LINNIK'S ERGODIC METHOD AND THE DISTRIBUTION OF INTEGER POINTS ON SPHERES

JORDAN S. ELLENBERG, PHILIPPE MICHEL AND AKSHAY VENKATESH

Abstract. We discuss the distribution of integral solutions to

$$x^2 + y^2 + z^2 = d, \text{ as } d \to \infty.$$

We give an exposition of Linnik's ergodic method, give a refinement of his results that makes use of large-deviation results for random walks on expander graphs, and indicate the connection of these with modern developments about ergodic theory on homogeneous spaces.

Contents

Part 1. Linnik equidistribution theorems 2
1. Introduction 2
2. An overview of the ergodic method 7

Part 2. Classical theory 15
3. The action of the class group on $\widetilde{\mathscr{H}_d}$ 15
4. Representations of binary quadratic forms by $x^2 + y^2 + z^2$. 24

Part 3. Adelization 26
5. Adelic interpretation of $\mathscr{H}_d(q)$. 28
6. Adelic interpretation of $\mathscr{H}_d$. 31
7. Adelic interpretation of $\mathrm{red}_q : \mathscr{H}_d \to \mathscr{H}_d(q)$. 34

Part 4. Graphs and Expanders 35
8. The graph structure on $\mathscr{H}_d(q)$ 35
9. Expander graphs and random walks 40

Part 5. Related topics 43
10. Miscellanea and extensions of the ergodic method. 43
11. Harmonic analysis and Linnik's method. 47
References 51

We thank agencies that have generously supported our research. J.E. was partially supported by NSF-CAREER Grant DMS-0448750 and a Sloan Research Fellowship; A.V. was partially supported by a Packard fellowship, a Sloan Research Fellowship, and an NSF grant; Ph.M. is partially supported by the Advanced research Grant 228304 from the European Research Council and the SNF grant 200021-125291.

Part 1. Linnik equidistribution theorems

1. Introduction

Let $d > 1$ be an integer which, for simplicity of exposition, we assume to be *squarefree*, and let $\mathscr{H}_d$ be the set of integer points on a 2-dimensional sphere of radius $d^{1/2}$:

$$\mathscr{H}_d := \{\mathbf{x} = (x, y, z) \in \mathbb{Z}^3,\ x^2 + y^2 + z^2 = d\}.$$

The study of $\mathscr{H}_d$ is a classical question of number theory. Surprisingly, there are interesting results about $\mathscr{H}_d$ that have been proved in the last two decades, and simply stated problems that remain unresolved. In increasing order of fineness, one may ask:

(1) When is $\mathscr{H}_d$ nonempty?
(2) If nonempty, how large is $\mathscr{H}_d$, and how can we generate points in $\mathscr{H}_d$?
(3) If $\mathscr{H}_d$ get large, how is it distributed on the sphere of radius $d^{1/2}$?

The first question was studied by Legendre and the answer is:

$\mathscr{H}_d$ is nonempty if, and only if, d is not of the form $4^a(8b-1)$ $(a, b \in \mathbb{N})$,

(or put in other terms, the quadratic equation $x^2 + y^2 + z^2 = d$ satisfies the *Hasse principle*.) Legendre's proof from 1798 however was incomplete[1] and the first complete proof was given by Gauss three years later in his *Disquisitiones arithmeticae* [Gau01].

The second question is somewhat subtler and its resolution is the consequence of the work of several people. The fundamental insight however come from Gauss work who showed that $\mathscr{H}_d$ is closely connected with the *set of classes of binary quadratic forms of discriminant $-d$*.

This relation amounts, in more modern terms, to the existence of a natural *action* of the ideal class group of the quadratic ring $\mathbb{Z}[\sqrt{-d}]$ on the quotient $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d$. This action is, in fact, *transitive* (at least if d is squarefree, which we assume here) and is faithful if and only if $d \equiv 3$ modulo 8. An exposition of these facts is given in §3. In particular, whereas $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d$ does itself not have a natural group structure (what would the identity be?) the notion of "arithmetic progression" makes sense on $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d$.

A first consequence of the existence of this action is an exact formula relating $|\mathscr{H}_d|$ to the class number of $\mathbb{Q}(\sqrt{-d})$. From there, Dirichlet's class number formula expresses $|\mathscr{H}_d|$ in terms of the residue at 1 of the Dedekind ζ-function of that field, and *Siegel's theorem* controls this value quite precisely, yielding

$$|\mathscr{H}_d| = d^{1/2+o(1)}. \tag{1.1}$$

The third question is the main focus of this paper; whereas it is not at first clear that is a worthy successor to the first two, its investigation has proved

[1] Legendre assumed the existence of primes in arithmetic progressions, which was proven about 40 years later by Dirichlet.

very rich. Progress on it has been entwined with the study of modular L-functions, as well as to the study of dynamics on homogeneous spaces. The first significant answer regarding this question are due to Y. V. Linnik who, in the late 50's, proved amongst other results the following,

Theorem 1.1 (Linnik)**.** *As $d \to +\infty$ amongst the squarefree integers satisfying $d \equiv \pm 1(5)$, the set*

$$\{\frac{\mathbf{x}}{\sqrt{d}}, \mathbf{x} \in \mathscr{H}_d\} \subset S^2$$

becomes equidistributed on the unit sphere S^2 with respect to the Lebesgue probability measure.

In explicit terms, that means that if

$$\mathrm{red}_\infty : \mathscr{H}_d \mapsto S^2$$

denotes the "scaling" map $\mathbf{x} \mapsto d^{-1/2}.\mathbf{x}$, then for any measurable subset $\Omega \subset S^2$ whose boundary has Lebesgue measure zero

$$\frac{|\mathrm{red}_\infty^{-1}(\Omega)|}{|\mathscr{H}_d|} = \mathrm{area}(\Omega)(1 + o(1)),\ d \to +\infty;$$

(we take the normalization $\mathrm{area}(S^2) = 1$).

Linnik obtained this by an ingenious technique which he called the "ergodic method." This method was generalized later (notably by Linnik's student, Skubenko) to establish several remarkable results about the distribution of the representations of large integers by (integral) ternary quadratic forms [Lin68]. Until recently, Linnik's ergodic method remained surprisingly little-known, although simplified treatments were given by a number of authors [Tet83, Mal84]; one possible reason is that the method did not fit into ergodic theory as the term is now usually understood, i.e. dynamics of a measure-preserving transformation.

Removing the constraint $d \equiv \pm 1$ modulo 5 proved to be very difficult, and was resolved only thirty years later by W. Duke [Duk88]; see §11.1 for discussion.

The aim of the present paper is to revisit and explain in a slightly different language Linnik's original approach and to present further refinements which do not seem present in Linnik's work and do not seem accessible to the other known methods. We have endeavoured to simplify the original proofs of Linnik as far as possible, and have favoured explicitness and directness over generality and (more regrettably) over more conceptual approaches.

1.2. **A refinement of Theorem 1.1.** Consider, for $x \in S^2$ and $\rho > 0$, the spherical cap $\Omega(x, \rho)$ of center x and radius $\rho > 0$ (i.e. the ball of center x and radius ρ with respect to the standard Riemannian metric on S^2). Define the *deviation* of Ω to be

$$\mathrm{dev}_d(\Omega) = \frac{1}{\mathrm{area}(\Omega)} \frac{|\mathrm{red}_\infty^{-1}(\Omega)|}{|\mathscr{H}_d|} - 1.$$

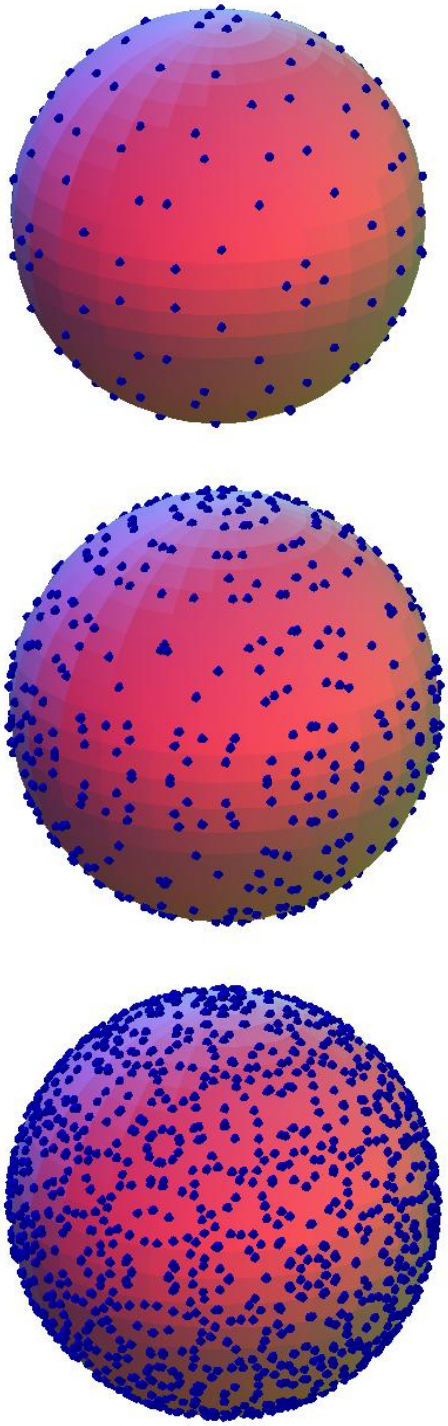

FIGURE 1. $d^{-1/2}\mathscr{H}_d \subset S^2$ for $d = 101,\ 8011,\ 104851$

Theorem 1.1 is equivalent to the fact that for any $x \in S^2$, $\rho > 0$,

$$\mathrm{dev}_d(\Omega(x,\rho)) \to 0 \text{ as } d \to \infty.$$

We shall establish the following refinement of the prior result:

Theorem. *Fix $\delta, \eta > 0$. For any $\rho \geqslant d^{-1/4+\delta}$, the measure of $x \in S^2$ for which $|\mathrm{dev}_d(\Omega(x,\rho))| \geqslant \eta$, tends to 0, as $d \to \infty$ as above (i.e. squarefree and $\equiv \pm 1(5)$).*

Note that this theorem implies Theorem 1.1. Moreover, in view of (1.1), it is easy to see that the exponent $1/4$ in the above statement is *optimal*. It would, of course, be desirable to replace "*tends to* 0 *as* $d \to \infty$" by "*equals* 0 *for sufficiently large* d." This seems extremely difficult. On the other hand, by different methods, it has been shown that this holds in a restricted range $\rho \geqslant d^{-\kappa}$ for $\kappa > 0$ some absolute constant (cf. §11).

1.3. **A modular analog.** It is also possible to prove versions of the above results where, instead of studying the position of $\mathscr{H}_d$ on the sphere, we study the congruence properties of points in $\mathscr{H}_d$. For q an integer coprime with d, let $\mathscr{H}_d(q)$ denote the "sphere modulo q"

$$\mathscr{H}_d(q) := \{\overline{\mathbf{x}} = (\overline{x}, \overline{y}, \overline{z}) \in (\mathbb{Z}/q\mathbb{Z})^3,\ \overline{x}^2 + \overline{y}^2 + \overline{z}^2 \equiv d \bmod q\}.$$

Theorem (Linnik)**.** *Let q be a fixed integer, coprime with 30. As $d \to +\infty$ amongst the squarefree integers satisfying $d \equiv \pm 1(5)$, $(d,q) = 1$, the multiset*

$$\{\mathbf{x} \pmod q,\ \mathbf{x} \in \mathscr{H}_d\} \subset \mathscr{H}_d(q)$$

becomes equidistributed on $\mathscr{H}_d(q)$ with respect to the uniform measure.

In explicit terms, that means that if

$$\mathrm{red}_q : \mathscr{H}_d \mapsto \mathscr{H}_d(q)$$

denotes the reduction modulo q map, then, for any $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$,

$$\frac{|\mathrm{red}_q^{-1}(\overline{\mathbf{x}})|}{|\mathscr{H}_d|} = \frac{1}{|\mathscr{H}_d(q)|}(1 + o(1)),\ \text{as } d \to +\infty.$$

We prove also the following refinement: define the *deviation* at $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ to be

$$\mathrm{dev}_d(\overline{\mathbf{x}}) = \frac{|\mathrm{red}_q^{-1}(\overline{\mathbf{x}})|}{|\mathscr{H}_d|/|\mathscr{H}_d(q)|} - 1$$

Theorem 1.4. *Fix $\nu, \delta > 0$ and suppose that $q \leqslant d^{1/4-\nu}$ and $(q, 30) = 1$. The fraction of $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ for which $|\mathrm{dev}_d(\overline{\mathbf{x}})| > \delta$ tends to zero as $d \to \infty$ with $d \equiv \pm 1(5)$.*

A qualitative consequence of this Theorem is that:
As long as $q \leqslant d^{1/4-\delta}$, almost any solution to $\bar{x}^2 + \bar{y}^2 + \bar{z}^2 = d$ with $(\bar{x}, \bar{y}, \bar{z}) \in (\mathbb{Z}/q\mathbb{Z})^3$, can be lifted to a solution $x^2+y^2+z^2 = d$, $(x,y,z) \in \mathbb{Z}^3$.

Again, it is natural to surmise that this is true for *all* solutions; but this appears to be a very difficult problem. It is also interesting to consider numerics related to this issue.

In this paper, we shall discuss the proof of Theorem 1.4 in some detail; all the other results may be obtained by modifying that proof; see §10 for a discussion of how to carry this out.

