Progress in Mathematics

Volume 186

Ernst Kleinert

Units in Skew Fields

Springer Basel AG

Author:

Ernst Kleinert
Mathematisches Seminar
Universität Hamburg
Bundesstr. 55
20146 Hamburg
Germany

1991 Mathematics Subject Classification 11R52

A CIP catalogue record for this book is available from the Library of Congress,
Washington D.C., USA

Deutsche Bibliothek Cataloging-in-Publication Data

Kleinert, Ernst:
Units in skew fields / Ernst Kleinert. – Boston ; Basel ; Berlin : Birkhäuser, 2000
(Progress in mathematics ; Vol. 186)

ISBN 978-3-0348-9555-2 ISBN 978-3-0348-8409-9 (eBook)
DOI 10.1007/978-3-0348-8409-9

9 8 7 6 5 4 3 2 1

Contents

Introduction and Survey of Results

Unit groups of orders (in finite dimensional semisimple algebras over the rational field) are the first examples of arithmetic groups, and thereby form a topic of obvious interest, on which I have reported in some length in [Kl]; in fact they are, by genesis, a natural meeting ground of group theory, number theory and geometry. However, it may safely be said that all research done so far in this field, one half is devoted to a single member of the class, namely the modular group $SL_2(\mathbb{Z})$, whereas the major part of the second half deals with other matrix groups, such as the Hilbert modular groups or other Chevalley groups. In the algebraic background, which is concentrated in the theory of Brauer groups, the skew fields form the core of the subject; in the unit theory, they have been neglected almost completely, with the only exception of the quaternion case. The reason, of course, is plain: the underlying $\mathbb{Q}$-group is anisotropic over $\mathbb{Q}$; this makes it hard to write down elements of them, so that even the finite generation of the unit groups is far from obvious; also, it prevents one from attacking other questions, like reduction modulo primes, Strong Approximation or the congruence problem, in a straightforward manner. I found this state of affairs highly unsatisfactory, and my intention in writing this monograph was to stimulate the interest in the topic by presenting a synopsis of methods and results, including the proofs which could not be given in [Kl]. I have deliberately restricted the treatment to the simplest case, namely that of a skew field central over $\mathbb{Q}$ and splitting over $\mathbb{R}$; this keeps the technicalities at the absolute minimum, while the leading ideas become, perhaps, more visible. One can safely expect that new results in this case will generalize, general versions of those proved here can be found in the original sources, or partly in [Kl].

We now present in some detail the results we shall deal with. Let D be a skew field with center $\mathbb{Q}$ and of index d; we assume that D splits over $\mathbb{R}$. The kernel of the reduced norm $nr : D \to \mathbb{Q}$ defines a $\mathbb{Q}$-group G, which can be defined over $\mathbb{Z}$ such that $G(\mathbb{Z})$, also denoted Γ, is the norm-one group in some maximal order Λ.

§0 begins with a very explicit parametrization of the set of skew fields which are our theme (Corollary 0.3). This rests on deep theorems on simple algebras over global fields which we state without proof. We also obtain, essentially as a consequence of Eichler's Norm Theorem, that all maximal orders in D are conjugate: further, $|\Lambda^\times : \Gamma| = 2$.

§1 contains Hey's celebrated theorem that $G(\mathbb{R})/G(\mathbb{Z}) = SL_d(\mathbb{R})/\Gamma$ is compact, plus some immediate consequences.

In §2 we present the classical reduction theory in a version due to Weyl. The group Γ operates on a space H_1^+ of quadratic forms which can be identified with $SO_d(\mathbb{R}) \setminus SL_d(\mathbb{R})$. It is shown that there exists a fundamental domain F

for this operation, which is the intersection of finitely many halfspaces, and that $H_1^+ = \bigcup_{\gamma \in \Gamma} F\gamma$ is a locally finite cover. This implies that Γ is finitely presentable.

§3 brings the definition and calculation of the Tamagawa number,

$$\tau(G) = \mathrm{vol}\,(G(\mathbb{A})/G(\mathbb{Q})) = 1,$$

following Weil. From this one obtains a formula for $\mathrm{vol}(SL_d(\mathbb{R})/\Gamma)$ in terms of the zeta function of D (3.7).

In §4 we show that Γ is, in various respects, a "large" subgroup of $G(\mathbb{R})$: it is Zariski-dense (4.1), an almost maximal discrete subgroup (4.8), the reduction modulo primes is surjective almost everywhere (4.3), Γ is dense in $G(\hat{\mathbb{Z}})$, where $\hat{\mathbb{Z}} = \Pi_p \mathbb{Z}_p$ (4.4). The last two results are easy consequences of the Strong Approximation which holds for G (4.9); we present a proof which is due to Swan.

§5 is devoted to a finiteness theorem due to Margulis: if $d \geq 3$, every noncentral normal subgroup of Γ has finite index. (Note that $C(G) = 1$ if d is odd and $C(G) = \{\pm 1\}$ if d is even.) This is false in the quaternion case $d = 2$. For the long and difficult proof of this result, we have to report about amenable groups and Kazhdan's property (T); an appendix contains some facts from measure theory which are needed in the proof.

In §6 we construct a Zariski-dense subgroup of Γ, using a suitable realization of D as a cyclic crossed product, as well as a nonabelian free subgroup, using arguments due to Tits. This material is new.

In §7 an example is constructed, presumably the simplest one apart from the quaternion case, and various methods are shown for finding units. To the best of my knowledge, this is the first time that a skew field with $d > 2$ is considered *explicitly* (with respect to the unit group).

The final §8 briefly discusses some major problems: why it is difficult to find a system of generators of Γ, the state of art in the problem of congruence groups and our modest knowledge of Betti numbers.

Clearly it was not possible to develop everything from scratch. The reader is assumed to have a working knowledge of number theory, the theory of algebras and of linear algebraic groups; the first and third being indispensable for arithmetic groups in general, the second for the special ones which are our theme. In deciding where to cut off the argument I have tried to pursue the following policy: to present in full those proofs which are directly connected with the relevant results, and to merely report on facts which are of more general nature (exception: Lemma 6.6). Inevitably, this is not well defined; but I have taken pains to give detailed references.

The difficulty of the subject stems not only from the difficulty of the single results but also from the fact that there is little interconnection between them, and that a large variety of methods has to be employed.

For each major theme, one has to make a fresh start. Perhaps there is some unifying point of view which still lies in the obscure; but revealing it can only succeed by the sort of overview I have tried to give here. If I succeed in

attracting others to these difficult and fascinating mathematical objects, I shall feel richly rewarded.

I am indebted to F. Grunewald and L. Vaserstein for the hints they gave me; to A. Rapinchuk and A. Potapchik for communicating their new results to me; and to Peter Slodowy who had an open ear and good advice for many problems. Last not least I thank Ines Köwing and Alice Günther, who turned old-fashioned handwriting into beautiful TEX.

0 Basic Facts

We begin by presenting a survey of the variety of skew fields to which the subsequent discussion applies. Such a survey is provided by the meanwhile classical theory of simple algebras over global fields, which is intimately connected with class field theory. The reader is referred to [Re], §32 for a most readable account of this theory.

Let, as above, D be a skew field with center $\mathbb{Q}$ and of index d (so that $g = \dim_{\mathbb{Q}} D = d^2$). For the moment, we make no assumptions concerning the splitting. Let v be a prime of $\mathbb{Q}$ and denote by $\mathbb{Q}_v$ the completion. Then $D_v = \mathbb{Q}_v \otimes D$ is a central simple algebra over $\mathbb{Q}_v$, hence has the form

$$D_v = M_{n(v)}(D^{(v)}),$$

where $D^{(v)}$ is a skew field with center $\mathbb{Q}_v$ and index $d^{(v)}$ (say), which is connected with d by the equation $d = n(v)d(v)$. We say that D is *ramified* at v if $d(v) > 1$, and that D *splits* at v if $d(v) = 1$.

If $v = \infty$, then $\mathbb{Q}_v = \mathbb{R}$ and either

$$D^{(\infty)} = \mathbb{R} \quad \text{or} \quad D^{(\infty)} = \mathbb{H},$$

where $\mathbb{H}$ denotes the Hamilton quaternions. One shows that $\mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} \cong M_4(\mathbb{R})$; this proves that the Brauer group $Br(\mathbb{R})$ of the real field is cyclic of order 2. We identify it with $\langle \frac{1}{2} + \mathbb{Z} \rangle \subset \mathbb{Q}/\mathbb{Z}$ and define

$$\mathrm{inv}_\infty(D) = 0 \quad \text{or} \quad \mathrm{inv}_\infty(D) = \frac{1}{2} + \mathbb{Z},$$

according to whether $D^{(\infty)} = \mathbb{R}$ or $D^{(\infty)} = \mathbb{H}$.

If $v = p$ is a finite prime then by Hasse's theory of local skewfields $D^{(p)}$ is a crossed product

$$D^{(p)} = (\mathbb{Q}_p(\zeta)|\mathbb{Q}_p, \pi),$$

where ζ denotes a $(p^{d(p)} - 1)$th root of unity over $\mathbb{Q}_p$, and the element π satisfies

$$\pi^{d(p)} = p \quad \text{and} \quad \pi\zeta\pi^{-1} = \zeta^{p^{r(p)}}$$

for some $r(p) \in \mathbb{N}$ with $1 \leq r(p) < d(p)$ and $gcd(r(p), d(p)) = 1$. It turns out that the *Hasse invariant*

$$\mathrm{inv}_p(D) = \frac{r(p)}{d(p)} + \mathbb{Z} \in \mathbb{Q}/\mathbb{Z}$$

is independent of the choices of ζ and π in $D^{(p)}$ and leads to an isomorphism

$$Br(\mathbb{Q}_p) \cong \mathbb{Q}/\mathbb{Z},$$

which we treat as an identification.

For all primes v of $\mathbb{Q}$, we have now defined local invariants $\mathrm{inv}_v D \in Br(\mathbb{Q}_v) \leq \mathbb{Q}/\mathbb{Z}$. The map inv_v extends to a homomorphism $Br(\mathbb{Q}) \to Br(\mathbb{Q}_v)$. We can now formulate a fundamental result:

Theorem 0.1. (1) *For almost all* $v, d(v) = 0$ (i.e. D *splits almost everywhere*).

(2) *Define* $\mathrm{inv} = (\mathrm{inv}_v)_v$ *and let* $s : \bigoplus_v Br(\mathbb{Q}_v) \to \mathbb{Q}/\mathbb{Z}$ *be the summation map. Then the sequence*

$$(S) \qquad 1 \to Br(\mathbb{Q}) \xrightarrow{\mathrm{inv}} \bigoplus_v Br(\mathbb{Q}_v) \xrightarrow{s} \mathbb{Q}/\mathbb{Z} \to 0$$

is exact.

It is worthwhile to state explictly what the exactness says: the injectivity of *inv* means that the central simple $\mathbb{Q}$-algebra A splits over $\mathbb{Q}$ (i.e. $A \cong M_n(\mathbb{Q})$) if and only if this holds locally everywhere; in other words, for the splitting of A a local-global principle is valid. The inclusion im inv $\subset$ ker s is a reciprocity theorem for the local invariants inv_v; and the inclusion ker $s \subset$ im inv is an existence theorem for global skewfields. Before using (S) to derive the desired survey, we need

Theorem 0.2. *The exponent of the class of* D *in* $Br(\mathbb{Q})$ *equals the index* d.

Note that it is an easy consequence of (S) and the description of inv that the exponent divides the index. Now we easily obtain

Corollary 0.3. *The set of skew fields* D *of index* d *over* $\mathbb{Q}$ *which split over* $\mathbb{R}$ *is parametrized by the following data:*

(1) *A finite set* $p_1, \ldots, p_s$ *of finite primes;*

(2) *for each* p_i, *natural numbers* $r_i, d_i, 1 \leq r_i < d_i$ *with* $gcd(r_i, d_i) = 1$ *and* d_i *a divisor of* d, *which satisfy*

(3) $lcm(d_i) = d$ *and*

(4) $\sum \frac{r_i}{d_i} \in \mathbb{Z}$.

Proof. We have seen that a skew field D of the required sort gives rise to data (1) and (2), satisfying (4) by reciprocity. Condition (3) follows from 0.2. Conversely, given (1)–(4), there exists a skew field D having local invariants $\mathrm{inv}_{p_i}(D) = \frac{r_i}{d_i}$ and 0 elsewhere. In particular, D splits over $\mathbb{R}$. By the injectivity of inv, the exponent of D is the *lcm* of the exponents of the $D^{(p)}$, hence equals d by (2) and (3). By 0.2, the index of D is d. This proves 0.3.

Note that at least two p's are required to satisfy (4). In the quaternion case ($d = 2$) (1)–(4) simply amount to prescribing an *even* number of ramified primes.

From now on, we assume that D splits over $\mathbb{R}$, that is, $\mathrm{inv}_\infty(D) = 0$.

Next we recall the notions of *reduced norm* and *trace*: choose a splitting field K of D (e.g., $K = \mathbb{R}$ or any maximal subfield of D) and an isomorphism

$$\varphi : K \otimes_{\mathbb{Q}} D \cong M_d(K).$$

One can show that the functions $nr = \det \circ \varphi$ and $tr = \text{trace} \circ \varphi$ are independent of the choice of K and φ, have values in $\mathbb{Q}$ and are related to the corresponding *regular* notions by $N = (nr)^d, Tr = d \cdot tr$. The reduced norm nr generates the group of rational homomorphisms $D^\times \to \mathbb{Q}^\times$. We quote

Theorem 0.4.
(1) $nr : D^\times \to \mathbb{Q}^\times$ *is onto.*

(2) *Every integer is the reduced norm of some integral element of* D.

These are special cases of theorems due to Hasse-Schilling-Maass and Eichler; see [Re], 33.15 and 34.8.

We denote by G the affine algebraic group (scheme) ker nr. $G(\mathbb{Q})$ is often denoted as $SL_1(D)$; we shall also use the notation $G = D^{(1)}$. By 0.4 (1), we have an exact sequence

$$1 \to G(\mathbb{Q}) \to D^\times \to \mathbb{Q}^\times \to 1.$$

Let $\{\alpha_1, \ldots, \alpha_g\}$ be a $\mathbb{Q}$-base of D consisting of integral elements. Then the function $nr(x_1\alpha_1 + \cdots + x_g\alpha_g)$, with the x_i considered as variables, is a homogeneous polynomial of degree d with coefficients in $\mathbb{Z}$. This shows that G can be defined over $\mathbb{Z}$.

More specifically, we fix a maximal order Λ, a $\mathbb{Z}$-base $\{\alpha_1, \ldots, \alpha_g\}$ of Λ and the corresponding definition of G. The group $G(\mathbb{Z})$, also denoted Γ, is our main object of interest. Since an integral element is a unit if and only if its norm is a unit, we have $G(\mathbb{Z}) \subset \Lambda^\times$; we shall see below that $|\Lambda^\times : G(\mathbb{Z})| = 2$. $G(\mathbb{Z})$ is an arithmetic subgroup of $G(\mathbb{R}) = SL_d(\mathbb{R})$. If $d \geq 3$, every discrete and cocompact subgroup of $SL_d(\mathbb{R})$ is commensurable to a (unique) $G(\mathbb{Z})$; this follows from Margulis' arithmeticity theorem ([Zi], Thm. 6.1.2). For $d = 2$, this is not true.

The (reduced) *discriminant* of Λ is defined as

$$d(\Lambda) = \det\left(tr(\alpha_i\alpha_j)\right)_{i,j}.$$

One has the formula

$$|d(\Lambda)| = \prod_p p^{d(d-n(p))}$$

(see [Re], §25), which makes the statement (1) in 0.1 more precise: the ramified p are exactly the divisors of $d(\Lambda)$. Note that the p-contribution to $d(\Lambda)$ depends

only on the ramification index $d(p)$; this is not so in the case of a number field. If d is a prime, then all ramified p have the same exponent $d(d-1)$ in $d(\Lambda)$.

The ideal theory of Λ is very easy:

Theorem 0.5. *Every left or right ideal of Λ is principal.*

This is an immediate consequence of Eichler's norm theorem ([Re] 34.9).

Now let Λ_1 and Λ_2 be two maximal orders and let A_{12} be any Λ_1-left and Λ_2-right ideal; e.g. $A_{12} = \Lambda_1\Lambda_2$. By the general theory of ideals of maximal orders ([Re], §22), we then have

$$A_{12}A_{12}^{-1} = \Lambda_1,\ A_{12}^{-1}A_{12} = \Lambda_2.$$

But viewing A_{12} as Λ_1-left ideal and writing $A_{12} = \Lambda_1 c, c \in D^\times$, by 0.5, we obtain

$$\Lambda_2 = c^{-1}\Lambda_1^{-1}\Lambda_1 c = c^{-1}\Lambda_1 c.$$

Corollary 0.6. *All maximal orders in D are conjugate.*

In this sense, statements on Λ are statements on D. We remark that generally, in simple algebras over number fields, there are finitely many conjugacy classes of maximal orders (see [D], p. 89).

Corollary 0.7. $|\Lambda^\times : \Gamma| = 2$.

Proof. We must show that Λ contains an element ε with $nr\varepsilon = -1$. Now there is, by 0.4(2), an integral $\mu \in D$ with $nr\,\mu = -1$. But every integral element is contained in some maximal order. Now Corollary 0.6 proves the claim.

1 Hey's Theorem and Consequences

Our first result, proved by Käte Hey in her doctoral thesis (Hamburg 1929), is surely the most popular theorem on D and Γ and has been reproduced at various places (cf. [Kl]). Meanwhile, it has been superseded by a very general criterion ([PR], 4.5); but we cannot possibly leave it out here.

Theorem 1.1. *$G(\mathbb{R})/\Gamma$ is compact.*

Proof ([Za]). Working with a $\mathbb{Z}$-base of Λ we identify $\Lambda = \mathbb{Z}^g \subset \mathbb{R}^g = D_{\mathbb{R}}$. Let C be any convex, compact, 0-symmetric subset of $\mathbb{R}^g$ with $\operatorname{vol} C > 2^g$. By Minkowski's lattice point theorem C contains a nonzero $a \in \Lambda$. If $x \in G(\mathbb{R})$, then $\operatorname{vol} Cx = \operatorname{vol} C$, and Cx contains a nonzero $a_x \in \Lambda$.

Now let (x_n) be a sequence in $G(\mathbb{R})$. Then there are $a_n \in \Lambda \backslash \{0\}$ such that

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

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

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

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

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

hence

$$a_k = a_1 \varepsilon_k, \quad \varepsilon_k \in \Lambda^\times.$$

Further,

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

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

$$x_e \varepsilon_e^{-1} = c_e^{-1} a_1$$

we now read off that $G(\mathbb{R})/\Gamma$ is (sequentially) compact. Note that we have used only a qualitative form of the lattice point theorem in that there was no need to specify C.

We remark that we have not used the fact that $C(D) = \mathbb{Q}$; the proof is valid for arbitrary skewfields of finite dimension over $\mathbb{Q}$. If D is a number field, the result includes the difficult part of Dirichlet's unit theorem; see [Kl], p. 215.

Now let $\Gamma_0 < \Gamma$ be a torsionfree subgroup of finite index (e.g. a congruence group, or $\Gamma_0 = \Gamma$ if d is odd), and put $C = SO_d(\mathbb{R})$, a maximal compact subgroup. Then $\Gamma_0 \cap C = 1$, and it follows that Γ_0 operates without fixed points on

$$Y := C\backslash G(\mathbb{R}).$$

(In the next paragraph we will see that this operation is also discontinuous; see 2.10.) This implies that

$$Y \to Y(\Gamma_0) := Y/\Gamma_0$$

is an unramified covering of manifolds. Since (as is well known) Y is contractible, in particular simply connected, this is the universal covering of $Y(\Gamma_0)$, and Γ_0 identifies with the fundamental group $\pi_1(Y(\Gamma_0))$. By 1.1, $Y(\Gamma_0)$ is compact.

Proposition 1.2. *The fundamental group of a compact connected manifold X is finitely generated.*

Proof. Write $\pi = \pi_1(X)$ and let

$$p : \tilde{X} \to X$$

be the universal covering. We can view π as the group of deck transformations of p, so that $X = \tilde{X}/\pi$.

For each $x \in X$ choose a simply connected open neighborhood V_x such that $p^{-1}(V_x)$ is a disjoint union $\bigcup_\alpha U_x^{(\alpha)}$ and p induces homeomorphisms $U_x^{(\alpha)} \tilde{\to} V_x$. Choose one U_x among the $U_x^{(\alpha)}$. Since X is compact, we can find a finite set I such that

$$X = \bigcup_{i \in I} V_i,$$

where we have written $V_i = V_{x_i}$. Likewise, we write $U_i = U_{x_i}$.

It now follows from the construction of $\tilde{X}$ and the operation of π on $\tilde{X}$ that, whenever $V_i \cap V_j \neq \emptyset$, there is a *unique* $\gamma_{ij} \in \pi$ such that $U_i\gamma_{ij} \cap U_j \neq \emptyset$. We show that the γ_{ij} generate π.

Fix an $i_0 \in I$ and let $\gamma \in \pi$ be given. Since

$$\tilde{X} = \bigcup_i U_i\pi$$

is connected, we can join $U_{i_0}\gamma$ and U_{i_0} by a chain of subsets $U_{i_r}\gamma_r$, in other words, we can find a chain

$$(i_0, \gamma_0), \ldots, (i_k, \gamma_k) \in I \times \pi$$

such that $i_k = i_0, \gamma_k = \gamma, \gamma_0 = id$ and

$$U_{i_r}\gamma_r \cap U_{i_{r+1}}\gamma_{r+1} \neq \emptyset.$$

It follows that

$$\gamma_{r+1}\gamma_r^{-1} = \gamma_{i_{r+1}i_r},\ 0 \leq r < k,$$

and

$$\gamma = \gamma_k = \gamma_{i_k i_{k-1}}\gamma_{i_{k-1}i_{k-2}} \cdots \gamma_{i_1 i_0}.$$

This proves 1.2. We remark that the data $\{(i,j)|V_i \cap V_j \neq \emptyset\}$ even allow the derivation of a finite presentation of π; see [Rag], Thm. 6.15.

Corollary 1.3. Γ *is finitely generated.*

Our second application of 1.1 concerns the cohomology. (The reader is referred to [Br] and [Se] for a general discussion.) Let Γ_0 be as above. Since $Y(\Gamma_0)$ is a (smooth) manifold, it can be endowed with the structure of a CW-complex. It follows that Y inherits a CW-structure in such a way that Γ_0 operates freely on the cells; see [Br], I.4.) This means that the augmented cellular chain complex of Y is a free resolution of the trivial Γ_0-module $\mathbb{Z}$, whence

Corollary 1.4. $H^*(\Gamma_0, M) = H^*(Y(\Gamma_0), M)$ *for every* Γ_0*-module* M *with trivial action. In particular,*

$$vcd\,\Gamma = cd\,\Gamma_0 = \dim Y(\Gamma_0) = \frac{d(d+1)}{2} - 1.$$

Here, $(v)cd$ means (virtual) cohomological dimension; $vcd\,\Gamma$ is *defined* to be $cd\,\Gamma_0$, and it is a fact that this number is independent of the choice of Γ_0; cf. [Se], 1.8.

The above is almost everything we know about $H^*(\Gamma_0)$, in case $d \geq 3$. In §8.3 we shall briefly address the question of the Betti numbers.

2 Siegel-Weyl Reduction Theory

In its simplest form, reduction theory is the study of the quotient $H(\mathbb{R})/H(\mathbb{Z})$, where H denotes a linear algebraic group defined over $\mathbb{Q}$. The basic content of this study is the construction of "good" fundamental domain for the operation of $H(\mathbb{Z})$ on $H(\mathbb{R})$ by multiplication, or rather the proof of the existence of such a domain; this yields certain finiteness results, among which the finiteness of $\mathrm{vol}(H(\mathbb{R})/H(\mathbb{Z}))$ and the finite presentability of $H(\mathbb{Z})$ are the most important ones. By the fundamental work of Borel and Harish-Chandra [BH], reduction theory has become a standardized topic in the arithmetic theory of algebraic groups, see [PR]. The case of units of orders had been settled before by Siegel [Si 1] and was treated again, in a somewhat different and more extensive manner, by Weyl [Wey]. The justification of the present chapter (which follows Weyl) lies in the considerable simplifications arising in our case; all one needs is some elementary linear algebra and, of course, Minkowski's lattice point theorem. This fortunate circumstance is due to the fact that the process of Jacobi transformation degenerates to identity.

The treatment is based on the appropriate concept of quadratic form, due to Siegel. Choose an isomorphism $\varphi : D_{\mathbb{R}} \cong M_d(\mathbb{R})$. (Recall that any two such φ differ by an inner automorphism.) Define, for $x \in D_{\mathbb{R}}$,

$$\overline{x} = \varphi^{-1}(\varphi(x)^t).$$

Clearly, $x \to \overline{x}$ is an involutory anti-automorphism of $D_{\mathbb{R}}$. Call x *symmetric* if $x = \overline{x}$. Now a *quadratic form in one variable* (!) *over* $D_{\mathbb{R}}$ is simply a symmetric $Q \in D_{\mathbb{R}}$. For $x \in D_{\mathbb{R}}$ we set

$$Q[x] = \overline{x}Qx,$$

a symmetric element of $D_{\mathbb{R}}$. Q is called *positive* ($Q > 0$) if $\varphi(Q)$ is a positive definite matrix. Now let R be the regular representation of $D, Tr = \mathrm{trace} \circ R$ its trace and define, for $x \in D_{\mathbb{R}}$,

$$t_Q[x] = TrQ[x] = d \cdot tr(\overline{x}Qx),$$

where the latter equality follows from the definition of the reduced trace. If $Q > 0$ and $x \neq 0$, then $t_Q[x] > 0$. Thus t_Q is a positive quadratic form, attached to the symmetric bilinear form

$$(x, y) \to Tr(\overline{x}Qy)$$

on the real vector space $D_{\mathbb{R}}$. Two remarks are in order. First, it seems unnatural to use Tr instead of tr, as is common in noncommutative arithmetic. Our reason is that the regular norm N will appear inevitably, and we shall have to apply the

arithmetic-geometric inequality to the eigenvalues; so the use of Tr simplifies the formulas. Second, we do not view φ as an identification. The reason is that we wish to identify Λ with the standard lattice $\mathbb{Z}^g$ later, but clearly we cannot identify Λ and $M_d(\mathbb{Z})$. Rather, we keep fixed the involutory antiautomorphism $x \to \overline{x}$ of $D_{\mathbb{R}}$. It is immediate that $Q > 0$ if and only if $Q = \overline{P}P$ for some $P \in D_{\mathbb{R}}^{\times}$. This shows that positiveness depends only on the involution. The positive Q form a space H^+, on which Γ operates from the right by the usual formula

$$(Q, \gamma) \to \overline{\gamma} Q \gamma, \ \gamma \in \Gamma.$$

It is clear that this space is identified via φ with the space of positive forms on $\mathbb{R}^d$, i.e. the space of symmetric positive $d \times d$-matrices, over $\mathbb{R}$; we recall that this space in turn, is identified with the homogeneous space $O_d(\mathbb{R})\backslash GL_d(\mathbb{R})$ via the map $g \to g^t g, g \in GL_d(\mathbb{R})$.

