)$. The canonical generator γ_u at $u = 0$ is a diagonal matrix, and the associated Möbius transformation is a rotation $z \mapsto e^{2\pi i/e_u} z$. The function $j(\gamma_u, z)$ is constant, and equal to the sign $e^{2\pi i\kappa_u}$, where $\kappa_u = \kappa_u(j)$ is the parameter. Hence γ_u acts on functions on $\mathfrak{E}$ by $(\phi \cdot_j \gamma_u)(z) = e^{-2\pi i\kappa_u}\phi(e^{2\pi i/e_u} z)$. It follows that the (Γ_u, j)-invariant functions are the functions of the form $\phi(z) = z^{\kappa_u e_u}\tilde{\phi}(w)$, where $w = z^{e_u}$. In particular, take $\tilde{\phi}(w) = w^l$ where l is an integer. Then $\phi(z) = z^{(\kappa_u+l)e_u}$. Therefore, if κ is any real number congruent to the parameter $\kappa_u(j)$ modulo $\mathbf{Z}$, then the following function is (Γ_u, j)-invariant:

$$\phi_{\Gamma,0,\kappa}(z) = \phi_\kappa(z) = z^{\kappa e_u}. \tag{1.4.1}$$

Assume next that $(\mathfrak{D}, u) = (\mathfrak{H}, \infty)$, where ∞ is Γ-parabolic. The canonical generator γ_u at $u = \infty$ is an upper triangular matrix with ± 1 in the diagonal, and the associated Möbius transformation is a translation $z \mapsto z + h$. The function $j(\gamma_u, z)$ is constant, and equal to the sign $e^{2\pi i\kappa_u}$, where $\kappa_u = \kappa_u(j)$ is the parameter. Hence γ_u acts on functions on $\mathfrak{H}$ by $(\phi \cdot_j \gamma_u)(z) = e^{-2\pi i\kappa_u}\phi(z+h)$. It follows that the (Γ_u, j)-invariant functions are the functions of the form $\phi(z) = e^{2\pi i\kappa_u z/h}\tilde{\phi}(q)$, where $q = e^{2\pi i z/h}$. In particular, take $\tilde{\phi}(q) = q^l$ where l is an integer. Then $\phi = e^{2\pi i(\kappa_u+l)z/h}$. Therefore, if κ is any real number congruent to the parameter $\kappa_u(j)$ modulo $\mathbf{Z}$, then the following function is (Γ_u, j)-invariant:

$$\phi_{\Gamma,\infty,\kappa}(z) = \phi_\kappa(z) = e^{2\pi i\kappa z/h}. \tag{1.4.2}$$

In general, if $u \in \mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, and κ is any real number congruent to the parameter $\kappa_u(j)$ modulo $\mathbf{Z}$, define a Γ_u-invariant function $\phi_{u,\kappa}$ as follows: in the two cases $u \in \mathfrak{D}$ and $u \in \partial_\Gamma \mathfrak{D}$ respectively, choose a Möbius transformation $\alpha\colon (\mathfrak{D}, u) \to (\mathfrak{E}, 0)$ and $\alpha\colon (\mathfrak{D}, u) \to (\mathfrak{H}, \infty)$. Set

$$\phi_{u,\kappa} := (\phi_\kappa)^\alpha,$$

where the right hand side is the weight-k conjugate of the function defined for the conjugate group ${}^\alpha\Gamma$ in (1.4.1) and (1.4.2) in the two cases respectively.

The Poincaré series $G(\Gamma, j, \Gamma_u, \phi_{u,\kappa})$ is denoted $G(\Gamma, j, u, \kappa)$, that is,

$$G(\Gamma, j, u, \kappa)(z) = \sum_{\gamma \in \Gamma_u \backslash \Gamma} \frac{\phi_{u,\kappa}(\gamma z)}{j(\gamma, z)}. \tag{1.4.3}$$

In terms of the chosen conjugation α,

$$G(\Gamma, j, u, \kappa)(z) = \varepsilon \sum \frac{\phi_\kappa(\alpha\gamma z)}{J(\alpha, \gamma z)^k j(\gamma, z)}. \tag{1.4.4}$$

It should be emphasized that the Poincaré series (1.4.3) is only defined when κ is congruent to the parameter $\kappa_u(j)$ of j at u. If u is a Γ-ordinary point, then the canonical generator γ_u is the identity, and the parameter $\kappa_u(j)$ is equal to 0. The parameter is also zero, when u is a j-regular cusp. Hence, in these two cases, the series is defined for integer values of κ. The series (1.4.3) for a j-regular cusp u and $\kappa = 0$ is called the *Eisenstein series* associated with the cusp u.

(1.5) Note. In spite of the notation on the left hand side of (1.4.3), the series does depend on choices. First, there is a sign involved in the definition of the weight-k conjugate under α (of course, when k is an integer, then there is a unique k'th power $J(\alpha, z)^k$, and we can take $\varepsilon = 1$). But the function $\phi_{u,\kappa} = (\phi_\kappa)^\alpha$, and hence also the Poincaré series, depends on the choice of α.

Compare the function $\phi_u = \phi_{u,\kappa}$ obtained from α with the function ϕ'_u obtained from a second choice α'. The isomorphisms α and α' differ by an automorphism δ of the target, $\alpha' = \delta\alpha$. By definition, $\phi_u = (\phi)^\alpha$, where ϕ is the function (1.4.1) or (1.4.2) respectively obtained from the conjugate group ${}^\alpha\Gamma$. Similarly, $\phi'_u = (\phi')^{\delta\alpha}$ where ϕ' is obtained from the conjugate group ${}^{\alpha'}\Gamma$. Hence it suffices to compare ϕ and $(\phi')^\delta$.

Assume first that u is in $\mathfrak{D}$. Then δ is a rotation in $\mathfrak{E}$ around 0, say $\delta z = \zeta z$ with $|\zeta| = 1$. The two functions ϕ and ϕ' are the same function $z^{\kappa e_u}$. Moreover, up to a complex sign, the conjugate function ϕ'^δ is equal to $\phi(\zeta z) = (\zeta)^{\kappa e_u}\phi(z)$ and hence, up to sign, equal to ϕ. Therefore, the function ϕ'_u is, up to sign, equal to ϕ_u. It follows that the series (1.4.3) for a point $u \in \mathfrak{D}$ is well defined up to a complex sign.

Assume next that u is a Γ-parabolic point. Then δ is a Möbius transformation in $\mathfrak{H}$ of the form $\delta z = rz + b$ with $r > 0$ and a real number b. The two functions ϕ and ϕ' are different: $\phi = e^{2\pi i\kappa z/h}$ where h is defined from ${}^\alpha\Gamma$ and $\phi' = e^{2\pi i\kappa z/h'}$ where h' is defined from ${}^{\alpha'}\Gamma$. Clearly, $h' = rh$. Hence, up to the sign $e^{2\pi i\kappa b/(rh)}$, the function $(\phi')(\delta z)$ is equal to $\phi(z)$. The weight-k conjugate ϕ'^δ is obtained from $\phi'(\delta z)$ by dividing by $J(\delta, z)^k$. The modulus of $J(\delta, z)^k$ is equal to $r^{-k/2}$. Hence ϕ'^δ is equal to ϕ multiplied with a nonzero complex number (of modulus $r^{k/2}$). Therefore, the function ϕ'_u is, up to multiplication by a nonzero number, equal to ϕ_u. It follows that the series (1.4.3) for a Γ-parabolic point u is well defined only up to multiplication by a non-zero number.

From (1.3.1) we obtain, for a general conjugation, the following equation up to multiplication by a nonzero number:

$$G(\Gamma, j, u, \kappa)^\alpha = G(\Gamma^\alpha, j^\alpha, \alpha^{-1}u, \kappa). \tag{1.5.1}$$

(1.6) Example. Take $(\mathfrak{D}, \Gamma, j) = (\mathfrak{H}, \Gamma(1), J^k)$ where k is an even integer. The cusp ∞ is J^k-regular since k is even. Hence the Eisenstein series $G_k = G(\Gamma(1), J^k, \infty, 0)$ is defined. The group Γ is homogeneous and the isotropy group at ∞ is dicyclic, generated by the matrix t. Obviously, the map,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto (c, d),$$

induces a bijection from $\Gamma_\infty\backslash\Gamma$ onto the set of pairs modulo ± 1 of relatively prime integers (c, d). Thus the Eisenstein series corresponding to the cusp ∞ and $\kappa = 0$ is the following,

$$G_k(\infty, 0)(z) = \frac{1}{2}\sum_{(c,d)=1} \frac{1}{(cz+d)^k}\,.$$

In other words, the series is the normalized Eisenstein series $G_k(z)$. It is normally convergent in $\mathfrak{H}$ for $k \geq 4$.

(1.7) Theorem G. *Assume that the weight k of the factor j is greater than 2. Let u be a point of $\mathfrak{D} \cup \partial_\Gamma \mathfrak{D}$, and let κ be a number congruent modulo $\mathbf{Z}$ to the parameter $\kappa_u(j)$. Then, in the setup of* (1.4)*, the Poincaré series $G = G(\Gamma, j, u, \kappa)$ defines a (Γ, j)-automorphic form. The function $G(z)$ is holomorphic at all points of $\mathfrak{D}$ that are not Γ-equivalent to u. In addition, it vanishes at all Γ-parabolic points that are not Γ-equivalent to u. At the given point u, the Γ-order of G is nonnegative if $\kappa \geq 0$; if $\kappa < 0$, then the Γ-order is equal to κ. Finally, if $\kappa = 0$ and u is a Γ-parabolic point (necessarily j-regular), then $G(z)$ does not vanish at u.*

Proof. By (1.5.1), we may, after conjugation, assume that the disk is the upper half plane $\mathfrak{H}$. In general, we think of the Poincaré series (1.4.1) as indexed by a chosen system of representatives γ for the cosets $\Gamma_u \backslash \Gamma$. The term corresponding to γ is the function,

$$g_\gamma(z) = \phi_{u,\kappa}(\gamma z)/j(\gamma, z),$$

and the Poincaré series is the series $G(z) = \sum_\gamma g_\gamma(z)$.

The proof to the Theorem will be given in three parts below. In part I, we prove that the series is normally convergent in $\mathfrak{H}$. It follows that the series defines a meromorphic (Γ, j)-invariant function $G(z)$ in $\mathfrak{H}$. By construction, the function ϕ_u is holomorphic in $\mathfrak{H}$, except possibly when $u \in \mathfrak{H}$ and $\kappa < 0$. In the exceptional case, the term $g_\gamma(z)$ in the series has a pole of order $-\kappa$ at $\gamma^{-1}u$.

Therefore, to finish the proof of the Theorem, it remains to study the behavior of $G(z)$ at the Γ-parabolic points: we have to prove that $G(z)$ is exponentially bounded and that the remaining assertions of the theorem about Γ-parabolic points hold. Clearly, to study the behavior of $G(z)$ at a given Γ-parabolic point we may, after conjugation, assume that the given Γ-parabolic point is the point ∞. Thus, in parts II and III, we assume that the point ∞ is Γ-parabolic.

The possible poles for the terms $g_\gamma(z)$ are the points in the orbit Γu (when $u \in \mathfrak{H}$ and $\kappa < 0$). Therefore, since ∞ is Γ-parabolic, there is a number $R > 0$ so that all the terms are holomorphic in the half plane $\mathfrak{H}_R : \Im z > R$. Hence $G(z)$ is holomorphic in $\mathfrak{H}_R$. In part II, the series $G(z)$ is broken up into partial sums $G_w(z)$ as follows: The group Γ acts on the right on the index set $\Gamma_u \backslash \Gamma$. In particular, the subgroup Γ_∞ acts. Hence, the index set is split into disjoint Γ_∞-orbits. Accordingly, the series $G(z)$ is split into a sum of partial series,

$$G(z) = \sum G_w(z), \tag{1.7.1}$$

where the sum is over all Γ_∞-orbits w of $\Gamma_u \backslash \Gamma$, and $G_w(z)$ is the series,

$$G_w(z) = \sum g_\gamma(z),$$

where the sum is over those representatives γ for which the right coset $\Gamma_u \gamma$ belongs to the orbit w. As a partial sum of the series $G(z)$, each series $G_w(z)$ is normally convergent in $\mathfrak{H}$; in particular, the function $G_w(z)$ is holomorphic in $\mathfrak{H}_R$. Moreover, the series (1.7.1) is normally convergent. As a sum over a Γ_∞-orbit, each function $G_w(z)$ is (Γ_∞, j)-invariant.

Consider the Γ_∞-orbit w containing a given right coset $\Gamma_u\beta$ where β is in Γ. Clearly, under the right action of Γ_∞, the isotropy group of $\Gamma_u\beta$ is the intersection,

$$\Gamma_\infty \cap \beta^{-1}\Gamma_u\beta = \Gamma_\infty \cap \Gamma_{\beta^{-1}u}.$$

A matrix in the intersection has ∞ and $\beta^{-1}u$ as fixed points. Hence, the intersection is the trivial subgroup 1 or ± 1 unless $\infty = \beta^{-1}u$. Of course, the exceptional case is only possible if u is Γ-parabolic and Γ-equivalent to ∞. In the exceptional case, where $\beta\infty = u$, the orbit w consists of the single right coset $\Gamma_u\beta$; the corresponding partial series will be denoted G_u. It consists of a single term,

$$G_u(z) = g_\beta(z) \qquad \text{where } \beta\infty = u. \tag{1.7.2}$$

In all other cases, the partial series corresponding to the orbit w containing $\Gamma_u\beta$ is the following:

$$G_w(z) = \sum_{\gamma \in \mathrm{P}\Gamma_\infty} g_{\beta\gamma}(z). \tag{1.7.3}$$

In part II we prove that each partial series $G_w(z)$, which is not the exceptional series (1.7.2), converges to 0 for $\Im z \to \infty$. The latter result is then used to finish the proof of the Theorem in part III. $\square$

(1.8) *Part I of the proof.* It will be convenient to consider the measure in $\mathfrak{H}$ defined by $d_k^{\mathfrak{H}}(z) = y^{k/2-2}dxdy$ (where $z = x + iy$). The measure $d_0^{\mathfrak{H}}(z) = y^{-2}dxdy$ is $\mathrm{SL}_2(\mathbf{R})$-invariant. It follows that the measure $d_k^{\mathfrak{H}}(z)$ respects weight-k conjugation as follows: if α is a matrix of $\mathrm{SL}_2(\mathbf{R})$ and g is continuous on the subset αM of $\mathfrak{H}$, then

$$\int_M |g^\alpha(z)| d_k^{\mathfrak{H}}(z) = \int_{\alpha M} |g(z)| d_k^{\mathfrak{H}}(z). \tag{1.8.1}$$

To prove that the Poincaré series converges normally, let K be a compact subset of $\mathfrak{H}$. Choose $\delta > 0$ and a compact subset M of $\mathfrak{H}$ such that, for any w in K, the closed disk: $|z-w| \le \delta$ is contained in M. Let C be the constant defined by $C := (\pi\delta^2)^{-1}\max(\Im z)^{2-k/2}$ where the maximum is over $z \in M$. For any function $g(z)$ holomorphic in a neighborhood of M we have, for $w \in K$, the equation,

$$\pi\delta^2 g(w) = \int_{|z-w|\le\delta} g(z)dxdy. \tag{1.8.2}$$

Indeed, in the disk, $g(z)$ is the uniform limit of its power series expansion around w. Hence, using polar coordinates to evaluate the integral $\int (z-w)^n dxdy$, it suffices to note that $\int_0^{2\pi} e^{in\theta}d\theta = 0$ for $n > 0$. Thus the equation (1.8.2) holds. From the equation, we obtain the estimate $\pi\delta^2|g(w)| \le \int_M |g(z)|dxdy$. Hence, by the definition of C,

$$|g(w)| \le C\int_M |g(z)| d_k^{\mathfrak{H}}(z). \tag{1.8.3}$$

The term $g_\gamma(z)$ in the Poincaré series is equal to $\phi(\gamma z)/j(\gamma, z)$ where $\phi = \phi_{u,\kappa}$. Hence $g_\gamma(z)$ is holomorphic in $\mathfrak{H}$ except possibly for a pole in the point $\gamma^{-1}u$ if $u \in \mathfrak{H}$. The compact set M meets only a finite number of points in the orbit Γu. Therefore, except for a finite number of terms, the term $g_\gamma(z)$ is holomorphic in a neighborhood of M. The term g_γ is equal to $\phi \cdot_j \gamma$, and in particular, it is a weight-k conjugate of $\phi(z)$. Therefore, by (1.8.1) and (1.8.3), we have, except for a finite number of terms in the series, the estimate for $w \in K$,

$$|g_\gamma(w)| \leq C \int_{\gamma M} |\phi(z)| d_k^{\mathfrak{H}}(z). \tag{1.8.4}$$

Hence, to prove that the Poincaré series converges normally, it suffices to prove the following assertion: the sum of the integrals in (1.8.4), over the representatives γ for which $u \notin \gamma M$, is finite.

