3. Finite generation: classical reduction theory

For $n \in \mathbf{N}$ let

$$\Gamma(n) = \text{kernel of } (\Gamma \to (\Lambda/n\Lambda)^\times)$$

the congruence group mod n; this is a normal subgroup of finite index. Obviously $\Gamma(n)$ is torsion free for $n > N$. With more effort, one can do much better: the regular representation injects $\Gamma(n)$ into the congruence group mod n in $GL_m(\mathbf{Z})$, $m = \dim A$, and Minkowski has shown that this is torsion free for $n > 2$ [Mi].

(4) Γ *contains only finitely many isomorphism classes of finite subgroups.*

Proof. If $\Gamma_0 < \Gamma$ is torsion free and normal of finite index, then every finite subgroup of Γ is isomorphic to a subgroup of Γ/Γ_0.

Later, we will show more: Γ contains only finitely many conjugacy classes of finite subgroups.

(5) Γ *is residually finite, that is, for every* $x \in \Gamma, x \neq 1$, *there is a normal subgroup* Γ_0 *of finite index such that* $x \notin \Gamma_0$.

Of course, almost all $\Gamma(n)$ will do. It follows that Γ is hopfian, that is, not isomorphic to a proper factor group (see [MKS], p. 116).

(6) Finally, let us mention here the following result due to Zassenhaus [Z2] (although it is not entirely elementary): Γ contains a solvable subgroup of finite index if and only if the Wedderburn components of A are number fields or definite quaternions over $\mathbf{Q}$.

Sketch of proof: the problem is readily reduced to simple A. The "If" part is then trivial.

Conversely, if matrices are involved, one knows that Γ has infinitely many subfactor groups of the form $SL_n(F)$, where F is a finite field. The same is therefore true of any subgroup of finite index. In the skew field case, the argument is more intricate; we refer to [Z2].

3. Finite generation: classical reduction theory

The most basic fact about Γ is that it is finitely generated; this is even valid for arbitrary arithmetic groups, as has been proved by A. Borel and Harish-Chandra in the fundamental paper [BHC]. Here I shall describe the classical approach, carried out by Siegel [S1], who completed earlier work of

Minkowski, Humbert, Weyl and Eichler. The leading idea is to make Γ operate on a suitable topological space; if this operation is "good enough", then generators can be read off form it, even, as we shall see in the next section, defining relations. Let us begin with the basic definitions.

Let the group H operate on the topological space T as a group of homeomorphisms. For a non-empty subset $F \subset T$ define

$$E(F) = \{h \in H \mid F \cap Fh \neq \varnothing\} .$$

If we think of F as a fundamental domain, then $E(F)$ consists of those elements which carry F to a "neighbor". The following basic observation occurs in [S1, section 9].

BASIC LEMMA. *Assume that*

(i) $FH = T$;

(ii) $FE(F)$ *is a neighborhood of* F; *and*

(iii) T *is connected.*

Then $E = E(F)$ *generates* H.

Proof. Let H_0 be the subgroup generated by E and $\{h_i\}$ be a set of right coset representatives of H mod H_0. Then the sets $X_i = FH_0h_i$ are disjoint, open and form a cover of T. Since T is connected, there can be only one of them.

Let us illustrate this at once with the most classical case of $H = \Gamma = SL_n(\mathbf{Z})$. In accordance with previous terminology, this is half the unit group. In order to obtain finite generation one has to find T and F such that F is not too small (otherwise (i) or (ii) might fail) and not too large (otherwise E might be infinite).

A plausible condition for E being finite is that H operates discontinuously, that is, no H-orbit has a cluster point. (If x is a cluster point of fH, write $x = f'h, f' \in F$; if there is a neighborhood $f' \in U \subset F$, then Uh contains infinitely many fh_i and $h_ih^{-1} \in E$). This rules out the most near-at-hand choice of T, the natural space $\mathbf{R}^n$. (Convince yourself for $n = 2$, that Γ does not operate discontinuously on $\mathbf{R}^2$!) A possible choice, however, is $T = G = SL_n(\mathbf{R})$, Γ operating by right multiplication. Γ is a discrete subgroup of G. For $t > 0$, $w > 0$ define

$$D_t = \{\mathrm{diag}(a_1, \ldots, a_n) \in G \mid 0 < a_i \leqslant ta_{i+1}\}$$
$$N_w = \{(u_{ij}) \text{ strict upper triangular with } |n_{ij}| \leqslant w\}$$

and

$$S_{t,w} = SO(n)D_tN_w \subset G .$$

This is called a "Siegel domain". One now proves two things:

(1) for $t \geqslant 2/\sqrt{3}$, $w \geqslant \frac{1}{2}$ we have $S_{t,w}\Gamma = G$;

(2) for all t, w the set

$$\{\gamma \in \Gamma \mid S_{t,w} \cap S_{t,w}\gamma \neq \varnothing\}$$

is finite.