Let now $Q > 0$. Since Λ is a (full) lattice in $D_{\mathbb{R}}$, t_Q takes a minimum on $\Lambda\backslash\{0\}$, say

$$t_Q[d] = \min\{t_Q[x] | 0 \neq x \in \Lambda\} =: t > 0.$$

Put $\tilde{Q} = \overline{d} Q d$ and consider the function

$$t_{\tilde{Q}}[x] = t_Q[dx]$$

on $D_{\mathbb{R}}$. It takes the minimum

$$t_{\tilde{Q}}[1] = Tr(\tilde{Q}) = t_Q[d] = t$$

on $d^{-1}\Lambda\backslash\{0\}$. If we consider d as given, we call such $t_{\tilde{Q}}$ $d^{-1}\Lambda$-*reduced.* This change of variables appears to be a mere technicality; but its simplifying power is enormous.

Theorem 2.1. (First finiteness theorem) *There is a constant* $c_1 = c_1(D) > 0$ *depending only on* D *such that* $|d^{-1}\Lambda : \Lambda| < c_1$ *if* $d^{-1}\Lambda$ *possesses reduced forms.*

Proof. We continue with the above notation and now identify Λ with the standard lattice $\mathbb{Z}^g$ in $D_{\mathbb{R}}$. The ellipsoid $S_{\tilde{Q},t}$ defined by

$$t^{-1} \cdot t_{\tilde{Q}}[y] < 1$$

contains no nonzero point from the lattice $d^{-1}\Lambda$, as is clear from the definitions. By the lattice point theorem

$$\operatorname{vol} S_{\tilde{Q},t} \cdot |d^{-1}\Lambda : \Lambda| \leq 2^g. \tag{1}$$

The volume is calculated by the following elementary result, whose proof is left to the reader:

Lemma 2.2. *Let* B *be a positive symmetric bilinear form on* $\mathbb{R}^g$, *and put*

$$S_B = \{x | B(x, x) < 1\}.$$

Then

$$\operatorname{vol} S_B = \frac{\alpha_g}{\sqrt{\Delta}},$$

where α_g *is the volume of the* g*-dimensional unit ball and*

$$\Delta = \det(B(e_i, e_j))$$

is the discriminant of B *with respect to the standard base.*

In our case, we have

$$B(x, y) = t^{-1} \cdot Tr(\overline{x}\tilde{Q}y)$$

and for the calculation of Δ we now can take any $\mathbb{Z}$-basis $\{\omega_1, \ldots, \omega_g\}$ of Λ. Then, by routine arguments,

$$\begin{aligned} \det\left(t^{-1}Tr(\overline{\omega}_i\tilde{Q}\omega_j)\right) &= t^{-g}\det(Tr(\overline{\omega}_i\omega_j))\det R(\tilde{Q}) \\ &= \pm t^{-g}\det(Tr(\omega_i\omega_j))N(\tilde{Q}) \\ &= \pm t^{-g}D(\Lambda)N(\tilde{Q}), \end{aligned}$$

where now $D(\Lambda)$ is the *regular* discriminant of Λ; we have used the fact that the linear map $x \to \overline{x}$ has determinant ± 1. All in all, we obtain

$$\operatorname{vol} S_{\tilde{Q},t} = \alpha_g \sqrt{\frac{t^g}{|D(\Lambda)|N(\tilde{Q})}}$$

Substituting this in (1) gives

$$|d^{-1}\Lambda : \Lambda|^2 \leq \frac{4^g}{\alpha_g^2} \frac{|D(\Lambda)|N(\tilde{Q})}{t^g} \tag{2}$$

Now recall that

$$t = Tr(\tilde{Q})$$

and apply the arithmetic-geometric inequality to the eigenvalues of $R(\tilde{Q})$. This gives

$$\frac{1}{g}t \geq \sqrt[g]{N(\tilde{Q})}$$

or

$$N(\tilde{Q}) \leq (\frac{1}{g}t)^g.$$

Thus we have from (2)

$$|d^{-1}\Lambda : \Lambda|^2 \leq |D(\Lambda)|(\frac{4}{g})^g \frac{1}{\alpha_g^2}. \tag{3}$$

Since $D(\Lambda)$ depends only on D, the theorem is proved.

We remark that a result of Blichfeldt ([Bl]) allows to replace α_g by the larger constant

$$\pi_g = \frac{\alpha_g}{g+2} 2^{\frac{g+2}{2}}.$$

More important is the following observation. $d^{-1}\Lambda$ is a (fractional) Λ-right ideal, and there are only finitely many such ideals with indices bounded as in (3). Moreover, $d_1^{-1}\Lambda = d_2^{-1}\Lambda$ if and only if $d_2 = \gamma d_1$ for some $\gamma \in \Lambda^\times$. Thus there are, up to $\Lambda^\times$-action, only finitely many $d_1, \ldots, d_r \in \Lambda$ at which the t_Q, $Q > 0$, attain their minimum on Λ. These may be called *critical*. It is near at hand to ask for further arithmetic significance of these elements.

Choose a full set of critical d_i; and attach to each of them the *cell* $Z(d_i)$ of positive Q, for which t_Q is Λ-minimal at d_i. If t_Q is Λ-minimal at d, then

$$d = \gamma d_i \quad \text{for some} \quad \gamma \in \Lambda^\times \quad \text{and} \quad i = 1, \ldots, r;$$

equivalently, $t_{\overline{\gamma}Q\gamma}$ is Λ-minimal at d_i, hence $\overline{\gamma}Q\gamma \in Z(d_i)$. Thus, the space H^+ of positive Q is the union

$$H^+ = \bigcup_{\gamma,i} \overline{\gamma} Z(d_i) \gamma,$$

or, if we set $Z = \bigcup_i Z(d_i)$,

$$H^+ = \bigcup_{\gamma} \overline{\gamma} Z \gamma. \tag{4}$$

Our next results will show that the $\Lambda^\times$-translates of Z overlap at most on boundaries, and that (4) is a locally finite cover of H^+.

Suppose that the cell $Z(d)$ is not empty. It is defined by infinitely many inequalities:

$$Z(d) = \{Q | t_Q[x] \geq t_Q[d], \quad \text{all} \quad 0 \neq x \in \Lambda\}.$$

For each x, this is a *linear* inequality on the space of forms, which we write

$$L(x)(Q) = t_Q[x] - t_Q[d] \geq 0.$$

We want to show that actually finitely many of them suffice to determine $Z(d)$. Call $x \in \Lambda$ *essential* if there is $Q \in Z(d)$ with

$$L(x)(Q) = 0, \quad \text{i.e.} \quad t_Q[x] = t_Q[d].$$

Lemma 2.3. *$Q \in Z(d)$ if and only if $L(x)(Q) \geq 0$ for the essential x.*

Proof. The "only if" part is trivial. So assume $Q \notin Z(d)$. We must show that Q violates at least one of the essential inequalities. There is $y \in \Lambda$ with

$$L(y)(Q) < 0, \quad \text{i.e.} \quad t_Q[y] < t_Q[d].$$

This is possible for only finitely many y; the hyperplanes $L(y) = 0$ separate Q from $Z(d)$. Take any $Q_0 \in Z(d)$. The segment

$$\{uQ + (1-u)Q_0 | u \in [0,1]\} \subset H^+$$

intersects the hyperplane $L(y) = 0$ in a point with parameter $u(y)$. Let $\tilde{u} = u(\tilde{y})$ be the minimum of these. Then

$$Q_1 = \tilde{u}Q + (1-\tilde{u})Q_0$$

is still $\in Z(d)$, because $\tilde{u}$ was minimal, and $\tilde{y}$ is essential, because $L(\tilde{y})(Q_1) = 0$. Since $L(\tilde{y})(Q) < 0$, the lemma is proved.

Theorem 2.4. (Second finiteness theorem) *There is a constant $c_2 = c_2(d)$ such that*

$$Tr(\overline{x}x) < c_2$$

for the essential x.

Since this leaves only finitely many possibilities for x, we have the

Corollary 2.5. *Every nonempty cell is a convex pyramid, i.e. the intersection of finitely many halfspaces.*

The proof of 2.4. depends on some more or less elementary facts on matrices:

Lemma 2.6. *Let Q be positive, and let $0 < r^{(1)} \leq \cdots \leq r^{(g)}$ be the eigenvalues of $R(Q)$. Then*

i) *for all x,*

$$r^{(1)}Tr(\overline{x}x) \leq Tr(\overline{x}Qx) \leq r^{(g)}Tr(\overline{x}x);$$

ii) *suppose there are $t > 0, 0 < c \leq 1$ such that*

$$\frac{1}{g}Tr(Q) \leq t,\ N(Q) \geq ct^g.$$

Then there are $b = b(c)$, $B = B(c)$ such that

$$bt \leq r^{(i)} \leq Bt,\ i = 1, \ldots, g.$$

Explicitly, b and B are the zeroes of

$$x\left(\frac{g-x}{g-1}\right)^{g-1} - c$$

in the interval $[0, g]$.

Proof. We leave i) to the reader. In ii), we may assume that $t = 1$. Then from $r^{(1)}r^{(2)} \dots r^{(g)} \geq c$ it follows that

$$r^{(1)}\left(\frac{r^{(2)} + \dots + r^{(g)}}{g-1}\right)^{g-1} \geq c.$$

Since

$$r^{(2)} + \dots + r^{(g)} \leq g - r^{(1)},$$

this implies that, if we define

$$f(x) = x\left(\frac{g-x}{g-1}\right)^{g-1},$$

we have

$$f(r^{(1)}) \geq c. \tag{5}$$

Now the logarithmic derivative

$$\frac{1}{f(x)}\frac{df}{dx} = \frac{g}{x}\frac{1-x}{g-x}$$

shows that f increases monotonely on the interval $[0,1]$ and decreases on $[1,g]$. Therefore, if b and B are the roots of $f(x) - c = 0$ in $[0,g]$, equation (5) implies $b \leq r^{(1)} \leq B$. Since we have not used the ordering of the $r^{(i)}$ in this argument, the same holds for all of them.

Proof of 2.4. As in the proof of 2.1 we transfer the situation to $d^{-1}\Lambda$ and write $\tilde{Q} = Q$ for simplicity. Thus, let $x \in d^{-1}\Lambda$ be essential and Q be $d^{-1}\Lambda$-reduced, i.e.

$$t_Q[x] = t_Q[1] = Tr(Q) =: t$$

is the minimum of t_Q on $d^{-1}\Lambda\backslash\{0\}$. Inequality (2) reads

$$N(Q) \geq c\left(\frac{Tr(Q)}{g}\right)^g$$

with a constant c depending on D. Applying 2.6 i) to Q, we have

$$t = Tr(\overline{x}\,Qx) \geq r^{(1)}Tr(\overline{x}x);$$

applying 2.6. ii), with t replaced by t/g, yields

$$r^{(1)} \geq \frac{b(c)}{g}\,t. \tag{6}$$

Together these imply

$$t \geq \frac{b(c)t}{g}\,Tr(\overline{x}x) \qquad \text{or} \qquad Tr(\overline{x}x) \leq g\,b(c)^{-1}.$$

This completes the proof.

An element $x \in D_{\mathbb{R}}$ is called *unitary* if $\overline{x}x = 1$ or $\overline{x} = x^{-1}$. The unitary elements in Λ from a finite subgroup U of $\Lambda^{\times}$. Since $t_Q(d) = t_Q(dx)$ for x unitary, we have $Z(d) = Z(dx)$. Suppose $Z(d) \neq \emptyset$. The *core* of $Z(d)$ is the set of $Q \in Z(d)$ such that

$$t_Q[y] > t_Q[d] \text{ if } y \notin dU;$$

in other words, Q takes its Λ-minimum on dU *only*. It is clear that such Q belong to no other cell. We now show that the interior $\overset{\circ}{Z}(d)$ of $Z(d)$ belongs to the core.

$Z(d)$ is defined by inequalities $L(x)(Q) \geq 0$, $x \in \Lambda$. It can happen that $L(x)$ vanishes identically on H^+.

Lemma 2.7. $L(x) \equiv 0$ *if and only if* $x \in dU$.

Proof. Again, we pass to $d^{-1}\Lambda$. Then the hypothesis is

$$Tr(\overline{x}Qx) = Tr(Q), \quad \text{all } Q \in H^+.$$

Taking $Q = 1$, we get

$$Tr(\overline{x}x) = g,$$

then $Q = x\overline{x}$ gives

$$Tr(\overline{x}x\overline{x}x) = Tr(x\overline{x}) = Tr(\overline{x}x) = g.$$

So $y = \overline{x}x - 1$ has $Tr(y^2) = 0$; but $y = \overline{y}$ and $Tr(y^2) = Tr(\overline{y}y) = 0$ implies $y = 0$, i.e. $\overline{x}x = 1$. This proves the lemma.

The lemma shows that

$$\overset{\circ}{Z}(d) = \{Q | L(x)(Q) > 0 \text{ for } x \notin dU\},$$

and we conclude

Corollary 2.8. *Different cells have disjoint sets of inner points.*

In order to prove that (4) is a locally finite cover of H^+, we define, for $c \in \Lambda$ and $p \geq 1$,

$$Z(c,p) = \{Q | t_Q[x] \geq \frac{1}{p} t_Q[c] \text{ for } 0 \neq x \in \Lambda\}.$$

Clearly $Z(c,1) = Z(c), Z(c,p) \subset Z(c,q)$ if $p < q$ and $H^+ = \bigcup_p Z(c,p)$.

Theorem 2.9. (Third finiteness theorem) *Given* $d, c \in \Lambda, p \geq 1$, *there are only finitely many* $s \in \Lambda^{\times}$ *such that*

$$Z^s(d) \cap Z(c,p) \neq \emptyset.$$

Proof. Once more, we work in $d^{-1}\Lambda$. Then $Q \in Z^s(c,p)$ means

$$Tr(\overline{x}\,\overline{s}Qsx) \geq \frac{1}{p} Tr(\overline{c}\,\overline{s}\,Qsc),\ 0 \neq x \in d^{-1}\Lambda.$$

Put $x = s^{-1}$, $y = sc$ to obtain

$$Tr(\overline{y}\,Qy) \leq p(TrQ). \tag{7}$$

If Q lies also in $Z(d)$, we know

$$r^{(1)} \geq g^{-1}bt \tag{8}$$

where $t = Tr(Q)$ and b is a constant. From (6), (7) and Lemma 2.6. i) we have

$$Tr(\overline{y}y)g^{-1}bt \leq r^{(1)}Tr(\overline{y}y) \leq Tr(\overline{y}\,Qy) \leq pt$$

or

$$Tr(\overline{y}y) \leq b^{-1}pg.$$

This leaves only finitely many possibilities for $y = sc$ and hence for s. 2.9. is proved.

Let $c \in \Lambda$ be given. The function which assigns to $Q \in H^+$ the minimal p such that $Q \in Z(c,p)$ is continuous. This implies that, if p is large enough, a neigborhood V of Q will be contained in $Z(c,p)$. 2.9 now shows that $V \cap Z(d) \neq \emptyset$ for only finitely many cells. This proves

Corollary 2.10. (4) *is a locally finite cover of* H^+.

Another consequence of 2.9 is that

$$E(Z) = \{s \in \Lambda^\times | Z \cap Z^s \neq \emptyset\}$$

is a finite set. Writing $\Delta := \Lambda^\times/\{\pm 1\}$, we now can resume our results as follows:

(a) The group Δ operates faithfully on the connected and simply connected space $X = H^+$;

(b) there is a connected closed set $Z \subset X$ such that $Z\Delta = X$;

(c) the set $E(Z) = \{\delta \in \Delta | Z \cap Z\delta \neq \emptyset\}$ is finite.

A well-known group theoretical mechanism now provides us with the "canonical" proof of

Corollary 2.11. Δ *is finitely presentable, with* $E(Z)$ *as a generating set.*

Proof. Put $\Delta_1 = \langle E(Z)\rangle$. Then $X = Z\Delta_1 \cup Z(\Delta\backslash\Delta_1)$; if $z\delta_1 = z'\delta$ for $z, z' \in Z, \delta_1 \in \Delta_1$, then $\delta\delta_1^{-1} \in \Delta_1$, hence $\delta \in \Delta_1$, and the above union is disjoint. Since both parts are closed and $\Delta_1 \neq \emptyset$, we conclude $\Delta\backslash\Delta_1 = \emptyset$.

A finite presentation can be constructed as follows: let $\tilde{\Delta}$ be the abstract group with generators x_δ for $\delta \in E(Z)$ and relations $x_{\delta_1} x_{\delta_2} (x_{\delta_1 \delta_2})^{-1} = 1$ if δ_1, δ_2 and $\delta_1 \delta_2 \in E(Z)$. Then there is a surjective homomorphism $\varphi : \tilde{\Delta} \to \Delta$, sending x_δ to δ. One can now show that (a) and (b) imply the injectivity of φ; the reader is referred to [Kl] and the references quoted there.

As another consequence of the reduction theory, we obtain

Corollary 2.12. $\Lambda^\times$ *contains only finitely many conjugacy classes of finite subgroups.*

Proof (Borel [Bo 1]). By linear algebra, we can identify

$$H_1^+ = C \backslash SL_d(\mathbb{R}),$$

where $C = SO_d(\mathbb{R})$. Let $H \subset \Lambda^\times$ be a finite subgroup. Then H is contained in a maximal compact subgroup C_0 of $SL_d(\mathbb{R})$, which is conjugate to $C, C_0 = gCg^{-1}$. It follows that $Cg^{-1}C_0 = Cg^{-1}$, so H fixes the point $P = Cg^{-1}$ of H_1^+. Let $\gamma \in \Gamma$ so that $P\gamma \in Z_1$. Then $P\gamma\gamma^{-1}H\gamma = P\gamma$, so $\gamma^{-1}H\gamma$ fixes $P\gamma$, in particular $\gamma^{-1}H\gamma \in E(Z_1)$, which is a finite set. This proves 2.12.

We have been working in the space H^+ of positive forms of arbitrary determinant because we used convexity in the proof of 2.3. Intersecting everything with the hypersurface H_1^+ of positive forms of determinant 1, we obtain a locally finite cover of H_1^+ by $\Lambda^\times$-translates of $Z_1 = Z \cap H_1^+$. By Theorem 1.1, Z_1 is *compact.*

In §8,1 it will be sketched how to extract an algorithm from the reduction theory presented above. It should be noted that the standard theory works with *Siegel domains* which have the advantage that they are explicitly given and of finite volume; the drawback is that they *contain* fundamental domains but are strictly larger.

3 The Tamagawa Number and the Volume of $G(\mathbb{R})/G(\mathbb{Z})$

3.1 Statement of the main result

In this chapter we present, following Weil [Wei], the definition and calculation of the Tamagawa number $\tau(G)$ of the algebraic $\mathbb{Q}$-group G. Let $\mathbb{A}$ be the adele ring of $\mathbb{Q}$. By definition,

$$\tau(G) = \mathrm{vol}\,(G(\mathbb{A})/G(\mathbb{Q}))$$

with respect to the Tamagawa measure, to be defined below. The main result, $\tau(G) = 1$, is easily translated into the relation

$$\mathrm{vol}(G(\mathbb{R})/G(\mathbb{Z})) = \Big(\prod_p \mathrm{vol}\,(G(\mathbb{Z}_p))\Big)^{-1};$$

it is not hard to calculate the right-hand side, whereas the left-hand side is inaccessible in the direct way, which would consist in constructing a fundamental domain and integrating over it. (For SL_n, this can be done, see [Si 2].) The purpose of this chapter is to make the reader understand the miracle that $\tau(G)$ can be determined in spite of this.

Weil's proof is, in some sense, indirect and involves other algebraic groups, namely $D^\times$, the multiplicative group of D, and $Z^\times$, the multiplicative group of the center of D, which is identified with G_m. For this reason we have to state the basic definitions and facts, as well as a number of results of a more technical nature, in some generality. For a fuller discussion, and for all proofs omitted here, the reader is referred to [Wei] and [PR]. *For the rest of this subsection, and the next one, we depart form our usual notation, writing $D^{(1)}$ instead of G and reserving G for a general group.*

Let V be a nonsingular variety defined over $\mathbb{Q}$. A *gauge form* on V is a differential form of degree $n = \dim V$, nonzero and without poles. A gauge form $\omega = f(x)dx_1 \cdots dx_n$ induces, for each place v of $\mathbb{Q}$, the measure

$$\omega_v = |f(x)|_v (dx_1)_v \cdots (dx_n)_v$$

on the $\mathbb{Q}_v$-variety V_v; here the measures $(dx)_v$ on $\mathbb{Q}_v$ are $(dx)_\infty = dx =$ Lebesgue measure, and $(dx)_p$, for p a finite prime, is normalized such that

$$\int_{\mathbb{Z}_p} (dx)_p = 1.$$

For almost all p, V reduces to a nonsingular variety over $\mathbb{F}_p$, and

$$\int_{V(\mathbb{Z}_p)} \omega_p = \frac{1}{p^n} |V(\mathbb{F}_p)| . \tag{1}$$

([Wei], 2.2.5) In the cases of interest for us, (1) will hold for all p.

If $V = G$ is a connected algebraic group, then there exists, up to scalars, a *unique* left invariant gauge form ω, again defined over $\mathbb{Q}$. ω_v is then a left Haar measure on $G(\mathbb{Q}_v)$. $G(\mathbb{Q}_v)$ is unimodular if and only if ω is also right-invariant; this holds in particular if G is reductive, as all our groups will be.

For $v = p$ finite define

$$\mu_p(G) = \int_{G(\mathbb{Z}_p)} \omega_p .$$

A set (λ_p) of positive real numbers is called a *set of convergence factors* for G if $\prod_p (\lambda_p^{-1} \mu_p(G))$ is absolutely convergent. The *Tamagawa measure* $(\omega, (\lambda_v))$ (where we insert $\lambda_\infty = 1$) is the measure on $G(\mathbb{A})$ which induces on each set

$$\prod_{v \in S} G(\mathbb{Q}_v) \times \prod_{p \notin S} G(\mathbb{Z}_p), \quad S \text{ a finite set of } v \text{ containing } \infty,$$

the product measure $\prod_v \lambda_v^{-1} \omega_v$. Because of the product formula, $(\omega, (\lambda_v))$ is independent of the choice of ω. If one can choose $\lambda_p = 1$, all $p, \omega = (\omega, 1)$ is called *the* Tamagawa measure; this is possible for G semisimple. In other cases there are "canonical" choices for the λ_p. This is illustrated by the two

Basic examples. 1) Let $G = GL_n$: Using (1), we obtain

$$\begin{aligned} \mu_p &= \int_{GL_n(\mathbb{Z}_p)} \omega_p = p^{-n^2} |GL_n(\mathbb{F}_p)| \\ &= p^{-n^2} (p^n - 1)(p^n - p) \cdots (p^n - p^{n-1}) \\ &= (1 - p^{-n})(1 - p^{-(n-1)}) \cdots (1 - p^{-1}) \end{aligned}$$

by a well-known counting argument. Since

$$\prod_p (1 - p^{-k}) = \zeta(k)^{-1},$$

ζ denoting the Riemann zeta function, we see that $\lambda_p = 1 - p^{-1}$ is a set of convergence factors. By the way, we can take

$$\omega = \frac{1}{(\det x)^n} \, dx_{11} \cdots dx_{nn}.$$

2) $G = SL_n$. This time

$$\begin{aligned}\mu_p &= p^{-n^2+1}|SL_n(\mathbb{F}_p)| \\ &= \frac{p}{p-1}\,p^{-n^2}|GL_n(\mathbb{F}_p)| \\ &= (1-p^{-n})\cdots(1-p^{-2}),\end{aligned}$$

and one can take $\lambda_p = 1$. In this case there is no easy expression for the invariant form ω; for $n = 2$, one can use

$$\omega = \frac{1}{x_{11}}\,dx_{11}\,dx_{12}\,dx_{21}$$

in a neighborhood of 1 ([PR], p. 167).

Now suppose that G is unimodular and (1) is a set of convergence factors. The *Tamagawa number* $\tau(G)$ is defined to be

$$\tau(G) = \int_{G(\mathbb{A})/G(\mathbb{Q})} \omega,$$

provided this number is finite. Here, $G(\mathbb{Q})$ is viewed as a subgroup of $G(\mathbb{A})$ via the diagonal embedding.

Theorem 3.1 $\tau(D^{(1)}) = 1$.

It is now known that $\tau(G) = 1$ for every simply connected group over a number field, as Weil had conjectured; $G = D^{(1)}$ was the first case to be settled. See [PR], p. 262 for a brief resumé.

Before embarking on the proof let us describe how $\operatorname{vol}(G(\mathbb{R})/G(\mathbb{Z}))$ can be derived from $\tau(G) = 1$. We assume that G has strong approximation . (See 4.8 below for a proof of this in our case.) Let

$$C = \prod_p G(\mathbb{Z}_p) \quad \text{and} \quad U = G(\mathbb{R}) \times C.$$

Then

$$G(\mathbb{Q}) \cap U = G(\mathbb{Z}) \quad \text{and} \quad G(\mathbb{A}) = G(\mathbb{Q})U,$$

the latter equality holding by strong approximation. This induces a homeomorphism of homogeneous spaces

$$G(\mathbb{A})/G(\mathbb{Q}) \cong U/G(\mathbb{Z}),$$

preserving the Tamagawa measure. The volume on the left is 1 by hypothesis. Now use the following easy

Lemma 3.2. *Let X, Y be locally compact groups and Z a subgroup of $X \times Y$ such that pr_X is injective on Z. Then*

$$\operatorname{vol}(X \times Y/Z) = \operatorname{vol} X/Z_X \cdot \operatorname{vol} Y,$$

where $Z_X = pr_X Z$, if the volumes exist.

For the proof simply note that, if F is a fundamental domain for Z_X on X, then $F \times Y$ is a fundamental domain for Z on $X \times Y$.

From 3.2. and the preceding discussion we obtain

$$\operatorname{vol}(G(\mathbb{R})/G(\mathbb{Z})) = (\operatorname{vol} C)^{-1}. \tag{2}$$

It is important to note that the volumes in (2) are taken with respect to *projections* of the Tamagawa measure, hence are no longer independent of the choice of ω. In any case, one has, using (1),

$$c \cdot \operatorname{vol}(G(\mathbb{R})/G(\mathbb{Z})) = \prod_p \frac{1}{p^n} |G(\mathbb{F}_p)|, \tag{3}$$

where $c \in \mathbb{Q}$. For $G = SL_n$ the right-hand side of (3) is $\zeta(2)\zeta(3)\cdots\zeta(n)$. The case of $D^{(1)}$ will be discussed later; we will show that $c = 1$ holds upon a natural choice of the measures.