To the latter assertion will be proved for an arbitrary compact subset M of $\mathfrak{H}$. We may assume that M is contained in a fundamental domain for Γ. Indeed, if F is a fundamental domain for Γ, then M meets only a finite number of transforms of F. Hence M decomposes into a finite number of pieces each of which is contained in a fundamental domain and, clearly, if the assertion holds for each piece, then it holds for M.

Since M is contained in a fundamental domain for Γ, the sum of the integrals in (1.8.4) is equal to the integral over the union,

$$M' := \bigcup \gamma M,$$

where the union is over the representatives γ of the cosets $\Gamma_u \backslash \Gamma$ for which $u \notin \gamma M$. Thus it suffices to prove that the following integral is finite:

$$\int_{M'} |\phi(z)| d_k^{\mathfrak{H}}(z). \tag{1.8.5}$$

Assume first that $u \in \mathfrak{H}$. A small neighborhood of u meets only a finite number of transforms of a fundamental domain, and hence only a finite number of transforms γM. It follows that the union M' is contained in the subdomain $\mathfrak{H}'$ obtained from $\mathfrak{H}$ by cutting away a small neighborhood of u. Hence, it suffices to prove that the following integral is finite:

$$\int_{\mathfrak{H}'} |\phi(z)| d_k^{\mathfrak{H}}(z).$$

The function ϕ is, by definition, the weight-k conjugate obtained from $\phi_{0,\kappa} = z^{e_u \kappa}$ by a Möbius transformation $\alpha \colon (\mathfrak{H}, u) \to (\mathfrak{E}, 0)$. Clearly, a suitable Möbius transformation is the map: $\alpha(z) = (z - u)/(z - \bar{u})$. Hence, up to a constant,

$$\phi(z) = (z - \bar{u})^{-k} \phi_{0,\kappa}(\alpha z).$$

The function $\phi_{0,\kappa}$ is bounded in the complement of a neighborhood of 0, and so the function $\phi(z)$ is bounded in $\mathfrak{H}'$. Hence, it suffices to prove that the following integral is finite:

$$\int_{\mathfrak{H}'} \frac{1}{|z-\bar{u}|^k} y^{k/2-2} dx dy.$$

The finiteness follows easily using polar coordinates: the integral $\int_\varepsilon^\infty r^{-k+k/2-2} r dr$ is finite because $k > 0$, and the integral $\int_0^\pi (\sin\theta)^{k/2-2} d\theta$ is finite, because $k > 2$.

Assume next that u is Γ-parabolic. Clearly, after a conjugation, we may assume that $u = \infty$. Let $z \mapsto z + h$ be the Möbius transformation associated to the canonical generator γ_∞. The union M' is over a system of representatives for the cosets $\Gamma_\infty\backslash\Gamma$. Since M is compact and ∞ is Γ-parabolic, there is a number R so that no transform γM intersects the half plane: $\Im z > R$. Consider the part of M' in the vertical strip: $nh \le \Re z \le (n+1)h$. The part is transformed, by the power γ_∞^{-n}, into the vertical strip $0 \le \Re z \le h$. Moreover, since $\phi(z)$ is (Γ_u, j)-invariant, it follows from (1.8.1) that the integral over the part is unchanged when the part is replaced by its transform. Therefore, we may assume that M' is contained in the part $\mathfrak{H}'$ of $\mathfrak{H}$ determined by the inequalities $0 < \Im z \le R$, $0 \le \Re z \le h$. Hence, it suffices to prove that the following integral is finite,

$$\int_{\mathfrak{H}'} |\phi(z)| d_k^{\mathfrak{H}}(z).$$

The function $\phi = \phi_{\infty,\kappa}$ is, by definition, the function

$$\phi(z) = e^{2\pi i\kappa z/h}.$$

It is bounded in the horizontal strip $0 < \Im z \le R$. In particular, it is bounded in $\mathfrak{H}'$. Hence, it suffices to prove that the following integral is finite:

$$\int_{\mathfrak{H}'} y^{k/2-2} dx dy.$$

The finiteness is obvious: the integral $\int_0^R y^{k/2-2} dy$ is finite, because $k > 2$.

Thus it has been proved in both cases that the integral (1.8.5) is finite. Hence the Poincaré series converges normally in $\mathfrak{H}$, and part I of the proof is complete. □

(1.9) *Part II of the proof.* By part I of the proof, we know that the Poincaré series $G(z)$ is a (Γ, j)-invariant meromorphic function in $\mathfrak{H}$. Hence Theorem G holds if there are no Γ-parabolic points. In the remaining parts of the proof we assume that ∞ is Γ-parabolic. Let $z \mapsto z + h$ be the Möbius transformation associated to the canonical generator at ∞. In this second part of the proof, we prove that the partial series $G_w(z)$ of (1.7.3), where $\beta \in \Gamma$ and $\beta\infty \ne u$, converges to 0 for $\Im z \to \infty$.

The term $g_{\beta\gamma}$ in the series (1.7.3) is equal to $(\phi_{u,\kappa} \cdot_j \beta) \cdot_j \gamma$. The function $\phi_{u,\kappa}$ is a weight-k conjugate of either $\phi_{0,\kappa}$ or $\phi_{\infty,\kappa}$, and $\phi_{u,\kappa} \cdot_j \beta$ is a weight-k conjugate of $\phi_{u,\kappa}$.

Therefore, replacing u by $\beta^{-1}u$, we may assume that $\beta = 1$. Then $u \neq \infty$. Therefore, for the conjugation α defining $\phi_{u,\kappa} = (\phi_\kappa)^\alpha$, we have that $J(\alpha, z) = cz + d$ where $c \neq 0$. In fact, if $u \in \mathfrak{H}$ we may assume that $J(\alpha, z) = z - \bar{u}$ and if $u \in \partial_\Gamma \mathfrak{H}$ we have c and d in $\mathbf{R}$. Hence, dividing by a nonzero number, we may assume that $J(\alpha, z) = z + d$, where d has nonnegative imaginary part. It follows that $|\phi_{u,\kappa}| = |z + d|^{-k}|\phi_\kappa(\alpha z)$. The function $\phi_\kappa(\alpha z)$ is equal to $(\alpha z)^{\kappa e_u}$ if $u \in \mathfrak{H}$ and equal to $e^{2\pi i\kappa(\alpha z)/h}$ if $u \in \partial_\Gamma \mathfrak{H}$. Since $u \neq \infty$, it follows in both cases that the function $\phi_\kappa(\alpha z)$ is bounded in $\mathfrak{H}_R$. Hence, with a constant $C > 0$, we have the estimate for $z \in \mathfrak{H}_R$:

$$|\phi_{u,\kappa}(z)| \leq C|z + d|^{-k}.$$

The translation $z \mapsto z+h$ associated to γ_∞ generates the group $\mathrm{P}\Gamma_\infty$. Moreover $|J(\gamma_\infty^n, z)| = 1$. Therefore, for $z \in \mathfrak{H}_R$,

$$|G_w(z)| \leq C \sum_n \frac{1}{|z + nh + d|^k}.$$

Clearly, the sum on the right side converges to 0 uniformly as $\Im z \to \infty$.

Thus we have proved that the partial series $G_w(z)$, except in the exceptional case, converges to 0 uniformly as $\Im z \to \infty$, and part II of the proof is complete. ∎

(1.10) *Part III of the proof.* Assume as in part II that the point ∞ is Γ-parabolic and that $\mathrm{P}\Gamma_\infty$ is generated by $z \mapsto z + h$. We have to study the behavior of $G(z)$ near ∞. It suffices to consider three cases:

(1) The point u is in $\mathfrak{H}$.
(2) The point u is Γ-parabolic and not Γ-equivalent to ∞.
(3) The point u is equal to ∞.

We have to prove in all cases that $G(z)$ is exponentially bounded at ∞. In fact, we have to prove in the cases (1) and (2) that the order of G at ∞ is positive; in case (3) we have to prove that the Γ-order is positive if $\kappa > 0$, and equal to κ is $\kappa \leq 0$.

By part II of the proof, the function $G(z)$ is a normally convergent series $G(z) = \sum G_w(z)$ of functions $G_w(z)$ holomorphic in a half plane $\mathfrak{H}_R$. One of the functions $G_w(z)$ may be the exceptional function G_u of (1.7.2). Denote by $\sum_w' G_w(z)$ the series of the remaining functions. Each function $G_w(z)$ is (Γ_∞, j)-invariant, and hence of the form,

$$G_w(z) = e^{2\pi i\kappa_\infty z/h}\tilde{G}_w(q), \quad q = e^{2\pi i z/h},$$

where $\tilde{G}(q)$ as a function of q is meromorphic in the punctured unit disk: $0 < |q| < 1$. The half plane $\mathfrak{H}_R$ is mapped by q onto a punctured disk $V - 0$ where V is an the open disk: $|q| < \varepsilon$. Since $G_w(z)$ is holomorphic in $\mathfrak{H}_R$, the function $\tilde{G}_w(q)$ is holomorphic in the pointed disk $V - 0$. Moreover, since $G_w(z) \to 0$ for $\Im z \to \infty$ by part II and $\kappa_\infty < 1$, it follows that $\tilde{G}_w(q)$ is also holomorphic for $q = 0$; in addition, $\tilde{G}_w(q)$ vanishes for $q = 0$ when $\kappa_\infty = 0$. As the series $\sum' G_w(z)$ converges normally in $\mathfrak{H}_R$, it follows that the series $\sum' \tilde{G}_w(q)$ converges normally in the pointed disk $V - 0$. Moreover, each term $G_w(q)$ in the series is holomorphic also at $q = 0$. Therefore, as is well known, the series $\sum' \tilde{G}_w(q)$ converges normally in the whole disk V. In particular, the sum $\tilde{G}(q) := \sum \tilde{G}_w(q)$ is holomorphic in V; in addition, it vanishes at $q = 0$ if $\kappa_\infty = 0$. Therefore, since

$$\sum' G_w(z) = e^{2\pi i \kappa_\infty z/h} \tilde{G}(q),$$

it follows that the series $\sum'_w G_w(z)$ has positive order at ∞.

Now, in the cases (1) and (2), we have $G(z) = \sum'_w G_w(z)$. Hence the assertions for the cases (1) and (2) hold. In case (3), the function $G(z)$ is the sum of two functions,

$$G(z) = G_u(z) + \sum' G_w(z).$$

As $u = \infty$, the first function is $\phi_\infty(z) = e^{2\pi i \kappa z/h}$, and so its Γ-order at ∞ is equal to κ. The second function has positive Γ-order. Clearly, the assertion for the case (3) is a consequence.

Thus part III of the proof is complete, and Theorem G has been proved. □

(1.11) Note. Consider, in the assumptions of Theorem G, the series $G(z) = G(\Gamma, j, u, \kappa)$ for $\kappa \geq 0$. It follows from the Theorem that the form $G(z)$ is an integral form. If $u \in \mathfrak{D}$, then $G(z)$ is a cusp form and if $u \in \partial_\Gamma \mathfrak{D}$, then $G(z)$ is a cusp form if an only if $\kappa > 0$. Note however that the form $G(z)$ may be the zero form. In fact, the examples in (Autm.7) contain several cases where $\mathcal{S}(\Gamma, j) = 0$ (for $k > 2$). In these cases, necessarily $G(z) = 0$.

It follows from the Theorem that the Eisenstein series $G(\Gamma, j, u, 0)$ associated with a j-regular Γ-parabolic point u is an integral non-vanishing form in $\mathcal{G}(\Gamma, j)$. It has order 0 at the cusp defined by u, and it vanishes at all other cusps. It follows that the evaluation map in (Autm.6.13) is surjective. In particular, we recover the result that the codimension of $\mathcal{S}(\Gamma, j)$ in $\mathcal{G}(\Gamma, j)$, for $k > 2$, is equal to the number of j-regular cusps.

(1.12) Note. It is a consequence of Theorem G that there are non-zero (Γ, j)-automorphic forms for a factor j of any weight k. Indeed, assume first that $k > 2$. Then the Poincaré series $G(z) = G(\Gamma, j, u, \kappa)$ obtained from, say, a Γ-ordinary point u and $\kappa < 0$ has Γ-order equal to κ. In particular, $G(z)$ is a nonzero function in $\mathcal{M}(\Gamma, j)$. For general k, choose an even integer $l > 2$ such that $k + l > 2$. Then there are nonzero functions $f \in \mathcal{M}(\Gamma, jJ^l)$ and $g \in \mathcal{M}(\Gamma, J^l)$. Whence the quotient f/g is a nonzero function in $\mathcal{M}(\Gamma, j)$.

2. A Fourier expansion of an Eisenstein Series.

(2.1). In this section, the disk is assumed to be the upper half plane $\mathfrak{H}$ and k is assumed to be an integer. We consider a level-N subgroup Γ of the modular group $\Gamma(1) = \mathrm{SL}_2(\mathbf{Z})$ and on Γ a factor on Γ of the following form,

$$j(\gamma, z) = J(\gamma, z)^k/\chi(\gamma), \tag{2.1.1}$$

where $\chi : \Gamma \to \mathbf{C}^*$ is a unitary character. We make a series of assumptions:

(i) By hypothesis, Γ is the preimage in $\Gamma(1)$, under the reduction map modulo N, of a subgroup $\overline{\Gamma}$ of $\mathrm{SL}_2(\mathbf{Z}/N)$. We assume that χ is a level-N character on Γ, that is, χ is the composition of the reduction map and a character $\chi : \overline{\Gamma} \to \mathbf{C}^*$.

(ii) We assume that $\overline{\Gamma}$ contains the subgroup of diagonal matrices in $\mathrm{SL}_2(\mathbf{Z}/N)$. Equivalently, since the subgroup of diagonal matrices corresponds to the subgroup $\Gamma_0^0(N)$, we assume that $\Gamma \supseteq \Gamma_0^0(N)$.

(iii) It follows from the assumption in (ii) that Γ is homogeneous. In order that the factor (2.1.1) is homogeneous, we assume that

$$\chi(-1) = (-1)^k. \tag{2.1.2}$$

(iv) The Γ-parabolic points consist of the rational numbers and the point ∞, since Γ is of finite index in $\Gamma(1)$. Since Γ is homogeneous, the canonical generator at ∞ is the matrix t_h for some positive integer h. We assume that the point ∞ is a j-regular cusp, that is, we assume that

$$\chi(t_h) = 1. \tag{2.1.3}$$

(2.2). Denote by Γ_+ the unipotent subgroup of Γ, that is, Γ_+ is the intersection of Γ and the subgroup of upper triangular matrices with 1 in the diagonal. Then Γ_+ is the cyclic subgroup generated by γ_∞, and it is of index 2 in Γ_∞. Clearly, the image of Γ_+ in $\overline{\Gamma}$ is the unipotent subgroup $\overline{\Gamma}_+$ of $\overline{\Gamma}$. By assumption (2.1)(iv), the character $\chi : \overline{\Gamma} \to \mathbf{C}^*$ is trivial on the subgroup $\overline{\Gamma}_+$.