Proofs can be found, e.g., in Borel's book [B2, § 1]. It is not difficult to apply the lemma, and hence Γ is finitely generated.

Remark on terminology: by a fundamental domain we mean a set containing a system of orbit representatives and such that H-translates of it intersect at most on the boundaries. It is equivalent to (i) of the Basic lemma that F contains a fundamental domain. Property (2) (and its generalizations) is called "Siegel's property" by Borel; (1) and (2) constitute what Borel calls "ensemble fondamental". Other authors require other properties or distinguish between "fundamental set" and "fundamental region". Note that a Siegel domain is not a fundamental domain in this sense! See [Te, 4.4] for Minkowski's classical fundamental domain of $SL_n(\mathbf{Z})$.

Let us briefly indicate (although this goes beyond our theme) how the argument generalizes to arithmetic groups. $SO(n)$ is a maximal compact subgroup of G, the set D of diagonal matrices in G is a maximal torus (a torus is a group isomorphic to a direct product of copies of $\mathbf{R}^\times$), and the set N of strict upper triangular matrices is a maximal unipotent subgroup of G. Such groups are reasonably unique, and one has the Iwasawa decomposition

$$\begin{aligned} SO(n) \times D \times N &\xrightarrow{\sim} G \\ (o, d, n) &\to odn\ , \end{aligned}$$

which is a diffeomorphism of manifolds. Let D operate by conjugation on the vector space $\mathfrak{g}$ consisting of n-by-n matrices of trace zero, the Lie algebra of G. The character group $\mathrm{Hom}(D, \mathbf{R}^\times)$ is generated, say, by the first $n-1$ coordinate functions and is isomorphic to $\mathbf{Z}^{n-1}$; for a character λ define

$$\mathfrak{g}^\lambda = \{x \in \mathfrak{g} \mid dxd^{-1} = \lambda(d)x, \quad \text{all} \quad d \in D\}$$

and call λ a root if $\mathfrak{g}^\lambda \neq 0$. Among the roots one can distinguish simple roots which can be choosen to be

$$\lambda_i : \mathrm{diag}(d_1, \ldots, d_n) \to d_i d_{i+1}^{-1}.$$

Thus,

$$D_t = \{d \in D \mid \lambda(d) \leqslant t, \lambda \text{ simple}, d \text{ positive}\}\ .$$

N_w is simply a "generic" compact subset of N. Now all of these concepts — maximal compact subgroups, tori, unipotent subgroups, Iwasawa decomposition, roots and simple roots — generalize to reductive real algebraic groups G. Hence Siegel domains can be defined completely analogously, and one can prove the analogues of (1) and (2) for arithmetic subgroups Γ of G; this has been done in [BHC]. By elementary property (1), this applies to unit groups of orders.

Secondly, let us pursue the connection of these concepts with reduction of quadratic forms. In applying the lemma it is natural to look for a manifold of least possible dimension which possesses a suitable F. In the case of $\Gamma = SL_n(\mathbf{Z})$, the observation that $SO(n) \cap \Gamma =$ compact and discrete, hence finite leads to the expectation that the operation of Γ on the coset space $SO(n)\backslash G$ still does the job. By linear algebra, the map

$$\pi: \begin{cases} G \to \text{symmetric positive matrices of determinant 1} \\ g \to g^t g \end{cases}$$

is surjective; this implies that the operation of G on these matrices, $(g, x) \to g^t x g$, is transitive. The stabilizer of 1_n is $SO(n)$; hence $SO(n)\backslash G$ identifies with that space, which in turn is identified with the space of positive definite quadratic forms of determinant 1. If $g = kdn$ is the Iwasawa decomposition, then from

$$\pi(g) = n^t d k^t k d n = n^t d^2 n$$

we see that $S_{t,w}$ is mapped to the Siegel domain

$$S'_{t^2,w} = \{n^t d n \mid d \in D_{t^2},\ n \in N_w\}$$

in the space of forms. Hence (1) translates to Minkowski's "reduction theorem" saying that every positive form of determinant 1 is a Γ-translate of an element of $S'_{4/3,1/2}$. It is clear that $E(S'_{t^2,w})$ is still finite.