3.2 Proof of 3.1

We proceed to the proof of 3.1. It consists in calculating integrals

$$I = \int_{D^\times(\mathbb{A})/D^\times(\mathbb{Q})} F(|nr(x)|)\omega'_{\mathbb{A}}$$

(the terms of which will be explained below) in two ways. 3.3 and 3.4 will provide a "direct" evaluation

$$I = c \cdot \int_0^\infty F(t)\,\frac{dt}{t}$$

with some $c > 0$; this depends on 1.1 and the uniqueness of the Haar measure on the multiplicative group $\mathbb{R}_+^\times$. Next, it is crucial that the constant c, which a *priori* depends on D, equals 1 (3.5.). The second evaluation, by the formal Lemma 3.6., then leads to the equation

$$I = \tau(D^{(1)})I.$$

Since $I \neq 0$ for suitable F, 3.1. follows. It will be appropriate to view D and all groups derived from D as *varieties*; so we have to write, e.g. $D^\times(\mathbb{Q})$ instead of $D^\times$, which denotes the variety. We denote by Z the center of $D^\times$; so $Z = G_m$ and $Z(\mathbb{Q}) = \mathbb{Q}^\times$. Since

$$D(\mathbb{Q})_p = D(\mathbb{Q}) \otimes \mathbb{Q}_p = M_d(\mathbb{Q}_p)$$

for almost all p, the discussion in Section 1 shows that we can take $\lambda_p = 1 - p^{-1}$ as convergence factors for $D^\times$ and $Z^\times$, and $\lambda_p = 1$ for $D^{(1)}$. The use of factors

$1 - p^{-1}$ will always be indicated by a dash at the symbols for the measures. Thus, $\omega'_{\mathbb{A}}$ is the Tamagawa measure on $D^{\times}(\mathbb{A})$ with factors $1 - p^{-1}$.

The reduced norm $nr : D \to Z$ extends (componentwise) to a norm

$$nr : D^{\times}(\mathbb{A}) \to I(\mathbb{Q}),$$

$I(\mathbb{Q})$ denoting the ideles. Let $|\cdot|$ be the idele module $I(\mathbb{Q}) \to \mathbb{R}^{+}$,

$$|(x_v)| = \prod_v |x_v|_v.$$

Let $I(\mathbb{Q})^{(1)}$ be the kernel of $|\cdot|$ and $D(\mathbb{A})^{(1)}$ the kernel of $|nr(\cdot)|$. Because of the product formula,

$$\mathbb{Q}^{\times} \subset I(\mathbb{Q})^{(1)} \quad \text{and} \quad D^{\times}(\mathbb{Q}) \subset D(\mathbb{A})^{(1)};$$

clearly $D^{(1)}(\mathbb{A}) \subset D(\mathbb{A})^{(1)}$.

Our first step is essentially an extension of 1.1.

Lemma 3.3. *Let $0 < m \leq M$ and $C = [m, M]$. Then the set of $x \in D^{\times}(\mathbb{A})/D^{\times}(\mathbb{Q})$ such that $|N(x)| \in C$ is compact.*

Proof. It suffices to show that $D(\mathbb{A})^{(1)}/D^{\times}(\mathbb{Q})$ is compact since then the set in question is a fibration over the compact space C with compact fibres, hence itself compact. Now nr induces a map

$$D(\mathbb{A})^{(1)}/D^{\times}(\mathbb{Q}) \to I(\mathbb{Q})^{(1)}/\mathbb{Q}^{\times}$$

which is surjective because $D(\mathbb{Q})$ splits over $\mathbb{R}$, and therefore nr is onto at all places. Further, this map has fibres $D^{(1)}(\mathbb{A})/D^{(1)}(\mathbb{Q})$. Now it is well known that, first $I(\mathbb{Q})^{(1)}/\mathbb{Q}^{\times}$ is compact (e.g. [N], p. 378), second, that the compactness of $D^{(1)}(\mathbb{A})/D^{(1)}(\mathbb{Q})$ follows from (in fact, is equivalent to) the compactness of $D^{(1)}(\mathbb{R})/D^{(1)}(\mathbb{Z})$ ([PR], p. 262), which is 1.1. This completes the proof.

Lemma 3.4. *There is a constant $c > 0$ such that*

$$\int_{D^{\times}(\mathbb{A})/D^{\times}(\mathbb{Q})} F(|nr(x)|)\omega'_{\mathbb{A}} = c \int_0^{\infty} F(t)\,\frac{dt}{t}$$

for all $F : \mathbb{R}^{+} \to \mathbb{C}$ for which the right-hand side is absolutely convergent.

Proof. If F has compact support on $\mathbb{R}^{+}$, then $F(|nr(\cdot)|)$ has compact support on $D^{\times}(\mathbb{A})/D^{\times}(\mathbb{Q})$ by the last lemma, so that the left-hand side makes sense. Since $|nr(\cdot)|$ is surjective and $\omega'_{\mathbb{A}}$ is a Haar measure, the left-hand side is unchanged

if F is replaced by the function $t \to F(t_0 t)$ for arbitrary $t_0 \in \mathbb{R}_+$. That is, the map

$$F \to \int_{D^\times(\mathbb{A})/D^\times/\mathbb{Q}} F(|nr(x)|)\omega'_{\mathbb{A}}$$

is translation invariant; it is clearly linear and monotone. It therefore determines an invariant integral on $\mathbb{R}_+^\times$; the lemma follows from the uniqueness of such integrals.

Lemma 3.5. *The constant c in 3.4 equals 1; in particular is independent of D.*

Actually, c is to be viewed as the residue of the Riemann zeta function at 1. The lemma is a by-product of establishing the basics of the theory of zeta functions of D - analytic continuation and the functional equation. Since these procedures are fairly well known by now (the arguments are those of Tate's thesis, which work equally well for division algebras; see also [BR]), and we shall not pursue the analytic theory any further, we give only a sketch, reminding the reader of what's going on.

The *standard character* on $\mathbb{A}$ is the function $\chi : \mathbb{A} \to S^1$ defined by

$$\chi((x_v)) = \exp(2\pi(\sum_p \langle x_p \rangle - x_\infty)),$$

where $\langle x_p \rangle$ denotes the p-adic "principal part" of x_p. Put $\chi_D = \chi \circ tr : D(\mathbb{A}) \to S^1$; this is a character which identifies $D(\mathbb{A})$ with its dual, and $D(\mathbb{Q})$ is self-orthogonal, i.e. $x \in D(\mathbb{Q})$ if and only if $\chi_D(xy) = 1$ for all $y \in D(\mathbb{Q})$. Defining, for suitable functions Φ on $D(\mathbb{A})$, the *Fourier transform* by

$$\hat{\Phi}(y) = \int_{D(\mathbb{A})} \Phi(x)\chi_D(xy)dx_{\mathbb{A}},$$

one has the inversion formula

$$\Phi(x) = \int_{D(\mathbb{A})} \hat{\Phi}(y)\overline{\chi_D(xy)}dy_{\mathbb{A}}$$

as well as the Poisson summation formula

$$\sum_{x \in D(\mathbb{Q})} \Phi(x) = \sum_{x \in D(\mathbb{Q})} \hat{\Phi}(x).$$

Most important are the Φ of *standard type*: these have the property that there exists a finite set S of primes including ∞ and a function Φ_S on $\prod_{v \in S} D(\mathbb{Q}_v)$ such that

$$\Phi((x_v)) = \Phi_S(x_S) \prod_{p \notin S} \Phi_p(x_p),$$

where Φ_p is the characteristic function of $D(\mathbb{Z}_p)$ on $D(\mathbb{Q}_p)$; of course, Φ_S has to satisfy suitable integrability conditions. For such Φ, the *zeta function of D with respect to Φ* is defined as

$$Z^{\Phi}(s) = \int_{D^{\times}(\mathbb{A})} |nr(x)|^s \Phi(x) \omega'_{\mathbb{A}}.$$

By the above-mentioned property of Φ, the essential part of $Z^{\Phi}(s)$ is the product

$$\prod_{p \notin S} \int_{D^{\times}(\mathbb{Q}_p)} |nr(x_p)|^s \Phi_p(x_p) \omega'_p$$

which is evaluated by a calculation already in [H]; see Section 3.3. It turns out that $Z^{\Phi}(s)$ is absolutely convergent for $Re\, s > d$, having at $s = d$ a simple pole with residue

$$\lim_{s \to d} (s - d) Z^{\Phi}(s) = \int_{D(\mathbb{A})} \Phi(x) dx_{\mathbb{A}}\,. \tag{1}$$

The analytic continuation is accomplished by splitting off the zeta-integral in two summands, which can be related by Poisson summation. Define, for $t > 0$ real, $\lambda(t)$ by

$$\lambda(t) = \begin{cases} 1 & \text{if } 0 < t < 1 \\ 1/2 & t = 1 \\ 0 & t > 1. \end{cases}$$

For $x \in D^{\times}(\mathbb{A})$ write

$$f_+(x) = \lambda(|nr(x)|^{-1}),\ f_-(x) = \lambda(|nr(x)|).$$

Evidently $f_+ + f_- = 1$. Now put

$$Z^{\Phi}_{\pm}(s) = \int_{D^{\times}(\mathbb{A})} f_{\pm}(x) |nr(x)|^s \Phi(x) \omega'_{\mathbb{A}},$$

so that

$$Z^{\Phi}(s) = Z^{\Phi}_+(s) + Z^{\Phi}_-(s).$$

Clearly, if $Z^{\Phi}(s)$ is convergent, then so is $Z^{\Phi}_+(s)$. On the other hand, if $Z^{\Phi}_+(s_0)$ is convergent, then so is $Z^{\Phi}_+(s)$ if $Res \leq Res_0$ (by simple majorization); and since $Z^{\Phi}(s)$ converges for $Res > d$, we conclude that $Z^{\Phi}_+(s)$ is an entire function on the complex plane, whereas $Z^{\Phi}_-(s)$ converges for $Res > d$. Split this integral off as

$$\int_{D^{\times}(\mathbb{A})} = \int_{D^{\times}(\mathbb{A})/D^{\times}(\mathbb{Q})} \sum_{D^{\times}(\mathbb{Q})}$$

so that

$$Z_-^{\Phi}(s) = \int_{D^\times(\mathbb{A})/D^\times(\mathbb{Q})} f_-(x)|nr(x)|^s \Big(\sum_{\alpha \in D(\mathbb{Q})} \Phi(x\alpha) - \Phi(0) \Big) \omega'_{\mathbb{A}} . \tag{2}$$

By Poisson summation,

$$\sum_{\alpha \in D(\mathbb{Q})} \Phi(x\alpha) = |nr(x)|^{-d} \sum_{\beta \in D(\mathbb{Q})} \hat{\Phi}(\beta x^{-1}) . \tag{3}$$

In the integral $Z_+^{\Phi}(s)$, replace s by $d-s, x$ by x^{-1}, Φ by $\hat{\Phi}$ and note that $f_+(x^{-1}) = f_-(x)$; then

$$Z_+^{\phi}(d-s) = \int_{D^\times(\mathbb{A})/D^\times(\mathbb{Q})} f_-(x)|nr(x)|^{s-d} \Big(\sum_{\beta \in D^\times(\mathbb{Q})} \hat{\Phi}(\beta x^{-1}) \Big) \omega'_{\mathbb{A}} . \tag{4}$$

Now from (2), (3) and (4) one gets

$$Z_-^{\Phi}(s) - Z_+^{\hat{\Phi}}(d-s) = \int_{D^\times(\mathbb{A})/D^\times(\mathbb{Q})} \mu(|nr(x)|) \omega'_{\mathbb{A}}, \tag{5}$$

where the real function μ is

$$\mu(t) = (\hat{\Phi}(0) t^{s-d} - \Phi(0) t^s) \lambda(t).$$

By Lemma 3.4, the integral in (5) equals

$$Z_-^{\Phi}(s) - Z_+^{\hat{\Phi}}(d-s) = c \Big(\frac{\hat{\Phi}(0)}{s-d} - \frac{\Phi(0)}{s} \Big).$$

Now one reads off that Z_-^{Φ} and Z^{Φ} are meromorphic in the whole plane, having a simple pole of residue

$$c\hat{\Phi}(0) = c \int_{D(\mathbb{A})} \Phi(x) dx_{\mathbb{A}}$$

at $s = d$. Comparing this with (1) gives $c = 1$.

For the second computation of I we need a technical lemma of Fubini type. Let $g \subset G$ be algebraic $\mathbb{Q}$-groups with g normal. Let ϕ be the projection $G \to H := G/g$ and write

$$H'(\mathbb{A}) = \phi(G(\mathbb{A})), H'(\mathbb{Q}) = \phi(G(\mathbb{Q})).$$

Assume that $H'(\mathbb{A})$ is open in $H(\mathbb{A})$ and that (1) is a set of convergence factors for g. It is not hard to show that in this case one can use for G and H the

same set of convergence factors ([Wei], Thm. 2.4.3). Denote by $d_{\mathbb{A}}x, d_{\mathbb{A}}y$ the corresponding Tamagawa measures on $G(\mathbb{A})$ and $H(\mathbb{A})$. Write L^+ for the space of positive, continuous, integrable functions.

Lemma 3.6. *With the above notations and assumptions, for every* $f \in L^+(H'(\mathbb{A})/H'(\mathbb{Q}))$,

$$\int_{G(\mathbb{Q})} f(\phi(x))d_{\mathbb{A}}x = \tau(g) \int_{H'(\mathbb{A})/H'(\mathbb{Q})} f(y)d_{\mathbb{A}}y.$$

For the proof, see [Wei], Thm. 2.4.4.

Now we compute I with the aid of 3.6, taking

$$G = D^\times, g = D^{(1)}$$

in this lemma, so that $\phi = nr$; we have

$$H'(\mathbb{A}) = I(\mathbb{Q}) \text{ and } H'(\mathbb{Q}) = \mathbb{Q}^\times$$

by the Hasse-Schilling-Maas norm theorem (see §0). Taking $f = F(|\cdot|)$, we get from 3.6

$$I = \tau(D^{(1)}) \int_{I(\mathbb{Q})/\mathbb{Q}^\times} F(|x|)\omega'_{\mathbb{A}}.$$

But the integral here is again I, namely for the trivial case $D = Z$, by 3.5; so that we in fact have

$$I = \tau(D^{(1)})I,$$

whence $\tau(D^{(1)}) = 1$. Theorem 3.1 is proved.

We remark that there is a third way to calculate I which yields the Tamagawa number of the *projective group* $P = D^\times/Z^\times$; it turns out that $\tau(P) = d$. The reader is referred to [Wei], Thm. 3.2.1.

3.3 The volume of $G(\mathbb{R})/G(\mathbb{Z})$

In this subsection we calculate (returning to our standard notation $G = D^{(1)}$) the volume of $G(\mathbb{Z}_p)$ and relate the product of these values to the residue of the zeta function $\zeta(D, s)$ at $s = 1$. From the discussion at the end of Subsection 1 we thereby get an expression for $\mathrm{vol}\,(G(\mathbb{R})/G(\mathbb{Z}))$.

For the local calculation, let us write

$$D_p = \mathbb{Q}_p \otimes D = M_n(D^{(p)}),$$

where the skew field $D^{(p)}$ has index e, so that $d = ne$. Let $\mathcal{O}$ be the maximal order of $D^{(p)}$ will prime ideal $\mathcal{P}$; then $\mathcal{O}/\mathcal{P} = \mathbb{F}_q$, $q = p^e$. Since $\Lambda_p = \mathbb{Z}_p \otimes \Lambda$ is

a maximal order in D_p, and all maximal orders of this algebra are conjugated (see [Re], §17), we have

$$G(\mathbb{Z}_p) = \ker nr : GL_n(\mathcal{O}) \to \mathbb{Z}_p^\times.$$

On $GL_n(D_p)$ and $\mathbb{Q}_p^\times$ we use the "standard" measures

$$\mu = |nr(x)|_p^{-ne} dx \quad \text{and} \quad \vartheta = |t|_p^{-1} dt,$$

respectively; on $GL_n(\mathcal{O})$ and $\mathbb{Z}_p^\times$ these are simply the additive measures; the measure v on $G(\mathbb{Z}_p)$ is determined by

$$\mu = v\vartheta$$

(cf. [PR], prop. 3.25) so that

$$\operatorname{vol} G(\mathbb{Z}_p) = \frac{\mu(GL_n(\mathcal{O}))}{\vartheta(\mathbb{Z}_p^\times)}.$$

Clearly

$$\vartheta(\mathbb{Z}_p^\times) = 1 - p^{-1}.$$

For the calculation of $\mu(GL_n(\mathcal{O}))$ we use the exact sequence

$$1 \to GL_n(\mathcal{O}, \mathcal{P}) \to GL_n(\mathcal{O}) \to GL_n(\mathbb{F}_q) \to 1,$$

coming from reduction mod $\mathcal{P}$, which yields

$$\mu(GL_n(\mathcal{O})) = \mu(GL_n(\mathcal{O}, \mathcal{P})) \cdot |GL_n(\mathbb{F}_q)|.$$

Now

$$GL_n(\mathcal{O}, \mathcal{P}) = 1_n + M_n(\mathcal{P})$$

has volume

$$\int_{M_n(\mathcal{P})} dx = \frac{1}{|\Lambda_p : M_n(\mathcal{P})|} \operatorname{vol} \Lambda_p = q^{-n^2};$$

whence

$$\mu(GL_n(\mathcal{O})) = q^{-n^2}(q^n - 1) \cdots (q^n - q^{n-1}) = (1 - q^{-n}) \cdots (1 - q^{-1})$$

and

$$\operatorname{vol} G(\mathbb{Z}_p) = \frac{1}{1 - p^{-1}} (1 - q^{-n}) \cdots (1 - q^{-1}). \tag{1}$$

For almost all $p, q = p$; let us write

$$\operatorname{vol} G(\mathbb{Z}_p) = (1 - p^{-ne}) \cdots (1 - p^{-2}) \phi_p$$

with the correction factor

$$\phi_p = \prod_{\substack{0<i<ne \\ i\not\equiv 0 \bmod e}} (1-p^{-i})^{-1}$$

if $e > 1$. Writing $\Delta = d(\Lambda)$ for the discriminant and recalling that the ramified p are exactly the divisors of Δ, we get from (1)

$$\prod_p \operatorname{vol} G(\mathbb{Z}_p) = \prod_p (1-p^{-d})\cdots(1-p^{-2}) \prod_{p|\Delta} \phi_p = (\zeta(d)\cdots\zeta(2))^{-1} \prod_{p|\Delta} \phi_p . \quad (2)$$

Now let us turn to the zeta function. The local zeta factors, as determined by Hey, are

$$\zeta_p(D,s) = \prod_{\substack{0\le i<d \\ i\equiv 0 \bmod e(p)}} \frac{1}{1-p^{ds-i}} ;$$

in particular, for $A = M_d(\mathbb{Q})$,

$$\zeta_p(A,s) = \prod_{0\le i<d} \frac{1}{1-p^{ds-i}} .$$

Again, let us write

$$\zeta_p(D,s) = \zeta_p(A,s)\psi_p(s)$$

with the correction factor

$$\psi_p(s) = \prod_{\substack{0<i<d \\ i\not\equiv \bmod e(p)}} (1-p^{ds-i}).$$

Then the zeta function of D is

$$\begin{aligned}\zeta(D,s) &= \prod_p \zeta_p(D,s) = \prod_p \zeta_p(A,s) \prod_{p|\Delta} \psi_p(s) \\ &= \zeta(ds)\zeta(ds-1)\cdots\zeta(ds-(d-1)). \prod_{p|\Delta} \psi_p(s).\end{aligned}$$

All factors are holomorphic at $s = 1$ except $\zeta(ds-(d-1))$, which has a simple pole of residue 1. Hence

$$\operatorname{res} \zeta(D,s)_{s=1} = \zeta(d)\cdots\zeta(2) \prod_{p|\Delta} \psi_p(1),$$

and combined with (2) this yields

$$(\prod_p \operatorname{vol}(G(\mathbb{Z}_p)))^{-1} = \operatorname{res} \zeta(D,s)_{s=1} \frac{1}{\prod_{p|\Delta} \psi_p(1)\cdot\phi_p} . \quad (3)$$

Now, for $p|\Delta$,

$$\begin{aligned}\psi_p(1)\phi_p &= \prod_{\substack{0<i<d \\ i\not\equiv 0 \bmod e(p)}} \frac{1-p^{d-i}}{1-p^{-i}} \\ &= \prod_{\substack{0<i<d \\ i\not\equiv 0 \bmod e(p)}} \frac{1-p^{i}}{1-p^{-i}} \\ &= \prod_{\substack{0<i<d \\ i\not\equiv 0 \bmod e(p)}} (-p^i).\end{aligned}$$

We can ignore the factors (-1) since the other expressions in (3) are positive. Now

$$\begin{aligned}\sum_{\substack{0<i<d \\ i\not\equiv \bmod e(p)}} i &= \sum_{0<i<d} i - e(p) \sum_{0<j<\frac{d}{e(p)}=n(p)} j \\ &= \frac{d(d-1)}{2} - e(p)\frac{n(p)(n(p)-1)}{2} \\ &= \frac{d(d-n(p))}{2}.\end{aligned}$$

The p-contribution to Δ has exponent

$$(e(p)-1)n(p)d = d(d-n(p)). \tag{4}$$

Combining (3), (4) and the discussion in Subsection 1 we finally arrive at

Theorem 3.7.

$$\operatorname{vol}(G(\mathbb{R})/G(\mathbb{Z})) = \operatorname{res}\zeta(D,s)_{s=1}|\Delta|^{\frac{1}{2}}.$$

To round off our discussion, we now calculate $|G(\mathbb{F}_p)|$ and show that the equation

$$p^{\dim G}|G(\mathbb{F}_p)| = \operatorname{vol}(G(\mathbb{Z}_p)) \tag{5}$$

actually holds for *all* p, including the ramified ones.

By definition,

$$G(\mathbb{F}_p) = \ker \overline{nr} : (\Lambda/p\Lambda)^\times \to \mathbb{F}_p^\times.$$

Noting that

$$\Lambda/p\Lambda = \Lambda_p/p\Lambda_p,$$

we can bring in the local structure, which is sufficiently well known. Again we write

$$\Lambda_p = M_n(\mathcal{O}), \mathcal{O} \text{ the maximal order of } D^{(p)},$$

and

$$|D^{(p)} : \mathbb{Q}_p| = e^2, \quad \text{so} \quad d = en.$$

It is known ([Re], §14) that

$$\mathcal{O} = \mathbb{Z}_p[\zeta, \pi],$$

where ζ is a primitive $(p^e - 1)$-h root of unity and π is a prime element of $\mathcal{O}$, operating on $\langle\zeta\rangle$ by

$$\pi\zeta\pi^{-1} = \zeta^{p^r};$$

the class $\frac{r}{e} + \mathbb{Z} \in \mathbb{Q}/\mathbb{Z}$ is the Hasse invariant of $D^{(p)}$ in the Brauer group $Br(\mathbb{Q}_p)$. π can be chosen so that $\pi^e = p$. It follows that

$$A_p := \mathcal{O}/_{p\mathcal{O}} = \mathbb{F}_p[\overline{\zeta}, \overline{\pi}]$$

(the bar denoting passage to classes mod p). Write $q = p^e$. Then

$$A_p = \mathbb{F}_q \circ \langle\overline{\pi}\rangle$$

can be viewed as a twisted semigroup ring. It is easy to see that

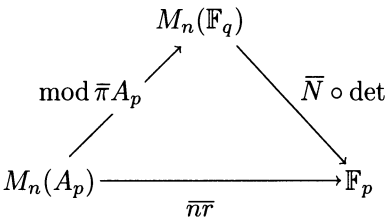
$$\operatorname{rad} A_p = \overline{\pi} A_p;$$

this implies

$$|\operatorname{rad} A_p| = q^{e-1}.$$

Let N be the field norm $\mathbb{Q}_p(\zeta) \to \mathbb{QR}_p$. Then $\overline{N}$ is the field norm $\mathbb{F}_q \to \mathbb{F}_p$.

Lemma 3.8. *The triangle*



commutes.

Proof. Take first $x \in M_n(\mathbb{Z}_p[\zeta]) \subset \Lambda_p$. Combining [Re], 9.15 and 9.29, we have

$$nr(x) = (N \circ \det)(x)$$

and the same equation mod p. Second, take $\overline{x} \in 1 + M_n(\overline{\pi} A_p)$. Then we must show that $\overline{nr}(\overline{x}) = 1$, in other words, for $x \in 1 + M_n(\pi(\mathcal{O})$ that $nr(x) \equiv 1 \bmod p$, which is clear. These two sorts of x generate $GL_n(A_p)$, and we are done.

Lemma 3.9. *Let*

$$A \xrightarrow[f]{} B \xrightarrow[g]{} C$$

be surjective homomorphisms of finite groups. Then

$$|\ker g \circ f| = |\ker f| \cdot |\ker g|.$$

The proof is obvious.

For every ring R, $R^\times \to (R/\operatorname{rad} R^\times)$ is onto; also, it is well known that $\overline{N}$ is onto. Hence we can combine 3.8. and 3.9. and obtain

$$|G(\mathbb{F}_p)| = |\ker \bmod \overline{\pi} A_p|_{GL_n(A_p)}| \cdot |\ker \overline{N}| \cdot |\ker \det|$$

The single terms are

$$|\ker \bmod \overline{\pi} A_p|_{GL_n(A_p)}| = |1 + M_n(\overline{\pi} A_p)| = q^{(e-1)n^2}, |\ker \overline{N}| = \frac{q-1}{p-1},$$

and

$$|\ker \det| = |SL_n(\mathbb{F}_q)| = (q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}) \frac{1}{q-1}.$$

Thus, using

$$\dim G = e^2 n^2 - 1,$$

we obtain

$$\begin{aligned} p^{-\dim G} |G(\mathbb{F}_p)| &= \frac{p}{p-1} q^{-n^2} (q^n - 1) \cdots (q^n - q^{n-1}) \\ &= \frac{1}{1 - p^{-1}} (1 - q^{-n}) \cdots (1 - q^{-1}), \end{aligned}$$

as promised.

4 The Size of Γ

4.1 Statement of results

In this section we present a number of results which indicate that, in various respects, Γ is "as large as it can be". The points of view are the Zariski closure, the image of reduction modulo a prime, the p-adic closure, and the maximality as a discrete subgroup of $G(\mathbb{R})$. Note that Theorem 1.1 could be listed here – Γ is "larger" than $SL_d(\mathbb{Z})$ in that it is cocompact but the latter is not. This is paradoxical inasmuch it always takes an effort to write down elements, let alone generators of Γ whereas generators of $SL_d(\mathbb{Z})$ or other Chevalley groups are easy to find. We shall return to this theme.

