Chapter 2
Arithmetic Hyperbolic Surfaces

As we recalled in the introduction, there are several ways to construct Fuchsian groups of the first kind. Of all of these groups, the most important from the number theoretic viewpoint are the arithmetic groups. The general definition of these groups is a bit technical so we content ourselves for the moment with describing a family of examples: the arithmetic groups coming from a quaternion algebra over $\mathbf{Q}$. Before describing them, we begin by proving several general results concerning the space of lattices.

2.1 The Space of Lattices

A *lattice* in $\mathbf{R}^n$ is a discrete subgroup isomorphic to $\mathbf{Z}^n$. In this section we study the set $\mathcal{L}_n$ of lattices in $\mathbf{R}^n$. We endow $\mathbf{R}^n$ with the Euclidean norm $|\cdot|$ for which the canonical basis $(e_1, \ldots, e_n)$ is orthonormal.

We define two fundamental invariants of a lattice $L \in \mathcal{L}_n$:

1. its *height* $H = H(L) = \min_{v \in L - \{0\}} \{|v|\}$, and
2. its *volume* $V = V(L)$, i.e., the Euclidean volume of the paralleletope subtended by a basis of L.

The group $\mathrm{GL}(n, \mathbf{R})$ acts transitively (on the left) on the set $\mathcal{L}_n$. Denote by $L_0 = \mathbf{Z}^n$ the standard lattice in $\mathbf{R}^n$ generated by the canonical basis. A lattice L generated by n (linearly independent) vectors $v_1, \ldots, v_n$ is the image of L_0 under the action of the invertible matrix whose columns are the v_i. The stabilizer of the $\mathbf{Z}$-module L_0 in $\mathrm{GL}(n, \mathbf{R})$ is the subgroup $\mathrm{GL}(n, \mathbf{Z})$ consisting of matrices that are invertible in $\mathcal{M}_n(\mathbf{Z})$. Note that the determinant of such a matrix is necessarily invertible in $\mathbf{Z}$ and is therefore equal to ± 1. Finally, the set $\mathcal{L}_n$ is naturally identified with the quotient $\mathrm{GL}(n, \mathbf{R})/\mathrm{GL}(n, \mathbf{Z})$. This identification induces a topology on $\mathcal{L}_n$: the quotient topology. By definition, a sequence L_m of lattices in $\mathbf{R}^n$ converges, in the quotient

N. Bergeron, *The Spectrum of Hyperbolic Surfaces*, Universitext,
DOI 10.1007/978-3-319-27666-3_2

topology, towards a lattice L of $\mathbf{R}^n$ if and only if there exists a basis $(f_1^m, \ldots, f_n^m)$ of L_m which converges towards a basis $(f_1, \ldots, f_n)$ of L. The lattice L_0 corresponds to the identity class in $\mathrm{GL}(n, \mathbf{R})/\mathrm{GL}(n, \mathbf{Z})$; it will serve as a convenient base point for the space $\mathcal{L}_n$. Recall that the determinant of n vectors $v_1, \ldots, v_n$ relative to the canonical basis is equal to the Euclidean volume of the paralleletope they subtend. Thus, if $L = g(L_0)$ with $g \in \mathrm{GL}(n, \mathbf{R})$, we have $V(L) = |\det g|$.

The set $\mathcal{L}_n$ is not compact: the volume can explode or the height can tend towards 0. The following theorem tells us that these are the only two ways that a family of lattices can degenerate.

Theorem 2.1 (Hermite-Mahler criterion) *A subset $M \subset \mathcal{L}_n$ is relatively compact if and only if there exist constants $\varepsilon > 0$ and $C > 0$ such that*

$$\begin{cases} H > \varepsilon \\ V < C \end{cases}$$

on M.

The main idea of the proof of Theorem 2.1 is to bring M into a fundamental domain for the action of $\mathrm{GL}(n, \mathbf{Z})$ on $\mathrm{GL}(n, \mathbf{R})$, and to study the ways of degenerating in this fundamental domain. When $n = 2$ the construction of such a fundamental domain can be deduced from that of the modular group acting on $\mathcal{H}$. In higher dimensions, the explicit knowledge of a fundamental domain is replaced by the Siegel sets that we now construct.

For g in $\mathrm{GL}(n, \mathbf{R})$, we write g_{ij} for the matrix entries of g. Let

$$\begin{aligned}
K &:= \mathrm{O}(n) = \{g \in \mathrm{GL}(n, \mathbf{R}) \mid {}^t g g = 1\}, \\
A &:= \{g \in \mathrm{GL}(n, \mathbf{R}) \mid g \text{ is diagonal with positive entries}\}, \\
A_s &:= \{a \in A \mid a_{ii} \leqslant s a_{i+1,i+1} \text{ for } i = 1, \ldots, n-1\} \quad \text{with } s \geqslant 1, \\
N &:= \{g \in \mathrm{GL}(n, \mathbf{R}) \mid g - I_n \text{ is strictly upper triangular}\} \text{ and}, \\
N_t &:= \{n \in N \mid |n_{ij}| \leqslant t \text{ for } 1 \leqslant i < j \leqslant n\} \quad \text{with } t \geqslant 0.
\end{aligned}$$

According to the *Iwasawa decomposition* of $\mathrm{GL}(n, \mathbf{R})$ – which boils down to the orthonormalization procedure of Schmidt – the map

$$(k, a, n) \longmapsto kan$$

is a homeomorphism of $K \times A \times N$ onto $\mathrm{GL}(n, \mathbf{R})$. We denote by $S_{s,t}$ the *Siegel domain*

$$S_{s,t} = K A_s N_t$$

and

$$\Gamma = \mathrm{SL}(n, \mathbf{Z}).$$

One knows that N is a closed subgroup of $\mathrm{GL}(n, \mathbf{R})$, homeomorphic to $\mathbf{R}^m$ ($m = n(n-1)/2$) by the map $\theta : n \mapsto (n_{ij})_{1 \leqslant i < j \leqslant n}$; as a result, N_t is compact.

We would like to prove the following

Lemma 2.2 *One has*

$$\mathrm{GL}(n, \mathbf{R}) = S_{s,t}\Gamma$$

as soon as $s \geqslant 2/\sqrt{3}$, $t \geqslant 1/2$.

Proof Let g be in $\mathrm{GL}(n, \mathbf{R})$ and $L = g(L_0)$. We shall proceed by induction on n.

We begin by remarking that given a linear subspace V of Euclidean space $\mathbf{R}^n$, the quotient space $\mathbf{R}^n/V$ can be identified with the orthogonal of V in $\mathbf{R}^n$ and, as such, is itself naturally a Euclidean vector space. One can then define, by induction on n, the notion of an admissible family of vectors in L. This is a family $(f_1, \ldots, f_n)$ of vectors in L such that

- f_1 is a vector in $L - \{0\}$ of minimal norm,
- the images $\dot{f}_2, \ldots, \dot{f}_n$ of $f_2, \ldots, f_n$ in the lattice $\dot{L} := L/\mathbf{Z}f_1$ of the Euclidean space $\mathbf{R}^n/\mathbf{R}f_1$ form an admissible family of $\dot{L}$, and
- the vectors f_i are of minimal norm among the vectors of L whose image in $\dot{L}$ is $\dot{f}_i$.

(When $n = 1$, we pay no attention to the last two conditions.)

It is a standard fact that the $\mathbf{Z}$-module L admits an admissible family $(f_1, \ldots, f_n)$ and that this family is in fact a $\mathbf{Z}$-basis. Multiplying g on the right by an element of Γ, if necessary, we can assume that for every $i = 1, \ldots, n$ we have $ge_i = f_i$.

We show by induction on n that if $g \in \mathrm{GL}(n, \mathbf{R})$ sends the canonical basis of $\mathbf{R}^n$ to an admissible family of vectors of the lattice $g(L_0)$ then $g \in S_{2/\sqrt{3},1/2}$.