Hence we can (in principle) obtain a finite set of generators from the Γ-operation on a space of dimension

$$n^2 - 1 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2} - 1 .$$

But now the attentive reader will object that this is somewhat like putting the cart before the horse because reduction theory doubtless has an interest in its own right whereas it is elementary to write down a finite set of generators

for $SL_n(\mathbf{Z})$. In fact, such a set can be given for $SL_n(R)$ if R is euclidean and finite over $\mathbf{Z}$ (see [Ne], p. 107) and for $SL_n(\mathbf{Z})$ one can do with

$$\begin{pmatrix} 1 & 1 & & & \\ & 1 & 0 & & \\ & & \ddots & & \\ 0 & & & & 1 \end{pmatrix} \text{ and } \begin{pmatrix} & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & 0 & 1 & \\ & & & & \ddots & \\ & & & & & 1 \\ (-1)^{n-1} & & & & & 0 \end{pmatrix},$$

as Hua and Reiner have shown [HR]. Hurwitz [H] treated $SL_2(R)$, where R is the integral domain of a number field, and remarked that the procedure can be generalized to $SL_n(R)$, giving a sketch for $n = 3$. The most general form of the argument was given by O'Meara [O'M]. The finite generation of $SL_n(R)$ can be derived directly from the finiteness conditions incorporated in the notion of number field, and there is no need to employ the geometry. This should also hold for the case in which skew fields are involved although a purely algebraic treatment of this case has — as far as I know — not been given.

The reply is that finite generation as such is a very weak information and gives hardly any insight into the structure of our unit groups. It is the *raison d'être* of groups to operate on sets having an internal structure, and it is by understanding the operation that we understand groups. With regard to units of orders, this has been stressed by Eichler [E1]:

„Von der Überzeugung ausgehend, daß die Begriffswelt der Zahlengeometrie die geeignete Grundlage für den Aufbau eines tragenden Gerüsts für die hyperkomplexe Einheitentheorie abgibt, beschäftige ich mich hier mit Darstellungen der Einheitengruppen durch affine Abbildungen eines Raumes auf sich. In dieser geometrischen Gestalt trat sie erstmals in der analytischen Zahlentheorie auf und führte auf geometrische Untersuchungen, die bis heute nicht in befriedigender Weise abgeschlossen werden konnten. Die Hauptaufgabe der Einheitentheorie sehe ich nun in der Auffindung von Invarianten dieser Abbildungsgruppen."

Needless to say, this is still the adequate view on units of orders. Furthermore, as we shall see later, the geometric method leads at least theoretically to defining relations among the generators thus found; in the only case where these can be derived purely algebraically ($SL_2(\mathbf{Z})$) this derivation has an artificial and a-posteriori character, and doubtless the most natural way to the presentation

$$SL_2(\mathbf{Z}) = C_4 *_{C_2} C_6$$

is by letting the group operate on a tree [Se1].

Moving towards general orders we first deal with the case where $A = D$ is a skewfield, in which the number geometric method works particularly smoothly. Put $D_{\mathbf{R}} = \mathbf{R} \otimes_{\mathbf{Q}} D$ and

$$G = \{x \in D_{\mathbf{R}}^{\times} \mid N(x)^2 = 1\} ,$$

N denoting the regular norm $D \to \mathbf{Q}$. Clearly, $\Gamma \subset G$ is a discrete subgroup. The following result was proved by Käte Hey in her doctoral thesis (Hamburg 1929) and reappears in [Sch], [E1], and [Z1].

THEOREM 1. *G/Γ is compact.*

Proof (according to [Z1]). We work with a $\mathbf{Z}$-basis of Λ, so that in $D_{\mathbf{R}} = \mathbf{R}^g$, $g = \dim D$, Λ appears as $\mathbf{Z}^g$.

Let C be any convex, 0-symmetric compact set in $\mathbf{R}^g$ such that $\mathrm{vol}(C) > 2^g$. By Minkowski's lattice point theorem, C contains a nonzero $a \in \Lambda$. If $x \in G$, then $\mathrm{vol}(Cx) = \mathrm{vol}(C)$ because of $|N(x)| = 1$, and Cx is still convex and 0-symmetric, hence contains a nonzero $a_x \in \Lambda$.

Now let (x_n) be a sequence of elements in G. Then there are $a_j \in \Lambda \setminus \{0\}$ such that

$$a_i = c_i x_i , \quad c_i \in C .$$

It follows that $|N(a_i)|$ is bounded because N is bounded on C. Because D is a skew field, we have

$$|N(a_i)| = |\Lambda : a_i \Lambda| \neq 0 .$$

Since there are only finitely many right ideals of bounded index, there is a subsequence (a_k) such that

$$a_k \Lambda = a_1 \Lambda \text{ (say)} ,$$

hence

$$a_k = a_1 \varepsilon_k , \quad \varepsilon_k \in \Gamma .$$

Further,

$$|N(c_k)| = |N(a_k)| = |N(a_1)| > 0 .$$

Since C is compact, (c_k) contains a convergent subsequence (c_l). The last inequality shows that (c_l^{-1}) is convergent. From

$$x_l \varepsilon_l^{-1} = c_l^{-1} a_1$$

we now read off that G/Γ is compact. Note that we have used, so to speak, only half of the lattice point theorem in that there was no need to specify C.