1.5. **Plan of the paper.** In §2 we describe the main steps of the proof of Theorems 1.1 and 1.4; for simplicity we focus on the second theorem and briefly indicate the modifications necessary to handle the first. Each of these steps is then explained in detail in the remainder of the paper:

- Part 1 concerns those parts of the proof that are most conveniently presented in classical language:
  - In §3, we describe the natural action of the class group on the quotient $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d$ and make it rather explicit. For that purpose, it is particularly useful to express everything in terms of quaternions. Presentations of similar material can be found in [Ven22, Ven29] and [She86]; our approach is slightly different.
  - In §4, we establish the important "basic lemma" of Linnik.
- Part 2 – comprised of §5, §6, and §7 is concerned with transferring some of the key objects – $\mathscr{H}_d, \mathscr{H}_d(q)$ – in terms of adelic quotients. This will be a key tool in proofs.
- Part 3 is concerned with the part of the proof that is related to expanders.
  - In §8 we prove that $\mathscr{H}_d(q)$ – endowed with a graph structure that we shall describe – is an expander graph.
  - In §9, we recall some basic facts from the theory of random walks on expander graphs. In particular we give a self-contained proof of a *large deviation estimate* (for non-backtracking paths) on such graphs.
- Part 4 is concerned with miscellaneous extensions of the main topics.
  - In §10, we discuss various possible extensions of Linnik's ergodic method.
  - In §11, we recall the alternative approach of Duke and others to these equidistribution problems, which is based on harmonic analysis and more precisely on the theory of automorphic forms. We also compare the two approaches.

1.6. **Acknowledgements.** The ideas of this paper are based on those of Linnik, and, indeed, this paper should be regarded in considerable part as an exposition of his work. This paper also draws on the ideas in the work of the latter two authors with M. Einsiedler and E. Lindenstrauss, as well as work of the first- and last- named author [ELMV09a, ELMV07, ELMV09b, EV08]. The novel results of the paper were obtained in 2005; we apologize for the delay in bringing them to print.

We also thank P. Sarnak for encouragement of the project, R. Masri for reading an early version and J.-P. Serre for his comments and criticism; finally special thanks are due to G. Harcos who read very carefully the whole manuscript and made numerous corrections and comments on that occasion.

2. An overview of the ergodic method

We now present an overview of the proof of Theorem 1.4. We will try to isolate the main steps of the proof, each one of which contains some key results of a more general mathematical interest. In the present section, we treat each of these results as a black box; in the latter part of the paper, we "open the black boxes" one by one and provide complete proofs.

2.1. **Assumptions and notation.** Throughout the paper we shall make the following assumptions:

(1) d will always denote a squarefree integer, not congruent to 7 modulo 8, and congruent to ± 1 modulo 5.
(2) q will always denote an integer prime to 30.

Remark. One could replace 5 by an arbitrary fixed prime p and the condition $d \equiv \pm 1$ modulo 5 by the assertion that p *is split in* $\mathbb{Q}(\sqrt{-d})$. The assumption that q is prime to 30 is for convenience and could be removed entirely.

Write $K = \mathbb{Q}(\sqrt{-d})$; we denote by $\mathscr{O}_K$ the ring of integers of K, and by $\mathrm{Pic}(\mathscr{O}_K)$ the ideal class group of $\mathscr{O}_K$. We also fix a square root of $-d$ in K, and denote it by $\sqrt{-d}$ without further commentary.

2.2. **Some natural quotients.** The problems considered in the introduction admit "obvious" *symmetries* owing to the evident action of $\mathrm{SO}_3(\mathbb{Z})$ on $\mathscr{H}_d$, S^2 or $\mathscr{H}_d(q)$. We denote the corresponding quotients by

$$\widetilde{\mathscr{H}_d} = \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d,\ \widetilde{\mathscr{H}_d}(q) = \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q),\ \widetilde{S^2} = \mathrm{SO}_3(\mathbb{Z})\backslash S^2.$$

and denote by $\tilde{\mathbf{x}}$ the orbit $\mathrm{SO}_3(\mathbb{Z})\mathbf{x}$ of any element in the above sets.

For the purposes of our main theorems, there is no essential difference between working with $\mathscr{H}_d$, S^2 or $\mathscr{H}_d(q)$ and working with $\widetilde{\mathscr{H}_d}$, $\widetilde{S^2}$ or $\widetilde{\mathscr{H}_d}(q)$ (since $\mathrm{SO}_3(\mathbb{Z})$ is finite). It is often conceptually clearer to consider the question of how $\widetilde{\mathscr{H}_d}$ becomes distributed in $\widetilde{\mathscr{H}_d}(q)$ or $\widetilde{S^2}$ than the similar question for $\mathscr{H}_d$ and S^2 or $\mathscr{H}_d(q)$; this becomes clear when considering these problems for other ternary quadratic forms, especially indefinite forms. However, the latter formulation being slightly more classical, we will make the (slight) extra effort required to state theorems on $\mathscr{H}_d$.

Recall that if G is a group, a *homogeneous space* for G is simply a set X on which G acts transitively (i.e. for any $x, x' \in X$, there is $g \in G$ such that $g.x = x'$). In such a case, the stabilizers of the elements of X under this action are all conjugate.

Proposition 2.3 ($\widetilde{\mathscr{H}_d}$ is a $\mathrm{Pic}(\mathscr{O}_K)$-homogeneous space)**.** *Let $d > 3$ be a squarefree integer not congruent to 7 mod 8. Then there exists a natural action of $\mathrm{Pic}(\mathscr{O}_K)$ on $\widetilde{\mathscr{H}_d}$, making $\widetilde{\mathscr{H}_d}$ into a homogeneous space for $\mathrm{Pic}(\mathscr{O}_K)$. The stabilizer of any point is trivial if $d \equiv 3$ modulo 4, and the order 2 subgroup generated by a prime above 2 if $d \equiv 1, 2$ modulo 4. In particular,*

$$|\mathscr{H}_d| = 24|\mathrm{Pic}(\mathscr{O}_K)| \text{ when } d \equiv 3 \ (8)$$

and

$$|\mathscr{H}_d| = 12|\mathrm{Pic}(\mathscr{O}_K)| \text{ when } d \equiv 1, 2 \ (4).$$

We prove a slightly more precise version of this statement in §3.1 and explain its adelic manifestation in §6.1.

Given $\mathfrak{a} \subset \mathscr{O}_K$ an ideal, we denote by $[\mathfrak{a}].\tilde{\mathbf{x}}$ the action of its corresponding ideal class on some element $\tilde{\mathbf{x}} \in \widetilde{\mathscr{H}_d}$.

2.4. **The $\mathrm{Pic}(\mathscr{O}_K)$-action.** While it is possible to describe explicitly the action of $\mathrm{Pic}(\mathscr{O}_K)$ on $\widetilde{\mathscr{H}_d}$ (or at least the action of the prime ideal classes), for the proof of the main results we will only need the action of the primes above 5. Since $d \equiv \pm 1(5)$, the prime 5 splits in K, which is to say that the principal ideal $5\mathscr{O}_K$ factors into a product of two prime ideals

$$5\mathscr{O}_K = \mathfrak{p}.\mathfrak{p}'.$$

We shall now realize explicitly the action of $[\mathfrak{p}]$ and the group it generates.

Let A, B, C be, respectively, rotations by angles $\mathrm{acos}(-4/5)$ around the x, y, z axes. They are described by the matrices:

$$A = \frac{1}{5}\begin{pmatrix} 5 & 0 & 0 \\ 0 & -4 & 3 \\ 0 & -3 & -4 \end{pmatrix}, B = \frac{1}{5}\begin{pmatrix} -4 & 0 & 3 \\ 0 & 5 & 0 \\ -3 & 0 & -4 \end{pmatrix}, C = \frac{1}{5}\begin{pmatrix} -4 & -3 & 0 \\ 3 & -4 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Proposition 2.5 (The action of a prime ideal)**.** *For $\mathbf{x} \in \mathscr{H}_d$, exactly two of*

$$\{A\mathbf{x}, A^{-1}\mathbf{x}, B\mathbf{x}, B^{-1}\mathbf{x}, C\mathbf{x}, C^{-1}\mathbf{x}\}$$

belong to $\mathscr{H}_d$. The classes of those two points in $\widetilde{\mathscr{H}_d}$ are $[\mathfrak{p}].\tilde{\mathbf{x}}$ and $[\mathfrak{p}'].\tilde{\mathbf{x}} = [\mathfrak{p}]^{-1}.\tilde{\mathbf{x}}$.

Proof. Since multiplication of an element of $\mathscr{H}_d$ by either of the matrices A, B, C or their inverse produce vector with rational coordinates with denominator equal to 1 or 5, the first part of the proposition can be verified by direct computation on the set $\mathscr{H}_d(5)$, which is to say, the set of solutions to $x^2 + y^2 + z^2 = d$ in $(\mathbb{Z}/5\mathbb{Z})^3$; one just has to check that for each $\bar{\mathbf{x}}$ in $\mathscr{H}_d(5)$, there are exactly two choices of $M \in \{5A, 5A^{-1}, 5B, 5B^{-1}, 5C, 5C^{-1}\}$ satisfying $M\overline{\mathbf{x}} = 0$. The second part will be proved in §3.12. □

Note that Proposition 2.5 in fact defines a *lifting* of the action of $[\mathfrak{p}]^{\mathbb{Z}}$ from $\widetilde{\mathscr{H}_d}$ to $\mathscr{H}_d$. This lifting is, in fact, rather less canonical than the $\mathrm{Pic}(\mathscr{O}_K)$-action on $\widetilde{\mathscr{H}_d}$.

2.6. **Trajectories.** We can use Proposition 2.5 to determine distinguished trajectories in $\mathscr{H}_d$. Write

$$\mathcal{A}_5 := \{A, A^{-1}, B, B^{-1}, C, C^{-1}\}.$$

To start with, pick any $\mathbf{x} \in \mathscr{H}_d$; now by Lemma 2.5, there are precisely two matrices in $\mathcal{A}_5$ – say $w_1,\ w_0$ – such that $w_1\mathbf{x}$ and $w_0^{-1}\mathbf{x}$ both belong to $\mathscr{H}_d$. Denote $w_1\mathbf{x}$ by $\mathbf{x}_1$, and $w_0^{-1}\mathbf{x}$ by $\mathbf{x}_{-1}$. Similarly, there is a unique choice of matrix $w_2 \in \mathcal{A}_5$ such that $w_2 \neq (w_1)^{-1}$ and $w_2\mathbf{x}_1$ belongs to $\mathscr{H}_d$; we denote $w_2\mathbf{x}_1$ by $\mathbf{x}_2$. In this way, repeated application of Lemma 2.5 gives rise to a sequence $(\mathbf{x}_i)_{i\in\mathbb{Z}}$ in $\mathscr{H}_d$ such that $\mathbf{x}_0 = \mathbf{x}$ and for any i,

$$\mathbf{x}_i \in \mathscr{H}_d,\ \mathbf{x}_{i+1} \in \{A\mathbf{x}_i, A^{-1}\mathbf{x}_i, B\mathbf{x}_i, B^{-1}\mathbf{x}_i, C\mathbf{x}_i, C^{-1}\mathbf{x}_i\} - \{\mathbf{x}_{i-1}\}.$$

Alternatively, this string can be represented by the data of $\mathbf{x} = \mathbf{x}_0$ and an infinite (necessarily periodic) word $W_{\mathbf{x}} = (w_i)_{i\in\mathbb{Z}}$, in the alphabet $\mathcal{A}_5$, satisfying, for any i,

$$w_{i+1} \neq (w_i)^{-1},\ \mathbf{x}_i = w_i\mathbf{x}_{i-1} \in \mathscr{H}_d. \tag{2.1}$$

A word satisfying the condition $w_{i+1} \neq (w_i)^{-1}$ is called *reduced*; it is easy to see that $W_{\mathbf{x}}$ is the unique reduced word satisfying (2.1), up to the "switch directions" transformation given by switching w_i and w_{1-i}^{-1}, or, equivalently, switching $\mathbf{x}_i$ and $\mathbf{x}_{-i}$. We refer to the sequence $(\mathbf{x}_i)_{i\in\mathbb{Z}}$ as the *trajectory* of $\mathbf{x}$.

The equivalence of this trajectory to the one defined by the action of $[\mathfrak{p}]^{\mathbb{Z}}$ is explained in §3.10.

2.7. **An example.** For the sake of concreteness, we include an explicit example. Take $d = 101$. In this case, $|\mathscr{H}_d| = 168$ and $|\operatorname{Pic}(\mathscr{O}_K)| = 14$. The points of $|\mathscr{H}_d|$, up to the action of $\mathrm{SO}_3(\mathbb{Z})$, are:

$$(10,1,0), \pm(9,4,2), \pm(8,6,1), \pm(7,6,4)$$

The *trajectory* containing $(10,1,0)$ is:

$$(10,1,0) \overset{B}{\to} (-8,1,-6) \overset{C^{-1}}{\to} (7,4,-6) \overset{B^{-1}}{\to} (-2,4,9) \overset{C^{-1}}{\to}$$
$$(4,-2,9) \overset{A}{\to} (4,7,-6) \overset{C^{-1}}{\to} (1,-8,-6) \overset{A^{-1}}{\to} (1,10,0)\dots$$

$$W_{(10,1,0)} = \dots B^\star C^{-1}B^{-1}C^{-1}AC^{-1}A^{-1}\dots$$

with $B^\star = w_1$.

We note that after seven steps (and thus, after any multiple of seven steps) the trajectory returns to the $\mathrm{SO}_3(\mathbb{Z})$-orbit of $(10,1,0)$. This periodicity of order 7 reflects the fact that the class of a prime ideal above 5 has order 7 in the class group of $\mathbb{Q}(\sqrt{-101})$, which is cyclic of order 14.

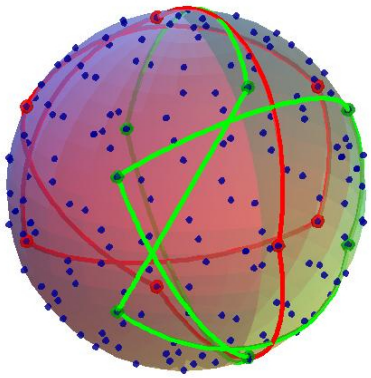

FIGURE 2. The trajectories of $(10, 1, 0)$ (red) and $(6, 1, 8)$ (green) within $\mathscr{H}_d$

2.8. **The lengths of the trajectories.** We should emphasize that it is certainly possible, and in some sense probably typical, for $[\mathfrak{p}]^{\mathbb{Z}}$ to be all or most of $\mathrm{Pic}(\mathscr{O}_K)$.