For any 2 by 2 matrix β, denote by ${}_2\beta$ the second row of β. Clearly, two matrices γ_1 and γ_2 with determinant 1 have the same second row if and only if $\gamma_1\gamma_2^{-1}$ is unipotent. It follows in particular that the map $\gamma \mapsto {}_2\gamma$ defines an injection,

$$\Gamma_+\backslash\Gamma \hookrightarrow \mathbf{Z}^2.$$

The image of the map will be denoted ${}_2\Gamma$. It consists of the pairs (c, d) of integers that occur as the second row of some matrix of Γ. Similarly, there is an injection,

$$\overline{\Gamma}_+\backslash\overline{\Gamma} \hookrightarrow (\mathbf{Z}/N)^2.$$

The image ${}_2\overline{\Gamma}$ consists of pairs of residue classes modulo N that occur as second row of a matrix of $\overline{\Gamma}$. The character $\chi : \overline{\Gamma} \to \mathbf{C}^*$ is trivial on the subgroup $\overline{\Gamma}_+$, and hence it induces a map from the set $\overline{\Gamma}_+\backslash\overline{\Gamma}$ to $\mathbf{C}^*$. Given the bijection above, the latter map may be viewed as a map from ${}_2\overline{\Gamma}$ to $\mathbf{C}^*$; we extend it with the value 0 to a map from $(\mathbf{Z}/N)^2$ to $\mathbf{C}$. The extended map,

$$\chi\colon (\mathbf{Z}/N)^2 \to \mathbf{C},$$

defines, by composition with the reduction map modulo N, a map $\chi\colon \mathbf{Z}^2 \to \mathbf{C}$. For $\alpha \in \mathrm{SL}_2(\mathbf{Z})$, denote by

$${}^{\alpha}\chi\colon \mathbf{Z}^2 \to \mathbf{C}$$

the map obtained by composition of $\chi\colon \mathbf{Z}^2 \to \mathbf{C}$ and right multiplication by α. Unwinding the definition, the value ${}^{\alpha}\chi(c,d)$ for integers c and d is determined as follows: if there exists a matrix γ of Γ such that (c,d) modulo N is the second row of $\gamma\alpha$, then ${}^{\alpha}\chi(c,d) = \chi(\gamma)$; otherwise ${}^{\alpha}\chi(c,d)$ is equal to 0.

By assumption (2.1)(ii), the group $\overline{\Gamma}$ contains the subgroup of diagonal matrices of $\mathrm{SL}_2(\mathbf{Z}/N)$. It follows that the row $(0,\bar{n})$ belongs to ${}_2\overline{\Gamma}$ if and only if $\bar{n}$ is invertible in $\mathbf{Z}/N$. Moreover, the group $(\mathbf{Z}/N)^*$ of invertible elements $\bar{n}$ is isomorphic to the group of diagonal matrices of $\mathrm{SL}_2(\mathbf{Z}/N)$. Hence χ defines, by restriction, a character $\chi\colon (\mathbf{Z}/N)^* \to \mathbf{C}^*$, and the function $\chi(0,\bar{n})$ is obtained from the character by extending it with the value 0 on residue classes that are not invertible.

The map $\chi(n) := \chi(0,n)$ is a residue character modulo N. From (2.1.2) we obtain the equation,

$$\chi(-n) = (-1)^k\chi(n). \tag{2.2.1}$$

In addition, for integers c and d we have the equation,

$${}^{\alpha}\chi(nc,nd) = \chi(n)\,{}^{\alpha}\chi(c,d). \tag{2.2.2}$$

Indeed, if n is not prime to N, then clearly both sides are equal to 0. Assume that n is prime to N, and let δ_n be a matrix of Γ whose reduction modulo N is a diagonal matrix with last row $(0,\bar{n})$. If γ is a matrix of Γ such that ${}_2\bar{\gamma}\bar{\alpha} = (\bar{c},\bar{d})$, then $\delta_n\gamma$ belongs to Γ and ${}_2\bar{\delta}_n\bar{\gamma}\bar{\alpha} = (\bar{n}\bar{c},\bar{n}\bar{d})$; hence

$${}^{\alpha}\chi(nc,nd) = \chi(\delta_n\gamma) = \chi(\delta_n)\chi(\gamma) = \chi(n)\,{}^{\alpha}\chi(c,d).$$

Thus the equation (2.2.2) holds if the right hand side is non-zero. It follows, replacing n by its inverse modulo N, that it holds if the left hand side is non-zero. Hence (2.2.2) holds.

(2.3). Consider, for $k \geq 3$, the Eisenstein series,

$$G(z) = G(\Gamma, j, \infty, 0)(z),$$

where j is the factor of (2.1.1). By Theorem G, the series is an integral j-automorphic form, and its order at ∞ is equal to 0. It vanishes at all cusps different from ∞. We will consider the Fourier expansion of $G(z)$ at any Γ-parabolic point v. So, let α be a matrix of $\mathrm{SL}_2(\mathbf{Z})$ such that $v = \alpha\infty$. Consider the weight-k conjugate series, $G^\alpha(z)$, that is,

$$G^\alpha(z) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \frac{1}{J(\alpha,z)^k j(\gamma,\alpha z)} = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \frac{\chi(\gamma)}{J(\gamma\alpha,z)^k},$$

where the last equation follows from the definition of j and the automorphy equations for $J(\gamma,z)$. Each right coset modulo Γ_∞ splits into two cosets modulo Γ_+. Hence, if we form the sum over the cosets modulo Γ_+, each term is repeated twice, and we obtain the equation,

$$2G^\alpha(z) = \sum_{\gamma\in\Gamma_+\backslash\Gamma} \frac{\chi(\gamma)}{J(\gamma\alpha,z)^k}\,.$$

By definition of the function ${}^\alpha\chi$, the equation may be rewritten as follows,

$$2G^\alpha(z) = \sum_{(c,d)\in{}_2\Gamma\alpha} \frac{{}^\alpha\chi(c,d)}{(cz+d)^k}\,. \tag{2.3.1}$$

The set of pairs (c,d) in ${}_2\Gamma\alpha$ is equal to the set of pairs (c,d) such that c and d are relatively prime and ${}^\alpha\chi(c,d)$ is non-zero. Indeed, obviously, if a pair (c,d) belongs to the first set, then it belongs to the second set. Conversely, assume that (c,d) belongs to the second set. Since c and d are prime, there is a matrix β in $\mathrm{SL}_2(\mathbf{Z})$ such that ${}_2\beta = (c,d)$. Since ${}^\alpha\chi(c,d)\neq 0$, there is a matrix γ in Γ such that, modulo N, ${}_2\gamma\alpha \equiv (c,d)$. The two matrices β and $\gamma\alpha$ have modulo N the same second row. Therefore, modulo N, the quotient $(\gamma\alpha)\beta^{-1}$ is a unipotent matrix of $\mathrm{SL}_2(\mathbf{Z}/N)$. The latter matrix can be lifted to a unipotent matrix τ of $\mathrm{SL}_2(\mathbf{Z})$. Replace β by $\tau\beta$. The replacement does not change the second row, so (c,d) is the second row of the new β. Moreover, for the new β, the quotient $(\gamma\alpha)\beta^{-1}$ is modulo N equal to 1. Therefore, the quotient belongs to $\Gamma(N)$. Since $\Gamma(N)$ is contained in Γ, it follows that the quotient $(\gamma\alpha)\beta^{-1}$ is in Γ. Thus $\beta = \gamma'\alpha$ with a matrix γ' of Γ. As $(c,d) = {}_2\beta$, it follows that (c,d) belongs to the first set ${}_2\Gamma\alpha$.

It follows that in the sum (2.3.1) is unchanged if it is formed over all pairs (c,d) of relatively prime integers. Consider the sum over all non-zero pairs (c,d) of integers. By grouping the terms according to the greatest common divisor, we obtain by (2.2.2) the equation,

$$\sum{}' \frac{{}^\alpha\chi(c,d)}{(cz+d)^k} = \sum_{n=1}^\infty \frac{\chi(n)}{n^k} \sum_{(c,d)=1} \frac{{}^\alpha\chi(c,d)}{(cz+d)^k}.$$

The third sum is, as noted above, equal to the sum in (2.3.1). Therefore, the following equation holds,

$$\sum{}' \frac{{}^\alpha\chi(c,d)}{(cz+d)^k} = \sum_{n=1}^\infty \frac{\chi(n)}{n^k}\bigl(2G^\alpha(z)\bigr). \tag{2.3.2}$$

By (2.2.1), the right hand side of (2.3.2) is unchanged if the factor 2 is omitted and the sum over $n \geq 1$ is replaced by the sum over $n \neq 0$. Whence,

$$\sum_{n\neq 0} \frac{\chi(n)}{n^k} G^\alpha(z) = \sum{}' \frac{{}^\alpha\chi(c,d)}{(cz+d)^k}\,. \tag{2.3.3}$$

Note that the factor $\sum_{n\neq 0} \chi(n)n^{-k}$ is non-zero. Indeed, in the sum over $n \geq 1$, the first term is equal to 1, and the sum of the remaining terms is, in absolute value, at most $\sum_{n\geq 2} n^{-k} \leq \int_1^\infty t^{-k}dt = 1/(k-1)$.

It follows from (2.3.3) that the series $G^\alpha(z)$, apart from the factor $\sum_{n\neq 0} \chi(n)n^{-k}$, is equal to the Eisenstein series $E_k^{{}^\alpha\chi}(z)$ considered in (App.2).

(2.4) Proposition. *The Eisenstein series $G(z)$ of* (2.3) *has at the cusp ∞ the Fourier expansion, with $q = e^{2\pi iz/N}$,*

$$G(z) = 1 + \frac{2}{A_k(\chi)}\Big(\frac{2i}{N}\Big)^k \sum_{r\geq 1} \sigma_{k-1}^\chi(r)q^r. \tag{2.4.1}$$

The constant $A_k(\chi)$ is the special number associated with the character $\chi(n)$, and $\sigma_{k-1}^\chi(r)$ is the weighted sum of $(k-1)$'st powers of divisors,

$$\sigma_{k-1}^\chi(r) = \sum_{d|r}\ \sum_{a \bmod N} \chi(r/d, a)e^{2\pi iad/N}d^{k-1}. \tag{2.4.2}$$

At a cusp $v = \alpha\infty$ which is not Γ-equivalent to ∞, the Fourier expansion is the following,

$$G^\alpha(z) = \frac{2}{A_k(\chi)}\Big(\frac{2i}{N}\Big)^k \sum_{r\geq 1} \sigma_{k-1}^{{}^\alpha\chi}(r)q^r\,, \tag{2.4.3}$$

where the definition of $\sigma_{k-1}^{{}^\alpha\chi}(r)$ is analogous to (2.4.2)*.*

Proof. By (App.1.10.2), for the constant in (2.3.3), we have the equation,

$$\sum_{n\neq 0} \frac{\chi(n)}{n^k} = \frac{(-1)^k\pi^k}{(k-1)!}A_k(\chi). \tag{2.4.4}$$

It follows from (2.1.2) that ${}^\alpha\chi(-c,-d) = (-1)^k\,{}^\alpha\chi(c,d)$. Hence, in the notation of (App.2), we have that $\big({}^\alpha\chi\big)^* = (-1)^k\,{}^\alpha\chi$.

Now, the series $G(z)$ is given by (2.3.3) with $\alpha = 1$. By definition, $\chi(n) = \chi(0,n)$. Hence the equation (2.4.1) follows directly from Corollary (App.2.5).

Assume that $v = \alpha\infty$ is not Γ-equivalent to ∞. If a matrix $\gamma\alpha$ has $(0, 1)$ as its second row, then it has ∞ as fixed point; thus $\gamma v = \gamma\alpha(\infty) = \infty$. Hence no matrix $\gamma\alpha$ with $\gamma \in \Gamma$ has $(0, 1)$ as its second row. Therefore, ${}^\alpha\chi(0,1) = 0$. It follows from (2.2.2) that ${}^\alpha\chi(0,n) = 0$ for all n. Consequently, the special number associated with the function ${}^\alpha\chi$ is equal to 0. Hence the equation (2.4.3) follows from the expansion for the Eisenstein series $E_k^{{}^\alpha\chi}(z)$ in Corollary (App.2.5) by dividing by the constant (2.4.4). ∎

(2.5) Note. The Fourier expansion of the Eisenstein series $G(z)$ in (2.4) is in terms of the parameter $q_N = e^{2\pi iz/N}$. On the other hand, $G(z)$ has period h, where t_h is the canonical generator of Γ at ∞, and so $G(z)$ has a Fourier expansion in terms of $q_h = e^{2\pi iz/h}$. Clearly, the matrix t_N belongs $\Gamma(N)$. Hence it belongs to Γ. Consequently, t_N is a power of t_h^l, that is, $N = lh$. It follows that the Fourier coefficient in (2.4.1) to q^r vanishes unless r is a multiple of l.

(2.6) Note. The Eisenstein series is not normally convergent for $k = 2$. However, the series on the right hand side of (2.3.3), when the summation is performed as in [App.2.2], is normally convergent, and if $G^\alpha(z)$ is defined by the equation (2.3.3) for $k = 2$, then the expansion of (2.4) holds. However, the resulting function $G^\alpha(z)$ can not be expected to be (Γ, j)-invariant in general.

When $k = 1$, the situation is even more complicated. In order that the summation described in [App.2.2] applies to the right hand side of (2.3.3), it is required that the sum $\sum_d {}^\alpha\chi(c, d)$ over d modulo N is equal to 0.

3. Example: Eisenstein series for the theta group.

(3.1). Recall that the θ-group Γ_θ is the subgroup of $\Gamma(1)$ formed by matrices that modulo 2 are congruent to either 1 or s. It is generated by the matrices t^2 and s. The θ-factor j_θ is a factor of weight $\frac{1}{2}$. It is determined on the generators by the equations,

$$j_\theta(t^2, z) = 1, \qquad j_\theta(s, z) = \sqrt{\frac{z}{i}}\,.$$

Let k be an integer. Then the power j_θ^{2k} is a factor of integral weight k. In particular, the square j_θ^2 is a factor of weight 1, and hence of the form,

$$j_\theta^2(\gamma, z) = J(\gamma, z)/\chi_\theta(\gamma),$$

where $\chi_\theta\colon \Gamma_\theta \to \mathbf{C}^*$ is a unitary character. The character is given on the generators as follows:

$$\chi_\theta(t^2) = 1 \qquad \chi_\theta(s) = i.$$

It follows that the character χ_θ is the character χ_θ considered in Exercise (Mdlar.3.8). It is given as follows,

$$\chi_\theta\begin{bmatrix} a & b \\ c & d\end{bmatrix} = \begin{cases} 1 & \text{for } d \equiv 1 \pmod 4, \\ i & \text{for } c \equiv 1 \pmod 4, \\ -1 & \text{for } d \equiv -1 \pmod 4, \\ -i & \text{for } c \equiv -1 \pmod 4. \end{cases}$$

The group Γ_θ is a level-4 group, and it follows from the description the χ_θ is a level-4 character. Clearly, the setup of (2.1) applies. The character $\chi_\theta(n)$ is a Dirichlet character modulo 4, usually denoted $\chi_4(n)$. The value $\chi_4(n)$ is 1 if $n \equiv 1$, it is -1 if $n \equiv -1$, and it is 0 if n is even. The function $\chi_\theta(c, d)$ is determined by the expression above. On a nonzero pair (c, d), the value is given as follows:

$$\chi_\theta(c, d) = \begin{cases} \chi_4(d) & \text{if } c \text{ is even and } d \text{ is odd}, \\ i\chi_4(c) & \text{if } c \text{ is odd and } d \text{ is even}, \\ 0 & \text{otherwise}. \end{cases}$$

(3.2). Assume that k is at least 3. Then the results of Sections 1 and 2 apply to the Eisenstein series,

$$G_k(z) := G(\Gamma_\theta, J^k/\chi_\theta^k, \infty, 0).$$

It follows that $G_k(z)$ is an integral form in $\mathcal{G}_k(\Gamma_\theta, \overline{\chi}_\theta^k)$. Moreover, its Fourier expansion is given by (2.4). Let us fix k and write χ for χ_θ^k. To determine the Fourier coefficients of the Eisenstein series, we need the special number $A_k(\chi)$ and the function $\sigma_{k-1}^\chi(r)$. The result is the following.