This is our first generalization of Dirichlet's unit theorem, the most classical result on units of orders, in that it contains what one calls the hard part of this theorem. In fact, let $D = K$ be a number field and write, in usual notations,

$$K_{\mathbf{R}} = \mathbf{R}^{r_1} \times \mathbf{C}^{r_2},\ r_1 + 2r_2 = g\ ;$$

we have

$$G = \{(x_1, \ldots, x_{r_1+r_2}) \in K_{\mathbf{R}} \mid (x_1 \ldots x_{r_1}) \mid x_{r_1+1} \mid^2 \ldots \mid x_{r_1+r_2} \mid^2 = 1\}\ .$$

The logarithm map

$$\log \begin{cases} G & \to \mathbf{R}^r,\ r = r_1 + r_2 - 1 \\ (x_i) & \to (\log \mid x_1 \mid, \ldots, \log \mid x_{r_1} \mid, 2\log \mid x_{r_1+1} \mid, \ldots, 2\log \mid x_{r_1+r_2-1} \mid) \end{cases}$$

is a homomorphism, continuous, surjective and has compact kernel. Since Γ is discrete in G, $\log \mid \Gamma$ has finite kernel, and $\log \Gamma$ is discrete in $\log G = \mathbf{R}^r$, hence a lattice. It follows that

$$\Gamma \cong W(K) \times \mathbf{Z}^{\tilde{r}},\ \tilde{r} = rk \log \Gamma \leqslant r\ ,$$

$W(K)$ denoting the roots of unity in K. This is the easy part of Dirichlet's theorem, the hard one being that $\tilde{r} = r$. In the standard presentations of the theorem, one now has to go through some unperspicous trickery (involving, of course, the lattice point theorem) in order to establish the existence of sufficiently many independent units. But clearly $\tilde{r} = r$ is equivalent to the compactness of $\log G / \log \Gamma$, which follows at once from Theorem 1.

The generalization of Theorem 1 to arithmetic groups is as follows: let $G \subset Gl_n$ be a reductive algebraic group defined over $\mathbf{Q}$, Γ an arithmetic subgroup. Then $G_{\mathbf{R}}/\Gamma$ is compact if and only if G^0 (= connected component of identity) has no nontrivial $\mathbf{Q}$-characters and all elements of $G_{\mathbf{Q}}$ are semisimple (see [B2], p. 55ff.). The reader might try to verify that the hypotheses of this result are satisfied if G is the algebraic group defined over $\mathbf{Q}$ by the norm-1-elements of a skew field.

The finite presentation of Γ can be extracted from Theorem 1. Let $K \subset G$ be a maximal compact subgroup; then $\Gamma \cap K$ is finite, hence Γ contains a subgroup Γ_0 of finite index such that $\Gamma_0 \cap K = 1$. Then $K \backslash G / \Gamma_0$ is a compact manifold, and since $K \backslash G$ is a homeomorphic to a Euclidean space, Γ_0 is its fundamental group. But the fundamental group of a compact manifold is always finitely presented (a proof of this fact can be found in [Ra], p. 95).

The two "extreme cases" $A = M_n(\mathbf{Q})$ and $A = D$ are comparatively easy; unfortunately, the general case offers difficulties which cannot be overcome

by a straightforward combination of these two. (Be sure to see clearly why the skew field property of D is indispensable in the proof of Theorem 1). However, Zassenhaus proves the following generalization in which A is not even required to be semisimple: there is a system F of right coset representatives of G mod Γ of the following form:

$$F = \{xW(x)VW(x)^{-1}\},$$

where the x run over a compact subset of G, $W(x) \in G$ is a function with finite range and V a torus with positive diagonal elements. Visibly, there is a resemblance to a Siegel domain. In the skewfield case, $V = 1$. From this one can derive the finite presentability of Γ along classical lines (see section 4).