Theorem 4.1. *Γ is Zariski dense in $G(\mathbb{C})$.*

This is simply a special case of the Borel Density Theorem ([Bo2]). A first consequence is

Corollary 4.2. *The subring $\mathbb{Z}(\Gamma)$ of Λ spanned by the elements of Γ is a $\mathbb{Z}$-order in D.*

The corresponding statement for SL_d is obvious. It is of interest (and the result will be used later) to examine the situation in the case of a number field K. Let R be the ring of integers of K and L the quotient field of $\mathbb{Z}(R^\times)$. Then $L \subseteq K$ are number fields having the same unit rank. This implies that $L = K$ or that L is totally real and K a totally imaginary quadratic extension.

(*Proof.* Let r be the number of embeddings $L \to \mathbb{R}$, s the number of pairs of embeddings $L \to \mathbb{C}$, and r', s' the corresponding numbers for K. Then $\dim L = r + 2s$ divides $\dim K = r' + 2s'$. By Dirichlet's unit theorem, the hypothesis is $r + s = r' + s'$. The natural number

$$\frac{r' + 2s'}{r + 2s} = \frac{r' + 2s'}{r' + s' + s} \text{ is } \leq 2;$$

if it equals 2, one easily derives $r' = s = 0$ as claimed.) Thus, the commutative analogue of 4.2 fails at most for CM-fields (of course not for all of them); and, vice versa, 4.2 usually fails for the noncommutative analogue of CM-fields, i.e., totally definite quaternions.

Next we turn to reduction modulo a prime p, that is, we ask for the image of the map $G(\mathbb{Z}) \to G(\mathbb{F}_p)$. In the case of SL_d, the surjectivity of the reduction map follows at once from the well-known fact that SL_d over a field is generated by elementary matrices plus the trivial fact that these can be lifted under surjective ring homomorphisms. The comparison of this with our case of

$G = D^{(1)}$ throws a light on the peculiarities of the skew field situation. The elementary matrices are the unipotent image of root subgroups; but $D^1(\mathbb{Q})$ consists of semisimple elements, and hence no argument like the one above is possible. On the other hand, locally, $G(\mathbb{Z}_p) = SL_d(\mathbb{Z}_p)$ almost everywhere. The following result, therefore, is not too surprising:

Theorem 4.3. *For p unramified, the reduction map $G(\mathbb{Z}) \to G(\mathbb{F}_p)$ is surjective.*

This will a straightforward consequence of the Strong Approximation theorem, which holds for G (see 4.9 below); we get, more generally, surjectivity of $G(\mathbb{Z}) \to G(\mathbb{Z}/m\mathbb{Z})$ if $(m, p) = 1$ for p ramified. An almost immediate consequence of Strong Approximation is

Theorem 4.4. *Let $\hat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$. Then $G(\mathbb{Z})$ is dense in $G(\hat{\mathbb{Z}})$. In particular, $G(\mathbb{Z})$ is dense in $G(\mathbb{Z}_p)$ $(=SL_d(\mathbb{Z}_p)$ for almost all $p)$.*

Theorem 4.3 is also a special case of a much more general result due to Matthews, Vaserstein, and Weisfeiler [MVW]:

Theorem 4.5. *Let G be a connected, simply connected, absolutely almost simple algebraic $\mathbb{Q}$-group, and Γ be a finitely generated Zariski dense subgroup. Then for almost all p, $\Gamma \bmod p = G(\mathbb{F}_p)$.*

Note that the statement makes sense because $G(\mathbb{F}_p)$ is defined for almost all p, and $\Gamma \to G(\mathbb{F}_p)$ is defined for the p not dividing the denominators of a set of generators of Γ. The crucial difference between 4.3 and 4.5 is, of course, that one cannot invoke Strong Approximation in the latter case, since Γ need not be of finite index in $G(\mathbb{Z})$. We will show in §6 how to construct such a subgroup in our case of $G = D^{(1)}$. (Margulis and Soifer [MS] have shown that linear groups like our Γ contain uncountably many maximal subgroups of infinite index!) It is worthwhile to note the following converse to 4.5:

Proposition 4.6. *With the notations of* 4.5, *let $\Delta \subset \Gamma$ be a subgroup such that $\Delta \bmod p = \Gamma \bmod p$ for infinitely many p. Then Δ is Zariski dense.*

In fact, otherwise there would exist a polynomial f over $\mathbb{Z}$ and $\gamma \in \Gamma$ such that $f|_\Delta \equiv 0$ and $f(\gamma) \neq 0$. From the hypothesis we can find a prime p with $\Delta \bmod p = \Gamma \bmod p$ and $p \nmid f(\gamma)$. But this is a contradiction since $f(\gamma) \equiv f(\delta) \equiv 0 \bmod p$ for some $\delta \in \Delta$. (The argument is due to Margulis-Soifer [MS], prop. 3.) Combining 4.5 and 4.6 we conclude that, if $\Delta \bmod p = \Gamma \bmod p$ holds infinitely often, it holds for almost all p.

The proof of 4.5 as given in [MVW] is a remarkable *tour de force*, but has the "fault" of using the classification of finite simple groups which, as everyone will feel, cannot be an indispensible requirement for such a "canonical" statement. In fact, M. Nori [No] has proved, by comparatively elementary means, a result which implies 4.5 and which is worth to be quoted here.

Let π be a finitely generated subgroup of $GL_n(\mathbb{Q})$ and S a finite set of primes such that, with $R = S^{-1}\mathbb{Z}$, $\pi \subset GL_n(R)$. Let A be the Zariski closure of π in GL_n, defined as a scheme over R by the ideal of polynomials in $R[x_{ij}, \det(x_{ij})^{-1}]$ which vanish identically on π. For a subgroup B of $GL_n(\mathbb{F}_p)$ let B^+ be the (characteristic) subgroup of B generated by the elements of B of order p.

Theorem 4.7. *For almost all p,*

$$A(\mathbb{F}_p)^+ \subset \pi \bmod p \subset A(\mathbb{F}_p).$$

If A is semisimple, then

$$A(\mathbb{F}_p) = A(\mathbb{F}_p)^+,$$

and this gives 4.5, after passing from G to a faithful representation defined over $\mathbb{Q}$ (in our case, we can take the regular representation of D).

Once more it is worthwhile to compare 4.3 and 4.4 with the corresponding statements for a number field K. If the prime p is unramified in K, then

$$\mathcal{O}_K/p\mathcal{O}_K = F_1 \times \cdots \times F_r$$

is a direct product of finite fields, and the norm-one torus T has

$$T(\mathbb{Z}) = \text{ norm-one elements of } \mathcal{O}_K^\times$$

and

$$T(\mathbb{F}_p) = \{(z_i) \in \prod F_i \mid \prod nr_{F_i|\mathbb{F}_p}(z_i) = 1\}.$$

Clearly, if $r > 1$, the projection of $T(\mathbb{F}_p)$ to any $F_i^\times$ (even to any subproduct with $(r-1)$ factors) is onto. On the other hand, the images of $\mathcal{O}_K^\times \to F_i^\times$ are far from being understood in general. As for 4.4, the key word is "Leopoldt Conjecture"; we refer to [Wa], §5.5 for a discussion related to the present one.

Finally, Γ is (almost) maximal in a naive sense:

Theorem 4.8. *There are only finitely many discrete subgroups H of $G(\mathbb{R})$ containing Γ.*

The obvious question of whether or not Γ is actually a maximal discrete subgroup seems to be much more delicate. A corresponding statement for SL_d is true; see [Bo2] and [Ram]; more generally, results in this direction have been obtained only for split arithmetical groups (see [Ro] and other references quoted there).

The results of this section hold for large classes of arithmetic groups. When specified to norm one groups of orders in simple $\mathbb{Q}$-algebras A (where "norm" means "reduced norm to the center"), the outcome is that Borel density and Strong Approximation hold if $A \otimes \mathbb{R}$ has no component isomorphic to $\mathbb{H}$, whereas 4.8 holds throughout.

4.2 Proofs

Proof of 4.1. We shall use without proof that $G(\mathbb{Q})$ is Zariski dense in $G(\mathbb{C})$. (This has been proved by Rosenlicht [Ros] and holds for every connected $\mathbb{Q}$-group H. According to the main result of that paper, the coordinate ring $\mathbb{Q}(H)$ can be embedded into a purely transcendental extension of $\mathbb{Q}$. This implies that there exists a generically surjective map of some affine space onto H, whence the claim.) For a subgroup $A \subset G(\mathbb{C})$ let $\overline{A}$ be the Zariski closure; if $A = \overline{A}$, A° denotes the connected component of the identity. If $A \subset B$ and $|B : A| < \infty$, then also $|\overline{B} : \overline{A}| < \infty$, and $\overline{A}^\circ = \overline{B}^\circ$. The latter, then, holds also if A and B are commensurable, i.e. $A \cap B$ has finite index in A as well as in B.

Now let $g \in G(\mathbb{Q})$; then Γ and $g\Gamma g^{-1}$ are commensurable, because $g\gamma g^{-1}$ is the norm-one group of the order $g\Lambda g^{-1}$ and any two orders have commensurable unit groups (this is easy; see [Kl]). Hence

$$\overline{\Gamma}^\circ = \overline{g\Gamma g^{-1}}^\circ = g\overline{\Gamma}^\circ g^{-1},$$

which shows that $\overline{\Gamma}^\circ$ is normalized by $G(\mathbb{Q})$. Since the normalizer is closed, and $G(\mathbb{Q})$ is Zariski dense, $\overline{\Gamma}^\circ$ is normalized by $G(\mathbb{C}) = SL_d(\mathbb{C})$. Since D is not a definite quaternion algebra, Γ is infinite (D contains maximal subfields other than imaginary quadratic ones, now apply Dirichlet's unit theorem). Therefore $\overline{\Gamma}$ has dimension > 0, and we conclude $\overline{\Gamma} = SL_d(\mathbb{C})$.

Proof of 4.2. Clearly $\mathbb{Z}(\Gamma)$ is a lattice in D invariant under left multiplication by elements of Γ. If its rank were strictly $< d^2$, $\mathbb{Z}(\Gamma) \otimes \mathbb{C}$ would be a proper subspace invariant under Γ, hence under $\overline{\Gamma} = G(\mathbb{C})$, hence under the $\mathbb{C}$-algebra spanned by $G(\mathbb{C})$, which clearly equals $M_d(\mathbb{C})$; so $\mathbb{Z}(\Gamma) \otimes \mathbb{C}$ would be a proper left ideal of $M_d(\mathbb{C})$, which is false since $1 \in \mathbb{Z}(\Gamma)$.

Proof of 4.3 and 4.4. Denote by $\mathbb{A}_f$ the $\mathbb{Q}$-adeles without the infinite component. Again we view $G(\mathbb{Q})$ as a subgroup of $G(\mathbb{A}_f)$ by the diagonal embedding.

Theorem 4.9. (Strong approximation) *$G(\mathbb{Q})$ is dense in $G(\mathbb{A}_f)$ with respect to the adele topology.*

This special case of a much more general theorem (see [PR], ch. 7) goes essentially back to Eichler [E]; the core of the argument is an approximation theorem for polynomials and was used by Eichler for his norm theorem. We postpone the proof and first show, following [MOV], how 4.3 and 4.4 are derived from it.

Let us first prove 4.4. Put $\hat{\mathbb{Z}} = \prod \mathbb{Z}_p$. $G(\hat{\mathbb{Z}})$ is an open subgroup of $G(\mathbb{A}_f)$, and $G(\mathbb{Z}) = G(\mathbb{Q}) \cap G(\hat{\mathbb{Z}})$. The intersection of a dense set and an open set is dense in the open set, so $G(\mathbb{Z})$ is dense in $G(\hat{\mathbb{Z}})$. *A fortiori*, its projections to the factors $G(\mathbb{Z}_p)$ are dense.

Now we turn to 4.3. We shall prove the following statement which is both sharper and more general than 4.3: let $m \in \mathbb{Z}$ be such that all its prime factors are unramified in D. Choose $q \in \mathbb{Z}$ with $(m, q) = 1$. *Then*

$$\Lambda/m\Lambda \cong M_d(\mathbb{Z}/m\mathbb{Z})$$

and the composite map

$$1 + q\Lambda \hookrightarrow \Lambda \to M_d(\mathbb{Z}/m\mathbb{Z})$$

induces a surjection

$$G(1 + q\Lambda) \to SL_d(\mathbb{Z}/m\mathbb{Z}).$$

Here, the left-hand side means, by abuse of notation, the congruence group mod q, i.e. the kernel of $G(\mathbb{Z}) \to G(\mathbb{Z}/q\mathbb{Z})$.

Let p be a prime divisor of m with exponent $e(p) = e$. Write $\Lambda_p = \Lambda \otimes \mathbb{Z}_p \cong M_d(\mathbb{Z}_p)$. Then

$$\Lambda/p^e\Lambda = \Lambda_p/p^e\Lambda_p = \Lambda \otimes \mathbb{Z}/p^e\mathbb{Z} = M_d(\mathbb{Z}/p^e\mathbb{Z}).$$

Now apply the Chinese Remainder theorem:

$$\begin{aligned} \Lambda/m\Lambda &= \Lambda \otimes \mathbb{Z}/m\mathbb{Z} &&= \Lambda \otimes \textstyle\prod_p \mathbb{Z}/p^{e(p)}\mathbb{Z} \\ &= \textstyle\prod \Lambda \otimes \mathbb{Z}/p^{e(p)}\mathbb{Z} &&= \textstyle\prod M_d(\mathbb{Z}/p^{e(p)}\mathbb{Z}) \\ &= M_d(\mathbb{Z}/m\mathbb{Z}). \end{aligned}$$

This proves the first statement. Now let S be the set of ramified primes and denote by $\hat{\mathbb{Z}}(S)$ the ideles without components in $S \cup \{\infty\}$. Consider the commutative diagram

$$\begin{array}{ccc} G(1 + q\Lambda) & \longrightarrow & G(\mathbb{Z}/m\mathbb{Z}) \\ \downarrow & & \downarrow \\ G\big(1 + q\Lambda \otimes \hat{\mathbb{Z}}(S)\big) & \xrightarrow{\varphi} & G\big(1 + q\Lambda \otimes \hat{\mathbb{Z}}(S)/m\hat{\mathbb{Z}}(S)\big), \end{array}$$

in which the vertical maps are embeddings and the horizontal maps are reductions mod m. We will show:

(i) $G(1 + q\Lambda)$ is dense in $G\big(1 + q\Lambda \otimes \hat{\mathbb{Z}}(S)\big)$;

(ii) the bottom arrow φ is surjective.

Since $G\big(\hat{\mathbb{Z}}(S)/m\hat{\mathbb{Z}}(S)\big)$ is finite, the cosets of $G\big(1 + q\Lambda \otimes \hat{\mathbb{Z}}(S)\big)$ modulo the kernel of φ are open and hence contain elements of $G(1 + q\Lambda)$, by (i). By (ii), then, $G(1 + q\Lambda) \to G\big(\hat{\mathbb{Z}}(S)/m\hat{\mathbb{Z}}(S)\big)$ is onto. Since the right vertical arrow is bijective, $G(1 + q\Lambda) \to G(\mathbb{Z}/m\mathbb{Z}) = SL_d(\mathbb{Z}/m\mathbb{Z})$ is surjective, too.

(i) is proved by the same argument we have used for 4.4: $G\left(1+q\Lambda\otimes\hat{\mathbb{Z}}(S)\right)$ is open in $G\left(\mathbb{A}(S)\right)$ (the idele group without S-components) and

$$G(1+q\Lambda) = G(\mathbb{Q}) \cap G\left(1+q\Lambda\otimes\hat{\mathbb{Z}}(S)\right).$$

To prove (ii), it suffices, again by the Chinese Remainder theorem, to show that, for $p \notin S$,

$$G(1+q\Lambda_p) \to G(\mathbb{Z}/p^e\mathbb{Z})$$

is onto. If $p \not| m$, i.e. $e=0$, the right-hand side is trivial. If $p|m$, then $p \not| q$ and $\Lambda + q\Lambda_p = \Lambda_p$, and since $p \notin S$, we have to show that

$$SL_d(\mathbb{Z}_p) \to SL_d(\mathbb{Z}/p^e\mathbb{Z})$$

is onto which is trivial.

Proof of 4.9. We now turn to the proof of Strong Approximation, combining [Sw] and [Re]. Let us begin with

Lemma 4.10. (Weak Approximation) *Let S be a finite set of primes. Then $G(\mathbb{Q})$ is dense in $\prod_{p\in S} G(\mathbb{Q}_p)$.*

Proof. We use that $G(\mathbb{Q}_p) = [D_p^\times, D_p^\times]$ and refer to [PR], Thm. 1.14 for a proof of this well-known fact. Let $p \in S$, $a_p \in G(\mathbb{Q}_p)$ and write $a_p = \prod_{i=1}^m \left[b_p^{(i)}, c_p^{(i)}\right]$, $b_p^{(i)}, c_p^{(i)} \in D_p^\times$. Inserting trivial factors if necessary we can assume this for all $p \in S$. By the usual approximation theorem, we can find $b^{(i)}, c^{(i)} \in D^\times$, approximating $b_p^{(i)}, c_p^{(i)}$, respectively, simultaneously for all $p \in S$. Then $a = \prod_i [b^{(i)}, c^{(i)}]$ lies in $G(\mathbb{Q})$ and approximates a_p for all $p \in S$.

The idea behind the proof of Strong Approximation is to replace approximation of elements by approximation of polynomials. This reduction works, in the last resort, by Krasner's lemma. The approximation of polynomials is simply accomplished by applying the usual Strong Approximation coefficientwise. We turn to the details of this program. Recall that $G(\mathbb{Q}_p)$ is identified with the subgroup of adeles having component 1 at primes $\neq p$. Denote by H the closure of $G(\mathbb{Q})$ in $G(\mathbb{A}_f)$. We must show that $H = G(\mathbb{A}_f)$.

Lemma 4.11. *If $G(\mathbb{Q}_p) \subset H$ for almost all p, then $H = G(\mathbb{A}_f)$.*

This is a very easy consequence of 4.10; the proof is left to the reader.

Now let $H' \subset H$ be the subgroup of $a \in H$ with $a_p = 1$ a.e.

Lemma 4.12. *H' is a normal subgroup of $G(\mathbb{A}_f)$.*

Proof. Let $a \in H'$, $a_p = 1$ for $p \notin S$, where S is some finite set of primes. Let $b \in G(\mathbb{A}_f)$. Approximate b by $c \in G(\mathbb{Q})$ at the $p \in S$, by 4.10. Then cac^{-1} is close to bab^{-1} at the $p \in S$, whereas the other components of these elements are 1. This shows that bab^{-1} is a limit of elements cac^{-1} in H'.

Now let p be unramified in D. Then $G(\mathbb{Q}_p) = SL_d(\mathbb{Q}_p)$, and $H' \cap G(\mathbb{Q}_p)$ is normal in this group, hence either equal to $G(\mathbb{Q}_p)$ or contained in $\mathbb{Q}_p^\times$. If we can exclude the latter possibility, 4.9 will follow from 4.11. What remains to be shown, therefore, is that $G(\mathbb{Q})$ contains elements not in $\mathbb{Q}_p^\times$ and arbitrarily near at 1 at all other places. A suitable limit of such elements then lies in H', but has noncentral p-component. The heart of the proof is

Lemma 4.13. *Let $f(x) = x^d + \cdots + (-1)^d \in \mathbb{Z}[x]$ and $a, b \in \mathbb{Z}$ with $(a,b) = 1$ be given. Assume that $f(x)$ is irreducible over $\mathbb{Q}_p$ at the ramified p, and no such p divides a. Then there is $\gamma \in \Gamma$ such that*

$$\gamma \equiv 1 \bmod a\Lambda \text{ and } f(\gamma) \equiv 0 \bmod b\Lambda.$$

Proof. Suppose we can find $g(x) = x^d + \cdots + (-1)^d \in \mathbb{Z}[x]$ satisfying

(i) $g(x)$ is irreducible at p ramified;
(ii) $g(x)$ is p-adically near to $f(x)$ for $p|b$;
(iii) $g(x)$ is p-adically near to $(x-1)^d$ for $p|a$.

Let $\alpha \in \mathbb{C}$ be a zero of $g(x)$. Then α is a unit and we claim that it can be embedded into D, by property (i). First, a number field K of degree d over $\mathbb{Q}$ can be embedded in D if and only if K is a splitting field of D ([Re], 28.10). Now K splits D if and only it it does so locally everywhere; i.e., for all primes p and primes $\mathfrak{p}$ of K, $\mathfrak{p}/p$, $K_\mathfrak{p}$ is a splitting field of $D^{(p)}$. But the local splitting is easy to determine: $|K_\mathfrak{p} : \mathbb{Q}_p|$ must be a multiple of $d(p)$ ([Re], 31.10). For $K = \mathbb{Q}(\alpha)$, this clearly follows from (i).

From (ii), $0 = g(\alpha)$ is near at $f(\alpha)$ for the p dividing b, and from (iii), 0 is near at $(\alpha-1)^d$ for the p dividing a. Since α is integral, it is contained in a maximal order, and some conjugate is in Γ and satisfies the required congruences. As for the existence of $g(x)$, (ii) and (iii) can be satisfied by using the Chinese Remainder theorem coefficientwise. (i) can be fulfilled if we know that, locally, a polynomial sufficiently close to an irreducible polynomial is also irreducible. This is Krasner's lemma; we refer to [Re], 33.8.

The last step in the proof of 4.9 is

Lemma 4.14. *Let $f(x)$ be as in* 4.13. *Let p be a fixed unramified prime. Then there is $\eta \in H$ having components* 1 *at all unramified $q \neq p$, and $f(\eta_p) = 0$.*

This suffices since we can choose $f(x)$ irreducible also at p (e.g. by using Eisenstein's lemma), and then $f(\eta_p) = 0$ implies $\eta_p \notin \mathbb{Q}_p$.

Proof of 4.14. Let $p_1, p_2, \ldots$ be the unramified primes other than p. Put $a_n = (p_1 \cdots p_n)^n$. Apply 4.13 with $a = a_n$ and $b = p^n$ to obtain $\gamma^{(n)} \in \Gamma$ such that

$$\gamma^{(n)} \equiv 1 \bmod a_n\Lambda,\ f(\gamma^{(n)}) \equiv 0 \bmod p^n\Lambda.$$

Clearly, $\gamma^{(n)} \to 1$ for all p_i. Since $G(\mathbb{Z}_p)$ is compact, we can choose a subsequence $\gamma^{(k)}$ converging to some $\eta_p \in G(\mathbb{Z}_p)$ satisfying $f(\eta_p) = 0$. Then $\eta = \lim \gamma^{(k)}$ is as required.

Proof of 4.8. We shall actually prove

Theorem 4.8′. *There are only finitely many subgroups of $G(\mathbb{C})$ which contain Γ as a subgroup of finite index.*

If $H \subset G(\mathbb{R})$ is discrete and contains Γ, then vol $\big(G(\mathbb{R})/H\big) > 0$, and since vol $\big(G(\mathbb{R})/\Gamma\big) < \infty$, the index $|H : \Gamma|$ must be finite; so 4.8 follows from 4.8′.

The following ad-hoc proof of 4.8′ combines ideas from [Ram] and [Bo2]. Let H be a group as in 4.8′. Let $H^{(t)}$ be the subgroup for H generated by all tth powers, where $t = |H : \Gamma|!$; thus, $H^{(t)} \subset \Gamma$. Now the tth power map $G(\mathbb{C}) \to G(\mathbb{C})$ is surjective and continuous; this implies that $H^{(t)}$ is still Zariski dense, since $\Gamma^{(t)}$ is and $\Gamma^{(t)} \subset H^{(t)}$. The argument from the proof of Cor. 4.2 now shows that $\mathbb{Q}(H^t) = D$. But clearly $H^{(t)}$ is normal in H, so we have shown that H normalizes D.

Consider the homomorphism

$$\pi : H \to \text{Aut } D^\times$$

arising from the conjugation. We have

$$H_0 := \ker \pi = H \cap \mathbb{C}^\times.$$

But $H_0^t \subset \Gamma$, so

$$H_0^t \subset \Gamma \cap \mathbb{Q}^\times \subset \{\pm 1\}.$$

This shows that H_0 consists of roots of unity, of order a divisor of $2d$. So it suffices to show that there are only finitely many possibilities for im π, and since

$$\text{Aut } D^\times = D^\times/\mathbb{Q}^\times$$

by the Skolem-Noether theorem, it suffices to show that there are only finitely many possibilities for H in $D^\times$.

Let $h \in H$. Since $h^t \in \Gamma$, h is integral. Let $\{\gamma_i\}$ be a $\mathbb{Q}$-base of D consisting of elements of Γ (Cor. 4.2). Write

$$h = \sum_i a_i \gamma_i, \; a_i \in \mathbb{Q};$$

then

$$h\gamma_j = \sum_i a_i \gamma_i \gamma_j \quad \text{and} \quad tr(h\gamma_j) = \sum_i a_i tr(\gamma_i \gamma_j) \in \mathbb{Z}, \text{ all } j.$$

Considering this as a system of linear equations for the a_i, and using $\delta := \det(tr\gamma_i\gamma_j) \neq 0$, we obtain, by Cramer's rule,

$$\delta H \subset \mathbb{Z}(\Gamma).$$

Put

$$L = \sum_{h \in H} h\mathbb{Z}(\Gamma).$$

Since $|H : \Gamma| < \infty$, this is a $\mathbb{Z}$-lattice in D, and from the preceding argument we see that

$$\mathbb{Z}(\Gamma) \subset L \subset \delta^{-1}\mathbb{Z}(\Gamma).$$

There are only finitely many lattices satisfying these inequalities, and H is contained in the stabilizer of one of them, which leaves only finitely many possibilities. Thereby 4.8′ is proved.

To round off our discussion, it is appropriate to introduce the *commensurability group* $C(\Gamma)$. More generally, if F is any algebraic $\mathbb{Q}$-group, the *commensurability group* of a subgroup $\Delta \subset F$ is defined as

$$C(\Delta) = \{x \in F(\mathbb{C}) \mid x\Delta x^{-1} \text{ and } \Delta \text{ are commensurable}\}.$$

It is easy to see that $C(\Delta)$ is actually a group, and that $C(\Delta) = C(\Delta')$ if Δ and Δ' are commensurable; so $C(F(\mathbb{Z}))$ depends only on F. In our case, we easily get

Proposition 4.15.

$$C(\Gamma) = \{(nr(g))^{-\frac{1}{d}} g \mid g \in D^\times\}.$$

Proof. As in the proof of 4.8′ we can show that $x \in C(\Gamma)$ normalizes D, and by Skolem-Noether, we have

$$x = \omega g,\ \omega \in \mathbb{C}^\times, g \in D.$$

From

$$1 = \det x = \omega^d nr(g)$$

we read off that x has the required form; the other inclusion is obvious.