(3.3) Proposition. *The Eisenstein series $G_k(z)$ of* (3.2)*, for $k \geq 3$, has the following Fourier expansion, with $q = e^{2\pi iz/2}$,*

$$G_k(z) = 1 + \sum_{r\geq 1} \beta_k(r) q^r,$$

where the coefficients $\beta_k(r)$ are given by the formula,

$$\beta_k(r) = \frac{4}{A_k}\big[2^{k-1}\sigma'_{k-1}(r) + \sigma''_{k-1}(r)\big];$$

the A_k are the numbers of (App.1.3) *and*

$$\sigma'_{k-1}(r) = \sum_{d|r} \chi_4\big(r/d\big)^k d^{k-1}, \quad \sigma''_{k-1}(r) = \sum_{d|r,\, d\equiv k \bmod 2} (-1)^{(d-k)/2} d^{k-1}.$$

Proof. The expansion is given by (2.4), however in terms of $q_4 = e^{2\pi iz/4}$. As $G_k(z)$ is periodic with period 2, we know a priori that the coefficients $\sigma^{\chi}_{k-1}(r)$ are only non-zero when r is even.

To determine the constant factor in the expansion (2.4.1), note that the function $\chi(n) = \chi_4(n)^k$ depends on the parity of k. Assume first that k is odd. Then $\chi(n) = \chi_4(n)$. Clearly, $\chi_4 = 2\delta - \chi_2$ where $\delta(n)$ is the function introduced in Example (App.1.13). Hence, $A_k(\chi) = 2A_k(\delta) - A_k(\chi_2)$. It follows from (App.1.13) and (App.1.12), for k odd, that $2A_k(\delta) = (-1)^k A_k/2^k$ and $A_k(\chi_2) = 0$. Therefore,

$$A_k(\chi) = (-1)^k A_k/2^k. \tag{3.3.1}$$

If k is even, then $\chi(n)$ is the function $\chi_2(n)$ considered in Example (App.1.12); hence the special number is given by the formula $A_k(\chi) = A_k/2^k$. Therefore, the formula (3.3.1) holds for all k.

From (3.3.1) we obtain, for the constant factor in the expansion (2.4.1), the equation, $(2/A_k(\chi))(2i/4)^k = 2(-i)^k/A_k$. Hence the expansion is the following,

$$G_k(z) = 1 + \frac{4}{A_k}\sum_{r\geq 1} \frac{1}{2i^k}\sigma^{\chi}_{k-1}(r)\, q_4^r.$$

Therefore, we have to prove that $\sigma^{\chi}_{k-1}(r)$ vanishes when r is odd and that $\sigma^{\chi}_{k-1}(2r) = 2i^k\beta_k(r)$.

The sum $\sigma^{\chi}_{k-1}(r)$ is over divisors d of r and over a modulo 4. Denote by $\sigma^{0,2}(r)$ the sum of the terms corresponding to $a = 0$ and $a = 2$ and by $\sigma^{1,3}(r)$ the sum of the terms corresponding to $a = 1$ and $a = 3$. Clearly, for the first sum,

$$\sigma^{0,2}(r) = \sum_{d|r}\big[\chi(r/d, 0) + \chi(r/d, 2)(-1)^d\big]d^{k-1} = \sum_{d|r} i^k \chi_4(r/d)^k\big[1 + (-1)^d\big]d^{k-1}.$$

Obviously, the terms in the sum corresponding to odd divisors d vanish. In particular, when r is odd, the sum is zero. For an argument of the form $2r$, the nonzero terms are obtained from the even divisors of $2r$, that is, for divisors of the form $2d$ for $d \mid r$. Whence,

$$\sigma^{0,2}(2r) = 2i^k \sum_{d|r} \chi_4(r/d)^k 2^{k-1} d^{k-1} = 2i^k 2^{k-1} \sigma'_{k-1}(r). \tag{3.3.2}$$

For the second sum,

$$\sigma^{1,3}(r) = \sum_{d|r} \big[\chi(r/d,1)i^d + \chi(r/d,3)(-i)^d\big] d^{k-1} = \sum_{d|r,\ \frac{r}{d} \text{ even}} \big[1^k i^d + (-1)^k(-i)^d\big] d^{k-1}.$$

Evidently, the sum is zero if r is odd. For an argument of the form $2r$, the nonzero terms are obtained from the divisors d of r. Hence,

$$\sigma^{1,3}(2r) = \sum_{d|r} i^d\big[1 + (-1)^{k+d}\big] d^{k-1} = 2i^k \sum_{d|r,\, d\equiv k} i^{d-k} d^{k-1} = 2i^k \sigma''_{k-1}(r). \tag{3.3.3}$$

Since $\sigma^{\chi}_{k-1}(r) = \sigma^{0,2}(r) + \sigma^{1,3}(r)$, it follows that $\sigma^{\chi}_{k-1}(r)$ vanishes for odd r, and it follows from (3.3.2) and (3.3.3) that $\sigma^{\chi}_{k-1}(2r) = 2i^k \beta_k(r)$. Hence $G_k(z)$ has the asserted Fourier expansion. ∎

(3.4) Note. Depending on the residue class of k modulo 4, the expression for $\beta_k(r)$ can be simplified as follows:

Assume first that k is even. Then $\chi_4(n)^k = \chi_2(n)$. Hence $\sigma'_{k-1}(r) = \sum'_{d|r} d^{k-1}$ where the prime indicates a sum over those divisors d for which r/d is odd. Clearly, $\sigma''_{k-1}(r)$ vanishes for odd r, and for even r,

$$\sigma''_{k-1}(r) = (-1)^{k/2} 2^{k-1} \sum_{d|\frac{r}{2}} (-1)^d d^{k-1}.$$

The sum splits into a difference of two: $\sum^{\text{ev}}_{d|\frac{r}{2}} d^{k-1} - \sum^{\text{odd}}_{d|\frac{r}{2}} d^{k-1}$, where the two sums are, respectively, over the even and odd divisors of $r/2$. It follows, for k even, that $\beta_k(r)$ is equal to $2^{k+1}/A_k$ times the following expression,

$$\sum_{d|r}{}' d^{k-1} + (-1)^{k/2} \sum^{\text{ev}}_{d|\frac{r}{2}} d^{k-1} - (-1)^{k/2} \sum^{\text{odd}}_{d|\frac{r}{2}} d^{k-1}. \tag{3.4.1}$$

If r is odd, then the last two sums vanish, and the first is equal to $\sigma_{k-1}(r)$. Assume that r is even. Clearly, the last sum is the sum over the odd divisors d of r. The first sum is over the (necessarily) even divisors d for which r/d is odd, and the second sum is over the even divisors d for which r/d is even. Thus every divisor d of r contributes with a non-zero term in exactly one of the three sums in the expression. It follows that the expression, for $k \equiv 0 \pmod 4$, is equal to $\sigma^{\rm ev}_{k-1}(r) - \sigma^{\rm odd}_{k-1}(r)$. Hence, for $k \equiv 0 \pmod 4$,

$$\beta_k(r) = \frac{2^{k+1}}{A_k} \times \begin{cases} \sigma_{k-1}(r) & \text{if } r \text{ is odd,} \\ \sigma^{\rm ev}_{k-1}(r) - \sigma^{\rm odd}_{k-1}(r) & \text{if } r \text{ is even.} \end{cases}$$

Similarly, for r even and $k \equiv 2 \pmod 4$, the expression (3.4.1) is equal to the difference $\sigma_{k-1}(r) - 2\sum^{\rm ev}_{d|\frac{r}{2}} d^{k-1}$. Now the even divisors of $\frac{r}{2}$ are of the form $2d$ where d is a divisor of $\frac{r}{4}$. Therefore, for $k \equiv 2 \pmod 4$,

$$\beta_k(r) = \frac{2^{k+1}}{A_k} \times \begin{cases} \sigma_{k-1}(r) & \text{if } r \text{ is odd,} \\ \sigma_{k-1}(r) - 2^k\sigma_{k-1}(r/4) & \text{if } r \text{ is even.} \end{cases}$$

The difference in the last expression equals $(1+2^{k-1})\sigma^{\rm odd}_{k-1}(r) + (4^{k-1}-2^k)\sigma_{k-1}(r)$. Indeed, if r is even, then a divisor in r is either odd, or it is of the form $2d$ where d is an odd divisor in r, or it is of the form $4d$ where d is a divisor of $\frac{r}{4}$. Hence, $\sigma_{k-1}(r) = (1+2^{k-1})\sigma^{\rm odd}_{k-1}(r) + 4^{k-1}\sigma_{k-1}(\frac{r}{4})$, and we obtain the alternative expression for the difference.

Assume that k is odd. Then $\chi_4(n)^k = \chi_4(n)$. The sum in $\sigma''_{k-1}(r)$ is over the odd divisors d of r. If d is odd, then $i^{d-k} = (-1)^{(d-1)/2}(-1)^{(k-1)/2}$. Moreover, $\chi_4(d) = (-1)^{(d-1)/2}$. Therefore, for odd k,

$$\beta_k(r) = \frac{4}{A_k}\sum_{d|r}\bigl[2^{k-1}\chi_4(r/d) + (-1)^{(k-1)/2}\chi_4(d)\bigr]d^{k-1}.$$

(3.5). The θ-function is (essentially) given by its Fourier expansion at the cusp ∞, that is, $\theta(z) = \sum_n q^{n^2}$ where $q = e^{2\pi i z/2}$. The θ-function belongs to $\mathcal{G}(\Gamma_\theta, j_\theta)$, and it vanishes at the second cusp represented by -1. Hence the power $\theta(z)^l$, for a positive integer l, belongs to $\mathcal{G}(\Gamma_\theta, j^l_\theta)$ and it vanishes at -1. Clearly, the power has the Fourier expansion at ∞,

$$\theta(z)^l = 1 + \sum_{r\geq 1} b_l(r)q^r,$$

where $b_l(r)$ is the number of solutions $(n_1, \dots, n_l)$ in $\mathbf{Z}^l$ to the equation,

$$r = n_1^2 + \cdots + n_l^2.$$

Assume that l is even, say $l = 2k$, where k is a positive integer. Then $j^{2k}_\theta = (j^2_\theta)^k = J^k/\chi^k_\theta$. Hence $\theta(z)^{2k}$ belongs to the space $\mathcal{G}(\Gamma_\theta, J^k/\chi^k_\theta)$ and it vanishes at -1. The Eisenstein series $G_k(z)$, for $k \geq 3$, has the same properties. Moreover, both forms have at ∞ a Fourier expansion with constant term 1. It follows that the difference $\theta(z)^{2k} - G_k(z)$ is a cusp form.

(3.6) Proposition. *Let $b_{2k}(r)$ be the number of solutions $(n_1,\dots,n_{2k})$ in $\mathbf{Z}^{2k}$ to the equation,*

$$r = n_1^2 + \cdots + n_{2k}^2. \tag{3.6.1}$$

Then, for $k \geq 3$, the following asymptotic formulas holds:

$$b_{2k}(r) - \beta_k(r) = \mathrm{O}(r^{k/2}).$$

Moreover, the equality $b_{2k}(r) = \beta_k(r)$, holds for $k = 2, 3, 4$, that is,

$$b_4(r) = 8 \times \begin{cases} \sigma(r) & \text{when } r \text{ is odd}, \\ 3\sigma^{\mathrm{odd}}(r) & \text{when } r \text{ is even}. \end{cases}$$
$$b_6(r) = 4\sum_{d|r}\bigl[4\chi_4(r/d) - \chi_4(d)\bigr]d^2.$$
$$b_8(r) = 16 \times \begin{cases} \sigma_3(r) & \text{when } r \text{ is odd}, \\ \sigma_3^{\mathrm{ev}}(r) - \sigma_3^{\mathrm{odd}}(r) & \text{when } r \text{ is even}. \end{cases}$$

Finally, for $k = 1$, we have the equation, $b_2(r) = \frac{1}{2}\beta_1(r) = 4\sum_{d|r}\chi_4(d)$.

Proof. The difference $b_{2k}(r) - \beta_k(r)$, for $k \geq r$, is the r'th Fourier coefficient in the cusp form $\theta(z)^{2k} - G_k(z)$. The asymptotic formula, for $k \geq 3$, follows from Theorem F (Autm.6.16) since the difference $\theta(z)^{2k} - G_k(z)$ is a cusp form.

Moreover, for $k \leq 4$, the space $\mathcal{S}_k(\Gamma_\theta, j_\theta^{2k})$ of cusp forms is equal to 0 by Example (Autm.7.8). Therefore, the equation $b_{2k}(r) = \beta_k(r)$ holds for $3 \leq k \leq 4$. It holds also for $k = 2$, because it is possible to prove that the Poincaré series $G_2(z)$ with a suitable summation does define an automorphic form. The proof is not easy, and it is not covered in these notes.

Finally, the explicit formula, $b_2(r) = \frac{1}{2}\sum_{d|r}\chi_4(r)$ is elementary. It can be proved using unique prime factorization in the ring $\mathbf{Z}[i]$ of Gaussian integers. ∎

(3.7) Exercise. Study $G_k(z)$ at the cusp -1 using conjugation by the transformation $u = st^{-1}$. Prove that

$${}^{u}\chi_\theta(c,d) = \begin{cases} i\chi_4(-d) & \text{if } c \text{ and } d \text{ are odd}, \\ i\chi_4(c-d) & \text{if } c \text{ is odd and } d \text{ is even}, \\ 0 & \text{otherwise}. \end{cases}$$

Deduce from (2.4.3) and (Autm.5.2) a formula for the number $b_{2k}^{\mathrm{odd}}(r)$ of decompositions of r into $2k$ odd squares.

Appendix

1. Bernoulli and Euler numbers.

(1.1) Setup. As is well known, two important sequences of numbers, the *Bernoulli numbers* $B_0, B_1, \dots$ and the *Euler numbers* $E_0, E_1, \dots$ are defined by the Taylor expansions,

$$\frac{z}{e^z - 1} = \sum_{k \geq 0} B_k \frac{z^k}{k!}, \qquad \frac{2e^z}{e^{2z} + 1} = \sum_{k \geq 0} E_k \frac{z^k}{k!}. \tag{1.1.1}$$

The first function has poles at the non-zero integral multiples of $2\pi i$. Hence the first series is convergent in the open disk: $|z| < 2\pi$. Similarly, the second series converges when $|z| < \pi$. Obviously, $B_0 = E_0 = 1$.

The numbers appear in the Taylor expansion of many important functions. The second function in (1.1.2) is equal to $1/\cosh z$. In particular, it is an even function of z. Hence all Euler numbers of odd index vanish. Moreover, evaluation at iz yields the expansion,

$$\frac{1}{\cos z} = \sum_{k \text{ even}} (-1)^{k/2} E_k \frac{z^k}{k!}. \tag{1.1.2}$$

In the first equation of (1.1.1), divide by z and add $\frac{1}{2}$ to obtain the following equation:

$$\frac{1}{2}\,\frac{e^z + 1}{e^z - 1} = \frac{1}{z} + (B_1 + \tfrac{1}{2}) + \sum_{k \geq 1} \frac{1}{k+1} B_{k+1} \frac{z^k}{k!}.$$

In the equation, the left side is an odd function of z. Therefore, the equation implies that $B_1 = -\frac{1}{2}$ and that all other Bernoulli numbers of odd index vanish. The equation is the expansion of the function $\frac{1}{2}\coth\frac{z}{2}$. Evaluate at $2iz$ and multiply by $2i$ to obtain,

$$\cot z = \frac{1}{z} + \sum_{k \text{ odd}} \frac{(-1)^{(k+1)/2} 2^{k+1}}{k+1} B_{k+1} \frac{z^k}{k!}. \tag{1.1.3}$$

Finally, from $\cot z - 2\cot 2z = \tan z$, we obtain the expansion,

$$\tan z = \sum_{k \text{ odd}} \frac{(-1)^{(k-1)/2} 2^{k+1}(2^{k+1} - 1)}{k+1} B_{k+1} \frac{z^k}{k!}. \tag{1.1.4}$$

(1.2) Definition. It will be convenient to introduce a third sequence of numbers $A_0, A_1, \dots,$ called the *Euler zigzag numbers* or *up/down numbers*. They are defined by the expansion,

$$\alpha(z) = \frac{1+\sin z}{\cos z} = \sum_{k\geq 0} A_k \frac{z^k}{k!}. \tag{1.2.1}$$

Obviously, $A_0 = 1$. The function $\alpha(z)$ is the sum of the even function $\sec z = 1/\cos z$ and the odd function $\tan z$. Hence (1.2.1) implies the expansions,

$$\frac{1}{\cos z} = \sum_{k \text{ even}} A_k \frac{z^k}{k!}, \qquad \tan z = \sum_{k \text{ odd}} A_k \frac{z^k}{k!}. \tag{1.2.2}$$

Accordingly, the A_k with even k are called *secant numbers*, and the A_k with odd k are called *tangent numbers*.