What we know in this direction is rather minimal. On the one hand, it is easy to see that

$$|[\mathfrak{p}]^{\mathbb{Z}}| \gg \log(d);$$

while this lower bound goes to ∞ with d, it remains quite small compared with the size of $|\mathrm{Pic}(\mathscr{O}_K)| = d^{1/2+o(1)}$. As far as we know, it is unknown whether there exist infinitely many squarefree $d \equiv 1(20)$ such that $|[\mathfrak{p}]^{\mathbb{Z}}|$ is greater than some positive power of d.

These considerations are, of course, related to the question:

for which d does the process described in Proposition 2.5 traverse all points in $\widetilde{\mathscr{H}_d}$?

This condition on d turns out to be equivalent to the condition that the prime ideal $\mathfrak{p}$ above 5 and the 2-torsion ideal class above 2 generate the ideal class group $\mathbb{Q}(\sqrt{-d})$. The question of *whether this happens for infinitely many d* is an interesting and seemingly difficult one; heuristically it is reasonable to suppose that this happens a positive fraction of the time. It is analogous to Gauss's famous question:

do there exist infinitely many $d > 0$ for which $\mathbb{Q}(\sqrt{d})$ has class number 1?

The prime above 5 plays the role in our situation that a real place does in Gauss's problem; the *period* of the process above is analogous to the regulator of a real quadratic field.

2.9. **Trajectories on the graph $\mathscr{H}_d(q)$.** We shall now begin to examine solutions to $x^2 + y^2 + z^2 = d$ modulo a positive integer q, as in Theorem 1.3. We shall suppose, as in that theorem, that q is relatively prime to 30.

Under this assumption, the matrices $\mathcal{A}_5 = \{A, A^{-1}, B, B^{-1}, C, C^{-1}\}$ define elements in $\mathrm{SO}_3(\mathbb{Z}/q\mathbb{Z})$ and act on $\mathscr{H}_d(q)$; this endows $\mathscr{H}_d(q)$ with a

structure of a 6-regular (undirected) graph by joining each $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ to

$$A\overline{\mathbf{x}},\ A^{-1}\overline{\mathbf{x}},\ B\overline{\mathbf{x}},\ B^{-1}\overline{\mathbf{x}},\ C\overline{\mathbf{x}},\ C^{-1}\overline{\mathbf{x}}.$$

By an abuse of notation we shall also refer to this graph as $\mathscr{H}_d(q)$. More precisely, it is a *multigraph* – we allow multiple edges between pairs of vertices, and edges joining a vertex to itself.

Now to each point $\mathbf{x} \in \mathscr{H}_d$, we may associate a well-defined *marked path* on the graph $\mathscr{H}_d(q)$ in the following way. Given $\mathbf{x} \in \mathscr{H}_d$, we constructed in §2.6 a sequence $(\mathbf{x}_i)_{i\in\mathbb{Z}}$ of points in $\mathscr{H}_d$, the *trajectory* of $\mathbf{x}$, which is well-defined up to the substitution $\mathbf{x}_i \mapsto \mathbf{x}_{-i}$. We denote by $\gamma_{\mathbf{x}}$ the reduction of this trajectory to $\mathscr{H}_d(q)$; the data of $\gamma_{\mathbf{x}}$ is the sequence of vertices $(\overline{\mathbf{x}}_i)_{i\in\mathbb{Z}}$ of $\mathscr{H}_d(q)$ ($\overline{\mathbf{x}}_i = \mathrm{red}_q(\mathbf{x}_i)$), together with a marked basepoint $\overline{\mathbf{x}}_0$ and a choice, for each i, of an edge joining $\overline{\mathbf{x}}_i$ to $\overline{\mathbf{x}}_{i+1}$. We denote by $\gamma_{\mathbf{x}}^{(\ell)}$ the segment of this path of length 2ℓ, centered at the marked vertex. In other words,

$$\gamma_{\mathbf{x}}^{\ell} = (\mathbf{x}_{-\ell}, \mathbf{x}_{-\ell+1}, \dots, \mathbf{x}_{\ell-1}, \mathbf{x}_{\ell}).$$

We may refer to it as the *truncated* trajectory of length 2ℓ.

Note that the trajectories arising as $\gamma_{\mathbf{x}}$ are not completely arbitrary paths on $\mathscr{H}_d(q)$; the condition $w_{i+1} \neq w_i^{-1}$ from (2.1) implies that $\gamma_{\mathbf{x}}$ never traverses the same edge twice in succession. (This does not, of course, forbid $\gamma_{\mathbf{x}}$ from traveling from $\overline{x}$ to $\overline{y}$ and then back to $\overline{x}$; it just has to use two distinct edges joining $\overline{x}$ and $\overline{y}$.) A *non-backtracking path* in $\mathscr{H}_d(q)$ is a (marked) path which never traverses the same edge twice in succession; equivalently, it can be defined by the data of a marked point $\bar{x}_0$, and a word $W_{\mathbf{x}} = (w_i)_{i\in\mathbb{Z}}$ satisfying $w_{i+1} \neq w_i^{-1}$; the vertices $\bar{x}_i$ can be determined inductively by the rule that $\overline{x}_{i+1}$ is the vertex arrived at by following the edge labeled w_{i+1} from $\overline{x}_i$.

For instance, if $d = 101, q = 7, P = (7, 4, -6)$, we have (see §2.7)

$$\gamma_P^{(2)} \ : \cdots\ [(3,1,0)] \overset{B}{\to} [(-1,1,1)] \overset{C^{-1}}{\to} [(0,4,1)]^{\star}$$
$$\overset{B^{-1}}{\to} [(-2,4,2)] \overset{C^{-1}}{\to} [(4,-2,2)] \cdots$$

Here $\star$ denotes the marked vertex, and we have used square brackets $[\dots]$ to denote reduction modulo 7.

Proposition 2.10 (Shadowing lemma)**.** *The sequences $W_{\mathbf{x}}$ and $W_{\mathbf{x}'}$ coincide for $-\ell + 1 \leqslant i \leqslant \ell$ if and only if $\mathbf{x} \equiv \pm\mathbf{x}'$ modulo 5^{ℓ}. In particular, $\gamma_{\mathbf{x}}^{(\ell)} = \gamma_{\mathbf{x}'}^{(\ell)}$ only if $\mathbf{x} \equiv \pm\mathbf{x}'$ modulo $q5^{\ell}$.*

This proposition states roughly that if two points have their trajectory agree for a long time then their initial points have to be 5-adically close as well as congruent modulo q. (In particular, the trajectories $\gamma_{\mathbf{x}}$, $\gamma_{\mathbf{x}'}$ are equal if and only if $\mathbf{x} = \pm\mathbf{x}'$.)

The "if" part is easy; the "only if" is not much harder, although this is less apparent from our definitions. Proposition 2.10 is proved in §8.3, using geometric properties of the Bruhat-Tits building of $\mathrm{SO}_3(\mathbb{Q}_5) \simeq \mathrm{PGL}_2(\mathbb{Q}_5)$.

Our aim is to show that the integral points $\mathscr{H}_d$ and their associated trajectories $\{\gamma_{\mathbf{x}}, \mathbf{x} \in \mathscr{H}_d\}$ are well distributed, in some sense, on the graph $\mathscr{H}_d(q)$. (In interpreting this statement, one should imagine that q is fixed, or not increasing too quickly, whereas $d \to \infty$.) To obtain this, we shall use Proposition 2.10 in conjunction with:

Proposition 2.11 (Linnik's basic Lemma)**.** *Let $e \in \mathbb{Z}$ such that $|e| < d$. The number of pairs $(\mathbf{x}_1, \mathbf{x}_2) \in \mathscr{H}_d^2$ with dot product $\mathbf{x}_1 \cdot \mathbf{x}_2 = e$ is $\ll_\varepsilon d^\varepsilon$.*

This is proved in §4. It corresponds to the "basic lemma" in Linnik's ergodic method and is in a sense, a generalization of the well known bounds

$$\tau(d) = \sum_{ab=d} 1 \ll_\varepsilon d^\varepsilon, \; r_2(d) = \sum_{a^2+b^2=d} 1 \ll_\varepsilon d^\varepsilon.$$

Indeed, bounding the divisor function $\tau(d)$, or the number of representations $r_2(d)$ of d as the sum of two squares, amount to bounding the number of representations of the rank 1 form dx^2 by the binary quadratic forms $Q(x, y) = xy$ and $Q(x, y) = x^2 + y^2$ respectively. By comparison, Proposition 2.11 concerns the number of ways to represent the rank two form $dx^2 + exy + dy^2$ by the rank three form $x^2 + y^2 + z^2$.

Proposition 2.12. *Let $\Sigma(d, \ell, q)$ be the number of pairs $(\mathbf{x}_1, \mathbf{x}_2) \in \mathscr{H}_d^2$ with $\gamma_{\mathbf{x}}^{(\ell)} = \gamma_{\mathbf{x}'}^{(\ell)}$. Then*

$$\Sigma(d, \ell, q) \ll_\varepsilon |\mathscr{H}_d| + d^\varepsilon \left(1 + \frac{d}{q^2 5^{2\ell}}\right),$$

for any $\varepsilon > 0$.

Proof. Indeed, $\gamma_{\mathbf{x}}^{(\ell)} = \gamma_{\mathbf{x}'}^{(\ell)}$ implies that $\mathbf{x} \pm \mathbf{x}' \equiv 0 \; (q5^\ell)$ for some choice of $\pm$; hence either $(\mathbf{x} + \mathbf{x}').(\mathbf{x} + \mathbf{x}') \equiv 0 \; (q^2 5^{2\ell})$ or $(\mathbf{x} - \mathbf{x}').(\mathbf{x} - \mathbf{x}') \equiv 0 \; (q^2 5^{2\ell})$ Since $\mathbf{x}.\mathbf{x} = \mathbf{x}'.\mathbf{x}' = d$, we have

$$2\mathbf{x}.\mathbf{x}' \equiv \pm 2d \bmod q^2 5^{2\ell}$$

in either case. Thus the number of pairs $(\mathbf{x}, \mathbf{x}')$ such that $\gamma_{\mathbf{x}}^{(\ell)} = \gamma_{\mathbf{x}'}^{(\ell)}$ is bounded by

$$|\mathscr{H}_d| + \sum_{\substack{|e|<d \\ e \equiv \pm d (q^2 5^{2\ell})}} |\{(\mathbf{x}, \mathbf{x}') \in \mathscr{H}_d^2, \; \mathbf{x}.\mathbf{x}' = e\}| \ll_\varepsilon |\mathscr{H}_d| + d^\varepsilon \left(1 + \frac{d}{q^2 5^{2\ell}}\right).$$

□

Proposition 2.13 ($\mathscr{H}_d(q)$ is an expander)**.** *For any q coprime with 30, the graph $\mathscr{H}_d(q)$ is connected and non-bipartite. In particular, its adjacency matrix, $A(\mathscr{H}_d(q))$, has a* spectral gap *and more precisely, if*

$$\lambda_1 = 6 \geqslant \lambda_2 \geqslant \ldots \lambda_{|\mathscr{H}_d(q)|} \geqslant -6$$

denote the eigenvalues of $A(\mathscr{H}_d(q))$, then

$$|\lambda_j| \leqslant 2\sqrt{5}, \quad j \neq 1.$$

This is proved in Section §8.4 using the adelic description of $\mathscr{H}_d(q)$. Indeed, the graphs $\mathscr{H}_d(q)$ are closely related to the original Ramanujan graphs of Lubotzky-Phillips-Sarnak[LPS88]. The content of this assertion is equivalent to the (optimal) Ramanujan bound on the Fourier coefficients of weight 2 holomorphic forms of level up to $18.q^2$. Thus, the existence of a spectral gap follows from any nontrivial bound for these Fourier coefficients – for instance, the works of Kloosterman or Rankin.

The bound on the spectral gap in Proposition 2.13 says that the family of 6-valent graphs $\left(\mathscr{H}_d(q)\right)_{(q,10)=1}$ form a family of $\frac{3-\sqrt{5}}{3}$ *expander* (in fact Ramanujan) graphs as $q \to \infty$. We refer to [Lub94] for an extensive and motivated discussion of expander graphs, their properties, applications, and their construction via automorphic forms.

We shall use the fact that $\mathscr{H}_d(q)$ is an expander through:

Proposition 2.14 (Large deviation estimates)**.** *Fix $\eta, \varepsilon > 0$. For any subset $\mathcal{B} \subset \mathscr{H}_d(q)$ with $|\mathcal{B}| \geqslant \eta \mathscr{H}_d(q)$, the fraction of non-backtracking paths γ of length 2ℓ satisfying:*

$$\left| \frac{|\gamma \cap \mathcal{B}|}{2\ell+1} - \frac{|\mathcal{B}|}{|\mathscr{H}_d(q)|} \right| \geqslant \varepsilon$$

is bounded by $c_1 \exp(-c_2 \ell)$, where c_1, c_2 depend only on ε, η.

By $|\gamma \cap \mathcal{B}|$ we mean the number of $i \in [-\ell, \dots, \ell]$ such that the vertex $\overline{\mathbf{x}}_i$ of γ is contained in $\mathcal{B}$; in other words, the "amount of time" γ spends in $\mathcal{B}$ if one imagines moving along the trajectory at a constant speed. It is then natural to compare the portion of time spent by a path in $\mathcal{B}$ (i.e. the ratio $|\gamma \cap \mathcal{B}|/(2\ell+1)$) with the probability of being in $\mathcal{B}$ (i.e. $|\mathcal{B}|/|\mathscr{H}_d(q)|$). Large deviation estimates show that with high probability, long paths spend the right amount of time in any large enough subset $|\mathcal{B}|$.