Clearly, if $|H : \Gamma| < \infty$, then $H \subset C(\Gamma)$; so $C(\Gamma)$ is a "universal reservoir" for all possible $H's$, and, more generally, for all subgroups of $G(\mathbb{C})$ commensurable with Γ.

The general result, due to Borel [Bo2], is as follows: *let N be the largest normal $\mathbb{Q}$-subgroup such that $N(\mathbb{R})$ is compact. Let $\pi : F \to F/N$ be the canonical map. Then*

$$C(F(\mathbb{Z})) = \pi^{-1}((F/N)(\mathbb{Q})).$$

In our case, $N = \mu_d$ is the group (scheme) of dth roots of 1. From the exact sequence

$$1 \to \mu_d \to G \to G/\mu_d \to 1$$

we obtain, by taking cohomology with respect to the absolute Galois group γ of $\mathbb{Q}$, the exact sequence

$$\begin{aligned} 1 \to \varepsilon \quad &\to G(\mathbb{Q}) \to (G/\mu_d)(\mathbb{Q}) \\ &\to H^1(\gamma, \mu_d) \to H^1(\gamma, G), \end{aligned}$$

where $\varepsilon = 1$ or $= \{\pm 1\}$. It is well known that

$$H^1(\gamma, \mu_d) \cong \mathbb{Q}^\times/\mathbb{Q}^{\times d} \quad \text{and} \quad H^1(\gamma, G) \cong \mathbb{Q}^\times/nrD^\times = 1$$

by our assumption on D (see [PR], ch. 2). Thus, Galois cohomology shows us $C(\Gamma)$ as an extension

$$1 \to G(\mathbb{Q})/\varepsilon \to C(\Gamma) \to \mathbb{Q}^\times/\mathbb{Q}^{\times d} \to 1.$$

The reader may convince himself that $C(\Gamma)$, as given by 4.15, fits into such a sequence via the homomorphism

$$\big(nr(g)\big)^{\frac{1}{d}} g \to nr(g) \cdot \mathbb{Q}^{\times d}.$$

5 Margulis' Finiteness Theorem

5.1 The Result

The purpose of this chapter is to prove a very special case of a general result due to Margulis [Ma 1]. It is surely the most difficult one of the results presented in this monograph, and it is the only one for which the fact that our Γ is essentially the unit group of an order seems to be of no help.

Theorem 5.1. *Assume $d \geq 3$. Then any noncentral normal subgroup of Γ has finite index.*

We note some immediate consequences.

Corollary 5.2.

(i) *The factor commutator group Γ_{ab} is finite.*

(ii) *If $\gamma \in \Gamma$, $\gamma \neq \pm 1$, then there are finitely many conjugates $\gamma^{t_1}, \ldots, \gamma^{t_r}$ which generate a subgroup of finite index of Γ.*

(iii) *If $w(\underline{x}) = w(x_1, \ldots, x_n)$ is any nontrivial word in a free group generated by $x_1, \ldots, x_n$, then there are finitely many specifications $\underline{x} \to \underline{\gamma}^{(i)}$ in Γ such that $\langle w(\underline{\gamma}^{(i)}) \rangle$ is cofinite.*

In fact, the normal subgroup generated by γ is noncentral, hence of finite index, hence finitely generated, and in a finite set of generators only finitely many conjugates of γ occur. This proves (ii); the proof of (iii) is similar, and (i) is obvious.

The property stated in 5.1 evidently carries over to groups commensurable with Γ. It is now easy to see that 5.1 fails for $d = 2$. In this case, any torsion free subgroup Δ of finite index is a surface group of genus $g \geq 1$, and Δ_{ab} is the first homology of the surface, isomorphic to $\mathbb{Z}^{2g}$. The crucial hypothesis for 5.1 is that, if $d \geq 3$, the rank of $G(\mathbb{R})$ is $d - 1 \geq 2$. The theorem therefore holds for unit groups of arbitrary semisimple orders with the exception of quaternion algebras over $\mathbb{Q}$ and over imaginary quadratic fields; one only has to extend the Kazhdan property (5.12) and the mechanism of parabolic subgroups (Section 5.5).

The scheme of the proof of 5.1 is as follows. We will introduce two group theoretical properties, (T) and (A). Γ has (T) (but not (A)), and it will be an immediate consequence of the definition that (T) descends to factor groups $\overline{\Gamma} = \Gamma/N$, $N < \Gamma$ normal. Now it is a general fact that a group having (T) and (A) is compact; if it is discrete, it must be finite. So it "remains" to show (and this will be the main bulk of the proof) that if $\overline{\Gamma}$ does not have (A), N must be central; it is in this form that Margulis stated his general theorem.

We shall first discuss (A) and (T), to follow with the essential part of the proof. Standard references are [Gr] and [Pa] for (A), [HV] for (T); I have also profited from Lubotzky's book [L1]. In the proof, I shall follow largely [Zi], without neglecting the original sources [Ma 1] and [Ma 2]. In the last section, I have collected some facts needed from measure theory.

5.2 Amenable groups

In this and the following section we abandon once more our standard notation and let G be a locally compact topological group, equipped with a (left) Haar measure. Let $L^\infty(G)$ denote the space of essentially bounded real-valued measurable functions on G; here, essentially bounded means bounded outside a null set, and two functions are identified if they agree outside a null set. A *mean* on G is a linear functional $m : L^\infty(G) \to \mathbb{R}$ satisfying

(i) $m(f) \geq 0$ if $f \geq 0$;
(ii) $m(\chi_G) = 1$, where $\chi_G \equiv 1$ on G.

m is called (left) *invariant* if

(iii) $m(g \cdot f) = m(f)$ for all $f \in L^\infty(G)$ and $g \in G$, where $g \cdot f(x) = f(g^{-1}x)$.

Finally, G is called *amenable* if a (left) invariant m exists.

Clearly, if G is compact, the Haar measure (or rather integral) provides an invariant mean, but not the only one; see [L], 2.2.11 for the construction of infinitely many different invariant means on the circle group S^1 (m becomes unique after restricting from $L^\infty(G)$ to *continuous* functions; cf. [Gr], p. 3/4).

Abelian groups are amenable ([Gr], 1.2.1). Amenability is inherited by closed subgroups and factor groups; conversely, if $N \lhd G$ is closed normal and N as well as G/N are amenable, then so is G ([Gr], p. 30). Consequently, a solvable, or compact-by-solvable group is amenable.

If G is amenable, then one obtains from m a left invariant, *finitely* additive measure μ on G by setting $\mu(E) = m(\chi_E)$, where E is a Borel set and χ_E its characteristic function. Using this, it is easy to see that a free group F on two generators a, b cannot be amenable: define, for $i \in \mathbb{Z}$, the subsets $H_i \subset F$ by

$$H_i = \{\text{reduced words } x = a^i b^j \cdots \text{ with } j \neq 0 \text{ if } x \neq a^i\}.$$

If λ_a, λ_b denote left multiplication by a, b, then $\lambda_a(H_i) = H_{i+1}$, all i, and $\lambda_b(H_i) \subset H_0$, $i \neq 0$. If the measure μ comes from the mean m as above, we have $\mu(F) = 1$, hence $\mu(H_i) = 0$, all i, by the first equality; by $\lambda_b(H_i) \subset H_0$, $\mu(H_0) \geq \mu\left(\bigcup_{i \neq 0} H_i\right)$, but $\mu(H_0) + \mu\left(\bigcup_{i \neq 0} H_i\right) = \mu(F) = 1$, hence $\mu(H_0) \geq \frac{1}{2}$, contradiction. (The argument can be carried over to free operations of F on sets X, and yields "paradoxical decompositions" of such X; see [L] for the relevance of this for the Hausdorff-Banach-Tarski paradox.)

The example settles the question of amenability for the groups we are mainly interested in, by the following well-known result due to Tits [T]:

Every subgroup of $GL_n(\mathbb{R})$ either contains a non-abelian free group or has a solvable subgroup of finite index (and hence is amenable).

Thus the presence of free subgroups is the only obstacle to amenability in the case of a linear group. For general (discrete) groups, this was an open question for a long time until a counterexample was found; see [Pa], ch. 3.

What we have given above is the "natural" definition of amenability, but there are many more characterizations of this property, e.g. in terms of representations, of group combinatorics, or by the Følner condition. We shall need a "fixed point property", of which the original definition is a special case. The set of means (functions satisfying (i) and (ii)) is a compact convex subspace of the unit ball in the dual space $L^\infty(G)^*$, with the weak-$*$-topology. G acts continuously on the means, and this action is *affine*, i.e.

$$g\big((1-\lambda)m_1 + \lambda m_2\big) = (1-\lambda)g(m_1) + \lambda g(m_2),$$

$0 \leq \lambda \leq 1$, since G acts linearly on the ambient space. An invariant mean is simply a fixed point of this action. One can show that this particular fixed point properly implies a whole family of similar ones:

Proposition 5.3. *G is amenable if and only if it has a fixed point whenever it operates by affine transformations on a compact convex set of a locally convex space, where the operation is continuous.*

See [Gr], 3.3.5. For brevity, let us call such a G-space an *affine* G-space. Let X be a compact metrizable G-space and write $M(X)$ for the space of probability measures on X. Then $M(X)$ is an affine G-space. This proves one half of

Proposition 5.4. *G is amenable if and only if for every continuous G-operation on a compact metrizable space X, there exists a G-invariant probability measure on X.*

For the other direction, one can use the barycenter construction to derive from an invariant probability measure the existence of a fixed point; see [Zi], p. 61.

The fixed point property can be used to extend the notion of amenability from G to G-spaces S. This extension preserves sufficiently many properties of amenability and allows, as an analogue of the reasoning leading to 5.4, to derive the existence of G-maps $S \to M(X)$, under suitable hypotheses; the point is that G itself need not be amenable.

To be precise, let S be a measure space (always of standard type; see Section 9) with right G-action, E a separable Banach space and $\pi : G \to \mathrm{Iso}(E)$ a representation. Then G operates on E^* by the adjoint representation π^*; let A

be a G-invariant, compact convex subset of the unit ball of E^*. Then G operates on the functions

$$F(S, A) \subset L^\infty(S, E^*)$$

(which form a compact convex G-space) by

$$(g \cdot \varphi)(s) = \pi^*(g^{-1})\varphi(sg).$$

This is the most straightforward case of what is called an *affine G-space over S*. To obtain the general notion one has to twist the operation by *cocycles* $S \times G \to \mathrm{Iso}(E)$ (π can be viewed as a cocycle not depending on S!), and the set A is allowed to vary with the argument $s \in S$. We won't need this general notion explicitly, so we don't define it.

Now the G-action on S is called *amenable* if every affine G-space over S has a fixed point. If G is amenable, then every G-space S is amenable. Also it is easy to see that G is amenable if the one-point set is an amenable G-space; more generally, G is amenable if it acts amenably on a space with finite invariant measure ([Zi], 4.3.3). If S is an amenable G-space and $\Gamma \subset G$ a closed subgroup, then S is an amenable Γ-space ([Zi], 4.3.5).

Important examples of amenable actions are furnished by

Proposition 5.5. *Let $H \subset G$ be a closed subgroup. Then G/H is an amenable G-space if and only if H is an amenable group.*

See [Zi], 4.3.2. We apply this to the following situation: G is a non-compact semisimple Lie group, $\Gamma \subset G$ a discrete subgroup and $P \subset G$ a parabolic subgroup. By Tits' theorem, P is amenable if and only if P is minimal parabolic, i.e. a Borel subgroup. Hence in this case G/P is an amenable Γ-space. We shall need the following property of amenable actions:

Proposition 5.6 ([Zi], 4.3.9): *Let S be an amenable G-space and X a compact metric G-space. Then there is a measurable G-map $S \to M(X)$ (possibly outside a Borel null set in S).*

Proof. Let $C(X)$ be the Banach space of continuous functions on X and π the representation of G on $C(X)$. Then $F\big(S, M(X)\big)$ is an affine G-space over S, which has a fixed point by amenability, i.e. a function $\varphi : S \to M(X)$ satisfying

$$\pi^*(g^{-1})\varphi(sg) = \varphi(s).$$

This holds for almost all s; hence after changing φ on a null set and switching to a right action of G on $M(x)$, we obtain a G-map as desired.

It should be remarked that 5.6 was first proved without using the notion of amenable spaces; see [Fu], 15.1 and [Ma2], 4.5.

5.3 Kazhdan's property (T)

By a *representation* of G we mean a homomorphism

$$\pi : G \to U(\mathcal{H})$$

into the unitary group of a complex Hilbert space $\mathcal{H}$ such that the resulting map $G \times \mathcal{H} \to \mathcal{H}$ is continuous. A *unit vector* of $\mathcal{H}$ is a vector of norm 1. The (left) *regular* representation is the operation on $\mathcal{H} = L^2(G)$, given by $(gf)(x) = f(g^{-1}x)$.

Let $K \subset G$ be a compact set and $\varepsilon > 0$. A unit vector $v \in \mathcal{H}$ is called (ε, K)-*invariant* if

$$\sup \{||\pi(g)v - v|| \mid g \in K\} < \varepsilon.$$

The representation π is said to *have almost invariant vectors* if there is an (ε, K)-invariant vector for every pair (ε, K).

The standard example for a π having almost invariant, but no invariant vectors, is the regular representation of $G = (\mathbb{R}, +)$. Let (ε, K) be given; $K \subset [-c, c]$ for c large enough. Let $a < b \in \mathbb{R}$; let χ be the characteristic function of $[a, b]$, and $v = (b-a)^{-\frac{1}{2}}\chi$. Then v is a unit vector and for $t \in \mathbb{R}$

$$\begin{aligned} ||\pi(t)v - v||^2 &= \int \big(v(x-t) - v(x)\big)^2 dx \\ &= 2 - \tfrac{2}{b-a}\int_a^b \chi(x-t)dx \leq \tfrac{2|c|}{b-a} \end{aligned}$$

if $|t| \leq c$. Choosing $b-a$ large enough, one sees that v is (ε, K)-invariant. Clearly π has no invariant vectors $\neq 0$; more generally, the regular representation of G has invariant vectors $\neq 0$ if and only if G is compact.

G is said to *have property* (T), or to be a *Kazhdan group*, if *every representation which has almost invariant vectors has (nonzero) invariant vectors.*

It is immediate from the definition that property (T) descends to factor groups. Also, it is easy to see that a compact group is Kazhdan, since then one can take $(\varepsilon, K) = \left(\frac{1}{2}, G\right)$, and if v is $\left(\frac{1}{2}, G\right)$-invariant, then $w = \int_G \pi(g)vdg$ is G-invariant and $\neq 0$.

We have seen that $(\mathbb{R}, +)$ is not Kazhdan. More generally:

Proposition 5.7. *G is amenable and Kazhdan if and only if G is compact.*

Proof. It is one of the characterizations of amenability (due to Hulanicki [Hu]) that the regular representation contains almost invariant vectors; so if G has property (T) in addition, it must be compact.

Corollary 5.8. *If G is a discrete Kazhdan group and the homomorphic image $\overline{G}$ of G is amenable, then $\overline{G}$ is finite. In particular, G_{ab} is finite.*

Another consequence of 5.7 is that G has no nontrivial homomorphism to $\mathbb{R}^n$ or $\mathbb{Z}^n$; in particular G is unimodular. Also, it follows that a free group is not Kazhdan.

We now present the main steps in the proof that $SL_n(\mathbb{R})$, $n \geq 3$, is Kazhdan, and begin with the concept of positive definite functions. A function $\varphi : G \to \mathbb{C}$ is called *positive definite* (positive, for short) if, for all $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ and $g_1, \ldots, g_n \in G$,

$$\sum_{i,j} \overline{\lambda}_i \lambda_j \varphi(g_i^{-1} g_j) \geq 0$$

(in particular, is real). The prototype of such φ comes from a unitary representation $\pi : G \to U(\mathcal{H})$ and any $v \in \mathcal{H}$: with $\varphi(g) = \langle v, \pi(g)v \rangle$ we have

$$\sum_{i,j} \overline{\lambda}_i \lambda_j \varphi(g_i^{-1} g_j) = \| \sum \lambda_i \pi(g_i) v \|^2 \geq 0.$$

A construction due to Gelfand, Neumark, and Segal shows that this example is universal. More precisely:

Proposition 5.9. *Suppose that φ is positive. Then there exists a unitary representation $\pi_\varphi : G \to U(\mathcal{H}_\varphi)$ and a vector $v_\varphi \in \mathcal{H}_\varphi$ such that*

$$\varphi(g) = \langle v_\varphi, \pi_\varphi(g) v_\varphi \rangle, \quad \text{all } g \in G;$$

moreover, the span of $\pi_\varphi(G)v_\varphi$ is dense in $\mathcal{H}_\varphi$. The triple $(\pi_\varphi, \mathcal{H}_\varphi, v_\varphi)$ is uniquely determined up to isomorphism. The construction has the following functioral properties:

(a) *if φ, ψ, χ are positive and $\varphi = \psi + \chi$, then π_ψ is contained in π_φ;*
(b) *if $\varphi = c > 0$ is constant, then $\pi_\varphi = \pi_0$ is the trivial representation;*
(c) *if φ is positive and $H \subset G$ is a closed subgroup, then $\pi_{\varphi/H}$ is contained in π_φ / H;*
(d) *if $\pi : G \to U(\mathcal{H})$ is a representation and $\varphi(g) = \langle v, \pi(g)v \rangle$ for some $v \in \mathcal{H}$, then π_φ is contained in π.*

For the proof see [HV], ch. 5. The following characterization of (T) in terms of positive functions is also proved there:

G is Kazhdan if and only if every sequence (φ_n) of positive functions with $\varphi_n(1) = 1$, which converges to 1 uniformly on every compact subset of G, does so on G.

We will use 5.9 for

Lemma 5.10. *Let $G = \mathbb{R}^n \rtimes SL_n(\mathbb{R})$, $n \geq 2$ with $SL_n(\mathbb{R})$ operating as usual on $H = \mathbb{R}^n$. If the representation π of G has almost invariant vectors, then its restriction to H has (nonzero) invariant vectors.*

Note that this is a "relative property (T)" for the pair (G, H); see [HV], 1.18.

Proof of 5.10 (sketch). Let $\{K_m\}$ be an increasing family of compact sets covering G. For every m choose a unit vector $v_m \in \mathcal{H}_\pi$ which is $(\frac{1}{m}, K_m)$-invariant. It is then easy to show that the positive functions

$$\varphi_m : g \to \langle v_m, \pi(g)v_m \rangle$$

converge to 1 uniformly on every compact subset of G.

The restriction of φ_m to $H = \mathbb{R}^n$ is a positive function and provides us with a probably measure μ_m on the dual $\hat{\mathbb{R}^n}$ of which φ_m is the Fourier transform (see [Ru], ex. 11.14). We identify $\hat{\mathbb{R}^n}$ with $\mathbb{R}^n$ and claim that $\mu_m(\{0\}) > 0$ for some m.

Otherwise, all μ_m would be probably measures on $\mathbb{R}^n \backslash \{0\}$, and if $p : \mathbb{R}^n \backslash \{0\} \to \mathbb{P}^{n-1}$ is the canonical projection, the $p_*(\mu_m)$ would be probability measures on $\mathbb{P}^{n-1}$. The space of such measures on $\mathbb{P}^{n-1}$ is weakly compact, so there would be a weakly convergent subsequence of $p_*(\mu_m)$. The same applies to $(g_*\mu_m)$ for $g \in SL_n(\mathbb{R})$; but since $(\mu_m - g_*\mu_m)$ weakly converges to 0, by the construction of the φ_m, we would obtain a $SL_n(\mathbb{R})$-invariant probability measure on $\mathbb{P}^{n-1}$, which, as is well known, cannot exist (see e.g. [Zi], 3.2.2).

Now choose m with $\mu_m(\{0\}) = c_m > 0$, write

$$\mu_m = c_m \delta_0 + \mu'_m$$

where δ_0 is the point measure supported at 0, and apply Fourier transformation to obtain

$$\varphi_m | \mathbb{R}^n = c_m + \varphi'_m,$$

where φ'_m is still positive on $\mathbb{R}^n$. Since c_m is the constant function, it follows from 5.9 that $\pi | \mathbb{R}^n$ contains the trivial representation, as was to be shown.

We need one more auxiliary statement. Let N_{n-1} be the subgroup of $SL_n(\mathbb{R})$ consisting of elements of the form

$$\begin{pmatrix} 1_{n-1} & x \\ 0 & 1 \end{pmatrix}, x \in \mathbb{R}^{n-1},$$

where 1_k is the k-dimensional unit matrix and x is a column; clearly $N_{n-1} \cong \mathbb{R}^{n-1}$.

Lemma 5.11. *Let π a representation of $SL_n(\mathbb{R})$ on $\mathcal{H}$ and $v \in \mathcal{H}$. If v is invariant under N_{n-1}, then v is invariant under $SL_n(\mathbb{R})$.*

Proof. Consider first the case $n = 2$. It suffices to show that $\varphi : g \to \langle v, \pi(g)v \rangle$ is constant.

Since v is fixed by N_1, φ is N_1-biinvariant, i.e. $\varphi(n_1 g n_2) = \varphi(g)$ for $n_1, n_2 \in N_1$. Now N_1 is the isotropy group of $(1,0)^t$ for the operation of $SL_2(\mathbb{R})$

on $\mathbb{R}^2$; so $SL_2(\mathbb{R})/N_1 \approx \mathbb{R}^2\backslash\{0\}$ and φ can be considered as a function on $\mathbb{R}^2\backslash\{0\}$, which is constant on the N_1-orbits; these are the lines parallel to the x-axis and the points of the x-axis. By continuity, φ is constant on the x-axis. This implies that v is invariant under the group P of upper triangular matrices, and hence φ is P-biinvariant.

Now φ is defined on $SL_2(\mathbb{R})/P = \mathbb{P}^1(\mathbb{R})$ and invariant on the P-orbits. This implies that φ is constant.

Now let $n \geq 2$. For $i = 1, \ldots, n-2$ put

$$h_i\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} 1_i & 0 & 0 & 0 \\ 0 & a & 0 & b \\ 0 & 0 & 1_{n-i-2} & 0 \\ 0 & c & 0 & d \end{pmatrix};$$

h_i is a homomorphism $SL_2(\mathbb{R}) \to SL_n(\mathbb{R})$, and $h_i\big(SL_2(\mathbb{R})\big) \cap N_{n-1} = h_i(N_1)$. If v is N_{n-1}-invariant, then it is also $h_i(N_1)$-invariant, hence $h_i\big(SL_2(\mathbb{R})\big)$-invariant for all i, by the case $n = 2$. But these images generate $SL_n(\mathbb{R})$, and we are done.

Now we can easily prove

Theorem 5.12. *For $n \geq 3$, $SL_n(\mathbb{R})$ is a Kazhdan group.*

Proof. Identify $\mathbb{R}^{n-1} \rtimes SL_{n-1}(\mathbb{R})$ with a subgroup of $SL_n(\mathbb{R})$ by the mapping

$$(x, g) \to \begin{pmatrix} g & x \\ 0 & 1 \end{pmatrix}.$$

If π has almost invariant vectors, then so has $\pi|\mathbb{R}^{n-1} \rtimes SL_{n-1}(\mathbb{R})$. By 5.10, π has a nonzero v invariant under N_{n-1}, and by 5.11 v is invariant under $SL_n(\mathbb{R})$. This completes the proof.

For our purposes, we still have to come down to the subgroup $D^{(1)}(\mathbb{Z})$ of $SL_d(\mathbb{R})$. It is clear that property (T) is not inherited by arbitrary closed subgroups (look at free subgroups!). But we do have

Theorem 5.13. *If $H \subset G$ is a closed subgroup of G of finite covolume, then H is Kazhdan if G is.*

Proof (sketch). Let π be a representation of H. The main step is to show:
if π has almost invariant vectors, then so has the induced representation $\operatorname{ind}_H^G \pi$.

This rests on the existence of a suitable Borel section of $G \to G/H$; see [HV], 3.3 for the details. This reduction accepted, the remainder of the proof is easy: suppose π has almost invariant vectors. Then so has $\operatorname{ind}_H^G \pi$, and since G is Kazhdan, there is a nonzero invariant f in the space of $\operatorname{ind}_H^G \pi$. By definition of ind, f is a measurable function $f : G \to \mathcal{H}_\pi$ such that

(i) $\quad f(xh) = \pi(h^{-1}) f(x)$, all $h \in H$, almost all $x \in G$,

(ii) $\quad f(gx) = f(x)$, all $g \in G$, almost all $x \in G$.

From (ii) we can choose $x_0 \in G$ such that

$$f(g) = f(x_0) =: v, \text{ almost all } g \in G.$$

Clearly $v \neq 0$ since $f \neq 0$. Let $h \in H$. We then have

$$f(xh) = v, \text{ almost all } x \in G;$$

but by (i),

$$f(xh) = \pi(h^{-1})v, \text{ almost all } x \in G,$$

which implies $\pi(h)v = v$ as desired.

5.13 shows that neither $SL_2(\mathbb{R})$ nor $SL_2(\mathbb{Z})$ are Kazhdan since it is well known that $SL_2(\mathbb{Z})$ contains free subgroups of finite index, and is of finite covolume in $SL_2(\mathbb{R})$.

We formally state the case of interest for us:

Corollary 5.14. *If $d \geq 3$, the group $\Gamma = D^{(1)}(\mathbb{Z})$ is Kazhdan.*

For $d = 2$ this is false since in this case Γ contains a subgroup Δ of finite index which is not Kazhdan, Δ_{ab} being infinite; cf. 5.8.

We have now established what we need for the proof of 5.1. In view of 5.14, however, it is clear that property (T) is of independent interest in the study of Γ. We therefore present some more consequences of (T).

It is one of the first results that a Kazhdan group is compactly generated; hence, in the discrete case, finitely generated (but not, in general, finitely presented; see [HV], ch. 3). Thereby 5.14 yields another proof of this fundamental result on our Γ. We also see that the rational and adelic groups, $D^{(1)}(\mathbb{Q})$ and $D^{(1)}(\mathbb{A})$, are not Kazhdan; this holds, more generally, for linear algebraic groups of dimension at least 1, defined over number fields; see [HV], 1.12.