By comparing (1.2.2) with the expansions (1.1.2) and (1.1.4) we obtain the following equations for $k \geq 0$:

$$A_k = \begin{cases} (-1)^{k/2} E_k & \text{if } k \text{ is even,} \\ (-1)^{(k-1)/2} 2^{k+1}(2^{k+1}-1)B_{k+1}/(k+1) & \text{if } k \text{ is odd,} \end{cases}$$

or,

$$E_k = (-1)^{k/2} A_k, \quad k \text{ even}, \qquad B_k = \frac{(-1)^{k/2-1}k}{2^k(2^k-1)} A_{k-1}, \quad k \geq 2 \text{ even}.$$

(1.3) Lemma. *The zigzag numbers A_k for $k \geq 0$ are positive. whole numbers. Moreover, if $k \geq 2$ is even, then kA_{k-1} is divisible by 2^{k-1}. Furthermore, the numbers are given by $A_0 = 1$ and the following recursion formula for $k \geq 0$:*

$$A_{k+1} = A_k + \binom{k}{2} A_{k-1} - \binom{k}{4} A_{k-3} + \binom{k}{6} A_{k-5} \pm \cdots.$$

Proof. The function $\alpha(z)$ of (1.2.1) is the sum of $\sec z = 1/\cos z$ and $\tan z$. To obtain the derivatives of $\alpha(z)$, note that $(\sec z)' = \sec z \tan z$ and $(\tan z)' = 1 + \tan^2 z$. It follows by induction on k that there are polynomials $F_0(t), F_1(t), \dots$ and $G_1(t), G_2(t), \dots$ and an equation,

$$\alpha^{(k)}(z) = \sec z \; F_k(\tan z) + G_{k+1}(\tan z).$$

The polynomials are given recursively: $F_0 = 1$, $G_1 = t$, and $F_{k+1} = tF_k + (1+t^2)F_k'$ and $G_{k+1} = (1+t^2)G_k'$. It follows easily that each of F_k and G_k is a polynomial of degree k, with positive integer coefficients in degrees $k, k-2, k-4, \dots$ and zero coefficients otherwise. In particular, $A_k = F_k(0) + G_{k+1}(0)$ is a positive integer. Hence the first part of the Lemma holds.

To prove the second part, consider the Taylor expansion,

$$\tau(z) = 2z\tan z = \sum C_k \frac{z^k}{k!}.$$

By (1.2.2), if $k \geq 2$ is even then C_k is the (even) integer $2kA_{k-1}$ and C_k vanishes otherwise. It suffices to prove that the integer C_k is divisible by 2^k.

Clearly, $\tau'(z) = 2\tan z + 2z(1+\tan^2 z)$. Multiplication by $2z$ yields the equation,

$$2z\tau'(z) - 2\tau(z) = 4z^2 + \tau(z)^2.$$

It follows by comparing the coefficients of $z^k/k!$ for $k > 2$ that

$$2(k-1)C_k = \sum_{i=0}^{k} \binom{k}{i} C_i C_{k-i}, \quad \text{for } k > 2. \tag{1.3.2}$$

As noted above, $C_k = 0$ if k is odd, or if $k = 0$. Moreover, $C_2 = 4$. Proceed by induction, and assume that $k > 2$ is even, $k = 2l$. In the sum (1.3.2), the extreme terms vanish, because $C_0 = 0$. In the middle term, the binomial coefficient $\binom{k}{l}$ is even, $\binom{k}{l} = 2\binom{k-1}{l-1}$. The remaining terms come in equal pairs, since $\binom{k}{i} = \binom{k}{k-i}$. Therefore, division by 2 in (1.3.2) yields:

$$(k-1)C_k = \sum_{0<i<l} \binom{k}{i} C_i C_{k-i} + \binom{k-1}{l-1} C_l^2. \tag{1.3.3}$$

By the induction hypothesis, each product $C_i C_{k-i}$ (including C_l^2) is divisible by 2^k. So the sum is divisible by 2^k. Hence, so is C_k, because the factor $k-1$ is odd.

To prove the recursion formula, note that $\alpha(z) = \alpha'(z)\cos z$. Insert the expansions $\alpha(z) = \sum A_k z^k/k!$ and $\alpha'(z) = \sum A_{k+1} z^k/k!$ and the well known expansion of $\cos z$. The recursion formula follows by comparing the coefficients of z^k. ∎

(1.4). A simple computation using the recursion formula gives the following table:

k	0	1	2	3	4	5	6	7	8	9	10	11
A_k	1	1	1	2	5	2^4	61	$2^4 \cdot 17$	1385	$2^8 \cdot 31$	50521	$2^9 \cdot 691$

(1.5) Definition. Let $\psi: \mathbf{Z} \to \mathbf{C}$ be a periodic function, say $\psi(n+N) = \psi(n)$ for all integers n. Associate with ψ the sequence of *transformed Bernoulli numbers* $A_k(\psi)$ for $k \geq 0$ defined by the expansion,

$$\alpha_\psi(z) = \sum_{b \mod N} \frac{\psi(b)}{N} \cot\frac{z-\pi b}{N} = \frac{1}{z}\sum_{k\geq 0} A_k(\psi)\frac{z^k}{k!}, \tag{1.5.1}$$

where the index b ranges over a system of representatives for the residue classes modulo N. At $z = 0$, only the factor $\frac{1}{N}\cot(z/N)$ in the term for $b = 0$ contributes with a pole; it is a simple pole with residue 1. Hence $A_0(\psi) = \psi(0)$. Clearly $\alpha_\psi(z)$ has period π, with poles at the numbers $z = a\pi$. In particular, the series in (1.5.1) converges for $|z| < \pi$.

The relation between that numbers $A_k(\psi)$ and the generalized Bernoulli numbers $B_k(\psi)$ is described in (1.15).

(1.6). The starting point of the following computations is the well known formula,

$$\frac{1}{z}+\sum_{n\geq 1}\Big(\frac{1}{z+n}+\frac{1}{z-n}\Big)=\pi\cot\pi z. \tag{1.6.1}$$

The series on the left is normally convergent, and so it defines a meromorphic function in $\mathbf{C}$. To prove the formula, note that the function on the left hand side has the following properties: it has poles exactly at the integers, with principal part around $z=n$ equal to $1/(z-n)$, it is periodic with period 1, and it is bounded in any domain $|\Im z|\geq \varepsilon$. Clearly, the function on the right has the same properties. It follows that the difference of the two functions is bounded in $\mathbf{C}$, and hence constant. As the two functions are odd functions, the difference is zero.

We shall rewrite the formula as follows:

$$\sum_n{}^{+}\frac{1}{z+n}=\pi\cot\pi z,$$

where the summation sign $\sum^{+}$ indicates the *symmetric summation* over all integers n, that is, the sum with the terms ordered as indicated in (1.6.1). Note that the sum $\sum_n (z+n)^{-k}$ is absolutely convergent when $k\geq 2$.

(1.7) Lemma. *Let $\psi\colon \mathbf{Z}\to\mathbf{C}$ be a periodic map. Then the following formula holds:*

$$\sum_n{}^{+}\frac{\psi(n)}{z+n}=-\pi\alpha_\psi(-\pi z). \tag{1.7.1}$$

Proof. Fix a system of representatives a for the residue classes modulo a period N. Then the symmetric summation on the left side of (1.7.1) can be replaced by the sum over a of the symmetric summation over n of the terms $\psi(a)/(z+a+Nn)$. For fixed a we have, by (1.6.1), the equation,

$$\sum_n{}^{+}\frac{1}{z+a+Nn}=\frac{1}{N}\pi\cot\frac{\pi z+\pi a}{N}=-\pi\frac{1}{N}\cot\frac{-\pi z-\pi a}{N}.$$

Multiply by $\psi(a)$, and form the sum over a. Clearly, the result is the right side of (1.7,1). ∎

(1.8) Proposition. *Let $\psi\colon\mathbf{Z}\to\mathbf{C}$ be a periodic map, and form the function $\alpha_\psi(z)$ in* (1.5.1). *Then, for $k\geq 1$,*

$$\sum_n{}^{+}\frac{\psi(n)}{(z+n)^k}=\frac{-\pi^k}{(k-1)!}\alpha_\psi^{(k-1)}(-\pi z). \tag{1.8.1}$$

As a consequence,

$$\sum_{n\neq 0}{}^{+}\frac{\psi(n)}{n^k}=\frac{-\pi^k A_k(\psi)}{k!}. \tag{1.8.2}$$

Proof. The first formula is obtained from (1.7.1) by applying the operator d^{k-1}/dz^{k-1} and multiplying the result by $(-1)^{k-1}/(k-1)!$. As is clear from the definition in (1.5.1), the function $\alpha_\psi(z)-\psi(0)/z$ has no pole at $z=0$, and the value at $z=0$ of its $(k-1)$'st derivative is A_k/k for $k\geq 1$. Hence, after subtracting from both sides of (1.8.1) the term $\psi(0)/z^k$, the formula (1.8.2) is obtained by evaluating the resulting equation at $z=0$. ∎

(1.9) Note. In the following examples it is convenient to obtain trigonometric relations using logarithmic derivation: the logarithmic derivative of $F(z)$ is the derivative of $\log F(z)$, equal to $F'(z)/F(z)$. Logarithmic derivation turns multiplication into addition.

For instance, the function $\alpha(z)$ of (1.2.1) is equal to $\cos z/(1 - \sin z)$, and hence equal to the logarithmic derivative of $A(z) = 1/(1 - \sin z)$. The function $\cot z$ is the logarithmic derivative of $\sin z$. Consequently, the function $\alpha_\psi(z)$ is the logarithmic derivative of the function,

$$F_\psi(z) := \prod_{b \bmod N} \left(\sin \frac{z - \pi b}{N}\right)^{\psi(b)}.$$

The exponents $\psi(b)$ are not assumed to be integers, so $F_\psi(z)$ is a multi valued function.

(1.10) Example. For the function $\chi \equiv 1$ ($N = 1$) we have $\alpha_1(z) = \cot z$. So the corresponding transformed Bernoulli numbers $A_k(1)$ are the Taylor coefficients of $z \cot z$. They may be obtained from (1.1.3). Alternatively, with $F_1 = \sin z$ and $A(z) = (1 - \sin z)^{-1}$ we have that

$$F_1(2z)/F_1(z) = 2\cos z = 2/\sqrt{A(z)A(-z)}.$$

Take the logarithmic derivative and multiply by z:

$$(2z)\alpha_1(2z) - z\alpha_1(z) = \tfrac{-z}{2}\big(\alpha(z) - \alpha(-z)\big).$$

Therefore, for the k'th Taylor coefficient: $(2^k - 1)A_k(1) = \frac{-1}{2}(1 + (-1)^k)kA_{k-1}$, that is,

$$A_0(1) = 1, \qquad A_k(1) = \frac{-k}{2^k - 1}A_{k-1}, \qquad \text{if } k \geq 2 \text{ even},$$

$A_0(1) = 1$, and $A_k(1) = 0$ otherwise.

Hence, by (1.8.2), if $k \geq 2$ is even, then $2\zeta(k) = \sum_{n \neq 0} n^{-k} = \pi^k A_{k-1}/(2^k - 1)(k-1)!$. Equivalently,

$$\frac{1}{1^k} + \frac{1}{2^k} + \frac{1}{3^k} + \cdots = \frac{\pi^k A_{k-1}}{2(2^k - 1)(k-1)!}, \qquad k \geq 2 \text{ even}. \tag{1.10.1}$$

(1.11) Example. Let $\psi = \psi_{1/2}$ be the parity character defined by $\psi_{1/2}(n) = 1$ when n is odd and $\psi_{1/2}(n) = 0$ if n is even ($N = 2$). The transformed Bernoulli numbers $A_k(\psi_{1/2})$ are determined from $F_{1/2} := \sin(\frac{z}{2} - \frac{\pi}{2})$. Obviously,

$$F_{1/2}(2z) = -\cos z = -1/\sqrt{A(z)A(-z)}.$$

Take the logarithmic derivative and multiply by z:

$$(2z)\alpha_{1/2}(2z) = \tfrac{-z}{2}\big(\alpha(z) + \alpha(-z)\big).$$

Therefore, by taking the k'th Taylor coefficient,

$$A_k(\psi_{1/2}) = \frac{-kA_{k-1}}{2^k} \quad \text{when } k \geq 2 \text{ is even,}$$

and $A_k(\psi_{1/2}) = 0$ otherwise. Hence, by (1.8.2),

$$\frac{1}{1^k} + \frac{1}{3^k} + \frac{1}{5^k} + \cdots = \frac{\pi^k A_{k-1}}{2^{k+1}(k-1)!}, \qquad k \geq 2 \text{ even.} \tag{1.11.1}$$

(1.12) Example. Let $\psi_{1/4}$ be the characteristic function of 1 modulo 4, defined by $\psi_{1/4}(n) = 1$ if $n \equiv 1 \pmod 4$ and $\psi_{1/4}(n) = 0$ otherwise ($N = 4$). The transformed Bernoulli numbers $A_k(\psi_{1/4})$ are determined from $F_{1/4} = \sin(\frac{z}{4} - \frac{\pi}{4})$. Clearly,

$$2F_{1/4}(2z)^2 = 2\sin^2(\tfrac{z}{2} - \tfrac{\pi}{4}) = 1 - \cos(z - \tfrac{\pi}{2}) = 1 - \sin z = A(z)^{-1}.$$

Take the logarithmic derivative and multiply by z:

$$2(2z)\alpha_{1/2}(2z) = -z\alpha(z).$$

Hence $A_k(\psi_{1/4}) = -kA_{k-1}/2^{k+1}$. As a consequence, for all $k \geq 1$, we have the formula,

$$\sum_{n \equiv 1 \pmod 4}^{+} \frac{1}{n^k} = \frac{\pi^k A_{k-1}}{2^{k+1}(k-1)!}. \tag{1.12.1}$$

If k is odd, it follows that

$$\frac{1}{1^k} - \frac{1}{3^k} + \frac{1}{5^k} \pm \cdots = \frac{\pi^k A_{k-1}}{2^{k+1}(k-1)!}, \qquad k \geq 1 \text{ odd.} \tag{1.12.2}$$

and if k is even, we obtain the result from Example (1.11).

(1.13) Exercise. Prove for $S(z) = 2\sin z$ the formula,

$$S(Nz) = \prod_{j=0}^{N-1} S(z + j\pi/N).$$

[Hint: use $X^N - 1 = \prod_{j=0}^{N-1}(X - e^{2\pi i j/N})$.] Deduce that the transformed Bernoulli numbers $A_k(\psi)$ are independent of the choice of period N entering in their definition.

(1.14) Exercise. Prove that A_k is equal to the number of *up-down permutations* of $(1,\dots,k)$, that is, permutations $(\sigma_1,\dots,\sigma_k)$ with $\sigma_1<\sigma_2>\sigma_3<\sigma_4>\sigma_5<\cdots$.

Hint: From the equation $2\alpha'(z)=\alpha(z)^2+1$, deduce the recursion formula,

$$2A_k=\sum_{i=0}^{k-1}\binom{k-1}{i}A_iA_{k-i-1},\qquad A_0=A_1=1.$$

Prove for the number a_k of up-down permutations the following formulas (for $k\ge 1$):

$$a_k=\sum_{\substack{i=0\\ i\text{ odd}}}^{k-1}\binom{k-1}{i}a_ia_{k-i-1}=\sum_{\substack{i=0\\ i\text{ even}}}^{k-1}\binom{k-1}{i}a_ia_{k-i-1}.$$

The i'th terms in the two sums count the numbers of up-down permutations having, respectively, the biggest element k at position $i+1$ and the smallest element 1 at position $i+1$.