Approaching the general case, now we could simply refer the reader to Borel's text [B2] since there is no point in reporting at length on the contents of a textbook which is standard since 25 years. On the other hand, even in a survey article the reader will expect to become acquainted more closely with the methods. Therefore let us consider in some detail Siegel's classical treatment. Actually, we follow Weyl [W] who found it necessary to provide a careful explanation of Siegel's "all too laconic" arguments. He divided the proof (of finite generation) into three "theorems of finiteness"; we will lead the discussion up to a point where the content and the rôle of these theorems become visible. Perhaps the clarity and elegance of Weyl's arguments is still of more than merely historical interest.

Let $A = M_n(D)$. A *lattice* N in D^n is a finitely generated $\mathbf{Z}$-module containing a D-basis of the right D-vector space D^n. Such a basis, $\mathscr{D} = \{d_1, \ldots, d_n\}$, is called a *semibasis* of N. Given $\mathscr{D}$, the set

$$L(\mathscr{D}) = \{(a_1, \ldots, a_n)^t \in D^n \mid \Sigma d_i a_i \in N\}$$

is another lattice, containing the standard basis vectors $e_1, \ldots, e_n$. $L(\mathscr{D})$ is called the representation of N in terms of $\mathscr{D}$, and all such $L(\mathscr{D})$ are called *admissible* lattices. The left order

$$O_l(N, A) = \{x \in A \mid xN \subseteq N\}$$

is our order Λ, and

$$\Gamma = \{x \in A \mid xN = N\} = \Lambda^\times$$

is the group which interests us; Weyl calls it the *lattice group*. If $\mathscr{D}$, $\mathscr{D}'$ are two semibases, then $L(\mathscr{D}) = L(\mathscr{D}')$ if and only if $\mathscr{D}' = s\mathscr{D}'$ for some $s \in \Gamma$.

An $\mathbf{R}$-basis of $D_{\mathbf{R}} = \mathbf{R} \otimes_{\mathbf{Q}} D$ is called *normal* if the regular representation R of $D_{\mathbf{R}}$ with respect to that basis has the property

$$R(D_{\mathbf{R}}) = R(D_{\mathbf{R}})^t \quad (t \text{ denoting transpose}).$$

It is not difficult to establish the existence of normal bases: let $K = Z(D)$ and write as before

$$\mathbf{R} \otimes_{\mathbf{Q}} K = \mathbf{R}^{r_1} \times \mathbf{C}^{r_2}, \; r_1 + 2r_2 = \dim_Q K .$$

Then

$$D_{\mathbf{R}} = \mathbf{R} \otimes_{\mathbf{Q}} D = (\mathbf{R} \otimes_{\mathbf{Q}} K) \otimes_K D = \prod_{i=1}^{r_1} \mathbf{R} \otimes_K^i D \times \prod_{j=1}^{r_2} \mathbf{C} \otimes_K^j D ,$$

where $\otimes^i (\otimes^j)$ indicates that the tensor product is to be formed with respect to the K-module structure of $\mathbf{R}(\mathbf{C})$ corresponding to the i-th (j-th) embedding of K into $\mathbf{R}(\mathbf{C})$. The $\mathbf{C} \otimes^j D$ are central simple $\mathbf{C}$-algebras and hence full matrix rings over $\mathbf{C}$. The $\mathbf{R} \otimes^i D$ are central simple $\mathbf{R}$-algebras and hence full matrix rings over $\mathbf{R}$ or $\mathbf{H}$, the quaternion skew field. More precisely, if $s^2 = \dim_K D$,

$$(3) \qquad \begin{cases} \mathbf{C} \otimes^j D \cong M_s(\mathbf{C}) & \\ \mathbf{R} \otimes^i D \cong M_s(\mathbf{R}), & \text{for } i = 1, \ldots, r_1' \text{ (say)} \\ \mathbf{R} \otimes^i D \cong M_{s/2}(\mathbf{H}), & \text{for } i = r_1' + 1, \ldots, r_1' + r_1'' = r_1 . \end{cases}$$

If we now replace the elements of $\mathbf{C}$ and $\mathbf{H}$ by their regular representations with respect to the standard bases, then the typical elements are

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in \mathbf{C}, \quad \begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix} \in \mathbf{H} ,$$

and transposing corresponds to the usual conjugation on $\mathbf{C}$ and $\mathbf{H}$. Combining this with the fact that for any skew field F, the regular representation of $M_n(F)$ over F is equivalent to n times the identity, we see that normal bases exist.