Write $g = kan$. As $(k^{-1}f_1, \ldots, k^{-1}f_n)$ is an admissible basis of $k^{-1}L$, we can assume that $k = I_n$. One then has $g = an$, so that

$$\begin{aligned} f_1 &= a_{11}e_1 \\ f_2 &= a_{22}e_2 + a_{11}n_{12}e_1 \\ &\cdots \\ f_i &= a_{ii}e_i + a_{i-1,i-1}n_{i-1,i}e_{i-1} + \cdots + a_{11}n_{1i}e_1. \end{aligned}$$

Denote by $\mathbf{R}^{n-1}$ the subspace of $\mathbf{R}^n$ orthogonal to $\mathbf{R}f_1 = \mathbf{R}e_1$. This is the Euclidean subspace of $\mathbf{R}^n$ generated by $e_2, \ldots, e_n$. The linear transformation $g : \mathbf{R}^n \to \mathbf{R}^n$ defines on the quotient by $\mathbf{R}f_1$ a linear transformation $\overline{g} : \mathbf{R}^{n-1} \to \mathbf{R}^{n-1}$ which, by the definition of an admissible family, sends the canonical basis $(e_2, \ldots, e_n)$ to an admissible family of the quotient lattice $L/\mathbf{Z}f_1$ obtained by the orthogonal projection of L onto $\mathbf{R}^{n-1}$. By the induction hypothesis, we therefore

have

$$|n_{ij}| \leqslant 1/2 \quad \text{for } 2 \leqslant i < j \leqslant n$$

and

$$a_{ii} \leqslant 2/\sqrt{3}\, a_{i+1,i+1} \quad \text{for } 2 \leqslant i \leqslant n-1.$$

It remains to show

$$|n_{1j}| \leqslant 1/2 \quad \text{for } 2 \leqslant j \leqslant n \text{ and } a_{11} \leqslant 2/\sqrt{3} a_{22}.$$

The first inequality is a consequence of

$$|f_j|^2 \leqslant |f_j + pf_1|^2, \quad \forall p \in \mathbf{Z}.$$

By subtracting the left-hand side from the expanded right-hand side and factoring out by a_{11}^2 (which is non-zero!), we deduce that for all $p \in \mathbf{Z}$,

$$p^2 + 2n_{1,j}p \geqslant 0.$$

This forces n_{1j} to be less than $1/2$ in absolute value.

The second inequality comes from $|f_1|^2 \leqslant |f_2|^2$, as this, when written out, is $a_{11}^2 \leqslant a_{22}^2 + a_{11}^2 n_{12}^2 \leqslant a_{22}^2 + 1/4 a_{11}^2$. □

Proof of Theorem 2.1 Fix $s \geqslant 2/\sqrt{3}$ and $t \geqslant 1/2$. It is clear that a subset M of $\mathcal{R}$ is relatively compact if and only if there exists a compact subset S of $S_{s,t}$ such that $M \subset S \cdot L_0 := \{gL_0 \mid g \in S\}$.

We first show the forward direction of the equivalence stated in Theorem 2.1. Let us fix $0 < r < R$ such that for every $g = kan$ in S and for every $i = 1, \ldots, n$ we have $r \leqslant a_{ii} \leqslant R$. Then

$$|\det g| = \prod_{i=1}^{n} a_{ii} \leqslant R^d$$

and

$$\min_{v \in \mathbf{Z}^n - \{0\}} |gv| \geqslant r,$$

for if $v = \sum_{i=1}^{\ell} m_i e_i \in \mathbf{Z}^n$ with $m_\ell \neq 0$, then

$$|gv| \geqslant |\langle ke_\ell, gv\rangle| = |\langle e_\ell, anv\rangle| = a_{\ell\ell}|m_\ell| \geqslant r.$$

We now show the converse direction. Let $S = \{g \in S_{s,t} \mid gL_0 \in \overline{M}\}$. It follows from Lemma 2.2 that $\overline{M}$ is contained in $S \cdot L_0$. But for every $g = kan$ in S we have

- $a_{11} = |ge_1| \geqslant \varepsilon$,
- $a_{ii} \leqslant s a_{i+1,i+1}$ for $i = 1, \ldots, n-1$,
- $\prod_{i=1}^{n} a_{ii} \leqslant C$.

We deduce that there exists $R > r > 0$ such that, for every $g = kan$ in S and for all $i = 1, \ldots, n$, we have $r \leqslant a_{ii} \leqslant R$. Thus S is compact, as is $\overline{M}$. □

2.2 Quaternion Algebras and Arithmetic Groups

Let F be an arbitrary field and let a and b be two non-zero elements of F. The corresponding *quaternion algebra* over F is the ring

$$D_{a,b}(F) = \{x_0 + x_1 i + x_2 j + x_3 k \mid x_0, \ldots, x_3 \in F\},$$

where

- addition is defined in the obvious way, so as to form a vector space of dimension 4 over F;
- the map $x \mapsto x + 0i + 0j + 0k$ is an (injective) ring homomorphism from F into $D_{a,b}(F)$ whose image – again written F – is contained in the center of $D_{a,b}(F)$. In other words,

$$x\alpha = \alpha x, \qquad \text{for all } x \in F,\ \alpha \in D_{a,b}(F);$$

- multiplication is determined by the relations

$$i^2 = a,\ j^2 = b,\ ij = k = -ji.$$

The *reduced norm* of $\alpha = x_0 + x_1 i + x_2 j + x_3 k \in D_{a,b}(F)$ is

$$\mathrm{N}_{\mathrm{red}}(\alpha) = x_0^2 - a x_1^2 - b x_2^2 + ab x_3^2. \tag{2.1}$$

The *conjugate* of α is

$$\overline{\alpha} = x_0 - x_1 i - x_2 j - x_3 k,$$

so that $\mathrm{N}_{\mathrm{red}}(\alpha) = \alpha\overline{\alpha} = \overline{\alpha}\alpha$. One defines the *trace* of α by

$$\mathrm{tr}(\alpha) = \alpha + \overline{\alpha} = 2x_0. \tag{2.2}$$

The terminology naturally comes from the famous example $D_{-1,-1}(\mathbf{R})$ of Hamilton's quaternions. On the other hand the map $D_{1,1}(\mathbf{R}) \to \mathcal{M}_2(\mathbf{R})$, which on basis vectors is defined by

$$1 \longmapsto I_2, \quad i \longmapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad j \longmapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad k \longmapsto \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

extends in a unique way to a ring isomorphism. This will be justified at the beginning of the proof of Theorem 2.3.

One calls a quaternion algebra D a *division algebra* if every non-zero element $\alpha \in D$ admits an inverse. This happens if and only if $\mathrm{N}_{\mathrm{red}}(\alpha) \neq 0$ for all $\alpha \neq 0$, in which case $\alpha^{-1} = \overline{\alpha}/\mathrm{N}_{\mathrm{red}}(\alpha)$.

Henceforth we fix two positive integers a and b. We can then naturally speak of the subring $\mathcal{O} := D_{a,b}(\mathbf{Z})$ in $D_{a,b}(\mathbf{Q})$.

Our goal is to prove the following theorem.

Theorem 2.3

1. *There exists a group isomorphism from*

$$D_{a,b}(\mathbf{R})^1 := \mathrm{SL}(1, D_{a,b}(\mathbf{R})) = \{g \in D_{a,b}(\mathbf{R}) \mid \mathrm{N}_{\mathrm{red}}(g) = 1\}$$

   to $G = \mathrm{SL}(2, \mathbf{R})$.
2. *The image* $\Gamma_{a,b}$ *in* G *of the group* $\mathcal{O}^1 = \mathrm{SL}(1, D_{a,b}(\mathbf{Z}))$ *via this isomorphism is a Fuchsian group of the first kind.*
3. *The following statements are equivalent:*
   a. $\Gamma_{a,b}$ *is cocompact in* G*;*
   b. $(0,0,0)$ *is the unique solution in integers of the Diophantine equation* $x^2 - ay^2 - bz^2 = 0$*;*
   c. $D_{a,b}(\mathbf{Q})$ *is a division algebra.*
4. *Two subgroups* $\Gamma_{a,b}$ *and* $\Gamma_{a',b'}$ *are commensurable in* G *if and only if the quadratic forms* $x^2 - ay^2 - bz^2$ *and* $x^2 - a'y^2 - b'z^2$ *are similar over* $\mathbf{Q}$.

Two subgroups Γ and Λ in G are said to be *commensurable* in G if, conjugating Γ in G if necessary, the group $\Gamma \cap \Lambda$ is of finite index in both Γ and Λ. Recall furthermore that two quadratic forms over a field of characteristics $\neq 2$ are said to be *equivalent* if a change of basis over the field transforms one into the other and are said to be *similar* if they are equivalent up to a non-zero scalar multiple (in the base field).