Such estimates, first proved by Chernoff for complete graphs, are a well-known and useful tool in different contexts; for instance it has been fruitfully applied in computer science (see [HLW06]). We were unable to find a reference in the existing literature for the particular version we need here (a large deviation estimate for non-backtracking walks), and so give a proof from first principles in §9.2 (Proposition 9.4).

2.15. **Conclusion of the proof.** We conclude this section by explaining how Propositions 2.10, 2.11 and 2.14 together imply Theorem 1.4.

Let $\mathcal{B}_\delta$ be the set of $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ such that

$$\mathrm{dev}_d(\overline{\mathbf{x}}) = \frac{|\mathrm{red}_q^{-1}(\overline{\mathbf{x}})|}{|\mathscr{H}_d|/|\mathscr{H}_d(q)|} - 1 > \delta. \tag{2.2}$$

Suppose that $|\mathcal{B}_\delta| \geqslant \eta |\mathscr{H}_d(q)|$. We will derive a contradiction for fixed δ, η and large enough d. A similar bound applies to the set of $\overline{\mathbf{x}}$ for which $\mathrm{dev}_d(\overline{\mathbf{x}}) < -\delta$; taken together these yield Theorem 1.4.

If $\mathcal{B}$ is a subset of $\mathscr{H}_d(q)$, and $\gamma_{\mathbf{x}}^{(\ell)}$ is the trajectory of an $\mathbf{x}$ chosen randomly – with respect to counting measure – from $\mathscr{H}_d$, the expected size of $\gamma_{\mathbf{x}}^{(\ell)} \cap \mathcal{B}$

is just the sum over $i \in [-\ell \dots \ell]$ of the probability that $\overline{\mathbf{x}}_i$ lies in $\mathcal{B}$; this probability is $|\mathrm{red}_q^{-1}(\mathcal{B})|/|\mathscr{H}_d|$, independently of i. In other words,

$$\frac{1}{|\mathscr{H}_d|}\sum_{\mathbf{x}\in\mathscr{H}_d}\frac{|\gamma_{\mathbf{x}}^{(\ell)}\cap\mathcal{B}|}{2\ell+1} = \frac{|\mathrm{red}_q^{-1}(\mathcal{B})|}{|\mathscr{H}_d|}. \tag{2.3}$$

We shall take $\mathcal{B} = \mathcal{B}_\delta$ and choose ℓ so that

$$\frac{1}{5}|\mathscr{H}_d| < q^2 5^{2\ell} \leqslant 5|\mathscr{H}_d|. \tag{2.4}$$

Note that we've used here the hypothesis that $q \leqslant d^{1/4-\nu}$.

From (2.2) and (2.3), the average value of $\dfrac{|\gamma_{\mathbf{x}}^{(\ell)}\cap\mathcal{B}_\delta|}{2\ell+1}$ as $\mathbf{x}$ ranges over $\mathscr{H}_d$ exceeds

$$\frac{|\mathcal{B}_\delta|}{|\mathscr{H}_d(q)|}(1+\delta) \geqslant \frac{|\mathcal{B}_\delta|}{|\mathscr{H}_d(q)|} + \delta\eta.$$

Since $\frac{|\gamma_{\mathbf{x}}^{(\ell)}\cap\mathcal{B}_\delta|}{2\ell+1} \leqslant 1$ for every $\mathbf{x}$, the number of $\mathbf{x} \in \mathscr{H}_d$ for which

$$\frac{|\gamma_{\mathbf{x}}^{(\ell)}\cap\mathcal{B}_\delta|}{2\ell+1} > \frac{|\mathcal{B}_\delta|}{|\mathscr{H}_d(q)|} + \delta\eta/2 \tag{2.5}$$

is at least

$$\frac{\delta\eta}{2}|\mathscr{H}_d| \gg_\varepsilon \delta\eta d^{1/2-\varepsilon}$$

for any $\varepsilon > 0$ (the last bound following from (1.1)).

Let M be the number of non-backtracking marked paths on $\mathscr{H}_d(q)$ satisfying (2.5). Given that ℓ is chosen so that (2.4) is valid, we see from Proposition 2.12 that the number of pairs $(\mathbf{x}, \mathbf{x}')$ yielding the same trajectory $\gamma_{\mathbf{x}}^{(\ell)}$ on $\mathscr{H}_d(q)$ is not much larger than the number of diagonal pairs $|\mathscr{H}_d|$: i.e. is bounded by $\ll_\varepsilon d^{1/2+\varepsilon}$. It follows that

$$M \gg_{\varepsilon,\delta,\eta} d^{1/2-\epsilon}. \tag{2.6}$$

for any $\epsilon > 0$.

On the other hand, we also have an upper bound for M. The total number of non-backtracking marked paths of length 2ℓ on $\mathscr{H}_d(q)$ is on order of $5^{2\ell-1}|\mathscr{H}_d(q)| \ll_\varepsilon d^{1/2+\varepsilon}$. Proposition 2.14 says that, of these, the proportion which satisfy (2.5) is at most $d^{-\tau}$ for some $\tau = \tau(\delta,\eta) > 0$; so

$$M \ll_{\varepsilon,\delta,\eta} d^{1/2-\tau} \tag{2.7}$$

which yields a contradiction for d sufficiently large.

2.16. **Hecke trees on the sphere: an idea of the proof of Theorems 1.1 and 1.2.** We briefly sketch the proof of Theorems 1.1 and 1.2. The archimedean analogue of Proposition 2.13 is the fact that the Hecke operator on $L^2(S^2)$

$$T_5.f:\ \mathbf{x}\mapsto \sum_{w\in\mathcal{A}_5} f(w.\mathbf{x})$$

has a spectral gap: the eigenvalue 6 has multiplicity one and any other eigenvalue λ satisfies

$$|\lambda|\leqslant 2\sqrt{5};$$

this again is a consequence of Deligne's bounds for Fourier coefficients of holomorphic modular forms.

Let $\mathcal{T}_5$ be the set of reduced words in the alphabet $\mathcal{A}_5$, which has the structure of a rooted 6-valent tree. To any $\mathbf{x}$ in S^2, one associates the infinite graph $\mathcal{T}_5.\mathbf{x}\subset S^2$, which is called *the Hecke tree*. The most direct consequence of the spectral gap property is that $\mathcal{T}_5.\mathbf{x}$ is dense in S^2 and even equidistributed: more precisely, uniformly on $\mathbf{x}\in S^2$, the set $\{W_\ell.\mathbf{x}\}$ where W_ℓ ranges over the reduced words in $\mathcal{T}_5$ of length ℓ becomes equidistributed as $\ell\to+\infty$ [LPS86]. A more refined consequence is a large deviation estimate for non-backtracking marked paths of length 2ℓ –uniform in the origin of the path–which is proven along the same lines as Proposition 2.14.

To prove Theorem 1.2, one proceeds from here as in the proof of Theorem 1.4, replacing conditions like "two trajectories stay equal mod q for a long time" with "two trajectories stay near each other on the sphere for a long time."

Part 2. Classical theory

3. The action of the class group on $\widetilde{\mathscr{H}}_d$

We present, in Proposition 3.5, a precise version of the homogeneous space structure on $\widetilde{\mathscr{H}}_d$ discussed in Proposition 2.3. As we have remarked, the basic ideas here go back to Venkov [Ven22, Ven29] and in some sense to Gauss.

It will be convenient to modify, slightly, the definition of $\widetilde{\mathscr{H}}_d$ in the case when $d\equiv 3$ modulo 4. Let $\mathrm{SO}_3(\mathbb{Z})^+$ be the index-2 subgroup of $\mathrm{SO}_3(\mathbb{Z})$ consisting of matrices which act on the coordinate lines via even permutations. Set

$$\widetilde{\mathscr{H}}_d^*=\begin{cases}\mathrm{SO}_3(\mathbb{Z})^+\backslash\mathscr{H}_d & \text{, if } d\equiv 1,2 \bmod 4\\ \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d & \text{, if } d\equiv 3 \bmod 4\end{cases}$$

Thus, $\widetilde{\mathscr{H}}_d^*$ and $\widetilde{\mathscr{H}}_d$ are equal when $d\equiv 3$ modulo 4; otherwise, the former is a double cover of the latter.

We shall show that in fact $\widetilde{\mathscr{H}}_d^*$ has the natural structure of torsor for $\mathrm{Pic}(\mathscr{O}_K)$. Recall that a space X homogeneous for the action of some group G and for which the stabilizer of some (hence any) point is trivial is called

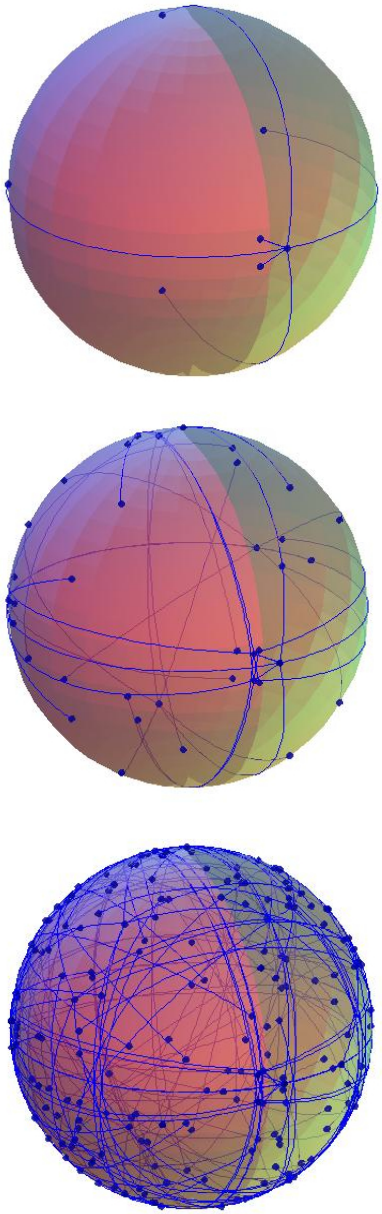

FIGURE 3. The first three layers of the Hecke tree at $(10, 1, 0)$

a *principal homogeneous space* or *torsor* for G. In that case, for any $x \in X$, the map

$$g \in G \mapsto g.x \in X$$

for any given $x \in X$ provides an identification of G with X as G-spaces. There is a priori no canonical way of choosing x; thus, one may think of a

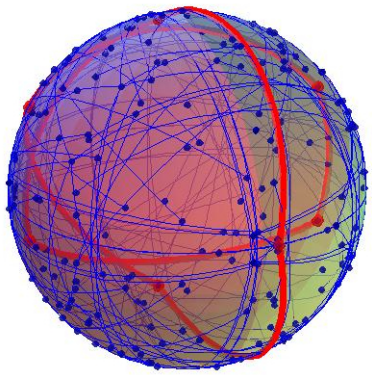

FIGURE 4. The trajectory of $(10, 1, 0)$ within its Hecke tree

torsor as a set endowed with many different identifications with G but, in general, with no *canonical* one. For instance, the set of nth roots of 2 is a torsor for the group μ_n of nth roots of unity.

3.1. **Quaternions.** We first recall the following classical facts.

Let B be the $\mathbb{Q}$-algebra of Hamilton quaternions. For $x = u + a.i + b.j + b.k \in \mathrm{B}$, the canonical involution is noted $\overline{x} = u - a.i - b.j - b.k$, the reduced trace $\mathrm{tr}(x) = x + \overline{x} = 2u$ and the reduced norm $\mathrm{Nr}(x) = x.\overline{x} = u^2 + a^2 + b^2 + c^2$. Let $\mathrm{B}^{(0)}$ denote the space of trace-free quaternions (the kernel of tr) also called the *pure* quaternions. The space $\mathrm{B}^{(0)}$ endowed with the reduced norm is a quadratic space and the map

$$(a, b, c) \mapsto a.i + b.j + b.k \tag{3.1}$$

is an isometry between the quadratic space $(\mathbb{Q}^3, a^2 + b^2 + c^2)$ and $(\mathrm{B}^{(0)}, \mathrm{Nr})$. In the sequel, we will freely identify $\mathbb{Q}^3$ with $\mathrm{B}^{(0)}$, and, in particular, consider the elements of $\mathscr{H}_d$ as trace-free quaternions.

We denote by $\mathrm{B}^\times$, B^1 and $\mathrm{PB}^\times = \mathrm{B}^\times / Z(\mathrm{B}^\times)$ respectively , the group of units of B, the subgroup of units of reduced norm one, and the projective group of units; these define $\mathbb{Q}$-algebraic groups, and the action of $\mathrm{B}^\times$ on $\mathrm{B}^{(0)}$ by conjugation induces a covering and an isomorphism of $\mathbb{Q}$-algebraic groups [Vig80, Th. 3.3]:

$$Z(\mathrm{B}^\times) \hookrightarrow \mathrm{B}^\times \twoheadrightarrow \mathrm{PB}^\times \xrightarrow{\sim} \mathrm{SO}(a^2 + b^2 + c^2). \tag{3.2}$$

3.2. **Integral structures.** Let $\mathrm{B}(\mathbb{Z})$ denote the ring of *Hurwitz quaternions*,

$$\mathrm{B}(\mathbb{Z}) = \mathbb{Z}[i, j, k, \frac{1 + i + j + k}{2}];$$

It is well known that the ring of Hurwitz quaternions, endowed with the reduced norm Nr, is *euclidean*: for any $y, q \in \mathrm{B}(\mathbb{Z}) - \{0\}$, there is $x, r \in \mathrm{B}(\mathbb{Z})$ such that $\mathrm{Nr}(r) < \mathrm{Nr}(q)$ and $y = qx + r$ ([Sam70, §5.7]). This implies that

any left (or right) $\mathrm{B}(\mathbb{Z})$-ideal is a principal ideal: any finitely generated left (resp. right) $\mathrm{B}(\mathbb{Z})$-module $I \subset \mathrm{B}(\mathbb{Q})$ is of the form $\mathrm{B}(\mathbb{Z})q$ (resp. $q\mathrm{B}(\mathbb{Z})$) for some $q \in \mathrm{B}(\mathbb{Q})$; moreover any subring of $\mathrm{B}(\mathbb{Q})$ which is finitely generated (as a $\mathbb{Z}$-module) is conjugate to a subring[2] of $\mathrm{B}(\mathbb{Z})$; in particular $\mathrm{B}(\mathbb{Z})$ is a maximal order of $\mathrm{B}(\mathbb{Q})$ and any maximal order in $\mathrm{B}(\mathbb{Q})$ is $\mathrm{B}(\mathbb{Q})^\times$-conjugate to it.