The definition of property (T), which might appear somewhat obscure at first sight, can be given a clear conceptual meaning in terms of the *Fell topology* on the unitary dual of G. Let $\tilde{G}$ denote the set of isomorphism classes of unitary representations $(\pi, \mathcal{H}_\pi)$ of G; the *unitary dual* is the subset $\hat{G}$ of irreducible π. Neighbourhoods of π in $\tilde{G}$ are parametrized by a compact set $K \subset G$, an $\varepsilon > 0$, and orthonormal vectors $v_1, \ldots, v_n \in \mathcal{H}_\pi$, and σ is in $V(\pi; K, \varepsilon, v_1, \ldots, v_n)$ if there are orthonormal $w_1, \ldots, w_n \in \mathcal{H}_\sigma$ such that

$$\sup\left\{|\langle w_i, \sigma(g)w_j\rangle - \langle v_i, \pi(g)v_j\rangle| \,\middle|\, g \in K, 1 \leq i, j \leq n\right\} \leq \varepsilon.$$

Specifying $\sigma = \pi_0$, the trivial representation, and $n = 1$, it is easy to see that π has almost invariant vectors if and only if $\pi_0 \in \overline{\{\pi\}}$, the closure of π in the Fell topology. We can now state:

Theorem. *The following properties are equivalent:*

(i) *G is Kazhdan.*
(ii) *π_0 is an isolated point in $\hat{G}$.*
(iii) *There is a neighbourhood V of π_0 in $\tilde{G}$ such that every $\pi \in V$ has nonzero invariant vectors.*
(iv) *Every irreducible π of finite dimension is isolated in $\hat{G}$.*

See [HV], 1.14.

Like amenability, property (T) can be characterized by a fixed point property. Let $\mathcal{H}$ be a real or complex Hilbert space and recall that the group of isometries of $\mathcal{H}$ is the semidirect product

$$\text{Is } (\mathcal{H}) = (\mathcal{H}, +) \rtimes U(\mathcal{H})$$

of the translation group and the unitary (orthogonal) group. An *affine action* of G on $\mathcal{H}$ is given by a continuous homomorphism $G \to \text{Is } (\mathcal{H})$. (This must be carefully distinguished from the sort of affine actions by which amenability was characterized in 5.3.)

Theorem. *G has property (T) if and only if every affine action of G on a real or complex Hilbert space has a fixed point.*

It is elementary that elements of Is $(\mathcal{H})$ having a fixed point are conjugate to elements in $U(\mathcal{H})$. Using this and the fact that the translation part of a homomorphism $G \to \text{Is } (\mathcal{H})$ is a 1-cocycle, it is easy to see that this fixed point property is equivalent to $H^1(G, \mathcal{H}) = 0$ for every affine action of G on $\mathcal{H}$; see [HV], 4.7. Another fixed point property (which however is *not* equivalent to (T)) is

Theorem. *Every action of a Kazhdan group on a tree has a fixed point.*

See [HV], 6.4.

Finally, we point out another aspect. Assume that G is a discrete Kazhdan group, $S = S^{-1}$ a finite symmetric set of generators and $\{N\}$ a family of finite indexed normal subgroups. Then one can show that the family $\{X(G/N); S\}$ of *Cayley graphs* of G with respect to S has strong connectivity properties, which are codified in the notion of "expanding families of graphs". Such families are needed in computer science, in particular, for the construction of effective networks; but whereas it is (comparatively) easy to show that they exist, it proved much harder to construct them explicitly. The above result is also due to Margulis, and the reader is strongly urged to consult [L] for this astonishing and unexpected application.

5.4 Proof of 5.1; beginning

We now specify the notation and put $G = D^{(1)}(\mathbb{R}) = SL_d(\mathbb{R})$, $\Gamma = D^{(1)}(\mathbb{Z})$. We assume that $d \geq 3$, and that $N \lhd \Gamma$ is a normal subgroup. From 5.8 and 5.14 we see that it suffices to prove

Theorem 5.1 a. *If Γ/N is not amenable, then $N \subset C(G)$*, i.e. $N \subset \{\pm 1\}$.

The first part of the proof consists in a reformulation of 5.1 a in terms of measure theory. By turning the non-amenability of Γ/N into a "positive" statement, the measure theory, which apparently is unrelated to the subject, comes in quite naturally – another striking example for the unity of mathematics.

Theorem 5.1 b. *Let $P \subset G$ be a minimal parabolic subgroup and suppose that we are given a measure class preserving Γ-map $\varphi : G/P \to X$ to a measurable Γ-space X. Then there is a parabolic subgroup $P' \supset P$ such that X and G/P' are isomorphic as Γ-spaces and φ corresponds (almost everywhere) to the natural map $G/P \to G/P'$.*

We recall that P is the stabilizer of a full flag in $\mathbb{R}^d$, hence is conjugate to the group of upper triangular matrices in G. A parabolic $P' \supset P$ then has the form

$$\left\{ \begin{array}{l} d_1\{ \\ d_2\{ \\ \\ d_k\{ \end{array} \left(\begin{array}{ccc} & & \\ & \ddots & \\ 0 & & \end{array} \right) \in G \right\}$$

and corresponds to the ordered partition $(d_1, \ldots, d_k)$ of d. Of course the measure on G/P' is the image of the Haar measure on G.

Proof of 5.1 a from 5.1 b. Since Γ/N is not amenable, there is a compact metric Γ/N-space X without an invariant probability measure (5.4). Now P is solvable, hence amenable, so by 5.5 G/P is an amenable G-space, and by the remark preceding 5.5, G/P is an amenable Γ-space. Viewing X as Γ-space, we now obtain from 5.6 that there is a measurable Γ-map $\varphi : G/P \to M(X)$. Since G/P is compact, there is a G-invariant probability measure μ on G/P, and $\varphi_*(\mu)$ is a Γ-invariant measure on $M(X)$, the space of probability measures on X; so $\varphi : (G/P, \mu) \to \big(M(X), \varphi_*(\mu)\big)$ is a measure class preserving Γ-map. From 5.1 b we then have that $\big(M(X), \varphi_*(\mu)\big) \cong G/P'$ as Γ-spaces, for a parabolic $P' \supset P$. If $P' = G$, then the one-point set $M(X)$ would be a Γ-invariant probability measure on X, contrary to the choice of X. So $G \neq P'$. Now N operates trivially on X, hence on $M(X)$, hence on G/P'. This means that

$$N \subset \bigcap_{g \in G} gP'g^{-1},$$

which implies 5.1 a since the intersection is a proper normal subgroup of G.

We can bring 5.1 b into a more plausible form:

Theorem 5.1 c. *Keep the notations of 5.1 b and let $B \subset B(G/P)$ be a subspace of $B(G/P)$, the Borel sets of G/P, which is closed in the topology as well as under the Boolean operations. If B is Γ-invariant, then B is G-invariant.*

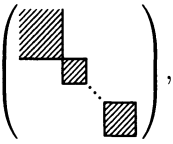
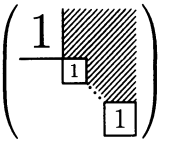
Proof of 5.1 b from 5.1 c. Let φ and X be as in 5.1 b. Let $B = \varphi^*\big(B(X)\big)$. Then B is Γ-invariant, hence G-invariant by 5.1 c. Since φ^* is a Boolean isomorphism, G acts on $B(X)$ by transport of structure, and, by the results quoted in the appenix to this chapter, G acts on X almost everywhere such that the restriction of this operation to Γ equals the original Γ-action a.e.. So we have made φ into a G-map a.e., which clearly preserves the measure class. The rest is easy: G is transitive on G/P, hence transitive up to a null set on X. This shows that $X \cong G/P'$ for a subgroup $P' \subset G$, and we can assume $P \subset P'$ since we have a surjection $G/P \to G/P'$.

Now things begin to look more promising – after all, Γ is Zariski dense in G. The strategy for the proof of 5.1 c is the following: Γ is enlarged to something *topologically* dense in G, so that the elements of G can be written as limits of elements of the enlarged Γ. But the enlargement will be as economical as possible, so that the transformations of B under the enlarged Γ remain under control. This will enable us to draw the desired conclusion.

We proceed to the details. The first step is to replace the homogeneous space G/P by a *group* $\overline{P}$ such that $B(G/P) = B(\overline{P})$ as G-spaces; more generally, we have to perform this step for any parabolic $P_0 \supset P$. Certain elements of G will operate as *automorphisms* on $\overline{P}_0$, and this enables us to apply results on *contracting automorphisms*, to be described below.

5.5 Interlude: parabolics and their opposites

This is a quite general concept, but we will only need the standard parabolics of $SL_d(\mathbb{R})$ which we have seen above (Thm. 5.1 b). Let P_0 correspond to the partition $(d_1, \ldots, d_k)$ of d. Then $P_0 = V_0 \rtimes R_0$ is a semidirect product, as visualized by

$R_0 =$  , $V_0 =$ 

More formally, V_0 is the unipotent radical, i.e. the maximal unipotent normal subgroup of P_0, and R_0 its reductive component. Evidently,

$$R_0 = \{(g_i) \in \prod_i GL_{d_i}(\mathbb{R}) \mid \prod \det g_i = 1\}.$$

The center $Z(R_0)$ consists of diagonal matrices which are blockwise scalar:

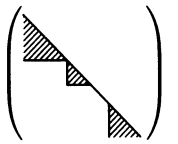
$$Z(R_0) = \{(\lambda_i 1_{d_i}) \mid \prod \lambda_i^{d_i} = 1\}.$$

The *opposite* groups are obtained by reflecting along the main diagonal. So $\overline{P}$ is the group of lower triangular, $\overline{V}$ the group of strict lower triangular matrices. Clearly $\overline{R}_0 = R_0$, and $\overline{P}_0 = \overline{V}_0 \rtimes R_0$. The following fact is standard (and intuitively plausible):

Proposition 5.15. *The restriction of $G \to G/P_0$ to $\overline{V}_0$ is a rational isomorphism of $\overline{V}_0$ onto an open conull set of G/P_0.*

For the proof, see [Zi], 5.1.4.

Define $\overline{L}_0 = R_0 \cap \overline{V} = P_0 \cap \overline{V}$; pictorially,

$\overline{L}_0 =$ 

Then $\overline{V} = \overline{V}_0 \rtimes \overline{L}_0$. It is now easy to see that the projection $\overline{V} \to \overline{V}_0$ in this semidirect decomposition makes the diagram

$$\begin{array}{ccc} \overline{V} & \longrightarrow & G/P \\ \downarrow & & \downarrow (*) \\ \overline{V}_0 & \longrightarrow & G/P_0 \end{array}$$

commute.

Denote by i the map $\overline{V} \to G/P$; i induces a measure space isomorphism. By transport of structure, G acts on $\overline{V}$ by measure space automorphisms; more precisely, for $g \in G$, $x \in \overline{V}$, $g \cdot x \in \overline{V}$ is defined by the equation

$$g\, i(x) = i(g \cdot x),$$

defined a.e. In two cases $g \cdot x$ can be made explicit:

(i) for $g \in \overline{V}$, $g \cdot x = g x$;
(ii) for $g \in P \cap \overline{P} =: A$, $g \cdot x = gxg^{-1}$.

The second equation follows from

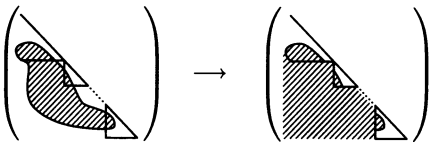
$$i(gxg^{-1}) = gxg^{-1}P = gxP = g\, i(x).$$

5.6 Continuation of the proof

We now make the following, odd-looking definition: for $E \subset \overline{V}$, put

$$\psi_0(E) = \overline{V}_0 \cdot (E \cap \overline{L}_0).$$

Pictorially,



$E \subset \overline{V}$ $\qquad$ $\psi_0(E)$

Later we shall understand how one is led to ψ_0. Let $B \subset B(\overline{V}) = B(G/P)$ be Γ-invariant. We want to show that, if $E \in B$, $g \in G$, then also $g \cdot E \in B$. This follows from the next two results:

Lemma 5.16. *$g \cdot E \in B$ if for all parabolics $G \neq P_0 \supset P$ and almost all $v \in \overline{V}$, $g \cdot \psi_0(vE) \in B$.*

Lemma 5.17. *For almost all $v \in \overline{V}$, and all parabolics $G \neq P_0$,*

$$g \cdot \psi_0(vE) = \lim_j \gamma_j \cdot E$$

for a sequence of $\gamma_j \in \Gamma$.

This will prove $g \cdot E \in B$ since B is closed and Γ-invariant.

Proof of 5.16. Let $B' \subset B \subset B(G/P)$ be the largest G-invariant subspace of B; we want to show that $B' = B$. By the correspondence between measure spaces and their Boolean algebras (used in the proof of 5.1 b from 5.1 c) we can identify $B' = B(G/P')$ for some parabolic $P' \supset P$. We are done if $P = P'$; so we assume $P \neq P'$ and derive a contradiction from the assumption that $B' \neq B$.

So choose $E \subset G/P$, $E \in B$, $E \notin B(G/P')$. Let $P_1, \ldots, P_m$ be the parabolics with $P' \supset P_i \supsetneq P$ and P_i minimal with that property. It is a standard fact that the P_i generate P'. Identify $B(G/P)$ and $B(G/P')$ with subspaces of $B(G)$ via the natural projection. Then $Eg = E$ for $g \in P$, but $Eg = E$ for all $g \in P'$ would imply $E \in B(G/P')$. So there is g in some P_i such that $Eg \neq E$. Then $E \notin B(G/P_i)$, which is embedded in $B(G/P)$ by $G/P \to G/P_i$. If we now view $E \in B(\overline{V})$, by the diagram $(*)$ above, in which P_0 is replaced by P_i, then $E\overline{L}_i \neq E$.

We have $\overline{V} = \overline{V}_i \rtimes \overline{L}_i$, and of course we assume $E \neq \emptyset$. For $w \in \overline{V}_i$ consider

$$L_w = \{y \in \overline{L}_i \mid wy \in E\} = \overline{L}_i \cap w^{-1}E.$$

If L_w were a null set for all w, then $E = \emptyset$; if L_w were conull for almost all w, then $E\overline{L}_i = E$. So for a set of $w \in \overline{V}_i$ of positive measure, L_w is neither null nor conull. For such $w, \overline{L}_i \cap v^{-1}E$ is neither null nor conull for $v \in w\overline{L}_i$, hence $\psi_i(vE)$ is not $\overline{L}_i$-invariant, a fortiori not P'-invariant. This means that, if we view $\psi_i(vE) \in B(G/P)$, it is not in the image of $B(G/P') = B'$. Since $d \geq 3$, and P_i is minimal with $P' \supset P_i \supsetneqq P$, we conclude $P_i \neq G$.

Now we can apply the hypothesis of 5.16 and find $v \in \overline{V}$ with $\psi_i(vE) \notin B'$, but $g \cdot \psi_i(vE) \in B$ for all $g \in G$. Then the Boolean subspace of B generated by all $g \cdot \psi_i(vE)$, $g \in G$, is strictly larger than B', but is G-invariant. This contradicts the maximality of B'.

In order to prove 5.17 we have to introduce two more allies.

5.7 Contracting automorphisms and the Moore Ergodicity theorem

Let H be a locally compact, compactly generated group. An automorphism α of H is called *contracting* if for any compact neighbourhood U of $1 \in H$ and any compact $K \subset H$ there is $N \in \mathbb{N}$ such that $\alpha^n(K) \subset U$ for $n \geq N$. In words, every compact K is eventually transported into arbitrary small U.

The most obvious example is $H = \mathbb{R}^n$ and $\alpha =$ scalar multiplication by $0 < a < 1$. The main example for us is as follows: Let $g = \operatorname{diag}(\lambda_1, \ldots, \lambda_d) \in GL_d(\mathbb{R})$ with $\lambda_i < \lambda_{i+1}$, and $\alpha = \operatorname{int} g$ be conjugation with g. Then α is contracting on P, and α^{-1} is contracting on $\overline{P}$. We shall need the following generalization: let $s = \operatorname{diag}(\lambda_i 1_{d_i})_i \in Z(R_0)$ (see Section 5) be such that $\lambda_i < \lambda_{i+1}$. Then $\operatorname{int} s$ is contracting on V_0, and $\operatorname{int} s^{-1}$ is contracting on $\overline{V}_0$.

The following is an easy consequence of the definition: *suppose α^{-1} is contracting. Let $E \subset H$ be a measurable set. Then*

(i) $\alpha^n(E) \to H$ *in measure if E contains a neighbourhood of 1,*
(ii) $\alpha^n(E) \to \emptyset$ *in measure if E misses a neighbourhood of 1.*

Margulis proved the following substantial extension of this, in which containment of neighbourhoods is reduced to containment of 1:

Theorem 5.18. *Suppose that α^{-1} is contracting. Then, after replacing α by some α^m if necessary, the following holds: if $E \subset H$ is measurable, then for almost all $h \in H$*

(i) $\alpha^n(hE) \to H$ *if* $1 \in hE$
(ii) $\alpha^n(hE) \to \emptyset$ *if* $1 \notin hE$.

See [Zi], 8.2.8 or [Ma2], IV.1.6 for the proof. We apply 5.18 to the following situation: assume that $H = V \rtimes L$ is a semidirect product, that α fixes V and

L and is identity on L. Given $E \subset H$ measurable, define, for $y \in L$,

$$E_y = \{x \in V \mid xy \in E\}.$$

Applying 5.18 to V, we have, for almost all $v \in V$,

$$\begin{aligned} \alpha^n(vE_y) &\to V \text{ if } 1 \in vE_y \Leftrightarrow y \in vE \\ \alpha^n(vE_y) &\to \emptyset \text{ if } 1 \notin vE_y \Leftrightarrow y \notin vE. \end{aligned}$$

Integrating over y, we get

$$\alpha^n(vE) \to V \cdot (vE \cap L) \text{ in measure}$$

for almost all $v \in V$. Using that α is trivial on L, and that L normalizes V, we obtain

Corollary 5.19. *Let $H = V \rtimes L$ and α as above. After replacing α by some α^m if necessary, we have for $E \subset H$ measurable and almost all $h \in H$, that*

$$\alpha^n(hE) \to V \cdot (hE \cap L).$$

Example 5.20. $P_0 = V_0 \rtimes R_0$ is a parabolic $\neq G$, $s \in Z(R_0)$ such that int s is contracting on V_0, and int s^{-1} is contracting on $\overline{V}_0$. Replacing, if necessary, s by some power, we have for $E \subset \overline{V}$ measurable and almost all $v \in \overline{V}$, that

$$(\text{int } s)^n(vE) \to \psi_0(vE).$$

This explains the definition of ψ_0.

Finally we need Moore's ergodicity theorem. (I am aware of the fact that it is against dramatic rules to introduce a protagonist in the second-last scene; but the result is used only here.) Let the (general) group G act on the measure space (S, μ), where it suffices to assume μ quasi-invariant, i.e. $\mu(A) = 0 \Leftrightarrow \mu(gA) = 0$ for $g \in G$. The operation is called *ergodic* if every G-invariant measurable subset is either null or conull. Clearly any transitive action is ergodic; more generally, any essentially transitive action (i.e. there is a conull orbit). An ergodic action which is not essentially transitive is called *properly ergodic*. A standard example for this is the $\mathbb{Z}$-action on the unit circle S^1 generated by $z \to e^{i\alpha}z$, where α is an irrational number.

This indicates that proper ergodicity means "complicated orbits". If S is a second countable space and G acts continuously, then almost all orbits are dense null sets ([Zi], 2.1.7).

The problem Moore's theorem deals with is: Let G be a semisimple Lie group and $H_1, H_2 \subset G$ are closed subgroups; when is H_1 ergodic on G/H_2? An elementary result is: H_1 is ergodic on G/H_2 if and only if H_2 is ergodic on G/H_1 ([Zi], 2.2.3).

Theorem 5.21. *Let G be an almost simple Lie group, Γ a lattice (a discrete subgroup of finite covolume) and H a closed noncompact subgroup. Then H is ergodic on G/Γ.*

For example, any lattice in $SL_n(\mathbb{R})$ (such as our $\Gamma = D^{(1)}(\mathbb{Z})$, or $SL_n(\mathbb{Z})$) acts ergodically on $\mathbb{P}^{n-1}$, on any Grassmann variety or flag variety, or on $\mathbb{R}^n \backslash \{0\}$. This follows by realizing these spaces ($\mathbb{R}^n$ for the last case) as homogeneous spaces of $SL_n(\mathbb{R})$.

5.8 End of proof

It still remains to show that, for almost all $v \in \overline{V}$, all $g \in G$, and $E \in B \subset B(\overline{V})$ we can write

$$g \cdot \psi_0(vE) = \lim_j \gamma_j \cdot E$$

for a suitable sequence $\{\gamma_j\} \subset \Gamma$.

Take $s \in Z(R_0)$ as in example 5.20. By 5.21 and the result mentioned before, the $\mathbb{Z}$-action on G/Γ via s is ergodic. This implies that for almost all $g \in G$, the set

$$\{\gamma g s^{-n} \mid \gamma \in \Gamma,\ n \in \mathbb{Z}\}$$

is dense in G. Actually, a sharper statement holds:

$$\{\gamma v s^{-n} \mid \gamma \in \Gamma,\ n \geq 0\}$$

is dense in G for almost all $v \in \overline{V}$. This is a sharpening in two respects: not only "one half" of $\langle s \rangle$ is taken off, but also the density is required for almost all $v \in \overline{V}$, and clearly $\overline{V}$ is of smaller dimension (a set containing almost all of G need not contain a single element of $\overline{V}$). However, since this improvement is of more technical character, the reader is spared the proof and is referred to [Zi], 8.3.3 or [Ma2], IV 2.3.

Now fix $g \in G$ and write

$$g = \lim_j \gamma_j v^{-1} s^{-n_j},\ n_j \geq 0.$$

Since we are free to replace s by powers, we can assume that $n_j \to \infty$. Write

$$g_j = \gamma_j v^{-1} s^{-n_j},\ \text{so } \gamma_j = g_j s^{n_j} v.$$

We have (recall the operation of G on $\overline{V}$)

$$\gamma_j \cdot E = g_j s^{n_j} v \cdot E = g_j \cdot (s^{n_j} v E s^{-n_j}).$$

If $j \to \infty$, $g_j \to g$ and $s^{n_j} v E s^{-n_j} \to \psi_0(vE)$ in measure, by example 5.20. Therefore, $\gamma_j \cdot E \to g \cdot \psi_0(vE)$. This finally proves 5.1.

At the end of this long journey it might be worthwile to look back to its main steps and to visualize the pattern of ideas leading to the proof. The interplay of amenability and property (T) provided the first reduction: if suffices to consider non-amenable factor groups $\overline{\Gamma}$ of Γ. Non-amenability of $\overline{\Gamma}$, together with amenability of the Γ-space G/P, brought in the measure theory: it suffices to show that, if $G/P \to X$ is a measure class preserving Γ-map, then $X \cong G/P'$ for some $P \subset P'$. This was reformulated in 5.1 c: if $B \subset B(G/P)$ is Γ-invariant, it is G-invariant.

To prove this, we viewed $B(G/P)$ as $B(\overline{V})$ and associated with $E \in B$ the family of the $\psi_0(vE)$, for all parabolics $P \subset P_0$, and almost all $v \in \overline{V}$. 5.16 showed that it suffices to prove $g \cdot \psi_0(vE) \in B$, and this follows from 5.17: $g \cdot \psi_0(vE) = \lim \gamma_j \cdot E$ for suitable $\gamma_j \in \Gamma$.

To prove 5.17 we had to introduce the auxiliary $s \in Z(R_0)$ to obtain the *dense* subsets $\{\gamma v s^{-n} | n \geq 0\}$ of G; density was secured (essentially) by the ergodicity theorem. We could then write $g = \lim \gamma_j v s^{-n_j}$ with $n_j \to \infty$, and the operation of the s^{n_j} deformed vE into $\psi_0(vE)$; in retrospect, the definition of ψ_0 can be seen as being motivated by this effect.

5.9 Appendix on measure theory

A *Borel space* is a set X together with a σ-algebra $B(X)$ of subsets of X, the *Borel sets* of X. If X is a topological space, $B(X)$ will be the algebra generated by the open (or closed) sets. A *Borel map* $f : X \to Y$ of Borel spaces is a map with $f^{-1}(B(Y)) \subset B(X)$. It is clear what is meant by isomorphism of Borel spaces. X is called a *standard Borel space* if it is isomorphic to a Borel subset of a complete separable metric space (a "Polish space").

A *measure* μ on X is a countably additive function on $B(X)$ with values in $\mathbb{R}_+ \cup \{\infty\}$. μ is *finite* if $\mu(X) < \infty$, and *σ-finite* if $X = \bigcup X_i$ with $\mu(X_i) < \infty$. μ is called a *probability measure* if $\mu(X) = 1$. The triple $(X, B(X), \mu)$ constitutes a *measure space*. A *null set* A is an $A \in B(X)$ with $\mu(A) = 0$. A is *conull* if $\mu(X \backslash A) = 0$. Two $A, B \in B(X)$ are often identified if they differ by a null set, i.e. $\mu((A \cup B) \backslash (A \cap B)) = 0$.

Two measures μ, ν are in the *same class* if $\mu(A) = 0$ if and only if $\nu(A) = 0$. A Borel map $f : (X, \mu) \to (Y, \nu)$ is *measure class preserving* if ν is in the same class as $f_*(\mu)$, the measure on Y defined by

$$f_*(\mu)(B) = \mu(f^{-1}(B)).$$

With the measure space $(X, B(X), \mu)$ there is associated the space $L^\infty(X)$ of integrable functions. $B(X)$ can be recovered from $L^\infty(X)$ as the set of (almost everywhere) idempotent functions. Thus $B(X)$ inherits a topology which depends only on the measure class of μ.

Every Borel map $f : X \to Y$ induces $f^* : B(Y) \to B(X)$, a morphism of σ-algebras. If f is measure class preserving, then f^* in *injective*; moreover, it is a *homeomorphism onto its image* with respect to the topologies of $B(X)$ and $B(Y)$ just mentioned. If X and Y are standard measure spaces, then a converse holds: if $\varphi : B(Y) \to B(X)$ is an injective continuous Boolean map, then there is a measure class preserving $f : X \to Y$, defined a.e., with $\varphi = f^*$.

Let the locally compact group G operate on X such that μ is G-quasi-invariant, i.e. $\mu(gA) = 0$ if and only if $\mu(A) = 0$. (In other words, $g : X \to X$ preserves the measure class.) Then G operates on $B(X)$. If $f : X \to Y$ preserves the measure class, then f^* is a G-map. From the last paragraph, one can deduce: if G acts continuously of $B(X)$, preserving the Boolean structure, then there is a measure class preserving G-action on X, defined a.e., which induces the given action on $B(X)$. This is used for the reduction to 5.1 c. For the proof, we refer to [Zi], Appendix B.

6 A Zariski Dense and a Free Subgroup of Γ

In this section we explicitly construct a Zariski dense subgroup and, semi-explicitly, a free subgroup of Γ. It suffices to construct such subgroups in the norm-one group of any order because, if Λ_1, Λ_2 are orders and $\Gamma_1 \subset \Lambda_1^\times$ is Zariski dense (free), then so is $\Gamma_1 \cap \Lambda_2^\times$. We shall use explicit matrix representations of Γ, which arise from the realization of D as a cyclic crossed product, and begin by recalling this construction.