According to Property 4 of Theorem 2.3, there exist infinitely many commensurability classes of discrete cocompact subgroups in G. Indeed, if p and q are two distinct prime numbers congruent to -1 modulo 4, the groups $\Gamma_{p,p}$ and $\Gamma_{q,q}$ are not commensurable, for otherwise there would exist $M \in \mathrm{GL}(3, \mathbf{Q})$ and $\lambda \in \mathbf{Q}$ such that $\Delta' = \lambda\, {}^tM \Delta M$, with $\Delta = \mathrm{diag}\{1, -p, -p\}$, $\Delta' = \mathrm{diag}\{1, -q, -q\}$. We can therefore assume that $\det M = \pm q/p$ (and $\lambda = 1$). Let $p^\alpha n$ be the least common multiple of

the denominators of the coefficients of M; we have $p^\alpha nM = (m_{ij})$ with $m_{ij} \in \mathbf{Z}$, and $\alpha \geqslant 1$. The first equation obtained in identifying the coefficients of Δ' and $\lambda\, {}^t M \Delta M$ is

$$(np^\alpha)^2 = m_{11}^2 - pm_{21}^2 - pm_{31}^2.$$

Thus $m_{11} \equiv 0 \pmod{p}$ and $m_{21}^2 + m_{31}^2 \equiv 0 \pmod{p}$. Since -1 is not a square modulo p, we necessarily have $m_{21} \equiv m_{31} \equiv 0 \pmod{p}$. Likewise, all of the coefficients of $p^\alpha nM$ are divisible by p. Therefore $p^{\alpha-1} nM \in \mathcal{M}_3(\mathbf{Z})$, in contradiction with the choice of $p^\alpha n$.

Proof of Theorem 2.3 The linear map $\Phi : D_{a,b}(\mathbf{R}) \to \mathcal{M}_2(\mathbf{R})$ defined on basis vectors by

$$\Phi(1) = I_2,\ \Phi(i) = \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix},\ \Phi(j) = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix},\ \Phi(k) = \begin{pmatrix} 0 & \sqrt{a} \\ -b\sqrt{a} & 0 \end{pmatrix}$$

is bijective. Moreover, Φ preserves multiplication, making it a ring isomorphism. Finally, if $\alpha = x_0 + x_1 i + x_2 j + x_3 k \in D_{a,b}(\mathbf{R})$, we have

$$\begin{aligned} \det(\Phi(\alpha)) &= (x_0 + x_1\sqrt{a})(x_0 - x_1\sqrt{a}) - (x_2 + x_3\sqrt{a})(bx_2 - bx_3\sqrt{a}) \\ &= x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = \mathrm{N}_{\mathrm{red}}(\alpha). \end{aligned}$$

Thus $\Phi(D_{a,b}(\mathbf{R})^1) = \mathrm{SL}(2, \mathbf{R})$, the first claim of the theorem is proved, and

$$\Gamma_{a,b} = \left\{ \begin{pmatrix} x_0 + x_1\sqrt{a} & x_2 + x_3\sqrt{a} \\ b(x_2 - x_3\sqrt{a}) & x_0 - x_1\sqrt{a} \end{pmatrix} \;\middle|\; \begin{array}{l} x_0, x_1, x_2, x_3 \in \mathbf{Z}, \\ x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 1 \end{array} \right\}.$$

The proofs of the other claims will be taken up in the following subsections. □

2.2.1 An Exceptional Isomorphism

From now on we denote by $D_{a,b}$ the real quaternion algebra $D_{a,b}(\mathbf{R})$, remembering that this algebra is defined over $\mathbf{Q}$ (and even over $\mathbf{Z}$) and that we can therefore speak of its rational (or integral) points. Let $P = \{\alpha \in D_{a,b} \mid \mathrm{tr}(\alpha) = 0\}$ be the set of pure quaternions in $D_{a,b}$. This is a subspace of $D_{a,b}$ defined over $\mathbf{Q}$ and isomorphic to $\mathbf{R}^3$. The reduced norm restricts to P as the quadratic form $q = -ax_1^2 - bx_2^2 + abx_3^2$.

Lemma 2.4 *There is a rational isomorphism*

$$D_{a,b}(\mathbf{Q})^*/\mathbf{Q}^* \xrightarrow{\simeq} \mathrm{SO}(q, \mathbf{Q})$$

which induces an isomorphism on the real groups[1]

$$D^1_{a,b}/\{\pm 1\} \longrightarrow \mathrm{SO}_0(q).$$

Proof It suffices to construct the first isomorphism. For that we consider the action of the group $D_{a,b}(\mathbf{Q})^*$ on $D_{a,b}(\mathbf{Q})$ by interior automorphisms. For $\alpha \in D_{a,b}(\mathbf{Q})^*$ and $\beta \in D_{a,b}(\mathbf{Q})$ we write

$$S_\alpha(\beta) = \alpha\beta\alpha^{-1}.$$

We verify immediately the following properties:

- S_α preserves the reduced norm,
- $(S_\alpha)_{|\mathbf{Q}} = \mathrm{Id}_{\mathbf{Q}}$, and
- $S_\alpha(P) = P$.

Let $s_\alpha = (S_\alpha)_{|P}$. Then $s : D_{a,b}(\mathbf{Q})^* \to \mathrm{O}(q)$ is a homomorphism whose kernel is $\mathbf{Q} \cap D_{a,b}(\mathbf{Q})^* = \mathbf{Q}^*$.

Before continuing we make a few preliminary calculations. For $\beta \in P$ such that $q(\beta) \neq 0$, the reflection τ_β associated with β is given by the formula

$$\tau_\beta(x) = x - 2\,\frac{q(x,\beta)}{q(\beta,\beta)}\,\beta \quad (x \in P).$$

From the definition of $q = (\mathrm{N}_{\mathrm{red}})_{|P}$ we deduce that

$$\tau_\beta(x) = x - \frac{x\overline{\beta} + \beta\overline{x}}{\beta\overline{\beta}}\,\beta = -\beta\overline{x}\,\overline{\beta}^{-1},$$

and since x and β are both pure quaternions we find

$$\tau_\beta(x) = -\beta x \beta^{-1} \quad (x \in P).$$

Already we see that if σ_β is the rotation of angle π with axis β, we have $\sigma_\beta = -\tau_\beta = s_\beta$.

Let $u \in \mathrm{O}(q, \mathbf{Q})$. We can write u as a product of reflections, $u = \tau_{\beta_1} \cdots \tau_{\beta_r}$ for $\beta_1, \ldots, \beta_r \in P$, with $q(\beta_i) \neq 0$, the integer r being even if u is in SO and odd otherwise. We therefore have

$$u(x) = (-1)^r \beta_1 \cdots \beta_r x (\beta_1 \cdots \beta_r)^{-1} \quad \text{for all } x \in P.$$

[1] We denote by $\mathrm{SO}_0(q) \cong \mathrm{SO}_0(2,1)$ the connected component of the identity of the special orthogonal group of the quadratic form q on P.

We now continue with the argument and show that s takes values in $\mathrm{SO}(q)$. Let $\alpha \in D_{a,b}(\mathbf{Q})^*$ and, arguing by contradiction, suppose that $s_\alpha \in \mathrm{O}_3^-(q)$. Under this hypothesis we would then have, for every $x \in P$,

$$s_\alpha(x) = \alpha x \alpha^{-1} = -\alpha' x (\alpha')^{-1}, \quad \text{where } \alpha' = \beta_1 \cdots \beta_r.$$

In other words, $-x = \beta x \beta^{-1}$, for every $x \in P$, with $\beta = \alpha^{-1}\alpha'$. As $\beta x \beta^{-1} = x$ for every $x \in \mathbf{Q}$, we would then have, for every x in $D_{a,b}(\mathbf{Q})$: $\overline{x} = \beta x \beta^{-1}$, with $\beta \in D_{a,b}(\mathbf{Q})^*$. But $x \mapsto \overline{x}$ is an anti-automorphism and $x \mapsto \beta x \beta^{-1}$ an automorphism – contradiction.