Finally, under (3.1), the lattice $\mathbb{Z}^3 \subset \mathbb{Q}^3$ becomes identified with the trace free integral quaternions,

$$\mathrm{B}^{(0)}(\mathbb{Z}) = \mathrm{B}^{(0)}(\mathbb{Q}) \cap \mathrm{B}(\mathbb{Z}).$$

In particular, we obtain a map $\mathrm{B}(\mathbb{Z})^\times \to \mathrm{SO}_3(\mathbb{Z})$ whose image is the index two subgroup $\mathrm{SO}_3(\mathbb{Z})^+$ and, similarly, $(\mathrm{B}(\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}_p)^\times \to \mathrm{SO}_3(\mathbb{Z}_p)$ (which is surjective unless $p = 2$).

While our primary interest will be in the Hamilton quaternions and sums of three squares, the reader will observe that much of what we state here and below is valid for more general quaternion algebras endowed with a maximal order. As this may be useful for readers interested in representations by other ternary quadratic forms, we will write several of the intermediate steps in more generality.

3.3. **Construction of representations using ideal classes.** As a first example of the usefulness of the quaternion, let us show how to deduce the Gauss-Legendre theorem from the *Hasse-Minkowski local-global principle*. If $d > 0$ is not of the form $4^a(8b-1)$, then, for any prime p, d is representable as a sum of three squares in $\mathbb{Z}_p^3$ and since d is positive, d is also representable over $\mathbb{R}$. By the *Hasse-Minkowski theorem* (cf. [Ser73][Thm. 8, p. 41]), there exists $x = (a, b, c) \in \mathbb{Q}^3$ such that $a^2 + b^2 + c^2 = d$. Let $x = ai + bj + ck$; then $x^2 = -d$ so the ring $\mathbb{Z}[x] = \mathbb{Z} + \mathbb{Z}x$ is finitely generated. By §3.2, there is $q \in \mathrm{B}(\mathbb{Q})$ such that $q\mathbb{Z}[x]q^{-1} \in \mathrm{B}(\mathbb{Z})$ so that $y := qxq^{-1} \in \mathrm{B}^{(0)}(\mathbb{Z})$ is integral and satisfies $y^2 = -d$.

More generally, the above scheme together with the group of $\mathscr{O}_K$-ideal classes makes it possible to generate plenty of *new* integral representations from a given one.

Any element $x \in \mathscr{H}_d$ yields an embedding of $\mathbb{Q}(\sqrt{-d})$ into $\mathrm{B}(\mathbb{Q})$: indeed $x^2 = -d$ and thus $\sqrt{-d} \mapsto x$ defines an embedding

$$\iota_x : K \mapsto \mathbb{Q}[x] \subset \mathrm{B}(\mathbb{Q}).$$

This embedding is *integral* in the following sense: let

$$\mathscr{O}_x = \mathrm{B}(\mathbb{Z}) \cap \mathbb{Q}[x]$$

then $\iota_x^{-1}(\mathscr{O}_x) = \mathscr{O}_K$ is the ring of integers[3] of K.

[2]Indeed, given R such a subring, $R\mathrm{B}(\mathbb{Z})$ is a right $\mathrm{B}(\mathbb{Z})$ ideal, so of the form $R\mathrm{B}(\mathbb{Z}) = q\mathrm{B}(\mathbb{Z})$ and $q^{-1}Rq \subset q^{-1}RR\mathrm{B}(\mathbb{Z}) = q^{-1}R\mathrm{B}(\mathbb{Z}) = \mathrm{B}(\mathbb{Z})$.

[3]It is an order containing $\mathbb{Z}[\sqrt{-d}]$, integrally closed at 2, because the local order $\mathrm{B}(\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}_2$ contains all elements of $\mathrm{B} \otimes_{\mathbb{Q}} \mathbb{Q}_2$ with norm in $\mathbb{Z}_2$.

Now, given such a x and given an $\mathscr{O}_K$-ideal, I, we can also construct a *new* integral representation $y \in \mathscr{H}_d$ from x and I: the finitely generated $\mathbb{Z}$-module $\mathrm{B}(\mathbb{Z})\iota_x(I)$ is a left $\mathrm{B}(\mathbb{Z})$-ideal, so of the form $\mathrm{B}(\mathbb{Z})q$ and then

$$y = qxq^{-1} \subset \mathrm{B}(\mathbb{Z})\iota_x(I)xq^{-1} = \mathrm{B}(\mathbb{Z})\iota_x(xI)q^{-1} \subset \mathrm{B}(\mathbb{Z})\iota_x(I)q^{-1} = \mathrm{B}(\mathbb{Z});$$

moreover if I is replaced by $\lambda.I = I.\lambda$ $\lambda \in K^\times$, q may be replaced by $q' = q\iota_x(\lambda)$ and $q'xq'^{-1} = q'\iota_x(\lambda)x\iota_x(\lambda^{-1})q'^{-1} = y$. Notice that q is defined only up to multiplication on the left by an element of $\mathrm{B}^\times(\mathbb{Z})$ this implies y is well defined up to $\mathrm{B}^\times(\mathbb{Z})$-conjugacy: in view of the isomorphism $\mathrm{B}^\times(\mathbb{Z})/\pm 1 \simeq \mathrm{SO}_3(\mathbb{Z})^+$, we obtain for $x \in \mathscr{H}_d$ fixed, a well defined map

$$[I] \in \mathrm{Pic}(\mathscr{O}_K) \mapsto \widetilde{\mathscr{H}}_d^*.$$

Let us see that this map is a $\mathrm{Pic}(\mathscr{O}_K)$-torsor structure on $\widetilde{\mathscr{H}}_d^*$.

3.4. For each pair of elements $x, y \in \mathscr{H}_d$ we define the abelian group

$$\Lambda_{x\to y} = \{\lambda \in \mathrm{B}(\mathbb{Z}) : x\lambda = \lambda y\}.$$

Then $\Lambda_{x\to y}$ has the structure of a module under $\mathscr{O}_x$ (via $\mu \in \mathscr{O}_x : \lambda \mapsto \mu\lambda$) and also under $\mathscr{O}_y$ (via $\nu \in \mathscr{O}_y : \lambda \mapsto \lambda\nu$). Both structures are torsion-free, and thus locally free.

In fact,in view of the isomorphism $\mathrm{SO}_3(\mathbb{Q}) \cong \mathrm{PB}^\times(\mathbb{Q})$, there exists a $q \in \mathrm{B}^\times(\mathbb{Q})$ such that $q^{-1}xq = y$. (By Witt's theorem, any two elements of $\mathbb{Q}^3$ of the same length can be rotated into one another.) One can then write $\Lambda_{x\to y}$ as $\mathbb{Q}[x]q \cap \mathrm{B}(\mathbb{Z})$. It follows that

$$\Lambda_{x\to y} \otimes_{\mathbb{Z}} \mathbb{Q} = \mathbb{Q}[x]q,$$

and thus that both module structures on $\Lambda_{x\to y}$ are, indeed, locally free of rank 1.

We note that the map $M \to M \otimes_{\mathscr{O}_x} \Lambda_{x\to y}$ gives a map from $\mathscr{O}_x$-modules to $\mathscr{O}_y$-modules; this induces an isomorphism $\mathrm{Pic}(\mathscr{O}_x) \to \mathrm{Pic}(\mathscr{O}_y)$, which is in fact the one induced by the isomorphism $\mathscr{O}_x \cong \mathscr{O}_y$ sending x to y. We denote by $[\Lambda_{x\to y}]$ the class of $\Lambda_{x\to y}$ in $\mathrm{Pic}(\mathscr{O}_x) = \mathrm{Pic}(\mathscr{O}_K)$.

Proposition 3.5 (Torsor structure on $\widetilde{\mathscr{H}}_d^*$)**.** *The set $\widetilde{\mathscr{H}}_d^*$ has the structure of a torsor for $\mathrm{Pic}(\mathscr{O}_K)$ in which the unique element of $\mathrm{Pic}(\mathscr{O}_K)$ mapping x to y is given by $[\Lambda_{x\to y}]$.*

This action of $\mathrm{Pic}(\mathscr{O}_K)$ descends to $\widetilde{\mathscr{H}}_d$ (obvious if $d \equiv 3 (\mathrm{mod}\, 4)$), and for $d \equiv 1, 2 (\mathrm{mod}\, 4)$ the stabilizer of any point of $\widetilde{\mathscr{H}}_d$ is the order 2 subgroup generated by a prime above 2.

More precisely, for any $x, y, z \in \mathscr{H}_d$,

A. *$\Lambda_{x\to y} \otimes_{\mathscr{O}_x} \Lambda_{y\to z}$ is isomorphic to $\Lambda_{x\to z}$ as an $(\mathscr{O}_x, \mathscr{O}_z)$-bimodule;*
B. *The class of $\Lambda_{x\to y}$ is trivial if and only if x and y are identified in $\widetilde{\mathscr{H}}_d^*$;*
C. *For every x and every $g \in \mathrm{Pic}(\mathscr{O}_K)$, there exists a y such $[\Lambda_{x\to y}] = g$.*

D. *If $d \equiv 1, 2$ modulo 4, and $x \neq y \in \widetilde{\mathscr{H}}_d^*$ project to the same element of $\widetilde{\mathscr{H}}_d$, then $[\Lambda_{x\to y}]$ is the ideal class of a prime above 2.*

Note that parts A–D imply the first statements.

In order to prove Proposition 3.5, we shall need a study of the corresponding *local* problem, which we undertake in §3.6.

3.6. **Some local analysis.** Let B be a quaternion algebra over $\mathbb{Q}_p$, let $\mathscr{O}$ be a maximal order of B and $\mathscr{O}^{(0)} \subset B^{(0)}$ be the lattice of trace zero elements in $\mathscr{O}$. As above, conjugation by element of $B^\times$ induces a surjective map

$$B^\times \twoheadrightarrow \mathrm{SO}(B^{(0)}, \mathrm{Nr})$$

under which $\mathscr{O}^\times$ maps to $\mathrm{SO}(\mathscr{O}^{(0)}, \mathrm{Nr})$. Let $\mathscr{O}^{(1)} \subset \mathscr{O}^\times$ be the group of norm-1 units. Let $d \in \mathbb{Z}_p$ be squarefree (that is, $\mathrm{ord}_p(d) \leqslant 1$) and let $\mathscr{O}^{(0,d)} \subset \mathscr{O}^{(0)}$ be the set of elements of $\mathscr{O}$ with trace 0 and norm d; $\mathscr{O}^{(0,d)}$ is obviously stable under the action of $\mathscr{O}^\times$ by conjugation.

We fix an element x of $\mathscr{O}^{(0,d)}$ and, as above, write $\Lambda_{x\to y}$ (or, when no confusion is likely, just Λ) for the set of $\lambda \in \mathscr{O}$ such that

$$x\lambda = \lambda y. \tag{3.3}$$

The solutions to (3.3) in B form a vector space of dimension 2, so Λ is a free $\mathbb{Z}_p$-module of rank 2.

Proposition 3.7. *The action of $\mathscr{O}^{(1)}$ and $\mathscr{O}^\times$ on $\mathscr{O}^{(0,d)}$ can be described as follows.*

- *Suppose B is a division algebra. Then there are two orbits of $\mathscr{O}^{(1)}$ on $\mathscr{O}^{(0,d)}$, which are interchanged by conjugation by any element of $\mathscr{O}^\times$ whose norm is not in $\mathrm{Nr}(\mathbb{Q}_p[x])$. In particular, the special orthogonal group of $\mathscr{O}^{(0)}$ acts transitively on $\mathscr{O}^{(0,d)}$.*
- *Suppose $B = M_2(\mathbb{Q}_p)$. Then:*
  - *If $p \neq 2$ and p does not divide d, the action of $\mathscr{O}^{(1)}$ on $\mathscr{O}^{(0,d)}$ is transitive.*
  - *Otherwise, there are two orbits of $\mathscr{O}^{(1)}$ on $\mathscr{O}^{(0,d)}$; they are interchanged by $\mathscr{O}^\times$, unless $p = 2$ and $d \equiv 3$ mod 4.*

Proof. Suppose B is a division algebra. Then $\mathscr{O}$ is the unique maximal order, and consists of all elements whose norm lies in $\mathbb{Z}_p$. Let λ be a nonzero element of Λ; then $y = \lambda^{-1}x\lambda$. Note that conjugation by λ is the desired isometry of $\mathscr{O}^{(0,d)}$ relating x to y.

If there is an element $\alpha \in \mathbb{Q}_p[x]$ with $\mathrm{Nr}(\alpha) = \mathrm{Nr}(\lambda)$, then $\alpha^{-1}\lambda$ is an element of Λ of norm 1; conversely, any element of Λ of norm 1 is $\alpha^{-1}\lambda$ for some $\alpha \in \mathbb{Q}_p[x]$ whose norm agrees with that of λ. This shows that the orbits of $\mathscr{O}^{(1)}$ on $\mathscr{O}^{(0,d)}$ are naturally identified with $\mathbb{Q}_p^\times/\mathrm{Nr}(\mathbb{Q}_p[x]^\times)$. This quotient is a group of order 2.

Now suppose that $B = M_2(\mathbb{Q}_p)$, so that we can take $\mathscr{O}^\times = \mathrm{GL}_2(\mathbb{Z}_p)$ and $\mathscr{O}^{(1)} = \mathrm{SL}_2(\mathbb{Z}_p)$.