As we have mentioned in the proof of 4.13, a number field K of degree $|K : \mathbb{Q}| = d$ can be embedded as a maximal subfield of D if and only if K splits D. K splits D if and only if $K_\mathfrak{p}$ splits D_p, where $\mathfrak{p}$ runs over the primes $\mathfrak{p}$ of $K, \mathfrak{p}/p$, and $K_\mathfrak{p}$ is the $\mathfrak{p}$-adic completion of K. Finally, $K_\mathfrak{p}$ splits D_p if and only if $|K_\mathfrak{p} : \mathbb{Q}_p|$ is divisible by the local index $e(p)$ of D_p. Let $p_1, \ldots, p_s$ be the ramified primes. So what we want is $K/\mathbb{Q}$ cyclic of degree d, and

$$e(p_i) \,|\, |K_{\mathfrak{p}_i} : \mathbb{Q}_{p_i}| \quad \text{for} \quad i = 1, \ldots, s \quad \text{and any} \quad \mathfrak{p}_i | p_i. \tag{1}$$

In addition we require K to be totally real, i.e.

$$K_\mathfrak{p} = \mathbb{R} \quad \text{for all} \quad \mathfrak{p}|\infty, \tag{2}$$

in order to get as many units as possible from the start. Such a K is provided by the Grunwald-Wang existence theorem ([Re], Thm. 32.18). Choosing additional primes $q_1, q_2, \ldots$, different from the p_i's, we can find cyclic K_j, $j = 2, 3 \ldots$ satisfying (1), (2) plus

$$\begin{aligned} (K_j)_{\mathfrak{q}_i} &= \mathbb{Q}_{q_i} \quad \text{for} \quad i = 1, \ldots, j-1 \quad \text{and} \quad \mathfrak{q}_i | q_i, \\ |(K_j)_{\mathfrak{q}_j} : \mathbb{Q}_{\mathfrak{q}_j}| &= d. \end{aligned} \tag{3}$$

From (3) it clearly follows that $K_j \cap K_l = \mathbb{Q}$ for $j \neq \ell$. We summarize:

Lemma 6.1. *There are infinitely many subfields of degree d of D, which are cyclic over $\mathbb{Q}$, totally real and pairwise disjoint over $\mathbb{Q}$.*

Let K be such a field. Let σ be a generator of $Gal(K|\mathbb{Q})$. By the Skolem-Noether theorem, the operation of σ on K arises from conjugation with a $u_\sigma \in D$, and $u_\sigma^d \in \mathbb{Q}^\times$. This gives us a realization of D as a cyclic crossed product

$$D = (K|\mathbb{Q}, a), \quad a \in \mathbb{Q}^\times,$$

which means that, as a $\mathbb{Q}$-algebra, D is generated by K and the additional element $u_\sigma = u$ satisfying

$$D = K \oplus Ku \oplus \cdots \oplus Ku^{d-1}, \; u\alpha u^{-1} = \sigma(\alpha), \; \alpha \in K, \; u^d = a.$$

We are free to multiply a by norm from K; in particular we may assume a integral. This implies that

$$\Lambda = (R|\mathbb{Z}, a),$$

where R is the integral domain of K, is an order in D. We remark that Λ is, in general, not maximal, not even hereditary; however, the latter can be enforced, under mild hypotheses on D, by a suitable choice of K and a. This was proved in the dissertation [Ka] of Kämpfer. One should note that, even in the quaternionic case $d = 2$, it takes quite an effort to construct a maximal order containing a given "standard order" Λ; see [Pe]. We now let Γ be the norm one group of $\Lambda^\times$.

The matrix representation of D, and thereby of Γ, arises as follows. We view D as a left K-vector space; then right multiplications with elements of D are K-automorphisms. Since the passage to right multiplication is an anti-homomorphism, we associate with $x \in D$ the transpose of the multiplication matrix with respect to the K-base $u^0 = 1, u, \ldots, u^{d-1}$ of D. This gives a matrix representation

$$M : D \to M_d(K).$$

If L is any field containing K, M provides an isomorphism

$$D_L \cong M_d(L),$$

since both sides are central simple L-algebras of dimension d^2. Explicitly,

$$M(\alpha) = \operatorname{diag}(\sigma^i(\alpha))_{i=0,\ldots,d-1} \quad \text{for} \quad \alpha \in K,$$

$$M(u) = \begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & 0 \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \\ & 0 & & & & 1 \\ a & & & & & 0 \end{pmatrix}$$

If L is any maximal subfield of D, we can construct an M as above; this shows that

$$nr_{D|\mathbb{Q}}|_L = N_{L|\mathbb{Q}}.$$

Hence the norm-one elements of L form a subtorus of $G(\mathbb{Q})$, which we denote T_L. If $S \subset L$ is a $\mathbb{Z}$-order, the norm-one elements of S form a group commensurable with $T_L(\mathbb{Z})$ (we think of T_L as defined over $\mathbb{Z}$ via an integral basis of L). The rank of $T_L(\mathbb{Z})$ is determined by the Dirichlet unit theorem, and if L is totally real, the Zariski closure $\overline{T}_L(\mathbb{Z})$ of $T_L(\mathbb{Z})$ in $G(\mathbb{R}) = SL_d(\mathbb{R})$ is a maximal torus. If we take $L = K$ and M as above, $M(\overline{T}_L(\mathbb{Z}))$ is the standard torus consisting of diagonal matrices. The other maximal subfield given by our construction is

$L = \mathbb{Q}(u) = \mathbb{Q}(\sqrt[d]{a})$. We can assume that $a > 0$; then the unit rank of L is easily calculated and equals

$$\frac{d-1}{2} \quad \text{if } 2 \nmid d, \qquad \frac{d}{2} \quad \text{if } 2 | d;$$

so T_L has maximal rank only if $d = 2$. In any case, rank $T_L \geq 1$.

The order Λ is generated, as a ring, by the two commutative orders R and $\mathbb{Z}[\sqrt[d]{a}]$. Clearly $\Gamma^\times$ contains the norm one elements of R and of $\mathbb{Z}[\sqrt[d]{a}]$. Let H be the subgroup of Γ generated by them. It would be interesting to know in which cases H has finite index in Γ. In any case, we do have

Theorem 6.2. *H is Zariski dense in $G(\mathbb{R})$.*

Proof. We show that H cannot be contained in a maximal Zariski closed subgroup of $G(\mathbb{R})$. These subgroups are the maximal parabolics, and those which contain the standard torus are conjugate to the standard maximal parabolics

$$P(k, d-k) = \left\{ \begin{pmatrix} \text{▨} & \text{▨} \\ 0 & \text{▨} \end{pmatrix} \right\}$$

by elements of the Weyl group. All these groups have the property that, for certain (i, j), the coefficients a_{ij} are always zero. It therefore sufficies to show that $\mathbb{Z}[u]^\times$ contains elements ε having *all* coefficients $\neq 0$, i.e.

$$\varepsilon = a_0 + a_1 u + \cdots + a_{d-1} u^{d-1}$$

with all $a_i \neq 0$. Since $\mathbb{Q}(u)$ is no CM-field, $\mathbb{Q}(u) = \mathbb{Q}(\mathbb{Z}[u]^\times)$ and hence $\mathbb{Z}[u]^\times \cap \Gamma$ is Zariski-dense in the norm-one torus of this field, as in the proof of 4.2. (Note that $\mathbb{Z}[u]$ is not necessarily a maximal order, but this does not affect the argument.) But the elements having at least one $a_i = 0$ lie in the union of d Zariski-closed proper subsets, which is a closed subset of strictly smaller dimension. This proves 6.2.

Applying 4.5, we obtain

Corollary 6.3. *for almost all p,*

$$H \bmod p = G(\mathbb{F}_p) \cong SL_d(\mathbb{F}_p).$$

Next we turn to the construction of free subgroups of Γ after Tits [T]. Our result is semi-explicit in the following way:

Theorem 6.4. *There are $\varepsilon \in R^\times$, $\gamma \in \Gamma$ and $N \in \mathbb{N}$, such that, for $n \geq N$, ε^n and $(\gamma \varepsilon \gamma^{-1})^n$ generate a free subgroup of Γ.*

We shall see (below) that the use of Tits' construction is greatly facilitated if we can assume that $M(\varepsilon)$ has pairwise different eigenvalues, equivalently, ε has pairwise different conjugates, equivalently, $K = \mathbb{Q}(\varepsilon)$. To this end, we show

Lemma 6.5. *There is $\varepsilon \in R^\times$ such that $K = \mathbb{Q}(\varepsilon)$. Moreover, ε can be chosen to be totally positive.*

Proof. First we assume that $d = |K : \mathbb{Q}|$ is a prime power. Then K has a unique proper maximal subfield L. Let A be the group generated by all $\varepsilon^2, \varepsilon \in R^\times$. Then $\mathbb{Q}(A) = K$, because $rk\, A = rk\, R^\times$, and we have seen already (4.2) that no proper subfield of K can have the same unit rank as K. If $\mathbb{Q}(\varepsilon) \neq K$, then $\mathbb{Q}(\varepsilon) \subset L$, and if this would hold for all $\varepsilon \in A$, it would follow that $\mathbb{Q}(A) \subset L$, contradiction. Clearly the elements of A are totally positive.

For the general case, let

$$\langle \sigma \rangle = C_1 \times \cdots \times C_s$$

be the primary decomposition of $Gal(K, \mathbb{Q})$ and let K_i be the fixed field of $\prod_{j \neq i} C_j$, so that $C_i = Gal(K_i, \mathbb{Q})$, and

$$K_i \cap \prod_{j \neq i} K_j = \mathbb{Q}.$$

Choose totally positive units $\varepsilon_i \in K_i$ such that $K_i = \mathbb{Q}(\varepsilon_i)$ and put $\varepsilon = \varepsilon_1 \cdots \varepsilon_s$. Let $\tau = (\tau_1, \ldots, \tau_s) \in \langle \sigma \rangle$ and assume, say, $\tau_1 \neq id$. Then $\varepsilon^\tau = \varepsilon$ would imply

$$1 = (\varepsilon_1^{\tau_1}/\varepsilon_1) \cdots (\varepsilon_s^{\tau_s}/\varepsilon_s),$$

whence

$$\varepsilon_1^{\tau_1}/\varepsilon_1 \in K_1 \cap \prod_{i \neq 1} K_i = \mathbb{Q}$$

and $\varepsilon_1^{\tau_1} = \pm\varepsilon_1$. This is impossible since ε_1 is totally positive and generates K_1. It follows that $\mathbb{Q}(\varepsilon) = K$, as claimed.

We now report from [T] what is needed for our purpose. Let $g \in GL_d(\mathbb{R})$ and suppose that g has real eigenvalues $\lambda_1, \ldots, \lambda_d$. Let

$$\Omega = \{\lambda_i \mid |\lambda_i| = \sup_j\{|\lambda_j|\}\}$$

and define

$$f_1(t) = \prod_{\lambda_i \in \Omega} (t - \lambda_i),\ f_2(t) = \prod_{\lambda_i \notin \Omega} (t - \lambda_i),$$

so that $f_1(t)f_2(t)$ is the characteristic polynomial of g. Further, define

$$V_i = \ker f_i(g),\ i = 1, 2$$

and let $A(g), A'(g)$ be the images of V_1, V_2 in the projective space $\mathbb{P}(\mathbb{R}^d)$. With these notations, we have

Lemma 6.6. *Suppose that $Y \subset GL_d(\mathbb{R})$ is a finite set of semisimple elements with real eigenvalues. Suppose further that*

(i) *for all* $x \in Y, A(x)$ *and* $A(x^{-1})$ *are points;*
(ii) *for all* $x \neq y$ *in* Y

$$\{A(x), A(x^{-1})\} \cap (A'(y) \cup A'(y^{-1})) = \emptyset.$$

Then there exists $N \in \mathbb{N}$ *such that, for* $m \geq N$, *the images of the* $x^m, x \in Y$, *in* $PGL_d(\mathbb{R})$ *generate a free group.*

This is a special case of [T], prop. 3.12. We proceed to verify the hypotheses of 6.6 for a two-element set $\{\varepsilon, \gamma\varepsilon\gamma^{-1}\}$ of Γ with suitable ε, γ as in 6.4; of course, we identify $\Gamma = M(\Gamma)$. Clearly, if ε is as in 6.5, it has norm 1, and its eigenvalues are positive and pairwise different. Hence $\Omega = \{\text{largest eigenvalue}\}$, $V_1 =$ the corresponding one-dimensional eigenspace, and $A(\varepsilon)$ is a point; the same holds for $\varepsilon^{-1}, \gamma\varepsilon\gamma$, and $\gamma\varepsilon^{-1}\gamma$, where $\gamma \in \Gamma$ is arbitrary. We now show that, given ε, condition (ii) of 6.6 comes down to a condition on γ which defines a nonempty Zariski open subset of $G(\mathbb{R})$; since Γ is Zariski dense, the condition can be fulfilled by suitable γ.

But this is easy: since ε and $\gamma\varepsilon\gamma^{-1}$ have the same eigenvalues, the polynomials f_1 and f_2 are the same for both of them, and

$$\begin{aligned} V_i(\gamma\varepsilon\gamma^{-1}) &= \ker f_i(\gamma\varepsilon\gamma^{-1}) = \ker \gamma f_i(\varepsilon)\gamma^{-1} \\ &= \gamma V_i(\varepsilon); \end{aligned}$$

so for example

$$A(\varepsilon) \notin A'(\gamma\varepsilon\gamma^{-1})$$

is equivalent to

$$V_1(\varepsilon) \not\subset \gamma V_2(\varepsilon),$$

which is Zariski-open on γ. It is clear that finitely many conditions of this kind define a nonempty set of $G(\mathbb{R})$, as claimed.

So far we have shown that, given ε, one can find $\gamma \in \Gamma$ and $N \in \mathbb{N}$, such that for $m \geq N$, the image of $H_0 := \langle \varepsilon^m, \gamma\varepsilon^m\gamma^{-1} \rangle$ in $PGL_d(\mathbb{R})$ is free. Now the kernel of $\Gamma \to PGL_d(\mathbb{R})$ is trivial if d is odd and $\langle -1 \rangle$ if d is even. In the latter case we are still free to replace ε by any power, so that H_0 lies in the congruence group, say, mod 3, not containing -1. Then H_0 is free, and 6.4 is proved.

We conclude this section by presenting another construction in the case $d = 2$, which relies on a result of M. Newman [Ne]. Generally it seems to be a formidable task to decide whether or not a given set of elements of a linear group generates a free subgroup. In the case of $SL_2(\mathbb{R})$, Newman has proved the following sufficient criterion: *let* $A, B \in SL_2(\mathbb{R})$ *two matrices with sign pattern*

$$A = \begin{pmatrix} - & - \\ + & + \end{pmatrix}, \quad B = \begin{pmatrix} - & + \\ - & + \end{pmatrix}$$

and tr $A \geq 2$, tr $B \geq 2$. Then A and B generate a free nonabelian group. (Here, "+" means a coefficient $\geq, 0$, likewise for "−".) This criterion lends itself, in a quite straightforward manner, to the construction of free unit groups in quaternion orders.

We continue with the previous notation. The typical element of $M(\Gamma)$ has the form

$$M(x,y) := \begin{pmatrix} x & y \\ \sigma(y)a & \sigma(x) \end{pmatrix},$$

where $x, y \in \mathbb{Z}[\sqrt{d}], d$ and $a \in \mathbb{N}$, and

$$N(x) - aN(y) = 1,$$

N denoting the norm from $\mathbb{Q}(\sqrt{a})$. In order to apply the criterion, we need $M(x,y), M(v,w)$ satisfying

(i) $Tr(x) \geq 2$, $Tr(v) \geq 2$,
(ii) $x \leq 0$; $\sigma(x) \geq 0$, $\sigma(y) \geq 0$,
(iii) $v \leq 0$, $w \geq 0$; $\sigma(v) \geq 0$, $\sigma(w) \leq 0$.

First note that if $M(x,y)$ satisfies (i) and (ii), we can take $(v,w) = (x,-y)$. If furthermore $N(x) < 0$, then also $N(y) < 0$, and (ii) holds after, if necessary, replacing x by $\sigma(x)$ and/or y by $\sigma(y)$. So all we have to do is to construct $M(x,y) \in \Gamma$ with $Tr(x) \geq 2$ and $N(x) < 0$.

Start with any $M(x,y) \in \Gamma$ having $x, y \in \mathbb{Z}$ and $xy \neq 0$. (Note that these come from $\mathbb{Z}[u]$ which is an order in a real quadratic field.) Let ε be a unit in $\mathbb{Z}[\sqrt{d}]$ with $N(\varepsilon) = 1$. Then $M(x\varepsilon^n, y) \in \Gamma$; form

$$M(x,y)M(x\varepsilon^n,y)^{-1} = \begin{pmatrix} x^2\sigma(\varepsilon)^n - ay^2 & * \\ * & x^2\varepsilon^n - ay^2 \end{pmatrix}.$$

Since there is ε with $\varepsilon \gg 1$ and, consequently, $0 < \sigma(\varepsilon) \ll 1$, we can arrange

$$x^2\sigma(\varepsilon)^n - ay^2 < 0,\ x^2\varepsilon^n - ay^2 > 0$$

and thereby (ii); condition (i) becomes

$$x^2 Tr(\varepsilon^n) - 2ay^2 \geq 2.$$

We are still free to replace $\varepsilon^n = \eta$ by powers; it is easy to see that $Tr(\eta^m)$ can be made arbitrarily large, and (i) can be fulfilled.

Note that a free subgroup of Γ cannot be of finite index (as in $SL_2(\mathbb{Z})$) since the virtual cohomological dimension $vcd(\Gamma) = 2$ (1.4). Conversely, it is known that subgroups of infinite index of a surface group are free [HKS]; since torsionfree, finite indexed subgroups of Γ are surface groups, the same holds for Γ. On the other hand, the argument from 6.2 shows that the free groups so constructed are Zariski dense.

7 An Example

In this section we present a skew field of index 3 over $\mathbb{Q}$; of course, we construct it as a cyclic crossed product.

Let $\zeta = \exp(2\pi i/7)$ and $\alpha = \zeta + \zeta^{-1}$. The minimal polynomial of α is

$$f(x) = x^3 + x^2 - 2x - 1;$$

clearly $K = \mathbb{Q}(\alpha)$ is the real subfield of the cyclotomic field $\mathbb{Q}(\zeta)$. The Galois group $G(K/\mathbb{Q})$ is generated by the restriction of $\zeta \to \zeta^2$, which we call σ.

Since 7 is the only ramified prime, and is tamely ramified, the discriminant is 7^2. This is also the discriminant of f; hence $R = \mathbb{Z}[\alpha]$ is the integral domain of K. Using the Minkowski bound, one easily checks that R has class number one. The decomposition of rational primes p is as follows: $p = 7$ is ramified (and $(\zeta - 1)(\zeta^{-1} - 1) = 2 - \alpha$ is a prime element over 7), and $p \neq 7$ is decomposed if $p \equiv \pm 1 \bmod 7$ and inert otherwise.

Claim. 2 *is not a norm from* K.

First proof by local classfield theory. $\mathbb{Q}_7(\alpha)$ is purely and tamely ramified over $\mathbb{Q}_7$, so 7 and all 1-units are norms; if 2 were a norm, the norm index would be ≤ 2, which contradicts the reciprocity law.

Second proof. First we whow for primes p:

$$p \in N(K) \Leftrightarrow p \in N(R) \Leftrightarrow p \text{ splits.}$$

Suppose $p = N(x)$, $x \in K$. Write x as a product of prime elements π of R. Since $N(\pi)$ is a prime or the cube of a prime, one sees that for every prime in the denominator of x there must be one of the *same* norm in the numerator, so these can be omitted without affecting $N(x)$. The rest is clear. (The argument extends to Galois extensions of $\mathbb{Q}$ with class number one.)

Since R is a principal ideal domain, the ideal norms are norms of integers. By the decomposition law, 2 is inert, hence cannot be a norm.

We now form the cyclic algebra

$$D = (K/\mathbb{Q}, 2)$$

which is a division ring by the above. The subring

$$\Lambda = (R/\mathbb{Z}, 2)$$

is a $\mathbb{Z}$-order in D. As a $\mathbb{Z}$-algebra, Λ has generators α and u_σ, satisfying

$$\alpha^3 + \alpha^2 - 2\alpha - 1 = 0,\ u_\sigma^3 = 2,\ u_\sigma \alpha = \alpha' u_\sigma,$$

where we write $a' = \sigma(a)$ for $a \in K$. One can check that

$$d(\Lambda) = 2^6 \cdot 7^6.$$

(There is some work to do. Fortunately, a formula for the reduced discriminant of a product order has been established in [Ka], 2.1.) Since a ramified prime appears with exponent 6 in the discriminent of a maximal order, and there must be at least two such primes, we conclude that Λ is maximal.

2-adically, $\mathbb{Q}_2(\zeta) = \mathbb{Q}_2(\alpha)$, σ operating by $\zeta \to \zeta^2$, and the given construction of D is the canonical one ([Re], 14.5). We read off that the Hasse invariant is $inv_2(D) = \frac{1}{3} + \mathbb{Z} \in \mathbb{Q}/\mathbb{Z} = Br(\mathbb{Q}_2)$. By the reciprocity law for the invariants, $inv_7(D) = \frac{2}{3} + \mathbb{Z}$.

Now we take up the norm-one group Γ. Clearly $\Lambda^\times = \langle -1 \rangle \Gamma$, and Γ is torsionfree (this holds always if d is odd). If $x \in \Lambda^\times$, then x or $-x \in \Gamma$; for the sake of convenience, let us look for elements of $\Lambda^\times$ rather than Γ. The subgroup of 6.2 becomes the group generated by $R^\times$ and $S^\times$ where

$$S = \mathbb{Z}[u_\sigma] = \mathbb{Z}[\sqrt[3]{2}]$$

(which is the maximal order in its quotient field). From the Dirichlet Unit Theorem, $R^\times$ mod $\langle -1 \rangle$ is free on two generators, and in fact can be generated by the real cyclotomic units. This follows from the fact that $\mathbb{Z}[\zeta]$ has also class number one; see [Wa]. Being content with a finite-indexed subgroup, we may take α and $\alpha + 1$ as generators: the logarithm map $l : K^\times \to \mathbb{R}^3$ is

$$l(x) = (\log|x|, \log|x'|, \log|x''|),$$

and one checks that the sign patterns of the vectors for the elements in question are

$$l(\alpha) = (+, -, +),\ l(\alpha + 1) = (+, -, -);$$

this shows that these vectors are not proportional, hence α and $\alpha + 1$ are multiplicatively independent. $S^\times$ has rank one, and a nontrivial element is $-1 + \sqrt[3]{2} =: e$. The proof of 6.2 shows that $\langle \alpha, \alpha + 1, e \rangle$ is Zariski-dense in $\Lambda^\times$.

We now show how to construct more elements of $\Lambda^\times$. Let us abbreviate

$$a + bu_\sigma + cu_\sigma^2 \ \text{ by } \ (a, b, c).$$

Our standard representation is

$$M(a, b, c) = \begin{pmatrix} a & b & c \\ 2c' & a' & b' \\ 2b'' & 2c'' & a'' \end{pmatrix}.$$

The most obvious thing to do is to write down the norm equation

$$\begin{aligned} nr\,(a,b,c) &= \det M\,(a,b,c) \\ &= N\,(a) + 2N\,(b) + 4N\,(c) - 2Tr\,(ab'c'') \quad = \pm 1. \end{aligned}$$

As it stands, this is not very illuminating, but simplifies considerably if $bc = 0$. This leads to the equations

$$\begin{aligned} N\,(a) &= 2N\,(b) \pm 1, \\ N\,(a) &= 4N\,(c) \pm 1. \end{aligned} \tag{$*$}$$

(To be precise: if a, b satisfy the first of these, then $(a, -b, 0) \in \Lambda^\times$; similarly for a, c.) Note that if $(a, b, 0)$ or $(a, 0, c) \in \Lambda^\times$, then also $(\varepsilon a, \eta b, 0)$ or $(\varepsilon a, 0, \eta c) \in \Lambda^\times$, where ε, η vary independently over $R^\times \cap \Gamma$. Thus, every nontrivial solution of $(*)$ gives rise to a four-parameter family of elements of $\Lambda^\times$. Also, we may replace a, b or a, c by conjugates independently.

To exploit $(*)$ we use the fact that the norms of elements $N\,(a)$, $a \in R$, are (up to sign) precisely the ideal norms which are known from the decomposition law. We obtain that $m \in \mathbb{Z}$ *is a norm from* R *if and only if for all primes* p *dividing* m, $p \neq 7$, *such that* $p \not\equiv \pm 1$ *mod 7, the exponent of* p *in* m *is divisible by* 3. Next, for $p \equiv \pm 1 \bmod 7$, we have to find $a \in R$ with $N\,(a) = p$. A systematic procedure is provided by the polynomial decomposition law: if

$$f\,(x) \equiv (x - a_1)(x - a_2)(x - a_3) \bmod p,$$

where $a_i \in \mathbb{Z}$, then the prime ideals over p in R are given by $\mathfrak{p}_i = (p, \alpha - a_i)$ $i = 1, 2, 3$. (See e.g. [J], 7.6.) We know that $\mathfrak{p}_i$ is principal, so we have to find $\pi_i = lcm\,(p, \alpha - a_i)$, which are the non-associate elements a of R having $|N\,(a)| = p$. In practical cases, this is not too hard: e.g. taking $p = 13$, we obtain

$$f\,(x) \equiv (x + 3)(x + 5)(x + 6) \bmod 13,$$

whence $\pi = lcm(13, \alpha + 3)$ is a prime element over 13; one checks that $\alpha + 3$ divides 13, so $\pi = \alpha + 3$; the conjugates of π are the other prime divisors of 13. However, it seems to be much more efficient to simply calculate

$$N\,(\alpha - k) = (\alpha - k)(\alpha' - k)(\alpha'' - k) = -f\,(k)$$

for small $k \in \mathbb{Z}$ and to look what happens. We quickly find

$$\begin{aligned} f\,(-3) &= -13, \quad \text{so} \quad 13 = N\,(\alpha + 3), \\ f\,(-4) &= -41, \quad \text{so} \quad 41 = N\,(\alpha + 4). \end{aligned}$$

As another example, $f(8) = 13 \cdot 43$, so

$$43 = \frac{N\,(8 - \alpha)}{N\,(\alpha + 3)},$$

which implies that a conjugate of $8-\alpha$ must be divisible by $\alpha+3$. By mechanical calculation, one finds

$$43 = N(2-\alpha+\alpha^2);$$

the same sort of calculation, based on $f(10) = 13 \cdot 83$, yields

$$83 = N(3-2\alpha+\alpha^2).$$

Some nontrivial combinations, usable for $(*)$, are

$$\begin{array}{rcrcr} 13 & - & 2\cdot 7 & = & -1 \\ 83 & - & 2\cdot 41 & = & 1 \\ 27 & - & 4\cdot 7 & = & -1 \\ 29 & - & 4\cdot 7 & = & 1, \end{array}$$

from which we obtain the elements

$$\begin{array}{ll} (\alpha+3, \alpha-2, 0), & (3-2\alpha+\alpha^2, -4-\alpha, 0), \\ (3, 0, \alpha-2), & (3-\alpha, 0, \alpha-2) \end{array}$$

of $\Lambda^\times$, plus their variations, as indicated above. I have no proof that equations $(*)$ have infinitely many (essentially different) solutions, but this looks rather plausible.