Summarizing, we have a homomorphism $s : D_{a,b}(\mathbf{Q})^* \to \mathrm{SO}(q, \mathbf{Q})$, and since the rotations of angle π are in the image and these generate the special orthogonal group, s is surjective. □

Henceforth we denote the isomorphism of Lemma 2.4[2] by

$$\Psi : D^1_{a,b}/\{\pm 1\} \longrightarrow \mathrm{SO}_0(q).$$

The group $\mathrm{SO}_0(q)$ is naturally embedded in $\mathrm{GL}(3, \mathbf{R})$ (take the natural coordinates (x_1, x_2, x_3) of P). We let $\mathrm{SO}_0(q, \mathbf{Z})$ be the intersection $\mathrm{SO}_0(q) \cap \mathrm{GL}(3, \mathbf{Z})$. The proof of Lemma 2.4 implies that

$$\mathcal{O}^1/\{\pm 1\} = \Psi^{-1}(\mathrm{SO}_0(q, \mathbf{Z})). \tag{2.3}$$

The subgroup $\mathcal{O}^1$ is therefore discrete in $D^1_{a,b}$ and its image, $\Gamma_{a,b}$, by Φ is equally a discrete subgroup of G.

Let g_{ij}, $1 \leqslant i, j \leqslant 3$, be the matrix entries of an element $g \in \mathrm{O}(q) \subset \mathrm{GL}(3, \mathbf{R})$. By a *rational representation* of the linear group $\mathrm{O}(q)$ on $\mathbf{R}^n$ we mean a homomorphism $\rho : \mathrm{O}(q) \to \mathrm{GL}(n, \mathbf{R})$ such that for every pair of integers $(\mu, \nu) \in [1, n]^2$, there exists a polynomial $P_{\mu,\nu} \in \mathbf{Q}[X_{ij} | 1 \leqslant i, j \leqslant 3]$ such that for every $g \in \mathrm{O}(q)$,

$$\rho(g)_{\mu,\nu} = P_{\mu,\nu}(g_{ij}),$$

where $\rho(g)_{\mu,\nu}$ denotes the (μ, ν) matrix entry of $\rho(g) \in \mathrm{GL}(n, \mathbf{R})$.

Lemma 2.5 *Let ρ be a rational representation of the linear group $\mathrm{O}(q)$ on $\mathbf{R}^n$. Then there exists a finite index subgroup of $\mathrm{O}(q, \mathbf{Z})$ which, via the action induced by ρ, leaves invariant the standard lattice $\mathbf{Z}^n$.*

[2] For $a = b = 1$ the map obtained in composing Ψ and Φ^{-1} induces an isomorphism between $\mathrm{PSL}(2, \mathbf{R})$ and $\mathrm{SO}_0(2, 1)$. This is one of the isomorphisms – called *exceptional* – which exist only between certain "small" Lie groups (see [57, pp. 518–520]).

Proof The coefficients of the polynomials $P_{\mu,\nu}$ defined by $\rho(g)_{\mu,\nu} = P_{\mu,\nu}(g_{ij})$ are rational. The same is true for the polynomials $Q_{\mu,\nu}$ defined by

$$\rho(g)_{\mu,\nu} - \delta_{\mu,\nu} = Q_{\mu,\nu}(g_{ij} - \delta_{ij}).$$

But the latter have vanishing constant term, for $\rho(I_3) = I_n$. Thus if m is the common denominator of all their coefficients, we have $Q_{\mu,\nu}(g_{ij}-\delta_{ij}) \in \mathbf{Z}$ as soon as $g_{ij}-\delta_{ij} \equiv 0 \pmod{m}$, i.e., $g \equiv I_3 \pmod{m}$. For such a choice of g the coefficients $\rho(g)_{\mu,\nu}$ are integral, which assures the invariance of $\mathbf{Z}^n$ under $\rho(g)$. Finally, these elements form a finite index subgroup of $\mathrm{O}(q, \mathbf{Z})$ since the quotient injects in $\mathrm{GL}(n, \mathbf{Z}/m\mathbf{Z})$. □

2.2.2 A Subset of the Space of Lattices

The map Ψ passes to the quotient as an embedding (continuous for the quotient topologies) of $D^1_{a,b}/\mathcal{O}^1$ into the space of lattices $\mathcal{L}_3 = \mathrm{GL}(3, \mathbf{R})/\mathrm{GL}(3, \mathbf{Z})$. Denote by M the image of this embedding. The subset M is also the orbit of the standard lattice L_0 under the action by left multiplication of the subgroup $\mathrm{SO}_0(q) \subset \mathrm{GL}(3, \mathbf{R})$ on $\mathcal{L}_3$.

Lemma 2.6 *The subset M is closed in* $\mathcal{L}_3$.

Proof Let L be a lattice in $\mathcal{L}_3$ and assume that there exists a sequence $\{L_i\}$ of lattices in M which converge to L. Each L_i is the image by an element of the group $\mathrm{SO}_0(q)$ of the standard lattice $L_0 = \mathbf{Z}^3$. The quadratic form q therefore takes integer values on each one. Since it is continuous, the form q must also take integer values on L.

Consider now a basis (e_1, e_2, e_3) of L. Since each convergent sequence of integers is stationary, there must exist a lattice L_{i_0} in the sequence $\{L_i\}$ with a basis (f_1, f_2, f_3) sufficiently close to the basis (e_1, e_2, e_3) such that

- $q(f_i) = q(e_i)$ for $i = 1, 2, 3$, and
- $q(f_i \pm f_j) = q(e_i \pm e_j)$ for $i, j = 1, 2, 3$.

We thus have $B(e_i, e_j) = B(f_i, f_j)$, where B is the underlying symmetric bilinear form of the quadratic form q. Taking $g \in \mathrm{GL}(3, \mathbf{R})$ to satisfy $g(e_i) = f_i$, we see that the lattice L_{i_0} can be sent to L by a matrix in $\mathrm{GL}(3, \mathbf{R})$ which leaves the quadratic form q invariant. It follows that the lattice L is contained in the orbit of the standard lattice under the action of the group $\mathrm{O}(q)$. The group $\mathrm{SO}_0(q)$ being of finite index (equal to 4) in $\mathrm{O}(q)$, we conclude that $L \in M$. □

2.2.3 The Cocompactness Criterion

The reduced norm being multiplicative, the algebra $D_{a,b}(\mathbf{Q})$ is a division algebra if and only if $(0, 0, 0, 0)$ is the unique integral solution of the Diophantine equation

$x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 0$. In this case we shall show that the subset M of $\mathcal{L}_3$ is compact.

The Subset M Is Compact

According to the preceding subsection, it suffices to show that M is relatively compact. For this we shall apply the Hermite-Mahler criterion.

Let us begin by noticing that every lattice L in M is of volume 1. It suffices therefore to verify that the height is bounded uniformly from below on M. Now by hypothesis $(0, 0, 0)$ is the only integral solution to the Diophantine equation $q(x_1, x_2, x_3) = 0$. The set

$$U = \{x \in \mathbf{R}^3 \mid |q(x)| < 1\}$$

is therefore an open neighborhood of 0 in $\mathbf{R}^3$ intersecting the standard lattice $L_0 = \mathbf{Z}^3$ only at the origin. But M is equal to the orbit of L_0 in $\mathcal{L}_3$ under the action of the group $\mathrm{SO}_0(q)$, and since $\mathrm{SO}_0(q)$ preserves the quadratic form q, the latter takes the same set of values on each of the lattices contained in M. The intersection of the open set U with any lattice belonging to M is therefore always reduced to the point $0 \in \mathbf{R}^3$. In other words, the height stays uniformly bounded from below on M. We may then apply the criterion of Hermite-Mahler to deduce that the subset M of $\mathcal{L}_3$ is relatively compact.

We have thus shown that the lattice $\Gamma_{a,b}$ is cocompact in G whenever $D_{a,b}(\mathbf{Q})$ is a division algebra.

Proof of the converse direction Suppose now that there exists an integral solution $\neq (0, 0, 0)$ to the Diophantine equation $q(x_1, x_2, x_3) = -ax_1^2 - bx_2^2 + abx_3^2 = 0$. Let us show that the quadratic form q is similar over $\mathbf{Q}$ to the quadratic form $-x_1^2 - x_2^2 + x_3^2$.