We fix one of them and obtain a conjugation on $D_{\mathbf{R}}$ by

$$\alpha \to \bar{\alpha} = R^{-1}(R(\alpha)^t) .$$

Call α symmetric if $\alpha = \bar{\alpha}$, positive if $R(\alpha) > 0$ is positive definite. A *quadratic form* over $D_{\mathbf{R}}$ is now a matrix

$$F = (\gamma_{ij}) \in M_n(D_{\mathbf{R}}), \quad \text{with } \overline{\gamma_{ji}} = \gamma_{ij} .$$

For $x \in (D_{\mathbf{R}})^n$ put

$$F[x] = \bar{x}^t F x ,$$

a symmetric element of $D_{\mathbf{R}}$. F is called *positive* if the real matrix $(R(\gamma_{ij}))$ is positive definite. Important note: it would not work to define $F > 0$ by $F[x] > 0$, all $x \neq 0$. Choosing $x = (0, \ldots, x_i, \ldots, 0)$, we had to have $\bar{x}_i \gamma_{ii} x_i > 0$, all $x_i \neq 0$ which implies

$$N(x_i)^2 N(\gamma_{ii}) > 0 ;$$

but if $D_{\mathbf{R}}$ is not a skew field, there will be $x_i \neq 0$ with $N(x_i) = 0$. The positive F form an open convex cone in the space of all forms, in particular a manifold of the same dimension. We call it H^+. Weyl shows next that $F > 0$ if and only if $F = \bar{A}^t A, A \in Gl_n(D_{\mathbf{R}})$; this implies that, as long as the conjugation $\alpha \to \bar{\alpha}$ is fixed, positiveness does not depend on the choice of a normal basis.

Let $Tr : D \to \mathbf{Q}$ (or $D_{\mathbf{R}} \to \mathbf{R}$) denote the trace of the regular representation. It is not hard to show that, if F is positive in the above sense, one has

$$t_F(x) := TrF[x] > 0, \quad \text{for } x \neq 0 \text{ in } (D_{\mathbf{R}})^n .$$

This is the correct definition of "positive form over a skew field"; as Weyl points out, a crucial step in Siegel's proof.

So far we have been setting the stage; now we come to the first main step, the method of successive minima originally invented by Minkowski. Let the lattice N and the positive form $t = t_F$ be given. Since for any real $s > 0$ there are only finitely many $d \in N$ with $t[d] < s$, t takes a minimum on N, say $t[d_1] = s_1$. Inductively, we define a semibasis $\mathscr{D} = \{d_1, \ldots, d_n\}$ of N by the requirement

$$t[d_m] = \min\{t[d] \mid d \in N \setminus [d_1, \ldots, d_{m-1}]\} ,$$

where $[d, \ldots, d_{m-1}]$ denotes the D-span of $d_1, \ldots, d_{m-1}$. Write $t[d_m] = s_m$; then $s_1 \leqslant s_2 \leqslant \ldots \leqslant s_n$. We say that $\mathscr{D}$ is *reduced with respect to* t. Now we make the change of variables which transforms d_i to the unit vector e_i and N to $L(\mathscr{D})$; the new form is again denoted t. We then have

$$t[x] \geqslant t[e_m] = s_m$$

for $x \in L(\mathscr{D}) \setminus [e_1, \ldots, e_{m-1}]$ that is, $(x_m, \ldots, x_n) \neq 0$. An arbitrary form satisfying these inequalities is called $L(\mathscr{D})$-*reduced*. We have now reached a point where we can state the finiteness theorems.

I. *There exist L-reduced forms for only finitely many admissible lattices L.*

In other words: if we fix N, but run over all positive F, only finitely many lattices $L(\mathscr{D})$ are produced by the method of successive minima.

If L is admissible, call

$$Z(L) := \{F \in H^+ \mid t_F \text{ is } L - \text{reduced}\}$$

the *cell* of L. If $Z(L)$ is not empty, it is defined by infinitely many inequalities.

II. *$Z(L)$ can actually be defined by finitely many of them.*

The proof also shows that different cells have disjoint sets of inner points.

We now come to what Weyl calls the "pattern of cells"; it is only here that our lattice group Γ comes into play. Every semibasis $\mathscr{D} = \{d_1, \ldots, d_n\}$ of N determines a cell $Z(\mathscr{D})$ of reduced forms:

$$F \in Z(\mathscr{D}) \Leftrightarrow t_F[x] \geqslant t_F[d_m] \quad \text{for all } x \in N \setminus [d_1, \ldots, d_{m-1}] .$$