Let $x = \begin{bmatrix} b & a \\ c & -b \end{bmatrix}$ be an element of $\mathscr{O}^{(0,d)}$ (so that $b^2 + ac = -d$) and $y = \begin{bmatrix} 0 & 1 \\ -d & 0 \end{bmatrix}$; then

$$\Lambda_{x\to y} = \mathbb{Z}_p \begin{bmatrix} b & 1 \\ c & 0 \end{bmatrix} + \mathbb{Z}_p \begin{bmatrix} a & 0 \\ -b & 1 \end{bmatrix}$$

and the elements of $\mathrm{Nr}(\Lambda_{x\to y})$ are those elements of $\mathbb{Z}_p$ represented by the quadratic form $Q = aX^2 + 2bXY - cY^2$, which has discriminant $-4d$. Thus, x and y are in the same orbit of $\mathscr{O}^{(1)}$ if and only if Q represents 1 over $\mathbb{Z}_p$.

For all facts used below about isomorphism classes of binary quadratic forms over $\mathbb{Z}_p$, see [Jon50, §31].

First, suppose p is odd. If p does not divide d, then Q is equivalent to $X^2 + dY^2$, and in particular represents 1. So in this case, $\mathscr{O}^{(1)}$ acts transitively on $\mathscr{O}^{(0,d)}$. If p divides d, then Q is equivalent to either $X^2 + dY^2$ or $\varepsilon X^2 + \varepsilon^{-1} dY^2$, where $\varepsilon \in \mathbb{Z}_p^\times$ is a nonsquare. In the former case, y is in the orbit of x; in the latter case,

$$y' = \begin{bmatrix} 0 & \varepsilon \\ -\varepsilon^{-1} d & 0 \end{bmatrix}$$

is in the orbit of x. So there are two orbits, as claimed. In both cases, Q represents an element of $\mathbb{Z}_p^\times$, so $\mathscr{O}^\times$ acts transitively on $\mathscr{O}^{(0,d)}$.

Now take $p = 2$. In this case, there are always exactly two equivalence classes of binary forms of discriminant $4d$, one of which represents 1 over $\mathbb{Z}_2$ and the other of which does not. The non-representing forms are:

- $2X^2 + 2XY + (1/2)(d+1)Y^2$ $\qquad (d = 3 \mod 4)$
- $\varepsilon X^2 + \varepsilon^{-1} dY^2$ $\qquad (d = 1, 2 \mod 4)$

where, in the latter case, $\varepsilon \in \mathbb{Z}_2^\times$ is an element which is not a norm from $\mathbb{Q}_2(\sqrt{-d})^\times$. In case $d = 1$ or 2 mod 4, we again see that either y or $y' = \begin{bmatrix} 0 & \varepsilon \\ -\varepsilon^{-1} d & 0 \end{bmatrix}$ lies in the orbit of x, so there are two orbits of $\mathscr{O}^{(1)}$ on $\mathscr{O}^{(0,d)}$. And again Q represents an element of $\mathbb{Z}_2^\times$, so some element of $\mathscr{O}^\times$ sends x to y.

In case $d = 3$ mod 4, take

$$y' = \begin{bmatrix} 1 & -2 \\ (1/2)(d+1) & -1 \end{bmatrix}$$

A direct computation shows that the orbit of x under $\mathscr{O}^{(1)}$ contains either y or y', so again there are two orbits. In this case, the two orbits are not interchanged by $\mathscr{O}^\times$. □

3.8. **Proof of Proposition 3.5, A.** We define a map

$$m : \Lambda_{x\to y} \otimes_{\mathscr{O}_y} \Lambda_{y\to z} \to \Lambda_{x\to z}$$

by sending $\lambda \otimes \mu$ to $\lambda\mu$. Evidently it preserves the natural structures of $(\mathscr{O}_x, \mathscr{O}_z)$-bimodule on the two sides. We will show that the image of m

– call it Λ' – coincides with $\eta\Lambda_{x\to z}$, for some $\eta \in \mathscr{O}_x$. This will imply our conclusion, because "multiplication by η^{-1}" is also an isomorphism of $(\mathscr{O}_x, \mathscr{O}_z)$-bimodules.

Clearly $\Lambda' \subset \Lambda_{x\to z}$. The $\mathbb{Z}_p$-module $\Lambda_{x\to y} \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is explicitly described by Proposition 3.7. If p is odd, it follows from from the second case of Proposition 3.7 that $\Lambda_{x\to y}, \Lambda_{x\to z}, \Lambda_{y\to z}$ all include elements of $\mathrm{B}(\mathbb{Z}_p) = \mathrm{B}(\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Z}_p$ whose norm lies in $\mathbb{Z}_p^\times$. It follows that Λ' also contains such an element. If $p = 2$, the same conclusion holds unless $d \equiv 3 \bmod 4$; if $d \equiv 3 \bmod 4$, we may conclude only that $\Lambda_{x\to y}$ contains an element whose norm has valuation $\leqslant 1$.

Both Λ' and $\Lambda_{x\to z}$ are free rank 1 modules under $\mathscr{O}_x$; thus, for every p, there exists $\eta_p \in \mathscr{O}_x \otimes \mathbb{Z}_p$ such that $\Lambda' \otimes \mathbb{Z}_p = \eta_p(\Lambda_{x\to z} \otimes \mathbb{Z}_p)$. In order that both localizations contain an element whose norm belongs to $\mathbb{Z}_p^\times$, it is necessary that η_p belong to $(\mathscr{O}_x \otimes \mathbb{Z}_p)^\times$; this implies that $\Lambda' \otimes \mathbb{Z}_p = \Lambda_{x\to z} \otimes \mathbb{Z}_p$.

Thus, unless $d \equiv 3$ modulo 8, we indeed have an equality $\Lambda' = \Lambda_{x\to z}$. In the case where $d \equiv 3$ modulo 8, we may still draw the conclusion that there exists an ideal I in $\mathscr{O}_K$, divisible only by the prime above 2, so that $\Lambda' = I.\Lambda_{x\to z}$. This implies the stated conclusion, since the prime of $\mathbb{Q}(\sqrt{-d})$ above 2 is principal when $d \equiv 3$ modulo 8.

3.9. **Proof of Proposition 3.5, B.** Let us recall first of all the following: for $x, y \in \mathscr{H}_d$,

$$x \in \mathrm{SO}_3(\mathbb{Z})^+ y \iff uxu^{-1} = y, \text{ some } u \in \mathrm{B}(\mathbb{Z}), \mathrm{Nr}(u) = 1.$$
$$x \in \mathrm{SO}_3(\mathbb{Z}) y \iff uxu^{-1} = y, \text{ some } u \in \mathrm{B}(\mathbb{Z}), \mathrm{Nr}(u) \in \{1, 2\}.$$

These facts may be verified by direct computation.

Notice that if $\Lambda_{x\to y}$ is trivial there exists a global generator λ for $\Lambda_{x\to y}$ as an $\mathscr{O}_x$-module. By the conclusions of the prior paragraph §3.8, the norm of λ is a p-adic unit for all $p > 2$.

We separate two cases, as in the prior proof.

Case 1: $d = 1, 2(4)$. Suppose that $\Lambda_{x\to y}$ is trivial. In this case, the norm of λ is also a unit at $p = 2$. Thus, there exists a unit $u \in \mathrm{B}(\mathbb{Z})^\times$ such that $uxu^{-1} = y$.

Conversely, if there exists a unit u with $uxu^{-1} = y$, then u lies in $\Lambda_{x\to y}$. The span $u.\mathscr{O}_x$ is a $\mathscr{O}_x$-submodule of $\Lambda_{x\to y}$, and coincides with it by the same reasoning as used in §3.8. Therefore, $\Lambda_{x\to y}$ is principal as an $\mathscr{O}_x$-module.

Case 2: $d = 3(8)$. if $\Lambda_{x\to y}$ is trivial. In this case, the 2-valuation of the norm of λ is at most 1. Therefore, $y = \lambda x \lambda^{-1}$ with λ an element of $\mathrm{B}(\mathbb{Z})$ of norm 1 or 2.

Conversely, suppose that $x, y \in \mathscr{H}_d$ are related by an element of $\mathrm{SO}_3(\mathbb{Z})$. Equivalently, there exists $\lambda \in \Lambda_{x\to y}$ with $\mathrm{Nr}(\lambda) \in \{1, 2\}$. If $\mathrm{Nr}(\lambda) = 1$, then $[\Lambda_{x\to y}]$ is principal, exactly as above. Suppose $\mathrm{Nr}(\lambda) = 2$. Reasoning as before, $\lambda.\mathscr{O}_x$ is an $\mathscr{O}_x$-submodule of $\Lambda_{x\to y}$ of index $\leqslant 2$; since the ideal above 2 in $\mathscr{O}_x$ is principal, we deduce again that $[\Lambda_{x\to y}]$ is trivial.

3.10. **Proof of Proposition 3.5, C.** Write $\widehat{\mathscr{O}}_K$ for $\mathscr{O}_K \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}$. Given $\alpha \in \widehat{\mathscr{O}}_K$ everywhere locally invertible, one defines the $\mathscr{O}_K$-ideal $(\alpha) = \alpha\widehat{\mathscr{O}}_K \cap \mathscr{O}_K$. Choose α such that (α) is in the class of $g \in \mathrm{Pic}(\mathscr{O}_K)$.

Choosing an $x \in \mathscr{H}_d$, and thus an embedding ι_x of $K \simeq \mathbb{Q}[x]$ in B; using ι_x to identify α with an element of $\mathbb{Q}[x] \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}$ we consider the left $\mathrm{B}(\mathbb{Z})$-ideal $\mathrm{B}(\mathbb{Z})(\alpha)$ and argue as in §3.3; this ideal is principal (as $\mathrm{B}(\mathbb{Z})$ is euclidean, §3.2) so there exists $q \in \mathrm{B}^\times(\mathbb{Q})$ so that $\mathrm{B}(\mathbb{Z})q^{-1} = \mathrm{B}(\mathbb{Z})(\alpha)$. In particular, we have $q = \alpha^{-1}u$ with $u \in \widehat{\mathrm{B}(\mathbb{Z})}^\times$.

Now let $y = q^{-1}xq = u^{-1}\alpha x \alpha^{-1}u = u^{-1}xu$; then y lies in $\widehat{\mathrm{B}(\mathbb{Z})}$ and in $\mathrm{B}(\mathbb{Q})$, so $y \in \mathscr{H}_d$.

Now

$$\Lambda_{x\to y} = \mathbb{Q}[x]q \cap \mathrm{B}(\mathbb{Z}) \cong \mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})q^{-1} = \mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})(\alpha) = (\alpha). \tag{3.4}$$

So $[\Lambda_{x\to y}] = g$, which proves assertion (3).

3.11. **Proof of Proposition 3.5, D.** Notation as in the statement of Proposition 3.5D. The proof of 3.5, B shows that $[\Lambda_{x\to y}]$ is a prime ideal above 2. It is not principal since $x \neq y$. Since $d \equiv 1, 2$ modulo 4, the prime 2 is ramified in $\mathbb{Q}(\sqrt{-d})$, and therefore this uniquely specifies $[\Lambda_{x\to y}]$.

3.12. **The** $\mathrm{Pic}(\mathscr{O}_K)$**-action explicated.** The arguments of §3.10 enables us to describe the action of $\mathrm{Pic}(\mathscr{O}_K)$ on $\widetilde{\mathscr{H}}_d^*$ or $\widetilde{\mathscr{H}}_d$ quite explicitly. Let $\mathfrak{p}$ be an ideal of norm 5 in $\mathscr{O}_K$; we now verify that the algorithm prescribed in §2.4 is indeed a lifting of the action of $[\mathfrak{p}]$ from $\widetilde{\mathscr{H}}_d^*$ to $\mathscr{H}_d$.

Let $x \in \mathscr{H}_d$, thus giving an embedding of ι_x of K into B. Let α be an element of $\widehat{\mathscr{O}}_K$ which projects to a unit in $\mathscr{O}_{K_v}$ when v is prime to $\mathfrak{p}$, and to a uniformizer in $\mathscr{O}_{K_\mathfrak{p}}$. Choose $q \in \mathrm{B}^\times(\mathbb{Q})$ such that $\mathrm{B}(\mathbb{Z})q^{-1} = \mathrm{B}(\mathbb{Z})(\alpha)$; then q has norm 5^{-1}.

Set[4] $q_2 = q(1+i)^{-1}$, which has norm $1/10$. Taking $y = q_2^{-1}xq_2$, we have, as in (3.4),

$$\Lambda_{x\to y} \cong \mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})(1+i)(\alpha)$$

The right-hand side is an $\mathscr{O}_x$-submodule of $\mathfrak{p}$ of index 2. If d is 1 or 2 mod 4, this means that $\Lambda_{x\to y}$ is in the class of $\mathfrak{q}\mathfrak{p}$, where $\mathfrak{q}$ is the ideal of $\mathscr{O}_x$ lying over 2. If d is 3 mod 4, we have

$$\mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})(1+i)(\alpha) = \mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})(2\alpha) \cong \mathbb{Q}[x] \cap \mathrm{B}(\mathbb{Z})(\alpha)$$

so the class of $\Lambda_{x\to y}$ is just that of $\mathfrak{p}$. We note, however, that the action of $\mathfrak{p}$ and $\mathfrak{p}\mathfrak{q}$ on $\widetilde{\mathscr{H}}_d$ is the same, since the action of $[\mathfrak{q}]$ permutes the fibers of $\widetilde{\mathscr{H}}_d^* \to \widetilde{\mathscr{H}}_d$.

[4]The motivation for introducing this may become clearer later; the lifting of the action from $\widetilde{\mathscr{H}}_d^*$ to $\mathscr{H}_d$ involves imposing additional congruences modulo 3, and q_2 will have better mod 3 properties.