(1.15) Exercise. The transformed Bernoulli numbers $A_k(\psi)$ introduced here are closely related to the generalized Bernoulli numbers $B_k(\psi)$ of Leopoldt, defined by the following expansion,

$$\beta_\psi(z)=\sum_{a=1}^{N}\psi(a)\frac{ze^{az}}{e^{Nz}-1}=\sum_{k=0}^{\infty}B_k(\psi)\frac{z^k}{k!}\,.$$

Indeed, let $\psi_-(n)=\psi(-n)$ and let $\hat\psi$ be the *discrete Fourier transform* of ψ, defined by the equation,

$$\hat\psi(n)=\sum\psi(a)\zeta^{an},\quad \zeta=e^{2\pi i/N}.$$

Then the following equation holds:

$$\alpha_\psi(z)=-\tfrac{i}{N}\hat\psi(0)+\tfrac{2i}{N}\beta_{\hat\psi_-}(\tfrac{2iz}{N}).$$

Prove the equation and deduce the following equation by comparing coefficients to z^{k-1}:

$$A_k(\psi)=\left(\tfrac{2i}{N}\right)^kB_k(\hat\psi_-),\quad k\neq 1,\qquad A_1(\psi)=\tfrac{-i}{N}\hat\psi(0)+\tfrac{2i}{N}B_1(\hat\psi_-).$$

Prove for a primitive multiplicative character $\chi\colon(\mathbf{Z}/N)^*\to\mathbf{C}^*$, extended to all of $\mathbf{Z}$ with the value zero on integers n not prime to N, that

$$\hat\chi=\tau(\chi,\zeta)\overline{\chi},$$

where τ is the Gauss sum.

(1.16) Exercise. Prove for k even that $kA_k/2^{k-1}$ is odd. [Hint: Induction. Use (1.3.3) divided by 2^k. Rewrite the sum using the equation $\binom{k}{i}=\binom{k-1}{i-1}+\binom{k-1}{i}$. The resulting expression should not depend on the parity of l. Deduce modulo 2 that the sum equals $\sum_{1\le j<l}\binom{k-1}{j}=2^{k-2}-1$.]

2. Fourier expansion of Eisenstein series.

(2.1) Setup. Fix a natural number N, let $\zeta = \zeta_N$ be the canonical N'th root of unity, $\zeta := e^{2\pi i/N}$, and let $q = q_N$ be the function of z defined by $q(z) = e^{2\pi iz/N}$. In addition, let $\tau(\psi, \xi)$ be the *Gauss sum*, defined for any N'th root of unity ξ as the sum,

$$\tau(\psi, \xi) := \sum_b \psi(b)\xi^b, \tag{2.1.1}$$

over a system of representatives b of the residue classes modulo N.

In Section 1 we studied the function $\sum^+ \psi(n)/(z+n)$ where the sum is the symmetric summation. We proved the equation $\sum^+ \psi(n)/(z+n) = -\pi\alpha_\psi(-\pi z)$, and we used it to study the function around $z = 0$. By definition of the function $\alpha_\psi(z)$ in (1.5), the equation is the following:

$$\sum_n{}^+ \frac{\psi(n)}{z+n} = \frac{\pi}{N}\sum \psi(b)\cot\frac{\pi(z+b)}{N}. \tag{2.1.2}$$

To study the function in the upper half plane $\mathfrak{H}$, use the equation,

$$\cot z = i\,\frac{e^{iz}+e^{-iz}}{e^{iz}-e^{-iz}} = i\Big(1 - \frac{2}{1-e^{2iz}}\Big).$$

With $z := \pi(z+b)/N$ it follows that

$$\cot\frac{\pi(z+b)}{N} = i\Big(1 - \frac{2}{1-\zeta^b q}\Big), \quad \zeta = e^{2\pi i/N}, q = e^{2\pi iz/N}.$$

For $z \in \mathfrak{H}$ we have $|q| < 1$, and hence $1/(1-\zeta^b q) = \sum_{d\ge 0}\zeta^{bd}q^d$. Therefore, from (2.1.1) and (2.1.2) we obtain the expansion,

$$\sum_n{}^+ \frac{\psi(n)}{z+n} = \frac{\pi i}{N}\sum_b \psi(b) - \frac{2\pi i}{N}\sum_{d\ge 0}\tau(\psi, \zeta^d)q^d, \quad z \in \mathfrak{H}. \tag{2.1.3}$$

Apply the operator d^{k-1}/dz^{k-1} and multiply by $(-1)^{k-1}/(k-1)!$. Since $q'(z) = (2\pi i/N)q(z)$, we obtain, for $k \ge 2$ and $z \in \mathfrak{H}$, the formula,

$$\sum_n \frac{\chi(n)}{(z+n)^k} = \frac{(-1)^k}{(k-1)!}\Big(\frac{2\pi i}{N}\Big)^k \sum_{d=1}^{\infty}\tau(\chi, \zeta^d)d^{k-1}q^d. \tag{2.1.4}$$

(2.2) Definition. Let $\chi : \mathbf{Z}\times\mathbf{Z} \to \mathbf{C}$ be a function of 2 integer variables, and set $\chi_m(n) := \chi(m,n)$. Assume that the function χ is bounded. In addition, assume for all m in $\mathbf{Z}$ that $\chi_m(n)$, as a function of n, has period N. Consider for $k \ge 1$ the associated *Eisenstein series*,

$$E_k^\chi(z) := \sum{}' \frac{\chi(m,n)}{(mz+n)^k}, \tag{2.2.1}$$

where the sum is over $(m,n) \neq (0,0)$. The series is normally convergent for $k \geq 3$ and defines a holomorphic function in the upper half plane $\mathfrak{H}$. If the summation is arranged as follows:

$$E_k^\chi(z) := \sum_{n\neq 0} \frac{\chi(0,n)}{n^k} + \sum_{m\neq 0}\sum_n \frac{\chi(m,n)}{(mz+n)^k}\,, \tag{2.2.2}$$

then the series is normally convergent also for $k = 2$ and defines a holomorphic function $E_2^\chi(z)$. In fact, as we shall show below, if the function χ satisfies that $\sum_{a \bmod N} \chi(m,a) = 0$ for all m, and if the summations over n in (2.2.2) are interpreted as the symmetric summations, then the right hand side of (2.2.2) is even convergent for $k = 1$.

For every natural number r, consider the following linear combination of k'th powers of the divisors of r:

$$\sigma_k^\chi(r) = \sum_{d|r}\sum_{a \bmod N} \chi(r/d,a)e^{2\pi i ad/N}\, d^k,$$

where the inner sum is over a system of representatives for the residue classes modulo N. In terms of the Gauss sums of (2.1),

$$\sigma_k^\chi(r) = \sum_{d|r} \tau(\chi_{r/d},\zeta^d)\, d^k.$$

Finally, denote by $A_k(\chi)$ the k'th transformed Bernoulli number of (1.5) associated to the function $\chi_0(n) = \chi(0,n)$.

(2.3) Observation. Obviously, for all $k \geq 1$, it follows by a change of summation order in (2.2.2) that $E_k^\chi(z+1) = E_k^\phi(z)$ where $\phi(m,n) := \chi(m,n-1)$; in particular, it follows that $E_k^\chi(z+N) = E_k^\chi(z)$. If $k \geq 3$, then it follows by a change of summation order in (2.2.1) that $E_k^\chi(-1/z) = z^k E_k^\varphi(z)$, where $\varphi(m,n) := \chi(-n,m)$. The argument does not work for $k = 2$ or $k = 1$.

(2.4) Proposition. *In the setup of* (2.2)*, assume either that* $k \geq 2$*, or that* $k = 1$ *and* $\sum_a \chi(m,a) = 0$ *for all* m*. Then, with* $q = e^{2\pi i z/N}$,

$$\sum_{m\geq 1}\sum_n \frac{\chi(m,n)}{(mz+n)^k} = \frac{(-1)^k}{(k-1)!}\Big(\frac{2\pi i}{N}\Big)^k \sum_{r=1}^{\infty} \sigma_{k-1}^\chi(r) q^r. \tag{2.4.1}$$

On the left, the sum over n is the symmetric summation when $k = 1$*.*

Proof. Assume that $m \geq 1$. Then $q(mz) = q(z)^m$. Hence, by (2.1.4) applied to mz,

$$\sum_n \frac{\chi(m,n)}{(mz+n)^k} = \frac{(-1)^k}{(k-1)!}\Big(\frac{2\pi i}{N}\Big)^k \sum_{d=1}^{\infty} \tau(\chi_m,\zeta^d) d^{k-1} q^{md}. \tag{2.4.2}$$

On the right side, the Gauss sum is bounded as a function of m,n since χ is assumed to be bounded. Obviously, the double series $\sum_{m,d} d^{k-1}q^{md}$ is normally convergent when q is in the unit disk. Hence the sum over $m \geq 1$ of the functions in (2.4.2) converges normally. In the resulting equation, rearrange the summation on the right side according to the value of $r := md$. Clearly (2.4.1) results. ∎

(2.5) Corollary. *Assume either that $k \geq 2$, or that $k = 1$ and $\sum_a \chi(m, a) = 0$ for all m. Define $\chi^*(m, n) := \chi(-m, -n)$, and let $b_r = \sigma^{\chi}_{k-1}(r) + (-1)^k \sigma^{\chi^*}_{k-1}(r)$ for $r \geq 1$. Then the Fourier expansion of $E^{\chi}_k(z)$ in $\mathfrak{H}$ is the following, with $q = e^{2\pi i z/N}$,*

$$E^{\chi}_k(z) = \frac{-\pi^k A_k(\chi)}{k!} + \frac{(-1)^k}{(k-1)!}\Big(\frac{2\pi i}{N}\Big)^k \sum_{r\geq 1} b_r q^r,$$

where $b_r = \sigma^{\chi}_{k-1}(r) + (-1)^k \sigma^{\chi^*}_{k-1}(r)$. If $A_k(\chi) \neq 0$, then the normalized series G^{χ}_k obtained by dividing E^{χ}_k by its constant term has the following expansion,

$$G^{\chi}_k(z) = 1 + \frac{(-1)^{k-1} k}{A_k(\chi)}\Big(\frac{2i}{N}\Big)^k \sum_{r\geq 1} b_r q^r.$$

Proof. The left side is equal to the sum of the two series in (2.2.2). The first (constant) series is, by (1.8.2), equal to the constant term in the asserted Fourier expansion. The second series is separated into two: the sum of terms for $m \geq 1$ and the sum of terms for $m \leq -1$. For the first sum, use the equation (2.4.1). In the second sum, replace m by $-m$ to obtain the following sum:

$$\sum_{m\geq 1} \sum_n \frac{\chi(-m, n)}{(-mz + n)^k}.$$

Replace n by $-n$ in the inner sum and multiply by $(-1)^k$. The result is the left hand side of (2.4.1) with χ replaced by χ^*. Thus (2.4.1) applies to the second sum as well, and we obtain for the Fourier coefficient b_r the asserted sum.

The expression for the normalized series follows immediately from the expression for $E^{\chi}_k(z)$. ∎

(2.6) Example. Assume that χ is the constant function $\chi(m, n) = 1$. Then $N = 1$, and the corresponding function E_k is the Eisenstein series considered in (Autm.2.1) for $k \geq 3$,

$$E_k(z) = \sum{}' \frac{1}{(nz + m)^k}.$$

The divisor sum is the sum,

$$\sigma_{k-1}(r) = \sum_{d|r} d^{k-1}.$$

The series vanishes when k is odd. Assume that $k \geq 2$ is even. Then $b_r = 2\sigma_{k-1}(r)$. Hence we obtain the Fourier expansion,

$$E_k(z) = 2\zeta(k) + \frac{(-1)^{k/2} 2(2\pi)^k}{(k-1)!} \sum_{r\geq 1} \sigma_{k-1}(r) q^r, \quad q = e^{2\pi i z}. \tag{2.6.1}$$

The normalized series G_k is obtained by dividing by $2\zeta(k)$. In the notation of Section 1, $2\zeta(k) = \pi^k A_{k-1}/(2^k-1)(k-1)!$. Hence we obtain for G_k the Fourier series,

$$G_k(z) = 1 + \frac{(-1)^{k/2}2^{k+1}(2^k-1)}{A_{k-1}} \sum_{r\geq 1} \sigma_{k-1}(r)q^r. \tag{2.6.2}$$

In particular, the constant term is equal to 1 and all the Fourier coefficients are rational numbers. In particular,

$$\begin{aligned} G_2(z) &= 1 - 24\sum_{r\geq 1}\sigma_1(r)q^r, \\ G_4(z) &= 1 + 240\sum_{r\geq 1}\sigma_3(r)q^r, \\ G_6(z) &= 1 - 504\sum_{r\geq 1}\sigma_5(r)q^r. \end{aligned} \tag{2.6.3}$$

The number A_{k-1} for even k is expressed by Bernoulli numbers in (1.2). It follows that the factor in (2.6.2) is equal to $-2k/B_k$, where B_k is the usual Bernoulli number.

3. The function of Weierstrass.

(3.1) Setup. Fix in $\mathbf{C}$ a lattice Ω, say $\Omega = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$ where the complex numbers ω_1 and ω_2 are linearly independent over $\mathbf{R}$. Denote by $\mathcal{M}(\Omega)$ the field of meromorphic Ω-*periodic* functions, that is, a function f belongs to $\mathcal{M}(\Omega)$ if and only if f is meromorphic in $\mathbf{C}$ and has any ω in Ω as period:

$$f(z+\omega) = f(z).$$

The group Ω acts on $\mathbf{C}$ by translations, and the Ω-periodic functions are the functions invariant under the action. Obviously, a fundamental domain for the action is any lattice parallelogram,

$$R = \{z_0 + t_1\omega_1 + t_2\omega_2 \mid 0 \le t_i \le 1\}.$$

The equivalence on the boundary of R identifies the opposite sides of the parallelogram. It follows that the quotient $X := \mathbf{C}/\Omega$ is a torus. In particular, the quotient is a compact connected Riemann surface of genus 1, and $\mathcal{M}(\Omega)$ is the field of meromorphic functions on X.

(3.2) Lemma. *Let f be a non-zero function in $\mathcal{M}(\Omega)$. Then the following two formulas hold:*

$$\sum_{u \bmod \Omega} \operatorname{ord}_u f = 0, \qquad \sum_{u \bmod \Omega} (\operatorname{ord}_u f)u \equiv 0 \pmod{\Omega}. \tag{3.2.1}$$

Proof. The boundary of the fundamental domain R is formed by four line segments: the segment L_1 from z_0 to $z_0 + \omega_2$, the segment L_2 from z_0 to $z_0 + \omega_1$, and the two translated segments $L_j' := L_j + \omega_j$ for $j = 1, 2$. We may assume that the angle from ω_2 to ω_1 is positive, that is, the quotient ω_1/ω_2 has positive imaginary part. Then the boundary of R, as an oriented path, is the sum

$$\partial R = L_1 + L_2' - L_1' - L_2.$$

Clearly, we may choose z_0 such that f has no zeros or poles on the boundary of R. Consider the path integral,

$$\frac{1}{2\pi i}\int_{\partial R} \frac{f'(z)}{f(z)}dz. \tag{3.2.2}$$

Its value is, by the residue formula, equal to the left hand side of the first formula in (3.2.1). On the other hand, since f'/f is Ω-periodic, it follows that $\int_{L_j'}(f'/f)dz = \int_{L_j}(f'/f)dz$. Hence the path integral (3.2.2) is equal to 0. Therefore, the first formula in (3.2.1) holds.