Indeed, there exists $w \in P(\mathbf{Q})$ – a vector of P with rational coordinates – such that the orthogonal complement $w^\perp$ of w with respect to q contains a non-zero q-isotropic rational vector u. Write $\lambda = -1/q(w)$ and $q' = \lambda q$. We have $q'(w) = -1$ and $q'(u) = 0$. Since $q_{|w^\perp}$ is non-degenerate, there exists a rational vector $v \in w^\perp$ such that $\mu := q'(u, v) \neq 0$. Note that the expression

$$q'(v + tu) = q'(v) + 2tq'(u, v)$$

is linear in t and has a zero $t \in \mathbf{Q}$. Replacing v by $v + tu$ if necessary, we can suppose that $q(v) = 0$. If x and y are in $\mathbf{Q}$,

$$q'(xu + yv) = 2\mu xy = \left(\frac{x + 2\mu y}{2}\right)^2 - \left(\frac{x - 2\mu y}{2}\right)^2.$$

In the basis $(w, u - v/(2\mu), u + v/(2\mu))$ the quadratic form q' is therefore equal to $-x_1^2 - x_2^2 + x_3^2$.

Lemma 2.7 *If $D_{a,b}(\mathbf{Q})$ is not a division algebra, the group $\Gamma_{a,b}$ is commensurable with the group* $\mathrm{SL}(2,\mathbf{Z})$. *In particular, it is Fuchsian of the first kind and non-cocompact.*

Proof Assume that $D_{a,b}(\mathbf{Q})$ is not a division algebra. There exists an integral solution $(x_0,x_1,x_2,x_3) \neq (0,0,0,0)$ to the Diophantine equation $x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 0$. We first prove that there exists an integral solution $\neq (0,0,0)$ to $-ax_1^2 - bx_2^2 + abx_3^2 = 0$. If $x_0 = 0$, there is nothing to show. Suppose therefore that $x_0 \neq 0$. The integers ax_1^2 and bx_2^2 cannot be both the zero. Assume for example that ax_1^2 is non-zero. Since $x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 0$, we obtain $x_0^2 - bx_2^2 = a(x_1^2 - bx_3^2)$. Then a simple calculation shows that

$$(b(x_0x_3 + x_1x_2), a(x_1^2 - bx_3^2), x_0x_1 + bx_2x_3)$$

is an integer solution to $-ax_1^2 - bx_2^2 + abx_3^2 = 0$. If the latter is zero, then $-ax_1^2 + abx_3^2 = 0$ and $(x_1,0,x_3)$ is a non-zero integer solution (since $x_1 \neq 0$) to $-ax_1^2 - bx_2^2 + abx_3^2 = 0$.

According to the paragraph preceding Lemma 2.7, the quadratic forms $q = -ax_1^2 - bx_2^2 + abx_3^2$ and $q' = -x_1^2 - x_2^2 + x_3^3$ are similar over $\mathbf{Q}$. Since multiplying a quadratic form by a non-zero scalar does not change the associated orthogonal group, the groups $\mathrm{O}(q)$ and $\mathrm{O}(q')$ are conjugate by a matrix belonging to $\mathrm{GL}(3,\mathbf{Q})$. Conjugation by a rational invertible matrix being a rational representation, Lemma 2.5 implies that the image of the group $\mathrm{O}(q,\mathbf{Z})$ in $\mathrm{O}(q')$ is commensurable with the group $\mathrm{O}(q',\mathbf{Z})$. The isomorphism between $\mathrm{SO}_0(q')$ and $\mathrm{PSL}(2,\mathbf{R})$ then implies that the groups $\Gamma_{a,b}$ and $\Gamma_{1,1}$ are commensurable in G. The result follows by observing that $\Gamma_{1,1}$ is commensurable with the group $\mathrm{SL}(2,\mathbf{Z})$. □

The group $\Gamma_{a,b}$ is therefore commensurable with $\mathrm{SL}(2,\mathbf{Z})$ if and only if $-ax^2 - by^2 + abz^2$ is similar to $-x^2 - y^2 + z^2$. On the other hand

$$ab(-ax^2 - by^2 + abz^2) = -b(ax)^2 - a(by)^2 + (abz)^2,$$

so that the quadratic forms $-ax^2 - by^2 + abz^2$ and $x^2 - ay^2 - bz^2$ are themselves similar. The three first claims of Theorem 2.3 are thus proved.

2.2.4 Commensurability Classes

Although not directly used in the rest of the text, the fourth and final claim of Theorem 2.3 will be established in this subsection. In order to be brief, we freely use more advanced algebraic concepts and results; appropriate references are quoted at the end of this chapter. Proofs requiring more background have been put in small typeface.

We begin by remarking that it follows from the proof of Lemma 2.5 that if the quadratic forms $x^2 - ay^2 - bz^2$ and $x^2 - a'y^2 - b'z^2$ (and hence $-ax^2 - by^2 + abz^2$ and $-a'x^2 - b'y^2 + a'b'z^2$) are similar, the groups $\Gamma_{a,b}$ and $\Gamma_{a',b'}$ are commensurable. We must now prove the converse.

Assume then that the groups $\Gamma_{a,b}$ and $\Gamma_{a',b'}$ are commensurable. Then there exists an $\mathbf{R}$-linear map $\alpha : \mathbf{R}^3 \to \mathbf{R}^3$ which sends the quadratic form $q = -ax^2 - by^2 + abz^2$ to $q' = -a'x^2 - b'y^2 + a'b'z^2$ and, by conjugation, a finite index subgroup $\Gamma \subset \mathrm{SO}(q, \mathbf{Z})$ to a finite index subgroup $\Gamma' \subset \mathrm{SO}(q', \mathbf{Z})$. Let $\overline{\Gamma} \subset \mathrm{O}(q)$ (resp. $\overline{\Gamma'} \subset \mathrm{O}(q')$) be the extension by $\{\pm I_3\}$ of the group Γ (resp. Γ').

Lemma 2.8 *The* $\mathbf{Q}$*-vector subspace generated by* $\overline{\Gamma}$ *in* $\mathcal{M}_3(\mathbf{R})$ *is equal to* $\mathcal{M}_3(\mathbf{Q})$.

Proof The $\mathbf{Q}$-vector subspace generated by $\overline{\Gamma}$ is a subalgebra of $\mathcal{M}_3(\mathbf{Q})$. According to the Burnside lemma [76, Cor. 3.4] it suffices to show that the space $\mathbf{Q}^3$ is a simple $\mathbf{Q}\overline{\Gamma}$-module.

We first show that the group $\overline{\Gamma}$ is dense in $\mathrm{O}(q) \cong \mathrm{O}(1,2)$ with respect to the *Zariski topology*. In other words, every polynomial in the matrix entries of matrices in $\mathcal{M}_3(\mathbf{R})$ which vanishes on $\overline{\Gamma}$ should also vanish on $\mathrm{O}(q)$. This can be deduced from the following facts:

1. The closure (in the Zariski topology) of $\overline{\Gamma}$ in $\mathrm{O}(q)$ is a subgroup of $\mathrm{O}(1,2)$ which is not abelian.
2. The group $\mathrm{SO}(q)$ is minimal among the non-abelian subgroups of $\mathrm{O}(q)$ which are closed with respect to the Zariski topology.

The first statement is immediate. The second statement follows from the fact that the group $\mathrm{SO}(q)$ is a connected algebraic group. Its Lie algebra is isomorphic to the Lie algebra of $\mathrm{SL}(2, \mathbf{R})$ which does not contain any proper non-abelian sub Lie algebra.

Since $\overline{\Gamma}$ contains $\{\pm I_3\}$ it follows from the preceding facts that $\overline{\Gamma}$ is dense in $\mathrm{O}(q) \cong \mathrm{O}(1,2)$ with respect to the Zariski topology.

Now, since the natural representation of $\mathrm{O}(1,2)$ in $\mathbf{R}^3$ is irreducible over $\mathbf{C}$, the representation ρ of $\overline{\Gamma}$ in $\mathrm{GL}(3, \mathbf{C})$ is also irreducible. □

End of the proof of Theorem 2.3 Since α sends $\overline{\Gamma}$ onto $\overline{\Gamma'}$, α sends $\mathbf{Q}\overline{\Gamma}$ onto $\mathbf{Q}\overline{\Gamma'}$. But according to Lemma 2.8, we have $\mathbf{Q}\overline{\Gamma} = \mathcal{M}_3(\mathbf{Q})$, which implies that $\alpha = \lambda\alpha_0$, where α_0 is defined over $\mathbf{Q}$ and $\lambda \in \mathbf{R}^*$. Thus α_0 sends q to $\lambda^{-2}q'$ and, since $q \neq 0$, the scalar λ^{-2} is in $\mathbf{Q}$. This shows that the forms q and q' are similar over $\mathbf{Q}$, completing the proof of the proof of Theorem 2.3. □

Remark 2.9 Theorem 2.3 implies in particular that there always exists an infinity of solutions to the Diophantine equation $x_0^2 - ax_1^2 - bx_2^2 + abx_3^2 = 1$.