The integral quaternions of norm 10 can all be expressed in the form ru, where $u \in \mathrm{B}(\mathbb{Z})^\times$ and r is an element of

$$\mathcal{A}_5 = \{1 \pm 3i, 1 \pm 3j, 1 \pm 3k\}. \tag{3.5}$$

It follows that, for each x in $\mathscr{H}_d$, the class $[\mathfrak{p}]\tilde{x}$ in $\widetilde{\mathscr{H}_d}$ must be represented by $r^{-1}xr$ for some $r \in \mathcal{A}_5$. Now a direct computation shows that the action of conjugation by the six elements of $\mathcal{A}_5$ yields precisely the action of the six matrices appearing in §2.6. For example, conjugation by $1 - 3i$ acts on $\mathrm{B}^{(0)}$ via the rule

$$\begin{aligned} i &\mapsto i, \\ j &\mapsto \frac{1}{10}(1+3i)j(1-3i) = -\frac{4}{5}j + \frac{3}{5}k, \\ k &\mapsto \frac{1}{10}(1+3i)k(1-3i) = -\frac{3}{5}j - \frac{4}{5}k \end{aligned}$$

which corresponds to the matrix A in §2.4.

4. Representations of binary quadratic forms by $x^2 + y^2 + z^2$.

We now discuss the proof of Proposition 2.11, "Linnik's basic lemma." We shall, in fact, discuss two proofs; the first (§4.1) is simply quoting a result of G. Pall, which in turn rests on Siegel's mass formula; the second, presented in §4.3 after a preliminary discussion, is based on ideas related to §3.

4.1. **A result of Gordon Pall.** Let $(Q, \mathbb{Z}^m)$, $(R, \mathbb{Z}^n)$ be two non-degenerate integral quadratic forms with $m \geqslant n$. One says that Q *represents* R if there exists a $\mathbb{Z}$-linear map $\iota : \mathbb{Z}^n \to \mathbb{Z}^m$ such that, for any $\mathbf{x} \in \mathbb{Z}^n$, $Q(\iota(\mathbf{x})) = R(\mathbf{x})$. It follows that ι is an embedding.

In particular, given $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{H}_d$ with $\mathbf{x}_1.\mathbf{x}_2 = e$, the linear map

$$\iota : \mathbb{Z}^2 \to \mathbb{Z}\mathbf{x}_1 + \mathbb{Z}\mathbf{x}_2$$

defines a representation of the binary quadratic form $R(x, y) = dx^2 + 2exy + dy^2$ by the ternary form $Q(x, y, z) = x^2 + y^2 + z^2$.

Therefore, counting the number of pairs $(\mathbf{x}_1, \mathbf{x}_2) \in \mathscr{H}_d^2$ with $\mathbf{x}_1.\mathbf{x}_2 = e$ is essentially equivalent (up to the action of the finite group $\mathrm{SO}_3(\mathbb{Z})$) to counting the number of representations of R by Q.

The precise computation of the number of embeddings, $r(a, b, c)$ say, of a binary quadratic form $ax^2 + bxy + cy^2$ into the ternary quadratic $x^2 + y^2 + z^2$ in the *most general case* (no primitivity assumptions, etc.) was first carried out by Pall [Pal49, Theorem 4, page 359]. His theorem gives a formula

$$r(a, b, c) = 24 \cdot 2^\nu \cdot \prod_{p | 2(b^2 - 4ac)} r_p(a, b, c) \tag{4.1}$$

where it follows from Pall's result that:

(1) ν is the number of distinct odd primes dividing the discriminant b^2-4ac;
(2) $r_p(a,b,c)$ is bounded by an absolute constant unless $p^2|(a,b,c)$.

In particular, it follows that

$$|r(a,b,c)| \ll \max(a,b,c)^{\varepsilon}, \tag{4.2}$$

when (a,b,c) has no square factor.

One way to obtain a quantitative result like (4.2) is by use of Siegel's mass formula, combined with a computation of local densities. We will *sketch* in the rest of this section an alternate approach. We must first revisit the material of §3 and describe a "different" (although closely related, as we shall see) connection between $\mathscr{H}_d$ and the class group.

4.2. **The orthogonal complement construction.** We have seen that $\widetilde{\mathscr{H}}_d^*$ can be placed in bijection with the group $\mathrm{Pic}(\mathscr{O}_K)$; but there is no natural group structure on $\widetilde{\mathscr{H}}_d^*$, for there is no natural choice of an identity element of $\widetilde{\mathscr{H}}_d^*$. In other words, the torsor structure is natural but admits no natural trivialization. However, the orthogonal complement construction, which we shall now explain, gives a trivialization of the "square" $\widetilde{\mathscr{H}}_d^* \times_{\mathrm{Pic}(\mathscr{O}_K)} \widetilde{\mathscr{H}}_d^*$. [5]

Let $\mathcal{Q}_{-d}$ be the set of binary quadratic forms $aX^2+bXY+cY^2$ of discriminant

$$b^2-4ac = \mathrm{disc}(\mathscr{O}_K) = \begin{cases} -d, & d \equiv 3 \bmod 4 \\ -4d, & d \equiv 1,2 \bmod 4 \end{cases}$$

considered up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence. Then $\mathcal{Q}_{-d}$ parametrizes rank-2 quadratic lattices Λ of discriminant $\mathrm{disc}(\mathscr{O}_K) \in \{-d,-4d\}$, endowed with an orientation $\wedge^2\Lambda \cong \mathbb{Z}$.

There is a natural map $\widetilde{\mathscr{H}}_d^* \to \mathcal{Q}_{-d}$ sending $x \in \mathscr{H}_d$ to the induced quadratic form on $x^\perp$ (the rank 2 lattice in $\mathbb{Z}^3$ of vectors orthogonal to x); here we understand that, if $d \equiv 3$ modulo 8, then we scale the resulting quadratic form by $\frac{1}{2}$, and that a basis (λ,λ') of $x^\perp$ is oriented if $\lambda\wedge\lambda'\wedge x > 0$.

Composing with the classical identification of $\mathcal{Q}_{-d}$ with $\mathrm{Pic}(\mathscr{O}_K)$ now gives a map

$$\underline{\mathrm{Perp}} : \widetilde{\mathscr{H}}_d^* \to \mathrm{Pic}(\mathscr{O}_K) \tag{4.3}$$

which can be described concretely as $\underline{\mathrm{Perp}}(x) = [\Lambda_{x\to\bar{x}}]$. Both sides admit an action of $\mathrm{Pic}(\mathscr{O}_K)$; the left-hand side by means of the torsor structure described in section § 3.1, and the right-hand side by multiplication. But $\underline{\mathrm{Perp}}$ is not equivariant for this action; rather, it intertwines the action of $\overline{\mathrm{Pic}(\mathscr{O}_K)}$ on the left with the *square* of this action on the right, as we will

[5] The situation is similar to one familiar from geometry: If X is a smooth cubic plane curve over a field k with Jacobian E, then the points $X(k)$ form a torsor for the group $E(k)$. This torsor has no canonical trivialization, but the embedding of X into the plane gives a canonical trivialization of the "cube" $X \times_E X \times_E X$.

now see. Suppose given x, y in $\mathscr{H}_d$ such that $[\Lambda_{x\to y}] = \alpha \in \mathrm{Pic}(\mathscr{O}_K)$. One checks that $[\Lambda_{y\to x}] = [\Lambda_{\bar{x}\to\bar{y}}]$. Thus

$$\underline{\mathrm{Perp}}(y) = [\Lambda_{y\to\bar{y}}] = [\Lambda_{y\to x}][\Lambda_{x\to\bar{x}}][\Lambda_{\bar{x}\to\bar{y}}] = \alpha^2\underline{\mathrm{Perp}}(\bar{x}).$$

It follows that the image of $\underline{\mathrm{Perp}}$ is precisely one coset of $2\,\mathrm{Pic}(\mathscr{O}_K)$. [6] Moreover, the cardinality of a fiber of $\underline{\mathrm{Perp}}$ is just the order of the 2-torsion subgroup of $\mathrm{Pic}(\mathscr{O}_K)$.

What $\underline{\mathrm{Perp}}$ supplies is not a trivialization of the torsor $\widetilde{\mathscr{H}}_d^*$, but a trivialization of its *square* in the group of $\mathrm{Pic}(\mathscr{O}_K)$-torsors. The square is a torsor T which can be described explicitly as the set of equivalence classes of pairs (x, y) under the relation $(x, y) = (\alpha x, \alpha^{-1}y)$ for all $\alpha \in \mathrm{Pic}(\mathscr{O}_K)$. The action of $\alpha \in \mathrm{Pic}(\mathscr{O}_K)$ sends (x, y) to $(\alpha x, y)$ (or to $(x, \alpha y)$, which is the same.) Then the map sending (x, y) to $[\Lambda_{x\to\bar{y}}]$ is a canonical isomorphism between T and $\mathrm{Pic}(\mathscr{O}_K)$.

4.3. **Bounds for representations, revisited.** We now sketch how the "orthogonal complement construction" leads us to another proof of Proposition 2.11.

For simplicity, we consider only the case where $b^2 - 4ac$ is of the form $-4d$, with d squarefree. Given an embedding

$$\iota : (\mathbb{Z}^2, ax^2 + bxy + cy^2) \hookrightarrow (\mathbb{Z}^3, x^2 + y^2 + z^2).$$

consider the orthocomplement inside $\mathbb{Z}^3$ of $\iota(\mathbb{Z}^2)$. It is generated by a single vector $\mathbf{x} \in \mathbb{Z}^3$, unique up to sign. The quadratic lattice $\mathbf{x}^\perp$ is plainly the same as its sublattice $\iota(\mathbb{Z}^2)$ after tensoring with $\mathbb{Q}$; since d is squarefree, the two lattices are in fact identical. This implies that $\mathbf{x}.\mathbf{x} = d$.

The embedding ι is determined up to at most 6 possibilities (6 being the maximal number of automorphisms of a positive definite form in rank two) by $\mathbf{x}$. It suffices, therefore, to count the number of $\mathbf{x} \in \mathscr{H}_d$ so that the quadratic form induced on $\mathbf{x}^\perp$ is isomorphic to $(\mathbb{Z}^2, ax^2 + bxy + cy^2)$. By §4.2, this is at most $24|\,\mathrm{Pic}(\mathscr{O}_K)[2]|$, where [2] denotes 2-torsion. By genus theory we have $|\,\mathrm{Pic}(\mathscr{O}_K)[2]| \ll d^\varepsilon$, and the bound (4.2) follows.

Part 3. Adelization

In this part, we interpret, in adelic terms, the various classical arithmetic sets and structures discussed so far –i.e. $\mathscr{H}_d$, $\widetilde{\mathscr{H}}_d$, $\mathrm{Pic}(\mathscr{O}_K)$, $\mathscr{H}_d(q)$, $\widetilde{\mathscr{H}}_d(q)$ – and the various maps between them. This interpretation gives us an alternative approach to the results described in Part 2 and more importantly, will be key to the proof of Proposition 2.13. We refer to [Kna97] for a recent gentle introduction to the adèlic theory of algebraic groups (in relation with automorphic forms), and to [BM66, PR94] for more extensive treatments.

[6] It is irresistible to ask *which* coset. A quadratic form $Ax^2 + Bxy + Cy^2$ of discriminant d embeds, over $\mathbb{Q}$, into $\mathbb{Z}^3$ with the standard quadratic form only if $(A, d/A)_p = (d, d)_p$ for every p dividing $2d$; here $(a, b)_p$ is the *Hilbert symbol*. This condition in fact defines a coset of squares, and so describes exactly the image of $\underline{\mathrm{Perp}}$.

We denote by $\mathbb{A}_f$ the ring of finite adèles of $\mathbb{Q}$, i.e. the restricted product of $(\mathbb{Q}_p)_{p \text{ prime}}$ with respect to the sequence of maximal compact subgroups $(\mathbb{Z}_p)_{p \text{ prime}}$, and by $\mathbb{A} = \mathbb{R} \times \mathbb{A}_f$ the ring of adèles. We denote by $\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p \subset \mathbb{A}_f$ the maximal compact subgroup of $\mathbb{A}_f$; alternatively $\widehat{\mathbb{Z}}$ is the closure of $\mathbb{Z}$ in $\mathbb{A}_f$. Given G a $\mathbb{Q}$-algebraic group with a chosen model over $\mathbb{Z}$, we denote by $\mathrm{G}(\mathbb{A}_f)$, $\mathrm{G}(\mathbb{A})$, $\mathrm{G}(\widehat{\mathbb{Z}})$ the groups of points of G in the corresponding rings.

We begin by summarizing the contents of this part (§5 – §7) via a commutative diagram:

We set (see §5.2 below for the definition of $\mathrm{genus}_{\mathrm{SO}_3}(\mathbb{Z}^3)$)

$$\begin{aligned} \mathcal{P} &= \{(L, \mathbf{x}),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(\mathbb{Z}^3),\ \mathbf{x} \in L,\ \ \mathbf{x}.\mathbf{x} = d\}. \\ \mathcal{P}_{(q)} &:= \{(L, \overline{\mathbf{x}}),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(\mathbb{Z}^3),\ \overline{\mathbf{x}} \in L/qL,\ \overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d (\mathrm{mod}\, q)\}. \end{aligned}$$

For the sequel, we need to fix *base points* $\mathbf{x}_0 \in \mathscr{H}_d$ and $\overline{\mathbf{x}}_q \in \mathscr{H}_d(q)$; this being done, we define

$$K_f[q] := \ker(\mathrm{SO}_3(\widehat{\mathbb{Z}}) \to \mathrm{SO}_3(\mathbb{Z}/q\mathbb{Z})) \tag{4.4}$$

$$K'_f[q] := \{g \in \mathrm{SO}_3(\widehat{\mathbb{Z}}) : g.\overline{\mathbf{x}}_q = \overline{\mathbf{x}}_q\} \tag{4.5}$$

and denote by $\mathrm{SO}_{\mathbf{x}_0}$ the stabilizer of $\mathbf{x}_0$ in SO_3. We will show in the following sections how to describe $\widetilde{\mathscr{H}_d}$ and $\widetilde{\mathscr{H}_d}(q)$ adelically, by means of a commutative diagram

$$\begin{array}{ccccc} \widetilde{\mathscr{H}_d} & \xrightarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P} & \xleftarrow{\sim} & \mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}) \\ \downarrow{\scriptstyle \mathrm{red}_q} & & \downarrow{\scriptstyle \mathrm{red}_q} & & \downarrow \\ \widetilde{\mathscr{H}_d}(q) & \xrightarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)} & \xleftarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K'_f[q]. \end{array} \tag{4.6}$$

We will also explain how to describe $\mathscr{H}_d$ and $\mathscr{H}_d(q)$ adelically, which is slightly more involved technically (though no different conceptually.)