Another method for constructing elements of $\Lambda^\times$ is the following: if $\lambda \in \Lambda$, then $\mathbb{Z}[\lambda]^\times \subset \Lambda^\times$, and if $\lambda \notin \Lambda^\times$, we get at least one new element; two if we are lucky and λ is totally real. The point is simply that the search for units is much easier in three dimensions than in nine. The basis for the calculations in $\mathbb{Q}(\lambda)$ is the reduced characteristic (and minimal) polynomial

$$\text{char. pol. } (\lambda, t) = \det\big(t \cdot I_3 - M(\lambda)\big).$$

For example, one checks that, for $\lambda \in R$,

$$\text{char. pol. } (\lambda + u_\sigma, t) = \text{char. pol. } (\lambda, t) - 2.$$

Taking $\lambda = \alpha$, we obtain

$$t^3 + t^2 - 2t - 3.$$

Let $\beta = \alpha + u_\sigma$. One checks that the minimal polynomial has one real zero, hence $\mathbb{Q}(\beta)$ has unit rank one, and $\beta + 1$ is a unit (look at the regular representation of $\beta + 1$ in $\mathbb{Q}(\beta)$). Thereby we obtain

$$(1+\alpha, 1, 0) \in \Lambda^\times,$$

which, of course, was available from our first procedure, using the solution $1 - 2 \cdot 1 = -1$ of $(*)$, or varying $e = (-1, 1, 0)$ as indicated. Or,

$$\text{char. pol. } (\lambda + u_\sigma^2, t) = \text{char. pol. } (\lambda, t) - 4$$

for $\lambda \in R$. With $\lambda = \alpha$ one obtains as above that

$$(\alpha + u_\sigma^2)^2 - 3 = (\alpha^2 - 3, \alpha^2 + \alpha - 2, 1) \in \Lambda^\times.$$

As a variant of this method, check

$$\text{char. pol. } (\lambda e^{-1}, t) = \text{char. pol. } (\lambda, t) - 2S_2(\lambda)t,$$

where $e^{-1} = (1,1,1)$, $\lambda \in R$ and $S_2(\lambda)$ is the second elementary symmetric function of the conjugates of λ; in other words, the polynomial

$$t^3 + at^2 + bt + c \text{ becomes } t^3 + at^2 - bt + c.$$

For example, $\lambda = \alpha - 1$ produces

$$t^3 + 4t^2 - 3t - 1$$

which is totally real and generates a field $\neq K$ (look at the discriminant). Here, λe^{-1} is itself in $\Lambda^\times$, but $\lambda e^{-1} - 1$ is an independent unit. It is clear that by this method one can produce as many units as one wants, though calculations become awesome as coefficients increase; after all, there is no simple standard algorithm for calculating units in number fields of degree > 2.

(A special feature of these calculations will not have gone unnoticed. If $t^3 + at^2 + bt + c$ is the minimal polynomial of some $\lambda \in K$, then

$$t^3 + at^2 + bt + c - 2,\ t^3 + at^2 + bt + c - 4,\ t^3 + at^2 - bt + c$$

are again irreducible over $\mathbb{Q}_2$ and $\mathbb{Q}_7$. One would not normally expect that superficial manipulations of this kind – in fact every $d \in D \backslash \mathbb{Q}$ provides one – preserve any substantial property of polynomials. Perhaps another mystery of skew field arithmetic is waiting here to be brought to light.)

A third procedure for generating units is as follows: if $\gamma \in \Lambda^\times$ and $d \in D$, then $d\gamma d^{-1}) \in (d\Lambda d^{-1})^\times$. Since any two unit groups are commensurable, there is $m \in \mathbb{N}$ such that

$$(d\gamma d^{-1})^m = d\gamma^m d^{-1} \in \Lambda^\times.$$

Running over non-normalizers of Λ, one can produce new units out of given ones in a purely mechanical way. Example: let p be a rational prime split in K, and let π be one of three primes of R dividing p. Then, for $(a,b,c) \in \Lambda$,

$$\pi\,(a,b,c)\pi^{-1} \in \Lambda \Leftrightarrow \pi' | b \text{ and } \pi'' | c.$$

For instance, if $(a,b,c) = (-1,1,0) = e$, one can find a recursion formula for the coefficients of

$$e^N = \big(a\,(N),\, b\,(N),\, c\,(N)\big),$$

and one has to determine the smallest N with

$$b(N) \equiv c(N) \equiv 0 \bmod p.$$

This method is not very suitable for actual calculations, but might be of theoretical interest: if $\Gamma_0 \subset \Gamma$ is a subgroup closed under this operation, then it is normalized by Γ and therefore of finite index, by 5.1.

I hope that by now the reader is convinced that, although elements of Γ cannot be seen at the surface of things, as in the case of matrix groups, their calculation poses no unsurmountable difficulties. There remains, however, the problem of deciding whether or not a given set of units suffices to generate Γ or at least a cofinite subgroup; we shall deal with this question in the next section. Now it is plausible (and was suggested to me by Grunewald) that one should begin the search for generators with those coming from subfields of D of small discriminant. Let us conclude this section by pursuing this in the present case.

A list of cubic fields with small (absolute values of) discriminants is given in [DF], p. 112. Among the three fields of discriminant less than $|d(R)| = 49$ there is only one, of discriminant -31, which does not split over $\mathbb{Q}_2$ and $\mathbb{Q}_7$; a generating polynomial is

$$g(x) = x^3 + x + 1.$$

If $g(\beta) = 0$, then

$$\text{char. pol. } (\beta + 1, t) = t^3 - 3t^2 + 4t - 1.$$

One can check that this is also the polynomial of $(\alpha - 1)^{-1} e^{-1}$. Since $g(x)$ has only one real zero, we thereby possess the contributions to Γ coming from the two "discriminantally smallest" subfields of D.

A final remark: in view of practical calculation it would appear wiser to stick to the smallest possible discriminants of D, which are $(2 \cdot 3)^6$ and $(2 \cdot 5)^6$. One could replace K by the real subfield of the 9^{th} cyclotomic field, and again 2 is a non-norm, but the resulting crossed product order Λ would not be maximal because of the wild ramification. And, alas, there is no cyclic, real, cubic field of discriminant a power of 2 or 5. So, in a sense, our D is "nearest at hand". As for the realization of a D with prescribed discriminant, the following is proved in [Ka]: let $p_1, \ldots, p_s$ be the primes ramified in D. Put $c = p_1 \cdots p_s$. Then there exists a totally real cyclic field E, such that $|E : \mathbb{Q}| = d$, E is ramified outside the p_i, the ramification is tame, and $D \cong (E/\mathbb{Q}, c)$. Let R be the integral domain of E and put $\Lambda = (R/\mathbb{Z}, c)$. Then

$$d(\Lambda) = c^{d(d-1)} \text{ discr } (E)^d.$$

If D_{p_i} is a skewfield for all p_i (this is automatically so if d is a prime), D has discriminant $c^{d(d-1)}$, and the additional factor discr $(E)^d$ makes it clear that Λ is not maximal.

8 Problems

8.1 Generators

In the last paragraph we have constructed elements of Γ (for a special case only, but it is obvious how to generalize the procedures), but no attempt was made to prove that the elements so obtained actually generate Γ or at least a subgroup of finite index. To the best of my knowledge, this has been achieved for no such Γ, if $d \geq 3$. Paradoxically enough, generation of a cofinite subgroup becomes trivial if we pass from Λ to $M_n(\Lambda)$, $n \geq 2$: by results of K-theory, the elementary matrices will do the job; see e.g. [BRi] and the references quoted there. In the quaternion case $d = 2$, an algorithm is given in Fricke-Klein [FK], as well as many examples ([FK], p. 539 ss); marvellously drawn are the tesselations of the upper half plane resulting from the explicit fundamental domains and their translates.

There are at least two ways to devise an algorithm giving generators of Γ or a cofinite subgroup. The first one is derived from Section 2 and consists in the following main steps:

(1) Find the ideals $I < \Lambda$ with $|\Lambda : I| < c_1$, where c_1 is the constant of 2.1. Find generators $d_i \in \Lambda$ of these I.

(2) Find those d_i for which the cell $Z(d_i)$ is nonempty (the *critical* d_i).

(3) Put $Z = \bigcup Z(d_i)$, where d_i runs over the critical d_i, and find the $\gamma \in \Gamma$ such that $\overline{\gamma} Z \gamma \cap Z \neq \emptyset$. By 2.9, there are only finitely many such γ, and by 2.11 they generate Γ.

For step (2), take into account 2.3 plus 2.4 to obtain a finite set of $x \in \Lambda$ which contains the essential ones. Then the question of whether or not $Q \in Z(d_i)$ comes down to checking a finite set of linear inequalities for the coefficients of Q. For (3), it suffices to have an upper bound for $Tr(\overline{\gamma} \overline{d}_i d_i \gamma)$, for $\gamma \in \Gamma$ which are to satisfy $\overline{\gamma} Z(d_i) \gamma \cap Z(d_j) \neq \emptyset$, where $i \neq j$. The proof of 2.9 provides such a bound.

To describe the second algorithm, we start with the observation that $\Gamma_0 < \Gamma$ is cofinite if and only if $G(\mathbb{R})/\Gamma_0$, equivalently: H_1^+/Γ_0 is compact. To check for compactness, one can approximate a fundamental domain à la Dirichlet for the "growing" subgroup $\Gamma_0 = \langle \gamma_1, \ldots, \gamma_s \rangle$, generated by the γ's already constructed. The simple idea is the following: denote by $\varrho(\ ,\ )$ the $G(\mathbb{R})$-invariant hyperbolic metric on H_1^+. Take a $Q_0 \in H_1^+$ not fixed by any $\gamma \neq 1$. Define, for $\gamma \in \Gamma$,

$$F(\gamma) = \{ Q \in H_1^+ \mid \varrho(Q, Q_0) \leq \varrho(Q, \gamma Q_0) \},$$

a halfspace in H_1^+. It is well known (see [GGP], §1) that Γ_0 has the fundamental domain

$$\bigcap_{\gamma\in\Gamma_0} F(\gamma) \subset \bigcap_i F(\gamma_i);$$

so if the latter is compact, then Γ_0 is cocompact, hence cofinite in Γ. What is needed here is a criterion for $\bigcap_i F(\gamma_i)$ being compact, equivalently, bounded.

Let us consider the case $d = 3$. The natural operation of Γ on H^+ is easy to handle, but the det-1 condition cannot be taken care of effectively. This can be accomplished by the parametrization

$$\Phi : \mathbb{R}_+^2 \times \mathbb{R}^3 \xrightarrow{\sim} H_1^+$$

defined by

$$A = (x, y; u, v, w) \to B := \begin{pmatrix} x & u & v \\ & y & w \\ 0 & & (xy)^{-1} \end{pmatrix} \to B^t B.$$

It is easy to see that Φ is injective. To prove surjectivity, one has to solve

$$\Phi(A) = \begin{pmatrix} x^2 & xu & xv \\ & u^2+x^2 & uv+yw \\ * & & v^2+w^2+z^2 \end{pmatrix} = \begin{pmatrix} a & d & e \\ & b & f \\ * & & c \end{pmatrix}$$

for a given right-hand side in H_1^+, which comes down essentially to

$$a > 0,\ b - d^2/a > 0,$$

and these inequalities are satisfied by the Hurwitz criterion for positiveness. Thus, Φ is a diffeomorphism. But pulling back the Γ-action via

$$\gamma \circ A = \Phi^{-1}\left(\overline{\gamma}\Phi(A)\gamma\right)$$

one obtains horrible formulas. Comparing the operations of Γ on $G(\mathbb{R})$ (by multiplication), on H^+ and on $\mathbb{R}_+^2 \times \mathbb{R}^3$ via Φ, one is led to the conclusion that the price for shrinking the dimensions has to be paid by increasing complexity of the action.

8.2 The congruence problem

The *congruence group* $\Gamma(n)$ is defined as the kernel of the reduction map

$$\Gamma = G(\mathbb{Z}) \to G(\mathbb{Z} \bmod n).$$

Clearly $\Gamma(n)$ is a normal subgroup of finite index, and the congruence problem is the question: does every subgroup $N < \Gamma$ of finite index contain some $\Gamma(n)$? (It

is easy to see that no generality is lost by assuming N normal.) Let us say that (CP) holds if this is so. The following reformulation of the problem has turned out to be fruitful: Γ carries two topologies relevant to the question, the basis of the first one being the cofinite subgroups, of the second one the congruence subgroups; denoting $\tilde{\Gamma}, \hat{\Gamma}$ the respective completions, one has an exact sequence

$$1 \to C \to \tilde{\Gamma} \to \hat{\Gamma} \to 1,$$

and C is called the *congruence kernel*. It is easy to see that (CP) is equivalent to $C = 1$; the quantitative version of the (CP)-problem is the calculation of C. The advantage of this approach is that one can apply cohomological methods to the above sequence.

Of course, analogous questions can be asked for arbitrary arithmetical groups; we refer to [PR], 9.5 for a historical as well as methodical introduction, and to [Kl] for known results in the case of unit groups. In our special case, the output is meager. *Conjecturally*, C is finite for $d \geq 3$ and infinite for $d = 2$ ([PR], 9.45). As for (CP) itself, not even a conjecture has been made (in public at least). Paradoxically again, (CP) holds if we pass from Λ to $M_n(\Lambda)$, $n \geq 3$, by the results of [BRe] (these do not cover $M_2(\Lambda)$).

There is, however, a necessary condition for (CP), still a conjecture in general, which has been verified in some cases. It is concerned with the normal structure of $G(\mathbb{Q})$ rather than $G(\mathbb{Z})$. Let T be the set of primes p such that D_p is a skewfield (so T is a subset of the set of ramified p). The following was conjectured by Platonov and Margulis:

> (PM) *For any non-central normal subgroup* $N \subset G(\mathbb{Q})$ *there exists an open normal subgroup* W *of* $\prod_{p \in T} G(\mathbb{Q}_p)$ *such that* $N = G(\mathbb{Q}) \cap W$.

Here, $G(\mathbb{Q})$ is diagonally embedded into the product. Since the topology of $G(\mathbb{Q}_p)$ comes from the p-adic valuation, an open subgroup of $G(\mathbb{Q}_p)$ contains some $G(1 + p^r \mathbb{Z}_p)$, so is to be regarded as a congruence group. Note also that T may be empty, in which case the conjecture says that $G(\mathbb{Q})$ is *projectively simple*. If (PM) holds, one also says that the normal subgroups of $G(\mathbb{Q})$ have the *standard description*.

It is not too hard to show that (CP) implies (PM); see [PR], prop. 9.9. (PM) has been verified, by now, for d of the form $d = 2^r \cdot 3$; in particular, (PM) holds in our example. For more details, see [PoR]. There is no obvious way back: (PM) can be formulated for algebraic groups and holds for $SL_2(\mathbb{Q})$ (in a trivial way), whereas it is well known that (CP) fails for $SL_2(\mathbb{Z})$. But perhaps the rank-one case is exceptional in this respect, too.

Another development of more recent date is connected with the congruence question. A group A has *bounded generation* (BG), or is of *finite width at most*

t, if there are $x_1, \ldots, x_t \in A$ such that

$$A = \langle x_1 \rangle \cdots \langle x_t \rangle;$$

in words: the x_i generate A and satisfy relations which allow to transform every product of the x_i's into a product where they appear in the given order; in a sense, this is the opposite of freeness. The first (nontrivial) examples were matrix groups like $SL_n(\mathbb{Z})$, $n \geq 3$ [CK] (in spite of the fact that these contain free subgroups!). It is easy to see that (BG) carries over to groups commensurable with A; this shows that $SL_2(\mathbb{Z})$ does not have (BG). It also shows that (BG) for unit groups of orders only depends on the "algebraic background", the underlying algebra.

Of course, (BG) is of independent interest, but for our skew field units, nothing is known (or at least conjectured). However, the following has been proved in [PR2]: *if $G(\mathbb{Q})$ satisfies (PM), then (BG) for Γ implies that the congruence kernel C is finite.* The converse even holds unconditionally: *if C is finite, then (BG) holds.* So at least for $d = 2^r \cdot 3$, (BG) is equivalent to the finiteness to C.

Another group theoretical property related to these questions is "asymptotic growth of the number of subgroups of given index"; we refer to [PR2] for the definitions and for the connection with the congruence problem. The present state of art is found in Lubotzky's paper [L2].

8.3 Betti numbers

We only touch the boundary of the vast topic entitled "cohomology of arithmetic groups" and assume for simplicity that $\Delta \subset \Gamma$ is torsionfree of finite index (so we can take $\Delta = \Gamma$ if d is odd). From 1.14 we know that $cd(\Delta) = \frac{d(d+1)}{2} - 1 =: c(d)$ and

$$H^*(\Delta, \mathbb{R}) = H^*(Y, \mathbb{R}),\ Y = H_1^+/\Delta.$$

The *Betti numbers* of Δ are

$$b_i = \dim H^i(\Delta, \mathbb{R}).$$

By Poincaré duality, we have

$$b_i = b_{c(d)-i},\ i = 0, \ldots, c(d).$$

Clearly, $b_0 = b_{c(d)} = 1$. From 5.1 (or simply property (T)) plus the Hurewicz theorem we get $b_1 = b_{c(d)-1} = 0$ if $d \geq 3$. For $d = 2$, b_1 is twice the genus of Y, and is (essentially) a linear function of $\varphi(p_1 \cdots p_s)$, where $p_1, \ldots, p_s$ are the ramified primes and φ is the Euler function (note $p_1 \cdots p_s = |d(\Lambda)|^{\frac{1}{2}}$). Thus for $d = 3$, only $b_2 = b_3$ remains to be calculated, and perhaps it is not unreasonable

to expect that this number turns out to be a function of the discriminant, or other arithmetic invariants attached to D.

The *Euler characteristic* of Δ is the alternating sum

$$\chi(\Delta) = \sum (-1)^i b_i.$$

By the results in [Se], §3, the space H_1^+ carries, in a natural way, a measure μ such that

$$\chi(\Delta) = \int_{H_1^+/\Delta} \mu$$

for every torsionfree cofinite $\Delta \subset \Gamma$. By [Se], prop. 23 (see also the list in Exemples 2, p. 138) *this measure is identically zero for* $d \geq 3$, so that $\chi(\Delta) = 0$ in these cases. If $d \equiv 0, 3 \bmod 4$, then $c(d)$ is odd, and $\chi(\Delta) = 0$ holds trivially; but for $d \equiv 1, 2 \bmod 4$, $c(d)$ is even, and we obtain a nontrivial relation among the b_i's. In particular, b_j is even for $j = \frac{c(d)}{2} + 1$, and for $d = 5$, we obtain the inequality

$$1 + b_2 + b_4 + b_6 \geq b_3 + b_5.$$

To discover the laws by which these numbers are governed, would surely be of deep interest. But as with the congruence problem, the presently existing methods seem to give no clue to this problem.

References

[BH] Borel, A. and Harish-Chandra. Arithmetic Subgroups of Algebraic Groups, Math. Ann. 75 (1962) 485–535

[Bl] Blichfeldt, H.F. The Minimum Value of Quadratic Forms, And the Closest Packing of Spheres, Math. Ann. 101 (1929) 605–608

[Bo1] Borel, A. Arithmetic Properties of Linear Algebraic Groups, *Proc. Int. Congr. Stockholm* (1962), 10–22

[Bo2] Borel, A. Density and Maximality of Arithmetic Subgroups. Crelle 244 (1966) 78–89

[Br] Brown, K. *Cohomology of Groups*. Springer 1982

[BR] Bushnell, C.J. and I. Reiner. A survey of Analytic Methods in Noncommutative Number Theory, in: *Orders and their Applications*, Proceedings Oberwolfach 1984, Springer Lecture Notes in Math. 1142

[BRe] Bak, A. and U. Rehmann. The Congruence Subgroup and Metaplectic Problems for SL_n, $n \geq 2$, of Division Algebras. J. of Algebra 78 (1982) 475–547

[BRi] Bhandari, A. and J. Ritter. Large Subgroups in the Unit Groups of Arithmetic Orders. J. of Algebra 178 (1995) 512–529

[CK] Carter, D. and G. Keller. Bounded Elementary Generation of $SL_n(\mathcal{O})$, Amer. J. Math. 105 3 (1983) 673–687

[CR] Curtis, C. and I. Reiner, *Methods of Representation Theory* II. Wiley 1987

[D] Deuring, M. *Algebren*, 2. Aufl. Springer 1968

[DF] Delone, B.N. and D.K. Fadeev. Theory of Irrationalities of third Degree (Russian) Acad. Sci. URSS Steklov vol. 11 (1940)

[E] Eichler, M. Allgemeine Kongruenzklasseneinteilungen ... Crelle 179 (1938) 227–251

[FK] Fricke, R. and F. Klein. *Vorlesungen über die Theorie der Automorphen Functionen*. Bd 1, Leipzig 1897

[Fu] Furstenberg, H. Boundary Theory and Stochastic Processes on Homogeneous Spaces, in: Harmonic Analysis on Homogeneous Spaces, Sympos. Pure a. Appl. Math. 26 (1973) 193–229

[GGP] Gelfand, I.M., M.I. Graev and I. Piatetski-Shapiro. *Representation Theory and Automorphic Functions*, Saunders 1966

[Gr] Greenleaf, F.P. *Invariant Means on Topological Groups*, D. Van Nostrand 1969

[H] Hey, K. Analytische Zahlentheorie in Systemen hyperkomplexer Zahlen, Thesis Hamburg 1929

[HKS] Hoare, A.M., A. Karrass and D. Solitar. Subgroups of Infinite Index in Fuchsian Groups. Math. Zeitschr. 125 (1972) 59–69

[Hu] Hulanicki, A. Means and Følner Conditions on Locally Compact Groups. Studia Math. 27 (1966) 87–104

[HV] de la Harpe, P. and A. Valette. La propriété (T) de Kazhdan pour les groupes localement compacts, Astérisque 175 (1989)

[J] Janusz, G., *Algebraic Number Fields* Academic Press 1973

[Ka] Kämpfer, H.-H. Über die Existenz hereditärer zyklischer Ordnungen in zentral einfachen Algebren über algebraischen Zahlkörpern, Thesis Köln 1981

[Kl] Kleinert, E. Units of Classical Orders: A Survey, L'Enseignement Mathématique 40 (1994) 205–248

[L1] Lubotzky, A. *Discrete Groups, Expanding Graphs and Invariant Measures*, Birkhäuser 1994

[L2] Lubotzky, A. *Subgroup growth and Congruence Subgroups*, Inventiones Math. 119 (1995) 267–295

[Ma1] Margulis, G.A. Quotient Groups of Discrete Subgroups and Measure Theory, Funct. Anal. Appl. 13 (1979) 178–187

[Ma2] Margulis, G.A. *Discrete Subgroups of Semisimple Lie Groups*, Springer 1991

[MS] Margulis, G.A. and G.A. Soifer. Maximal Subgroups of infinite index in finitely generated linear groups, J. of Algebra 69 (1981) 1–23

[MOV] Magurn, B., R. Oliver and L. Vaserstein. Units in Whitehead Groups of Finite Groups, J. of Algebra 84 (1983) 324–360

[MVW] Matthews, C.R., L.N. Vaserstein and B. Weisfeiler. Congruence Properties of Zariski-Dense Subgroups I, Proc. London Math. Soc. (3) 48 (1984) 514–532

[N] Neukirch, J. *Algebraische Zahlentheorie* Springer 1992

[Ne] Newman, M. Pairs of Matrices Generating Discrete Free Groups and Free Products. Mich. Math. J. 15 (1968) 155–160

[No] Nori, M.V. On Subgroups of $GL_n(\mathbb{F}_p)$, Inventiones Mathematicae 88 (1987) 257–275

[Pa] Paterson, A.L.T. *Amenability*, AMS Mathematical Surveys and Monographs 29. Providence 1988

[Pe] Perlis, S. Maximal Orders in Rational Cyclic Algebras of Composite Degree. Transact. AMS 46 (1939) 82–96

[PR] Platonov, V. and A. Rapinchuk. *Algebraic Groups and Number Theory*, Academic Press 1994

[PR2] Platonov, V.P. and A. Rapinchuk. Abstract Properties of S-arithmetic Groups and the Congruence Problem, Russian Acad. Sci. Izv. Math. 40 (1993) No 3, 455–475

[PoR] Potapchik, A. and A. Rapinchuk. The Normal Subgroup Structure of SL_D and the Classification of Finite Simple Groups, C.R. Acad. Sci. Paris 320 (1995) 657–662

[Rg] Ragunathan, M.S. *Discrete Subgroups of Lie Groups*, Springer 1972

[Ram] Ramanathan, K.G. Discontinuous Groups II. Nachrichten a.d. Akad. d. Wissenschaften Göttingen, Mathem.-Physik. Klasse (1964) 145–164

[Re] Reiner, I. *Maximal Orders*, Academic Press 1975

[Ro] Rohlfs, J. Die maximalen arithmetisch definierten Untergruppen zerfallender einfacher Gruppen. Math. Annalen 244 (1979) 211–231

[Ros] Rosenlicht, M. Some Rationality Questions on Algebraic Groups, Annali di Matematica 43 (1957) 25–50

[Ru] Rudin, W. *Functional Analysis*, Mc Graw Hill 1973

[Se] Serre, J.-P., Cohomologie des groupes discrets. In: *Prospects in Mathematics*, Princeton UP 1971

[Si1] Siegel, C.L. Discontinuous Groups. Ann. of Math. 44 (1943) 674–689

[Si2] Siegel, C.L. A mean value theorem in the geometry of numbers, Annals of Math. 46 (1945) 340–347

[Sw] Swan, R.G. Strong Approximation and Locally Free Modules, in "Ring Theory and Algebra III, Proceedings of the Third Oklahoma Conference", pp. 153 223, Dekker 1980

[T] Tits, J. Free Subgroups in Linear Groups, J. of Algebra 20 (1972) 250–270

[V] Vignéras, M.-F. *Arithmétique des Algèbres de Quaternions*, Springer Lecture Notes in Math. 800 (1980)

[Wa] Washington, L.C. *Introduction to Cyclotomic Fields*, Springer 1982

[Wei] Weil, A. *Adeles and Algebraic Groups*, Birkhäuser 1982

[Wey] Weyl, H. Fundamental Domains for Lattice Groups in Division Algebras I, II (1944/45) Ges. Abh. Bd. IV, 232–264

[Za] Zassenhaus, H. On the Units of Orders, J. of Algebra 20 (1972) 368–395

[Zi] Zimmer, R.J. *Ergodic Theory and Semisimple Groups*, Birkhäuser 1984

Index