2.3 Arithmetic Hyperbolic Surfaces

Hyperbolic surfaces are associated with discrete subgroups of $\mathrm{SL}(2, \mathbf{R})$ that are torsion free – those containing no non-trivial element of finite order. We now explain how to eliminate torsion if we have it.

2.3.1 Eliminating Torsion

The process of eliminating torsion is based on the following theorem of Minkowski.

Theorem 2.10 *Let n be an integer $\geqslant 2$ and F a finite subgroup of* $\mathrm{GL}(n, \mathbf{Z})$. *Then F injects in* $\mathrm{GL}(n, \mathbf{Z}/\ell\mathbf{Z})$ *for every integer $\ell \geqslant 3$.*

Proof We consider the standard embedding $\mathbf{Z}^n \subset \mathbf{R}^n$, so that the group F identifies as a subgroup of $\mathrm{GL}(n, \mathbf{R})$. By averaging any choice of positive definite quadratic form on $\mathbf{R}^n$ over F we obtain a norm invariant under the action of F. We may normalize so that the non-zero points of $\mathbf{Z}^n$ closest to 0 are at distance 1; let $V \subset \mathbf{R}^n$ be the subspace generated by these points. By renormalizing the orthogonal complement $V^\perp$, we may again assume that the points of the lattice $\mathbf{Z}^n$ in $\mathbf{R}^n - V$ closest to the origin are at distance 1. Since F preserves V and $V^\perp$, the resulting quadratic form is still F-invariant. Continuing in this way we obtain an F-invariant norm on $\mathbf{R}^n$ such that

1. every vector in $\mathbf{Z}^n$ is of norm greater than or equal to 1, and
2. the points in $\mathbf{Z}^n$ of norm 1 generate the space $\mathbf{R}^n$.

Now let $\gamma \in F$ be such that γ is sent to the identity in $\mathrm{GL}(n, \mathbf{Z}/\ell\mathbf{Z})$ for some integer $\ell \geqslant 1$. Viewed as a matrix in $\mathcal{M}_n(\mathbf{Z})$, $\gamma - I_n \in \ell\mathcal{M}_n(\mathbf{Z})$. If $\gamma \neq I_n$, there exists an element $x \in \mathbf{Z}^n$ of norm 1 such that $\gamma x \neq x$. But $\gamma x - x$ is then a non-zero vector $\in \ell\mathbf{Z}^n$ and is thus of norm $\geqslant \ell$. On the other hand, γx and x both belong to the sphere[3] of radius 1. They are thus at distance at most 2 from one another and ℓ is necessarily less than or equal to 2. □

Let a and b be two positive integers. Given a positive integer N we define the N-th *principal congruence subgroup* of $\Gamma_{a,b}$ as the subgroup $\Gamma_{a,b}(N) \subset \Gamma_{a,b}$ given by the image in $\Gamma_{a,b}$ of the subgroup

$$\{x \in \mathcal{O}^1 \mid x - 1 \in ND_{a,b}(\mathbf{Z})\} \subset \mathcal{O}^1.$$

Note that $\Gamma_{a,b}(N)$ is contained in the image in $\mathrm{SL}(2, \mathbf{R})$ of the subgroup

$$\mathrm{SO}_0(q, \mathbf{Z}) \cap \mathrm{Ker}\left(\mathrm{GL}(3, \mathbf{Z}) \longrightarrow \mathrm{GL}(3, \mathbf{Z}/N\mathbf{Z})\right)$$

under the map $\Phi \circ \Psi^{-1}$. We keep the notation from the proof of Theorem 2.3: the quadratic form q is equal to $-ax^2 - by^2 + abz^2$.

Corollary 2.11 *For every pair of integers $(a, b) \geqslant 1$ and for every integer $N \geqslant 3$, the subgroup $\Gamma_{a,b}(N) \subset \mathrm{SL}(2, \mathbf{R})$ is torsion free. The quotient*

$$X_{a,b}(N) := \Gamma_{a,b}(N)\backslash\mathcal{H}$$

[3] Recall that the norm is F-invariant.

is therefore a hyperbolic surface of finite area. It is compact if and only if the quadratic form $x^2 - ay^2 - bz^2$ *does not represent* 0 *over* $\mathbf{Q}$.

Proof The subgroup $\mathrm{SO}_0(q, \mathbf{Z}) \cap \mathrm{Ker}\left(\mathrm{GL}(3, \mathbf{Z}) \to \mathrm{GL}(3, \mathbf{Z}/N\mathbf{Z})\right)$ is torsion free since, according to Theorem 2.10, the kernel

$$\mathrm{Ker}\left(\mathrm{GL}(3, \mathbf{Z}) \longrightarrow \mathrm{GL}(3, \mathbf{Z}/N\mathbf{Z})\right) \quad (N \geqslant 3)$$

contains no finite subgroup. The group $\Gamma_{a,b}(N)$ ($N \geqslant 3$) is therefore itself torsion free and is clearly of finite index in $\Gamma_{a,b}$. Theorem 2.3 then implies Corollary 2.11. □

An *arithmetic surface* is any surface admitting a cover isometric to a finite cover of one of the surfaces $X_{a,b}(N)$ defined in Corollary 2.11. Note that the fourth claim of Theorem 2.3 implies that $X_{a,b}(N)$ and $X_{a',b'}(N')$ admit two finite isometric covers if and only if $x^2 - ay^2 - bz^2$ and $x^2 - a'y^2 - b'z^2$ are similar as quadratic forms over $\mathbf{Q}$.

2.3.2 The Modular Surface and Its Covers

The fundamental example of an arithmetic surface is the modular surface. As we recalled in the introduction, the subset

$$D = \{z = x + iy \in \mathcal{H} \mid |x| < 1/2 \text{ and } |z| > 1\}$$

of the upper half-plane $\mathcal{H}$ is a fundamental domain for the action of the modular group $\Gamma(1) = \mathrm{SL}(2, \mathbf{Z})$. Moreover, i is an elliptic vertex of order 2, $\zeta = (1+i\sqrt{3})/2$ is an elliptic vertex of order 3, ∞ is the only cusp (up to equivalence) and the genus of $X(1)$ is 0.

We may add to the above example all of the congruence covers of the modular surface. Let N be an integer $\geqslant 1$. The *principal congruence subgroup of level* N, denoted $\Gamma(N)$, is the subgroup of the modular group made up of matrices congruent to the identity modulo N, i.e.,

$$\Gamma(N) = \left\{ \gamma \in \mathrm{SL}(2, \mathbf{Z}) \;\middle|\; \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{N} \right\}. \tag{2.4}$$

Theorem 2.12 *The group* $\Gamma(N)$ *is normal in* $\Gamma(1) = \mathrm{SL}(2, \mathbf{Z})$ *and of index*

$$\mu_N = [\Gamma(1) : \Gamma(N)] = N^3 \prod_{p|N} (1 - p^{-2}).$$

The group $\overline{\Gamma(N)}$ *is of index*

$$\overline{\mu_N} = \begin{cases} \mu_N/2 & \text{if } N > 2 \\ \mu_N & \text{if } N = 2 \end{cases}$$

in $\mathrm{PSL}(2, \mathbf{Z})$. *It is torsion free as soon as* $N > 1$ *and the number of cuspidal equivalence classes is*

$$h_N = \frac{\overline{\mu_N}}{N}.$$

Proof Let $N \geqslant 1$. We begin with the following result.

Lemma 2.13 *The homomorphism*

$$\psi : \mathrm{SL}(2, \mathbf{Z}) \longrightarrow \mathrm{SL}(2, \mathbf{Z}/N\mathbf{Z})$$

is surjective.

Proof Let us show more generally that the homomorphism

$$\psi : \mathrm{SL}(m, \mathbf{Z}) \longrightarrow \mathrm{SL}(m, \mathbf{Z}/N\mathbf{Z})$$

is surjective.