To prove the second formula, consider the path integral,

$$\frac{1}{2\pi i}\int_{\partial R} \frac{zf'(z)}{f(z)}dz. \tag{3.2.3}$$

The poles of the integrand are the poles and the zeros of $f(z)$. At a point u which is either a zero or a pole of f, the residue of $zf'(z)/f(z)$ is equal to ku, where $k = \operatorname{ord}_u f$. Hence, by the residue formula, the path integral is equal to the left hand side of the second formula in (3.2.1). On the other hand, since f'/f is Ω-periodic, it follows that $\int_{L'_j}(zf'/f)dz = \int_{L_j}(zf'/f)dz + \omega_j \int_{L_j}(f'/f)dz$. Hence the path integral (3.2.3) is equal to the following sum,

$$-\frac{\omega_1}{2\pi i}\int_{L_1}\frac{f'(z)}{f(z)}dz + \frac{\omega_2}{2\pi i}\int_{L_2}\frac{f'(z)}{f(z)}dz. \tag{3.2.4}$$

By the transformation formula for integrals, the path integral $\int_{L_j}(f'/f)dz$ is equal to the integral $\int_{fL_j}(1/w)dw$. Moreover, the image path fL_j is a closed path in $\mathbf{C}^*$. Therefore, the integral $(2\pi i)^{-1}\int_{L_j}(f'/f)dz$ is an integer. It follows that the sum (3.2.4) is an integral linear combination of ω_1 and ω_2. Hence the sum belongs to Ω. Thus the second formula of (3.2.1) holds. ∎

(3.3) Remark. The two formulas of (3.2) are trivial if f is constant. Assume that f is a non-constant function in $\mathcal{M}(\Omega)$. Then the formulas apply to $f(z) - \lambda$ for any λ in $\mathbf{C}$. It follows from the first formula that the number of times $f(u)$, for u modulo Ω, takes the value λ is equal to the number of poles; in particular, the number is independent of λ. Similarly, by the second formula, again modulo Ω, the sum of the complex numbers u for which $f(u) = \lambda$ is congruent to the sum of the poles of f; in particular, the sum is modulo Ω independent of λ.

As a special case, it follows that if a function f in $\mathcal{M}(\Omega)$ has no poles, then f is constant.

(3.4). The *Weierstrass $\wp$-function* is the function $\wp(u) = \wp_\Omega(u)$ defined by the series,

$$\wp(u) = \frac{1}{u^2} + \sum_{\omega\neq 0}\Big(\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2}\Big), \tag{3.4.1}$$

where the sum is over non-zero ω in Ω. The series is normally convergent in $\mathbf{C}$ and the function $\wp(u)$ is a meromorphic function with poles of order 2 at the points of Ω. Obviously, it is an even function: $\wp(-u) = \wp(u)$.

It is not hard to see directly that $\wp(u)$ is Ω-periodic. Alternatively, we may proceed as follows: consider the derivative:

$$\wp'(u) = \frac{-2}{u^3} + \sum_{\omega\neq 0}\frac{-2}{(u-\omega)^3} = \sum_{\omega}\frac{-2}{(u-\omega)^3}.$$

Clearly, the derivative $\wp'(u)$ is Ω-periodic. Therefore, for $\omega_0 \in \Omega$, there is an equation, $\wp(u+\omega_0) - \wp(u) = C$ with a constant $C = C(\omega_0)$. Take $u := -u - \omega_0$ in the equation. Since $\wp(u)$ is an even function, it follows that $C = 0$. Hence $\wp(u)$ is Ω-periodic.

To obtain the Laurent expansion of $\wp(u)$ at the origin, consider the difference $\wp(u) - 1/u^2$. It follows from (3.4.1) that the difference is holomorphic at the origin with the value 0. Moreover, by applying the operator $(d/du)^k$ for $k \geq 1$ to the difference, we obtain the series,

$$\sum_{\omega\neq 0}\frac{(-1)^k(k+1)!}{(u-\omega)^{k+2}}.$$

Clearly, the value of the series at the origin is equal to the number $(k+1)!E_{k+2}$ where $E_k = E_k(\Omega) = \sum_{\omega\neq 0}\omega^{-k}$ as defined in (Autm.2.1). Hence $\wp(u)$ has at the origin the Laurent expansion,

$$\wp(u) = \frac{1}{u^2} + \sum_{k\geq 1}(k+1)E_{k+2}u^k. \tag{3.4.2}$$

From the expansions for $\wp(u)$ and $\wp'(u)$,

$$\wp = u^{-2} + 3E_4u^2 + 5E_6u^4 + \cdots, \quad \wp' = -2u^{-3} + 6E_4u + 20E_6u^3 + \cdots,$$

we obtain easily the equation,

$$(\wp')^2 - 4\wp^3 + 60E_4\wp = -140E_6 + \cdots.$$

The left hand side is Ω-periodic and its possible poles belong to Ω. The right hand side is holomorphic at 0. Therefore, the left hand side has no poles. It follows that the left hand side is constant, equal to $-140E_6$. Hence, the $\wp$-function satisfies the following differential equation:

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3, \tag{3.4.3}$$

where $g_2 = g_2(\Omega) = 60E_4(\Omega)$ and $g_3 = g_3(\Omega) = 140E_6(\Omega)$.

(3.5). Modulo ω, the origin is the only pole of $\wp'(u)$, and it is a pole of order 3. Therefore, by (3.2), $\wp(u)$ has 3 zeros. Consider the three numbers $u_1 := \omega_1/2$, $u_2 = \omega_2/2$, and $u_3 := (\omega_1+\omega_2)/2$. They are inequivalent modulo Ω, and $u_j \equiv -u_j \pmod{\Omega}$. Since $\wp'(u)$ is an odd function, the numbers u_j are zeros of $\wp'(u)$. Hence the u_j are exactly the zeros of $\wp'(u)$.

The three values $\lambda_j = \wp(u_j)$ are different. Indeed, for $\lambda \in \mathbf{C}$, the function $\wp(u) - \lambda$ has a pole of order 2 at the origin. Hence, again by (3.2), the function $\wp(u) - \lambda$ has two zeros. For $\lambda = \lambda_j$, the point u_j is a zero for $\wp(u) - \lambda_j$, and of multiplicity 2, since $\wp'(u_j) = 0$. Hence u_j is the only zero of $\wp(u) - \lambda_j$.

By (3.4.3), the values $\lambda_j = \wp(u_j)$ are roots of the polynomial $4X^3 - g_2X - g_3$, and they are different. The latter polynomial has, for arbitrary g_2, g_3 in $\mathbf{C}$, three roots λ_j in $\mathbf{C}$, and the discriminant of the polynomial is the number,

$$D = 16(\lambda_1-\lambda_2)^2(\lambda_1-\lambda_3)^2(\lambda_2-\lambda_3)^2 = g_2^3 - 27g_3^2.$$

Hence, for $g_2 = g_2(\Omega)$ and $g_3 = g_3(\Omega)$, it follows that the discriminant $D(\Omega) = g_2^3 - 27g_3^2$ is non-zero.

(3.6) Proposition. *Let $g_2 = g_2(\Omega)$ and $g_3 = g_3(\Omega)$. Then the discriminant $D(\Omega) = g_2^3 - 27g_3^2$ is non-zero. Moreover, the map $u \mapsto (\wp(u), \wp'(u))$ induces a bijection from the set of points u in $\mathbf{C} - \Omega$ modulo Ω to the set of pairs $(x, y) \in \mathbf{C}^2$ satisfying the equation*

$$y^2 = 4x^3 - g_2x - g_3. \tag{3.6.1}$$

Proof. The first assertion was proved in (3.5). To prove the second, note first that, by (3.4.3), the pair $(x, y) := (\wp(u), \wp'(u))$ satisfies the equation (3.6.1).

Consider conversely a pair (x_0, y_0) satisfying the equation (3.6.1). The function $\wp(u) - x_0$ has, counted with multiplicity, two zeros. If u_0 is a zero, then so is $-u_0$. If $y_0 = 0$, then x_0 is one of the three roots λ_j in the polynomial $4x_0^3 - g_2 - g_3$ and u_0 is, modulo Ω, the corresponding unique zero u_j of $\wp(u) - \lambda_j$. If $y_0 \neq 0$, then u_0 is not one of the numbers u_j. Hence the two numbers u_0 and $-u_0$ are different modulo Ω. Moreover, the two values $\wp'(u_0)$ and $\wp'(-u_0) = -\wp'(u_0)$ are non-zero and hence different. Hence exactly one of the values is equal to y_0 and the other value is equal to $-y_0$.

Therefore, in both cases, there is a unique u_0 modulo Ω such that $(\wp(u_0), \wp'(u_0)) = (x_0, y_0)$. □

(3.7) Note. Let z be a point in the upper half plane $\mathfrak{H}$. Consider the lattice $\Omega_z := \mathbf{Z}z + \mathbf{Z}$. Then there is an associated $\wp$-function $\wp_{\Omega_z}(w)$. Moreover, the numbers $E_k(\Omega_z)$, for $k \geq 3$, and $g_2(\Omega_z)$ and $g_3(\Omega_z)$, are functions of z. In fact, they are the functions $E_k(z)$ of Example (Autm.2.1), and $g_2(z)$ and $g_3(z)$ of Example (Autm.2.5).

By Example (2.5), $E_k(z) = 2\zeta(k)G_k(z)$ and $2\zeta(k) = \pi^k A_k/(2^k-1)(k-1)!$. In particular, as $A_4 = 2$ and $A_6 = 2^4$,

$$g_2(z) = 60E_4(z) = \frac{(2\pi)^4}{2^2 \cdot 3} G_4(z), \quad g_3(z) = 140E_4(z) = \frac{(2\pi)^6}{2^3 \cdot 3^3} G_6(z).$$

As a consequence,

$$D(z) = g_2(z)^3 - 27g_3(z)^2 = (2\pi)^{12} \cdot \frac{G_4(z)^3 - G_6(z)^2}{12^3}.$$

Hence, except for the factor $(2\pi)^{12}$, the discriminant of (3.6), as a function $D(z)$ of $z \in \mathfrak{H}$, is equal to the discriminant $\Delta(z)$ of Example (Autm.2.3), cf. (Autm.7.1). Thus the non-vanishing of $D(\Omega)$ implies the non-vanishing of $\Delta(z)$. Similarly,

$$\frac{g_2(z)^3}{g_2(z)^3 - 27g_3(z)^2} = \frac{G_4(z)^3}{G_4(z)^3 - 27G_6(z)^2}. \tag{3.7.1}$$

It follows that the left hand side of (3.7.1) multiplied by 12^3 is equal to Klein's function $j(z)$ defined in (Autm.2.5).

In general, for a pair (a_2, a_3) of complex numbers such that $a_2^3 - 27a_3^2 \neq 0$, we will write

$$j(a_2, a_3) := \frac{12^3 a_2^3}{a_2^3 - 27a_3^2}.$$

If (a_2', a_3') is a second pair, then $j(a_2', a_3') = j(a_2, a_3)$ if and only if, for some non-zero number λ,

$$a_2' = \lambda^4 a_2, \quad a_3' = \lambda^6 a_3. \tag{3.7.2}$$

Indeed, the "if" part is obvious. Assume conversely the equality of the two numbers j. Clearly, if $a_2 = 0$, then $a_2' = 0$, and so (3.7.2) holds with 6 possible values of λ. Assume that $a_2 \neq 0$. Then the first equation of (3.7.2) can be solved with 4 values of λ. For any such λ, it follows from the equality of the two j's that $a_3^2 = \lambda^{12} a_3'^2$. Therefore, if $a_3 \neq 0$ then the two equations (3.7.2) hold for 4 values of λ, and if $a_3 = 0$ then the equations (3.7.2) hold for 2 values of λ.

(3.8) Proposition. *Given two complex numbers a_2 and a_3 such that $a_2^3 - 27a_3^2 \neq 0$. Then there exists a unique lattice Ω in $\mathbf{C}$ such that $a_2 = g_2(\Omega)$ and $a_3 = g_3(\Omega)$.*

Proof. Note first that any lattice Ω is of the form,

$$\Omega = \lambda \Omega_z, \tag{3.8.1}$$

for some $\lambda \in \mathbf{C}^*$ and $z \in \mathfrak{H}$. Indeed, $\Omega = \mathbf{Z}\omega_1 + \mathbf{Z}\omega_2$, where we may assume that $\omega_1/\omega_2 \in \mathfrak{H}$; hence it suffices to take $\lambda := \omega_2$ and $z := \omega_1/\omega_2$. Moreover, z is uniquely determined up to multiplication by a matrix in $\mathrm{SL}_2(\mathbf{Z})$.

Next, note that the functions $g_2(\Omega)$ and $g_3(\Omega)$ are homogeneous, respectively of degree -4 and -6:

$$g_2(\lambda\Omega) = \lambda^{-4} g_2(\Omega), \quad g_3(\lambda\Omega) = \lambda^{-6} g_3(\Omega). \tag{3.8.2}$$

Indeed, more generally, it is obvious that $E_k(\Omega)$ is homogeneous of degree $-k$.

Now, by (3.7), Klein's j-invariant is the function $j(z) = j(g_2(z), g_3(z))$. The j-invariant is, by (Autm.7.4), an isomorphism $\mathfrak{H}/\Gamma(1) \xrightarrow{\sim} \mathbf{C}$. In particular, there is a unique orbit $\Gamma(1)z$ in $\mathfrak{H}$ such that $j(a_2, a_3) = j(z)$. It follows, as noted in (3.7), that $a_2 = \lambda^{-4} g_2(z)$ and $a_3 = \lambda^{-6} g_3(z)$. Therefore, by (3.8.2), if Ω is defined by (3.8.1), then $a_2 = g_2(\Omega)$ and $a_3 = g_3(\Omega)$.

To prove that Ω is unique, assume for a second lattice Ω' the following two equations:

$$g_2(\Omega) = g_2(\Omega'), \qquad g_3(\Omega) = g_3(\Omega'). \tag{3.8.3}$$

The equations are preserved if Ω and Ω' are multiplied by the same factor. Hence we may assume that

$$\Omega = \lambda \Omega_z \qquad \Omega' = \Omega_{z'}, \tag{3.8.4}$$

for a non-zero λ and $z, z' \in \mathfrak{H}$. In fact we may assume that z and z' belong to a given system of representatives in the standard fundamental domain F for the action of $\mathrm{SL}_2(\mathbf{Z})$ on $\mathfrak{H}$.

From the equations (3.8.4), it follows that $j(z) = j(z')$. Therefore, since z and z' are assumed to belong to a system of representatives, it follows that $z = z'$. Hence,

$$\Omega = \lambda \Omega_z, \quad \Omega' = \Omega_z. \tag{3.8.5}$$

Therefore, by (3.8.2), (3.8.3), and (3.8.5),

$$g_2(z) = \lambda^4 g_2(z), \qquad g_3(z) = \lambda^6 g_3(z). \tag{3.8.6}$$

If $g_2(z)$ and $g_3(z)$ are non-zero, it follows that $\lambda = \pm 1$; in particular, then Ω_z is invariant under multiplication by λ. If $g_2(z) = 0$, then it follows from the second equation that λ is a sixth root of unity. Moreover, since z is a zero of g_2, it is a zero of G_4, and hence, by (Autm.7.3), $z = \rho$. Therefore, the lattice Ω_z is invariant under multiplication by λ. Finally, if $g_3(z)$ is equal to 0, it follows similarly that λ is a fourth root of unity and that $z = i$; hence Ω_z is invariant under multiplication by λ. Thus, in all cases, the lattice Ω_z is invariant under multiplication by λ. Therefore, by (3.8.5), $\Omega = \Omega'$, and the proof of the Proposition is complete. □

4. Notes on elliptic curves.

(4.1) Note. By definition, an *elliptic curve* is a pair (X, o) consisting of a (compact, connected) Riemann surface X and a distinguished point o of X. For instance, the torus $X := \mathbf{C}/\Omega$ associated with a lattice Ω in (3.1), with the image of 0 as distinguished point, is an elliptic curve. As a second example, let g_2 and g_3 be complex numbers such the $g_2^3 - 27g_3^2 \neq 0$. Consider the equation,

$$y^2 = 4x^3 - g_2x - g_3. \tag{4.1.1}$$

The equation defines an affine algebraic curve in the affine plane $\mathbf{C}^2$. The affine plane $\mathbf{C}^2$ is an open subset of the projective plane $\mathbb{P}^2(\mathbf{C})$. In fact, the points of $\mathbb{P}^2(\mathbf{C})$ are defined by homogeneous non-zero sets of coordinates (x, y, z) up to multiplication by a (non-zero) scalar. Among these sets of coordinates are the equivalence classes of coordinates (x, y, z) for which $z \neq 0$. Clearly, the latter equivalence classes are represented by coordinates of the form $(x, y, 1)$, and hence by points (x, y) of $\mathbf{C}^2$. The *projective curve* corresponding to the equation (4.1.1) is the set of points in $\mathbb{P}^2(\mathbf{C})$ whose homogeneous coordinates (x, y, z) satisfy the homogenized equation,

$$y^2z = 4x^3 - g_2xz^2 - g_3z^3. \tag{4.1.2}$$

Clearly, the homogenized equation (4.1.2) is satisfied for (x, y, z) with $z \neq 1$ if and only if (4.1.1) is satisfied for $(x/z, y/z)$. In addition, the homogeneous equation is satisfied for $(x, y, 0)$ if and only if $x = 0$. Hence the projective curve consists of the points of the affine curve and the point represented by $(0, 1, 0)$.