If we associate with $Z(\mathscr{D})$ the admissible lattice $L(Z) = L(\mathscr{D})$, then

$$\begin{aligned} L(Z) = L(Z') &\Leftrightarrow L(\mathscr{D}) = L(\mathscr{D}') \\ &\Leftrightarrow \mathscr{D}' = s\mathscr{D}, \text{ some } s \in \Gamma \\ &\Leftrightarrow Z' = sZ , \end{aligned}$$

where Γ operates on the forms in the usual manner:

$$st(x) = t(s^{-1}x), \; x \in (D_{\mathbf{R}})^n .$$

Fix once and for all finitely many semibases $\mathscr{D}_1, \ldots, \mathscr{D}_r$ such that $L(\mathscr{D}_1), \ldots, L(\mathscr{D}_r)$ are all the admissible lattices having reduced forms. If F is any form, F determines a semibasis $\mathscr{D}$ such that $L(\mathscr{D})$ has a reduced form. Hence there is $s \in \Gamma$ and i such that $\mathscr{D} = s\mathscr{D}_i$ and $sF \in Z(\mathscr{D}_i)$. In other words, the union

$$Z_0 = \bigcup_i Z(\mathscr{D}_i)$$

is a fundamental domain for the operation of Γ on the space H^+. The "Third Theorem of Finiteness", or the "Theorem of Discontinuity", shows that Z_0 has only finitely many neighbors. More precisely, Weyl defines, for any given semibasis $\mathscr{D}$ and real numbers $p \geqslant 1$, $w > 0$, a subset $H(\mathscr{D}, p, w)$ of H^+ with the following properties:

(i) for $p > 1$, $w > 0$, $H(\mathscr{D}, p, w)$ contains an open neighborhood of $Z(\mathscr{D})$;

(ii) if $p > p'$, $w > w'$, then

$$H(\mathscr{D}, p, w) \supset H(\mathscr{D}, p', w'), \quad \text{and} \quad H^+ = \bigcup_{p,w} H(\mathscr{D}, p, w) .$$

III. *Given any cell Z, $\mathscr{D}, p$ and w, the set*

$$\{s \in \Gamma \mid sZ \cap H(\mathscr{D}, p, w) \neq \varnothing\}$$

is finite.

The latter clearly implies that

$$E(Z_0) = \{s \in \Gamma \mid sZ_0 \cap Z_0 \neq \varnothing\}$$

is finite. Let us check condition (ii) of the basic lemma. There is a union $\tilde{H}$ of finitely many $H(\mathscr{D}, p, w)$ containing an open neighborhood U of Z_0. Then there are only finitely many $s \in \Gamma$ with $sZ_0 \cap \tilde{H} \neq \varnothing$. If all of these are in $E(Z_0)$, $U \subset E(Z_0)Z_0$ because every point of U is a Γ-translate of a point of Z_0. Let $s_1, \ldots, s_r$ be those not in $E(Z_0)$. Since $s_i Z_0$ and Z_0 are disjoint, closed, and H^+ is a normal space, there is an open $U_i \supset Z_0$ with $U_i \cap s_i Z_0 = \varnothing$. Then we can take $\bigcap_i U_i$.

To sum up: for the operation of Γ on H^+ there is a closed connected fundamental domain with finitely many neighbors, satisfying condition (ii) of the basic lemma. The finite generation of Γ is thereby proved; in the next section we will also extract finite presentability from the reduction theory.

We now turn to the question of minimal dimension mentioned earlier. Our space H^+ is the image of $GL_n(D_{\mathbf{R}})$ under the map $A \to \bar{A}^t A$. According to (3),

$$GL_n(D_{\mathbf{R}}) \cong GL_{ns}(\mathbf{R})^{r_1'} \times GL_{ns/2}(\mathbf{H})^{r_1''} \times GL_{ns}(\mathbf{C})^{r_2},$$

and H^+ arises by dividing out the product of the orthogonal, symplectic, and unitary groups, respectively, which are maximal compact. For $\mathbf{K} \in \{\mathbf{R}, \mathbf{H}, \mathbf{C}\}$, the real dimensions of the maximal compact subgroup of $GL_m(\mathbf{K})$ are

$$\frac{m(m-1)}{2}, \quad m(2m+1) \text{ and } m^2.$$

A simple calculation now shows that

$$\dim H^+ = r_1' \frac{k(k+1)}{2} + r_1'' \frac{k(k-1)}{2} + r_2 k^2 \tag{4}$$

$$=: r(A) + 1$$

where $k = ns$. In view of $N\Gamma \subset \{\pm 1\}$, the number $r(A)$ may be called the *geometric unit rank* of A; of course, for $A = K$, that is, $k = 1$, it coincides with the unit rank $r(K) = r_1 + r_2 - 1$ in the sense of number theory. Siegel shows that $r(A)$ is in fact the minimal dimension for a discontinuous action of Γ in a sense which we now explain.