We first assume that N is a prime power p^s. It suffices to show that the group $\mathrm{SL}(m, \mathbf{Z}/p^s\mathbf{Z})$ is generated by the transvection matrices $I_m + E_{ij}$, $i \neq j$, where E_{ij} is the $m \times m$ matrix all of whose coefficients are zero except in the i-th row and j-th column, where it is 1. Indeed, the transvections are clearly in the image of ψ.

We prove this by induction on m. The case $m = 1$ is trivial. Assume then that the statement is true for $m - 1$ where $m > 1$. Let $A \in \mathcal{M}_m(\mathbf{Z})$ be such that $\det(A) \equiv 1 \pmod{p^s}$. The first column of A contains at least one coefficient c which is invertible mod p^s; we shall use it as a pivot for row operations – these are realized by multiplying A by transvection matrices – until we are brought to a matrix whose first column is the vector $(1, 0, \ldots, 0)$. By operating next on the columns of this new matrix we are led to assume that

$$A = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & A' \end{array}\right)$$

with $A' \in \mathcal{M}_{m-1}(\mathbf{Z})$ and $\det(A') \equiv 1 \pmod{p^s}$. The result easily follows by induction.

We deduce the case for arbitrary N from the Chinese Remainder Theorem. If $N = \prod_k p_k^{s_k}$ is the prime power decomposition of N for distinct primes p_k, we have

$$\mathbf{Z}/N\mathbf{Z} \cong \prod_k (\mathbf{Z}/p_k^{s_k}\mathbf{Z}),$$

$$\mathrm{GL}(2, \mathbf{Z}/N\mathbf{Z}) \cong \prod_k \mathrm{GL}(2, \mathbf{Z}/p_k^{s_k}\mathbf{Z}),$$

$$\mathrm{SL}(2, \mathbf{Z}/N\mathbf{Z}) \cong \prod_k \mathrm{SL}(2, \mathbf{Z}/p_k^{s_k}\mathbf{Z}).$$

From this it follows that the transvection $I_m + (\prod_{k\neq\ell} p_k^{s_k})E_{ij}$ is congruent to the identity modulo $p_k^{s_k}$ for k different from ℓ and congruent to the transvection $I_m + E_{ij}$ modulo $p_\ell^{s_\ell}$. □

Consider now the kernel K of the surjective morphism

$$\mathrm{GL}(2, \mathbf{Z}/p_k^{s_k}\mathbf{Z}) \longrightarrow \mathrm{GL}(2, \mathbf{Z}/p_k\mathbf{Z}).$$

Since K is made up of matrices of $\mathcal{M}_2(\mathbf{Z}/p_k^{s_k}\mathbf{Z})$ which are congruent to the identity matrix I_2 mod p_k, the cardinality of K is $p_k^{4(s_k-1)}$.

Let us calculate the cardinality of $\mathrm{GL}(2, \mathbf{Z}/p_k\mathbf{Z})$. This group is made up of matrices with coefficients in $\mathbf{Z}/p_k\mathbf{Z}$, such that the column vectors are linearly independent. There are $p_k^2 - 1$ choices for the first column vector and, once the first has been fixed, there are $p_k^2 - p_k$ remaining choices for the second (we exclude the p_k multiples of the first vector). The cardinality of $\mathrm{GL}(2, \mathbf{Z}/p_k\mathbf{Z})$ is therefore equal to $(p_k^2 - 1)(p_k^2 - p_k)$.

The cardinality of $\mathrm{GL}(2, \mathbf{Z}/p_k^{s_k}\mathbf{Z})$ is therefore equal to $p_k^{4(s_k-1)}(p_k^2-p_k)(p_k^2-1) = p_k^{4s_k}(1-p_k^{-1})(1-p_k^{-2})$. Now $\mathrm{SL}(2, \mathbf{Z}/p_k^s\mathbf{Z})$ is the kernel of the surjective morphism $\det : \mathrm{GL}(2, \mathbf{Z}/p_k^{s_k}\mathbf{Z}) \to (\mathbf{Z}/p_k^{s_k}\mathbf{Z})^*$. The cardinality of $\mathrm{SL}(2, \mathbf{Z}/p_k^{s_k}/p_k^{s_k}\mathbf{Z})$ is then equal to $p_k^{3s_k}(1 - p_k^{-2})$. As a consequence

$$[\Gamma(1) : \Gamma(N)] = N^3 \prod_{p|N} (1 - p^{-2}).$$

The matrix $-I_2$ belongs to the group $\Gamma(N)$ if and only if $N = 2$. We deduce

$$[\overline{\Gamma(1)} : \overline{\Gamma(N)}] = \begin{cases} \mu_N/2, & \text{if } N > 2, \\ \mu_N, & \text{if } N = 2. \end{cases}$$

Let us now show that if $N > 1$ the group $\overline{\Gamma(N)}$ has no elliptic elements. The only fixed points of $\overline{\Gamma(1)} = \mathrm{PSL}(2, \mathbf{Z})$ in the fundamental hyperbolic triangle D are the two vertices of angle $\pi/3$ and the point i. The elliptic conjugacy classes in

$\mathrm{PSL}(2,\mathbf{Z})$ are therefore represented by

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}.$$

None of these elements is congruent to the identity modulo N if $N > 1$. Since $\overline{\Gamma(N)}$ is normal in $\overline{\Gamma(1)}$ we obtain, as claimed, that $\overline{\Gamma(N)}$ is torsion free.

Now, if s is a cusp, s is $\overline{\Gamma(1)}$-equivalent to ∞. But

$$\Gamma(1)_\infty = \left\{ \pm \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \;\middle|\; m \in \mathbf{Z} \right\},$$

$$\Gamma(N)_\infty = \Gamma(N) \cap \Gamma(1)_\infty = \left\{ \pm \begin{pmatrix} 1 & mN \\ 0 & 1 \end{pmatrix} \;\middle|\; m \in \mathbf{Z} \right\},$$

so that $[\Gamma(1)_\infty : \Gamma(N)_\infty] = N$. In this way $\Gamma(N)$ has exactly $\overline{\mu_N}/N$ equivalence classes of cusps. □

The surfaces $X(N) = \Gamma(N)\backslash\mathcal{H}$ are therefore "true" hyperbolic surfaces as soon as $N > 1$. These are the covers – ramified over the elliptic points – of the modular surface $X(1)$. The degree of this cover is equal to $\overline{\mu_N}$. The area of $X(N)$ is then equal to $\overline{\mu}_N$ times the area of $X(1)$, namely

$$\mathrm{area}(X(N)) = \begin{cases} \dfrac{\pi N^3}{6} \prod_{p|N} (1 - p^{-2}), & \text{if } N > 2, \\ 2\pi, & \text{if } N = 2. \end{cases} \tag{2.5}$$

The surface $X(N)$ $(N > 1)$ has $h_N = \overline{\mu_N}/N$ cusps. We obtain a fundamental domain for the action of $\Gamma(N)$ on $\mathcal{H}$ by taking the union of $\overline{\mu_N}$ translates of the triangle D. This induces a triangulation of $X(N)$ by $\overline{\mu_N}$ triangles. Each of these triangles has two vertices of angle $\pi/3$ around which are glued 6 triangles and then another vertex of angle 0 around which are glued N triangles. There are therefore $\overline{\mu_N}(\frac{1}{3} + \frac{1}{N})$ vertices, $\frac{3}{2}\overline{\mu_N}$ sides and $\overline{\mu_N}$ faces. The Euler characteristic of $X(N)$ is equal to

$$\overline{\mu_N}\left(\frac{1}{N} - \frac{1}{6}\right).$$

Now the Euler characteristic is also equal to $2-2g_N$, where g_N is the genus of $X(N)$. So we obtain

$$g_N = \begin{cases} 1 + \mu_N \dfrac{N-6}{24N}, & \text{if } N > 2, \\ 0, & \text{if } N = 2. \end{cases} \tag{2.6}$$

Every subgroup of the modular group containing $\Gamma(N)$ is called a *congruence subgroup*. The following example is particularly important,

$$\Gamma_0(N) = \left\{ \gamma \in \mathrm{SL}(2, \mathbf{Z}) \;\middle|\; \gamma \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\} .$$

2.4 Commentary and References

§ 2.1

The Iwasawa decomposition is discussed in detail in [93, p. 44].