The projective curve defined by (4.1.2) will be denoted $X(g_2, g_3)$. It is not hard to prove that $X(g_2, g_3)$ is a Riemann surface.

By Proposition (3.8), there is a lattice Ω such that $g_2 = g_2(\Omega)$ and $g_3 = g_3(\Omega)$. Let $\wp(u) = \wp_\Omega(u)$ be the function of Weierstrass. It follows from Proposition (3.6) that the map $u \mapsto (\wp(u), \wp'(u), 1)$ for $u \notin \Omega$ and $u \mapsto (0, 1, 0)$ for $u \in \Omega$ induces a bijection,

$$\mathbf{C}/\Omega \xrightarrow{\sim} X(g_2, g_3).$$

Clearly, the map is a map of Riemann surfaces. Hence it is an isomorphism. In particular, therefore $X(g_2, g_3)$ is a Riemann surface of genus 1, and with $o := (0, 1, 0)$ as distinguished point, it is an elliptic curve.

(4.2) Note. Let X be a Riemann surface. The divisors on X, see (Autm.8.3), form an abelian group, freely generated by the points of X. Two linearly equivalent divisors have the same degree. The group of linear equivalence classes of divisors is the *Picard group* $\operatorname{Pic} X$. Denote by $\operatorname{Pic}^n X$ the set of equivalence classes of degree n. Then each $\operatorname{Pic}^n X$ is a coset modulo the subgroup $\operatorname{Pic}^0 X$.

For any divisor D of X, denote by $\operatorname{cl}(D)$ the equivalence class of D in $\operatorname{Pic}(X)$. A point p of X defines a divisor $1.p$ of degree 1; its class $\operatorname{cl}(1.p)$ belongs to $\operatorname{Pic}^1 X$.

Lemma I. *Let X be a Riemann surface of genus* 1*. Then the map $p \mapsto \mathrm{cl}(1.p)$ is a bijection,*

$$X \xrightarrow{\sim} \mathrm{Pic}^1 X.$$

Proof. Since the genus is equal to 1, it follows from Riemann-Roch's Theorem that

$$\dim H^0(D) = \deg D \quad \text{if } \deg D > 0. \tag{4.2.1}$$

We have to prove the following assertion: for any divisor D of degree 1, there is a unique point p of X such that D and $1.p$ are linearly equivalent, that is, there is a unique point p of X such that, for some non-zero $\varphi \in \mathcal{M}(X)$,

$$\operatorname{div}\varphi + D = 1.p\,. \tag{4.2.2}$$

Now, the left hand side $\operatorname{div}\varphi + D$ is a divisor of degree 1. Hence it is of the form $1.p$ if and only if it is positive: $\operatorname{div}\varphi + D \geq 0$. By (4.2.1), the functions φ for which $\operatorname{div}\varphi + D \geq 0$ form a one-dimensional subspace of $\mathcal{M}(X)$. Therefore, the assertion holds. ∎

(4.3) Note. Consider an elliptic curve (X, o). Clearly, the map $D \mapsto D + 1.o$ induces a bijection $\mathrm{Pic}^0 X \xrightarrow{\sim} \mathrm{Pic}^1 X$. By composing with the bijection of Lemma I, we obtain a bijection from X to the commutative group $\mathrm{Pic}^0 X$. Hence there is a unique structure as an abelian group on the points of X such that the bijection is an isomorphism of groups. The structure will be called the *elliptic addition* on X, and it will be denoted additively. Clearly, under the bijection, the distinguished point o of X corresponds to the zero element of $\mathrm{Pic}^0 X$. Hence o is the zero element of the elliptic addition.

Unwinding the definition, the elliptic addition of two points p and q on X is the unique point $r = p + q$ of X such that we have the linear equivalence of divisors,

$$(1.p - 1.o) + (1.q - 1.o) \equiv 1.r - 1.o\,,$$

or, equivalently,

$$1.p + 1.q \equiv 1.r + 1.o\,. \tag{4.2.3}$$

Lemma II. *The elliptic addition, as a group structure on (X, o), is characterized by the following property: Let φ be a non-zero meromorphic function on X. Then the sum, $\sum_{p\in X}(\mathrm{ord}_p\,\varphi)p$, with respect to the elliptic addition, is equal o.*

Proof. Let us first prove that the property characterizes the elliptic addition. Clearly, by the property applied to a constant function, the element o is the zero element. If p and q are points of X, let r be the (unique) point defined by the relation (4.2.3). The latter relation means that for some non-zero function φ we have that $\operatorname{div}\varphi = 1.r + 1.o - 1.p - 1.q$. Therefore, by the property, the equation $r + o - p - q = o$ holds in the group X. It follows that $r = p + q$.

Next we prove that the property holds: By construction of the elliptic addition, the map $p \mapsto \mathrm{cl}(1.p - 1.o)$ is a group isomorphism from X to $\mathrm{Pic}^0 X$. Thus is suffices to prove that the sum, $\sum_{p\in X}(\mathrm{ord}_p\,\varphi)p$, with respect to the elliptic addition, under this isomorphism goes to the zero class of $\mathrm{Pic}^0 X$. Clearly, the sum goes to the class of the divisor,

$$\sum_{p\in X}(\mathrm{ord}_p\,\varphi)(1.p - 1.o) = \operatorname{div}\varphi - (\deg\operatorname{div}\varphi).o\,.$$

The divisor $\operatorname{div}\varphi$ is principal and its degree is equal zero. Hence the class of the divisor is equal to the zero class. Thus the elliptic addition has the property. ∎

(4.4) Note. Consider the elliptic curve $X = \mathbf{C}/\Omega$ of (4.1). It has a natural group structure as a quotient of the additive group of $\mathbf{C}$ modulo the subgroup Ω. It follows from Lemma II above and Lemma (3.2) that the addition induced on X by the addition of $\mathbf{C}$ is equal to the elliptic addition.

Consider the elliptic curve $X(g_2, g_3)$ of (4.1). It has a group structure defined by the well known addition on a smooth cubic. The latter addition is defined as follows: consider, to simplify, two different points $p = (x_1, y_1)$ and $q = (x_2, y_2)$ on the affine part of the curve. Let $L = 0$ be the equation of the line through p and q. It intersects the curve in a third point s on the curve. Let $M = 0$ be the equation of the line through o and s (it is simply the vertical line through s). By definition, the third point r of intersection of $M = 0$ and the curve is the composition of p and q. Now, the quotient L/M may be viewed as a meromorphic function of the curve. It is not hard to see that

$$\operatorname{div}(L/M) = 1.p + 1.q + 1.s - (1.o + 1.s + 1.r) = 1.p + 1.q - 1.o - 1.r\,.$$

Therefore, the point r is also the elliptic sum of p and q.

(4.5) Note. Let (X, o) be an elliptic curve. We will sketch the proof that (X, o) is isomorphic to a curve of the form $X(g_2, g_3)$ (the Weierstrass normal form).

First, since X is of genus 1, it follows from the Theorem of Riemann-Roch that $H^0(n.o)$ is of dimension n for $n > 0$. In particular, the following relations hold:

$$H^0(0.o) = H^1(1.o) \subset H^0(2.o) \subset H^0(3.o) \subset H^0(4.o) \subset H^0(5.o) \subset H^0(6.o).$$

(The first equality holds because $H^0(0)$ is the 1-dimensional space of constant functions and contained in the 1-dimensional space $H^0(1.o)$.)

Now choose a function x in $H^0(2.o)$ and not in $H^0(1.o)$ and a function y in $H^0(3.o)$ and not in $H^0(2.o)$. The functions x and y are holomorphic except at o; at o they have, respectively, a pole of order at most 2 and at most 3. Since $x \notin H^0(1.o)$, it follows that $\operatorname{ord}_o x = -2$. Similarly, $\operatorname{ord}_o y = -3$. Clearly, the function x^2 belongs to $H^0(4.o)$ and since $\operatorname{ord}_o x^2 = -4$ it does not belong to $H^0(3.o)$. Similarly, xy belongs to $H^0(5.o)$ and it does not belong to $H^0(4.o)$, and the two functions x^3 and y^2 belong to $H^0(6.o)$ and not to $H^0(5.o)$. It follows that the 5 functions $1, x, y, x^2, xy$, form a basis for $H^0(5.o)$, and that they supplemented with any of the two functions y^2 or x^3 form a basis of $H^0(6.o)$. Hence, there is a relation,

$$y^2 = ax^3 + bxy + cx^2 + dy + ex + f,$$

with $a \neq 0$. Now, in the choice of x and y, we could replace x by $\alpha x + \beta$ for $\alpha \neq 0$, and we could replace y by $\gamma y + \delta x + \varepsilon$ for $\gamma \neq 0$. It is not hard to see that with suitable replacements, we can obtain an equation in the Weierstrass normal form,

$$y^2 = 4x^3 - g_2 x - g_3, \tag{4.5.1}$$

and, moreover, the coefficients g_2 and g_3 are uniquely determined up to a choice of the form $g_2 \mapsto \lambda^4 g_2$ and $g_3 \mapsto \lambda^6 g_3$. The two functions x and y are holomorphic on the complement of o. Hence they define a map from $X - \{o\}$ into the subset of $\mathbf{C}^2$ defined by the equation (4.5.1).

The function y has a pole of order 3 at o and no more poles. Therefore, the function y has three zeros p_1, p_2, p_3, not necessarily different. By the property of Lemma II, in the group of X,

$$p_1 + p_2 + p_3 = o\,. \tag{4.5.2}$$

Consider the polynomial on the right side of Equation (4.5.1). It follows from the equation that the value $\lambda_1 := x(p_1)$ is a root of the polynomial. The function $x - \lambda_1$ has a pole of order 2 at o and no more poles. Hence it has two zeros, one of which is of course p_1. Let p_1' be the other zero (not necessarily different form p_1). Then, by the property of Lemma II,

$$p_1 + p_1' = o\,. \tag{4.5.3}$$

By comparing (4.5.3) and (4.5.2), it follows that p_1' is different from p_2 and p_3. On the other hand, since $\lambda_1 = x(p_1')$ is root of the polynomial, it follows from (4.5.2) that p_1' is a zero of y. Thus p_1' is one of the points p_1, p_2, p_3 and since p_1' is different from p_2 and p_3, we have necessarily $p_1' = p_1$. Hence p_1 is different from p_2 and p_3. We conclude that the 3 points p_1, p_2 and p_3 are different and that the function x takes 3 different values on these three points. As the latter values are roots of the polynomial, it follows that the discriminant $g_2^3 - 27g_3^2$ of the polynomial is non-zero.

Hence the elliptic curve $X(g_2, g_3)$ is defined, and we have obtained a map from X to $X(g_2, g_3)$. It follows as in the proof of (3.6) that the map is bijective. Hence it is an isomorphism of Riemann surfaces and of elliptic curves.

(4.6) Note. Given an elliptic curve (X, o). By (4.5), (X, o) is isomorphic to the elliptic curve $X(g_2, g_3)$. Moreover, (g_2, g_3) are unique up to the transformation defined in (4.5). It follows that the following complex number,

$$j(X, o) := \frac{12^3 g_2^3}{g_2^3 - 27g_3^2}\,.$$

is a well defined invariant of the elliptic curve. The invariant characterizes the elliptic curve, that is, two curves (X, o) and (X', o') are isomorphic if and only if $j(X, o) = j(X', o')$. Indeed, assume that the two invariants are the same. Then, by (3.7), with an obvious notation, we have the equations $g_2' = \lambda^4 g_2$ and $g_3' = \lambda^6 g_3$ with a non-zero scalar λ. Hence the two curves are isomorphic.

As noted in (4.1), every elliptic curve $X(g_2, g_3)$ has a parameterization $x = \wp(u)$, $y = \wp'(u)$ with a Weierstrass $\wp$-function defined by a suitable lattice Ω. Therefore, it follows from (4.5) that every abstract elliptic curve (X, o) is isomorphic to torus $\mathbf{C}/\Omega$.

(4.7) Exercise. Prove the addition formula for the $\wp$-function: Let u, v and w be complex numbers in the complement of Ω. Assume that $u + v + w$ belongs to Ω. Then,

$$\begin{vmatrix} \wp(u) & \wp'(u) & 1 \\ \wp(v) & \wp'(v) & 1 \\ \wp(w) & \wp'(w) & 1 \end{vmatrix} = 0.$$

I. Index.

Special symbols:

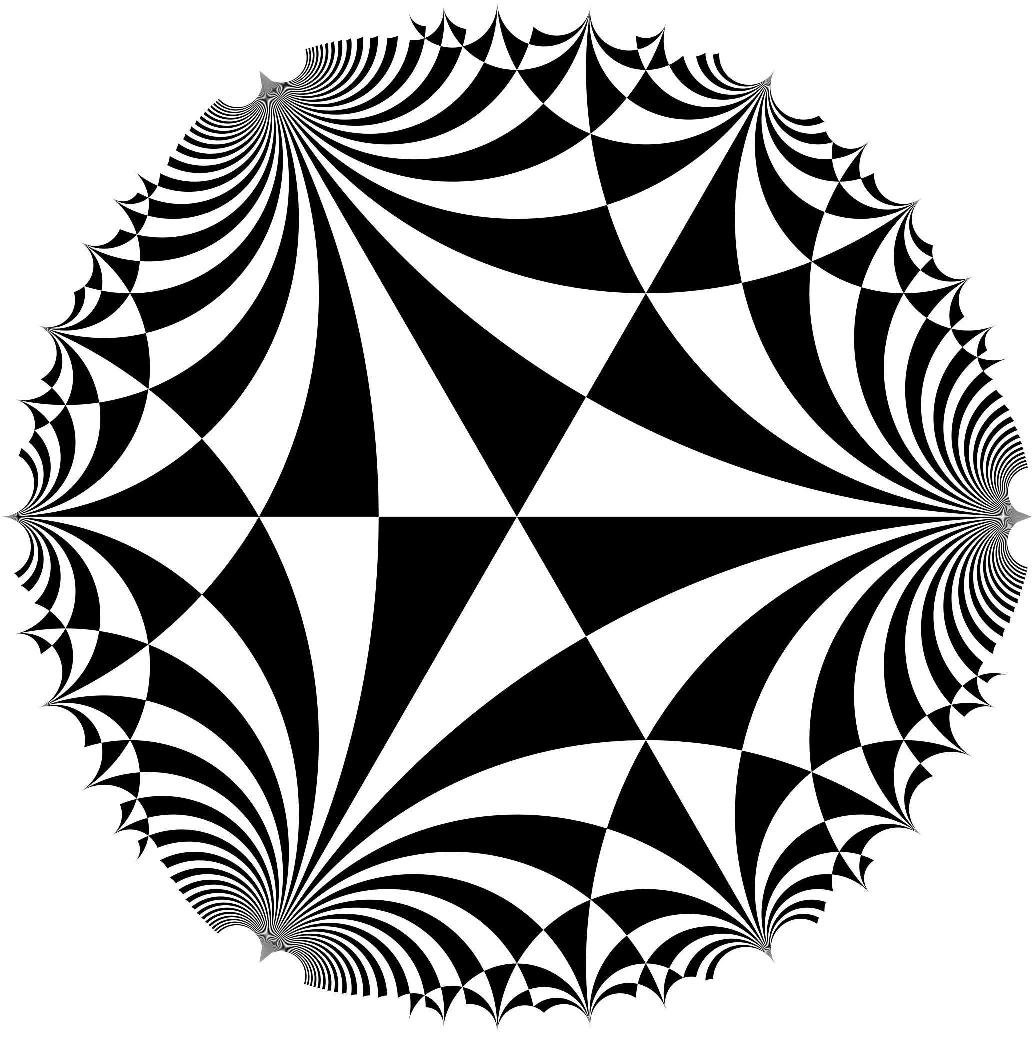


Image of front page under the Cayley tranformation.