Let more generally G be a locally compact topological group with a countable basis for the topology, $H < G$ a discrete subgroup and ν a Haar measure. Suppose that F is a set of coset representatives of G/H such that (a) F is a Borel set, and (b) $\nu(F) < \infty$. Siegel's first main result is

THEOREM. *In this situation, H operates discontinuously on the homogeneous space $C\backslash G$ if and only if C is a compact subgroup of G.*

First we have to check the hypotheses. By what has been said about the cells, (a) is easy; (b) by no means. We only sketch the proof in the case of $SL_n(\mathbf{Z})$ (see [B2], 1.11). Of course, it suffices to show that the Siegel domain

$$S_{t,w} = SO(n) \cdot D_t \cdot N_w$$

has finite volume in the Haar measure. Transferring the Haar measure to the factors of the Iwasawa decomposition, this comes down to the finiteness of

$$\int_{D_t} p(a)\,da\,,$$

where da is the Haar measure on the torus and

$$p((a_i)) = \prod_{i<j} a_i/a_j\,;$$

and this is not hard.

REMARKS

(1) The general finiteness criterion for the fundamental domain of arithmetic groups is that the underlying algebraic group has no $\mathbf{Q}$-characters ([Bo2], 12.5); that is, "half" the compactness criterion.

(2) It seems that the exact value of the volume has not yet been calculated in the general case although Weyl ([W], p. 263) hints at the possibility. It is of course known for $SL_n(\mathbf{Z})$ and some other cases; we refer to [Te, 4.4.4].

The theorem now shows that Γ cannot operate discontinuously on homogeneous spaces of $GL_n(D_{\mathbf{R}})$ of smaller dimension; a result stated already by Eichler [E2]. Of course, this does not rule out Γ-operations on spaces of smaller dimension which do not extend to the surrounding Lie group. In fact, such operations may be viewed as the basis of the cohomological results to which we come later.

The following simplification, however, is near at hand. Let R be the integral domain of the central field K and $S\Gamma$ be kernel of the reduced norm map $Nr: A^\times \to K^\times$, restricted to Γ (we will recall the definition of Nr in section 9). Then $(R^{\times ns}) = NrR^\times \subset Nr\Gamma$, and one deduces that $S\Gamma \cdot R^\times$, an almost direct product, has finite index in Γ. Since we don't care about finite indices and consider $R^\times$ as known by Dirichlet's theorem, we may concentrate on $S\Gamma$. In our previous notation (3), $S\Gamma$ is a discrete subgroup of

$$\prod^{r_1'} SL_{ns}(\mathbf{R}) \times \prod^{r_1''} SL_{ns/2}(\mathbf{H}) \times \prod^{r_2} SL_{ns}(\mathbf{C})$$

(where for $\mathbf{H}$, SL denotes elements of GL of reduced norm 1). Dividing out the maximal compact subgroups, we find that $S\Gamma$ operates discontinuously on a homogeneous space of dimension

$$r(SA) := r(A) - r(K) ,$$

which may be called the "*reduced geometric unit rank of* A". Explicitly, inferring

$$r(K) = r_1' + r_1'' + r_2 - 1 ,$$

we obtain from (4) the formula

$$(5) \qquad r(SA) = r_1' \frac{(k+2)(k-1)}{2} + r_1'' \frac{(k-2)(k+1)}{2} + r_2(k-1)(k+1) .$$

We will go through the cases of small $r(SA)$ in the concluding section.

It is surprising how easily the existence of a fundamental domain with finitely many neighbors implies another finiteness theorem, which has already been mentioned:

THEOREM 2. Γ *contains only finitely many conjugacy classes of finite subgroups.*

Proof [B1]. Let $G = Gl_n(D_{\mathbf{R}})$ and C be the maximal compact subgroup used above. Let $H < \Gamma$ be a finite subgroup. Then H is contained in a maximal compact $\tilde{C}$, which is conjugate to C: $\tilde{C} = gCg^{-1}$. Then $Cg^{-1}\tilde{C} = Cg^{-1}$, so H fixes the point $P = Cg^{-1}$ of $C\backslash G = H^+$. Let $\gamma \in \Gamma$ be such that $P\gamma \in Z_0$, the fundamental domain. It follows that $P\gamma\gamma^{-1}H\gamma = P\gamma$, so $\gamma^{-1}H\gamma \subset E(Z_0)$, which is finite. (This proof holds for arbitrary arithmetic groups.)

4. PRESENTATIONS I: THE THEORY OF TRANSFORMATION GROUPS

We have already indicated that not only generators but also defining relations can be extracted from a "good" operation of Γ on a "good" space and that reduction theory provides us with both. The basic idea is already inherent in Poincaré's treatment of Fuchsian groups (see e.g. [F], p. 168 ff.). Gerstenhaber [G] established the abstract setting; later contributions are due