§ 2.2

In general there are two possibilities for a quaternion algebra D over a field F of characteristic zero: either D is a division algebra, or $D \cong \mathcal{M}_2(F)$, the algebra of 2×2 matrices over F. See for example [88, Th. 2.1.7] for a proof of this result which we do not use here.

The general notion of arithmetic group is a bit technical; the reader can consult [99]. The classification of arithmetic groups described in this reference implies, in particular, that every arithmetic subgroup of $\mathrm{SL}(2, \mathbf{R})$ defined over $\mathbf{Q}$ is commensurable to one of the subgroups constructed in Theorem 2.3. One can more generally replace $\mathbf{Q}$ by a totally real number field. The construction is similar and treated in [67, 88].

The proof of Theorem 2.3 that we give here is slightly different from that which one usually finds in the literature, as for example in [67]. We reduce the argument to the study of arithmetic subgroups of $\mathrm{SO}(2, 1)$ via the exceptional isomorphism between $\mathrm{PSL}(2, \mathbf{R})$ and $\mathrm{SO}_0(2, 1)$. This way of constructing Fuchsian groups is essentially the same used by Poincaré in his article "Les fonctions fuchsiennes et l'arithmétique," which appeared in 1887. The groups obtained in this way are in fact the first examples of Fuchsian groups that Poincaré succeeded in constructing. He also used the geometry of the space of lattices $\mathcal{L}_3$ previously studied by Hermite, to whom we owe the proof of Theorem 2.1 when $n = 3$. Before that the geometry of the space $\mathcal{L}_2$ had been completely understood by Gauß in his *Disquisitiones arithmeticae* which contains a very clear exposition of the action of the modular group on the upper half-plane $\mathcal{H}$.

The construction of infinitely many commensurability classes of discrete cocompact subgroups in $\mathrm{SL}(2, \mathbf{R})$ is borrowed from [88, pp. 87–88].

Through the exceptional isomorphism the geometry of the orthogonal group of a non-degenerate quadratic form comes into play. A convenient reference on this subject is the book [96]. In particular, [96, Th. 2.6] gives a proof of the fact that the rotations of angle π generate the special orthogonal group.

The classification of commensurability classes relies on notions in algebraic geometry and algebraic groups. The books [97] and [16] can serve as good introductions to these notions.

§ 2.3

Our study of congruence subgroups of the modular group follows Shimura's exposition of the topic in [121, §1.6]. One can also have a look at the excellent book of Miyake [92]. These two references contain general computations on the cardinality of equivalence classes of elliptic fixed points of order 2 and 3 for the groups $\Gamma_0(N)$.

We have assumed familiarity with the Euler characteristic of a surface and the Euler phi function. For an introduction to these notions the reader can refer to [92, 121] and Serre's Course in Arithmetic [118], respectively.

2.5 Exercises

Exercise 2.14 Show that the group $\mathrm{SL}(2, \mathbf{Z}[\sqrt{2}])$ is not a Fuchsian group.

Exercise 2.15

1. Show that the quotient $\mathrm{SL}(2, \mathbf{Z})\backslash\mathrm{SL}(2, \mathbf{R})$, endowed with the quotient topology, is homeomorphic to the unit tangent bundle of the modular surface $\mathrm{SL}(2, \mathbf{Z})\backslash\mathcal{H}$. (This question is answered in § 9.2.)
2. Using the explicit description of the fundamental domain for $\mathrm{SL}(2, \mathbf{Z})$ given in § 1.4.1, deduce from the preceding question that a subset of volume 1 lattices in $\mathbf{R}^2$ is relatively compact if and only if the height is uniformly bounded from below on this set by a positive constant.

Exercise 2.16 When $n = 2$, compare the set D of § 1.4.1 and the orbit $S_{s,t}^{-1} i$ of the point $i \in \mathcal{H}$ under the action of a Siegel set. Deduce that $D(1) \subset S_{s,t}^{-1} i$ as soon as $s \geqslant 2/\sqrt{3}, t \geqslant 1/2$.

Exercise 2.17 Taking inspiration from the proof of Theorem 2.3, show that there exist infinitely many integer solutions (m, n) to the "Pell-Fermat" Diophantine equation

$$m^2 - an^2 = 1,$$

where a is an integer such that $\sqrt{a} \notin \mathbf{N}$.

Exercise 2.18

1. Let $\alpha = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}$. Show that $-I_2 \in \Gamma_0(N)$ and that

$$\Gamma_0(N) = \alpha^{-1} \Gamma(1) \alpha \cap \Gamma(1).$$

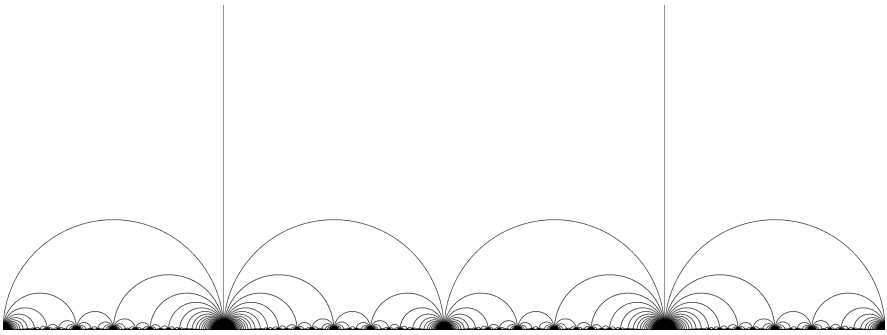

Fig. 2.1 Degree 6 cover of the modular surface

2. Show that the group $\Gamma_0(N)/\Gamma(N)$ is isomorphic to the subgroup of upper triangular matrices in $\mathrm{SL}(2, \mathbf{Z}/N\mathbf{Z})$. Deduce that it is of order $N\phi(N)$, where ϕ is the *Euler totient function* which associates with a positive integer N the cardinality of the group of invertible elements in $\mathbf{Z}/N\mathbf{Z}$.
3. Deduce from Question 2 that the index of $\Gamma_0(N)$ in $\Gamma(1)$ is given by

$$[\Gamma(1) : \Gamma_0(N)] = N \prod_{p|N} (1 + p^{-1}).$$

4. Show that the number h of equivalence classes of cusps for $\Gamma_0(N)$ is equal to the number of double classes in $\Gamma_0(N)\backslash\Gamma(1)/\Gamma(1)_\infty$.
5. Let M_N be the set of elements $(a, c) \in (\mathbf{Z}/N\mathbf{Z})^2$ of order N. Show that the map from $\Gamma(1)$ to the set M_N defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto (a, c)$$

induces a bijective map from $\Gamma_0(N)\backslash\Gamma(1)/\Gamma(1)_\infty$ onto $M_N/\sim$, where $\sim$ is the equivalence relation on M_N given by $(a, c) \sim (a', c')$ if

$$(a', c') \equiv \pm(ma + nc, m^{-1}c) \quad (m \in (\mathbf{Z}/N\mathbf{Z})^*,\ n \in \mathbf{Z}/N\mathbf{Z}).$$

6. Deduce from the two preceding questions that

$$h = \sum_{d|N,\ d>0} \phi\left(\gcd\left(d, (N/d)\right)\right),$$

where ϕ is the Euler phi function.

Exercise 2.19

1. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(8)$.
   a. Show that $a + d \neq 0$ (use the fact that -1 is not a square modulo 8).
   b. Show that a and d are odd and thus that $a + d \neq \pm 1$.
   c. Deduce that $|a + d| \geqslant 2$ and thus that the image of the group $\Gamma_0(8)$ in $\mathrm{PSL}(2, \mathbf{R})$ is a torsion free subgroup.
2. Deduce from the first question and the preceding exercise that the quotient $\Gamma_0(8)\backslash\mathcal{H}$ is a non-singular hyperbolic surface of genus 0 with 4 cusps, which is a degree 12 cover of the modular surface. The union of two adjacent polygons in Fig. 2.1, generated by Arnaud Chéritat, forms a fundamental domain for the group $\Gamma_0(8)$.

http://www.springer.com/978-3-319-27664-9

The Spectrum of Hyperbolic Surfaces

Bergeron, N.

2016, XIII, 370 p. 8 illus. in color., Softcover
ISBN: 978-3-319-27664-9