For this we define

$$\begin{aligned} \mathcal{Q} &= \{(L, \mathbf{x}, \theta),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(\mathbb{Z}^3),\ \mathbf{x} \in L,\ \mathbf{x}.\mathbf{x} = d, \theta : L/3L \simeq (\mathbb{Z}/3\mathbb{Z})^3\}. \\ \mathcal{P}_{(3,q)} &:= \{(L, \overline{\mathbf{x}}, \theta),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(\mathbb{Z}^3),\ \overline{\mathbf{x}} \in L/qL,\ \overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d (\mathrm{mod}\, q), \\ &\qquad\qquad \theta : L/3L \simeq (\mathbb{Z}/3\mathbb{Z})^3\}, \end{aligned}$$

and, for any integer a,

$$K_f[a, q] := K_f[a] \cap K'_f[q] \subset K'_f[q]. \tag{4.7}$$

Then we will construct a diagram

(4.8)

$$\begin{array}{ccccc}
\mathscr{H}_d & \xrightarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{Q} & \xleftarrow{\sim} & \mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)\times \mathrm{SO}_3(\widehat{\mathbb{Z}}/3\widehat{\mathbb{Z}})/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}) \\
\downarrow{\scriptstyle \mathrm{red}_q} & & \downarrow{\scriptstyle \mathrm{red}_q} & & \downarrow \\
\mathscr{H}_d(q) & \xrightarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(3,q)} & \xleftarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K_f[3,q] \\
\downarrow\downarrow & & \downarrow\downarrow & & \downarrow\downarrow \\
\widetilde{\mathscr{H}_d}(q) & \xrightarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)} & \xleftarrow{\sim} & \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K'_f[q];
\end{array}$$

where the vertical arrows at the bottom are the evident surjective maps. The horizontal arrows of these diagrams are described in §5 and §6, while the vertical arrows are discussed in §7.

4.4. **Convention: orthogonal groups vs. unit group of quaternions.** We have already noted in §3.1 that the quadratic spaces $(\mathbb{Q}^3, a^2+b^2+c^2)$, $(\mathrm{B}^{(0)}, \mathrm{Nr})$ are isometric and that the action –by conjugation– of the units of Hamilton quaternions, $\mathrm{B}^\times$, on $\mathrm{B}^{(0)}$ induces and isomorphism of $\mathbb{Q}$-algebraic groups

$$\mathrm{PB}^\times \simeq \mathrm{SO}_3$$

with $\mathrm{PB}^\times = Z(\mathrm{B}^\times)\backslash\mathrm{B}^\times$ the projective group of units. In particular the group of adelic points $\mathrm{PB}^\times(\mathbb{A})$ and $\mathrm{SO}_3(\mathbb{A})$ get naturally identified as are their various respective subgroups. In the sequel, we shall freely use this identification, moreover we will use the same notations for various subgroups $K_f[q]$, $K'_f[q]$ etc. to denote either some subgroup in $\mathrm{SO}_3(\mathbb{A})$ or its image in $\mathrm{PB}^\times$ under the above isomorphism.

5. Adelic interpretation of $\mathscr{H}_d(q)$.

In this section we identify $\mathscr{H}_d(q)$ (as well as the sphere S^2), with an adelic quotient of SO_3, verifying the second lines of (4.6) and (4.8).

5.1. **Adelic actions on rational lattices.** The adelic group $\mathrm{GL}_3(\mathbb{A}_f)$ acts on the space of lattices in $\mathbb{Q}^3$ as follows. If $L \subset \mathbb{Q}^3$ is a lattice, and $g = \prod_p g_p$ is an element of $\mathrm{GL}_3(\mathbb{A}_f)$, we define

$$g.L = \{\alpha \in \mathbb{Q}^3 : \iota_p(\alpha) \in g_p.L_p \text{ for all } p\},$$

where $\iota_p : \mathbb{Q}^3 \hookrightarrow \mathbb{Q}_p^3$ is the inclusion. This action is transitive, and the stabilizer of $L_0 = \mathbb{Z}^3$ under this action is $\mathrm{GL}_3(\widehat{\mathbb{Z}})$.

5.2. The SO_3-*genus* of the lattice L_0 is, by definition, the orbit of L_0 under the action of $\mathrm{SO}_3(\mathbb{A}_f) \subset \mathrm{GL}_3(\mathbb{A}_f)$: in other words, this is the set of all rational lattices which are everywhere locally isometric to L_0:

$$\mathrm{genus}_{\mathrm{SO}_3}(L_0) = \mathrm{SO}_3(\mathbb{A}_f).L_0 \simeq \mathrm{SO}_3(\mathbb{A}_f)/\mathrm{SO}_3(\widehat{\mathbb{Z}}).$$

The set of *genus classes* of L_0 is the set of $\mathrm{SO}_3(\mathbb{Q})$-orbits in $\mathrm{genus}_{\mathrm{SO}_3}(L_0)$.

Proposition. *There is only one* SO_3*-genus class – that is, the quadratic form* $x^2+y^2+z^2$ *has genus one. Equivalently,*

$$\mathrm{SO}_3(\mathbb{A}_f) = \mathrm{SO}_3(\mathbb{Q})\mathrm{SO}_3(\widehat{\mathbb{Z}}). \tag{5.1}$$

Proof. Let $L \in \mathrm{genus}_{\mathrm{SO}_3}(\mathbb{A}_f).L_0$, in particular the covolume of L equals the covolume of L_0 which is one. We identify L_0 with the traceless integral quaternions $\mathrm{B}^{(0)}(\mathbb{Z})$ and SO_3 with $\mathrm{PB}^\times$; in these terms, we need to show that there is $q \in \mathrm{B}^\times(\mathbb{Q})$ such that $qLq^{-1} = \mathrm{B}^{(0)}(\mathbb{Z})$. By definition, there is $q_f \in \mathrm{B}^\times(\mathbb{A}_f)$ such that $L = q_f\mathrm{B}^{(0)}(\mathbb{Z})q_f^{-1}$ and so $\mathscr{O} := \mathbb{Z}+L = q_f(\mathbb{Z}+\mathrm{B}^{(0)}(\mathbb{Z}))q_f^{-1}$ is a lattice in $\mathrm{B}(\mathbb{Q})$ containing the identity and stable by multiplication, i.e. an order; hence, by §3.2, there is $q \in \mathrm{B}^\times(\mathbb{Q})$ such that $q\mathscr{O}q^{-1} \subset \mathrm{B}(\mathbb{Z})$ and so $qLq^{-1} \subset \mathrm{B}^{(0)}(\mathbb{Z})$; since qLq^{-1} and $\mathrm{B}^{(0)}(\mathbb{Z})$ have the same covolume they are equal. □

If we replace $x^2+y^2+z^2$ by a different ternary quadratic form Q, this will in general not be so.

5.3. **Level structure.** By construction, the quadratic form $x^2+y^2+z^2$ takes integral values on any lattice in $\mathrm{genus}_{\mathrm{SO}_3}(L_0)$. In particular, given $q \geqslant 1$ an integer, the quotient lattice $L/qL \simeq (\mathbb{Z}/q\mathbb{Z})^3$ is naturally a quadratic space for the form $x^2+y^2+z^2$. *A* q*-level structure* on such a lattice is an additional datum related to L/qL. Here, we will consider two type of level q-structures:

- the principal q-structure: this is the datum of an isomorphism of quadratic spaces $\theta : L/qL \simeq (\mathbb{Z}/q\mathbb{Z})^3$. We will use this only for $q=3$ and mainly for cosmetic purposes.
- a weak q-structure: this is the datum of a point $\overline{\mathbf{x}} \in L/qL$ such that $\overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d(\mathrm{mod}\, q)$ when $(d,q)=1$. This will be the main structure considered in the present paper.

Related to these level structures, are the open compact subgroups of $\mathrm{SO}_3(\widehat{\mathbb{Z}})$,

$$K_f[q],\ K'_f[q],\ K_f[3,q]$$

defined by (4.4), (4.5), (4.7) relative to the choice of some base point $\overline{\mathbf{x}}_q \in \mathscr{H}_d(q)$.

5.4. $\widetilde{\mathscr{H}_d(q)}$ **as an adelic quotient.** Let us consider first the set of pairs

$$\mathcal{P}_{(q)} := \{(L,\overline{\mathbf{x}}),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(L_0),\ \overline{\mathbf{x}} \in L/qL,\ \overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d(\mathrm{mod}\, q)\}.$$

The group $\mathrm{SO}_3(\mathbb{Q})$ acts diagonally on $\mathcal{P}_{(q)}$. From (5.1) any $\mathrm{SO}_3(\mathbb{Q})$-orbit in $\mathcal{P}_{(q)}$ contains a pair of the form $(L_0,\overline{\mathbf{x}})$ with $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$. Moreover if two pairs $(L_0,\overline{\mathbf{x}})$ and $(L_0,\overline{\mathbf{x}}')$ give rise to the same $\mathrm{SO}_3(\mathbb{Q})$-orbit then $\overline{\mathbf{x}}$ and $\overline{\mathbf{x}}'$ differ by an element of $\mathrm{SO}_3(\mathbb{Q}) \cap \mathrm{SO}_3(\widehat{\mathbb{Z}}) = \mathrm{SO}_3(\mathbb{Z})$. It follows that the map $\overline{\mathbf{x}} \in \mathscr{H}_d(q) \mapsto (L_0,\overline{\mathbf{x}}) \in \mathcal{P}_{(q)}$ induces a bijection:

$$\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q) \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)}.$$

In the sequel, we will denote by $[L,\overline{\mathbf{x}}]$ the $\mathrm{SO}_3(\mathbb{Q})$-orbit of the pair $(L,\overline{\mathbf{x}})$.

In fact, since $L/qL \simeq (L \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}})/q(L \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}})$, the whole group $\mathrm{SO}_3(\mathbb{A}_f)$ acts diagonally on $\mathcal{P}_{(q)}$.

Lemma. *This action is transitive: fixing $\overline{\mathbf{x}}_q \in \mathscr{H}_d(q)$, we have*

$$\mathcal{P}_{(q)} = \mathrm{SO}_3(\mathbb{A}_f).(L_0, \overline{\mathbf{x}}_q).$$

Proof. Indeed any $(L, \overline{\mathbf{x}}) \in \mathcal{P}_{(q)}$ is in the orbit of a pair of the form $(L_0, \overline{\mathbf{x}}')$ for some $\overline{\mathbf{x}}' \in \mathscr{H}_d(q)$. It follows from the Lemma below that $\mathrm{SO}_3(\widehat{\mathbb{Z}})$ (through its projection to $\mathrm{SO}_3(\mathbb{Z}/q\mathbb{Z})$) acts transitively on $\mathscr{H}_d(q)$. Taking $k_{\mathbf{x}'} \in \mathrm{SO}_3(\widehat{\mathbb{Z}})$ such that $k_{\overline{\mathbf{x}}'}\overline{\mathbf{x}}_q = \overline{\mathbf{x}}'$, we have $k_{\overline{\mathbf{x}}'}^{-1}(L_0, \overline{\mathbf{x}}') = (L_0, \overline{\mathbf{x}}_q)$. □

Lemma 5.4.1. *For any prime p, the group $\mathrm{SO}_3(\mathbb{Z}_p)$ acts transitively on $\mathscr{H}_d(\mathbb{Z}_p)$ (defined in the vident way). Consequently, $\mathrm{SO}_3(\widehat{\mathbb{Z}})$ acts transitively on $\mathscr{H}_d(\widehat{\mathbb{Z}})$.*

Proof. This – which can be thought of as an analogue of Witt's theorem for rank 3 lattices over $\mathbb{Z}_p$ – is immediate from Proposition 3.7, using the first case if $p = 2$ and the second case if p is odd. □

We thus have $\mathcal{P}_{(q)} \simeq \mathrm{SO}_3(\mathbb{A}_f)/K_f'[q]$. From the above discussion, we deduce that

$$\widetilde{\mathscr{H}_d}(q) \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)} \xleftarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K_f'[q]. \tag{5.2}$$

5.4.2. *Lifting to $\mathscr{H}_d(q)$.* This adelic realization of $\widetilde{\mathscr{H}_d}(q)$ may, in fact, be lifted to an adelic realization of $\mathscr{H}_d(q)$ itself, at least when q is coprime with 3. For this we add the additional data of the principal level 3-structure discussed in §5.3. This is a little bit artificial, relying on the special fact that the reduction modulo 3 maps yields an isomorphism $\mathrm{SO}_3(\mathbb{Z}) \simeq \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$.

Consider the set of triples

$$\mathcal{P}_{(3,q)} :=$$
$$\{(L, \overline{\mathbf{x}}, \theta),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(L_0),\ \overline{\mathbf{x}} \in L/qL,\ \overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d(q),\ \theta : L/3L \simeq (\mathbb{Z}/3\mathbb{Z})^3\}.$$

As above, $\mathrm{SO}_3(\mathbb{A}_f)$ (hence its subgroup $\mathrm{SO}_3(\mathbb{Q})$) acts diagonally on $\mathcal{P}_{(3,q)}$: explicitly for $g \in \mathrm{SO}_3(\mathbb{A}_f)$, we have

$$g.(L, \overline{\mathbf{x}}, \theta) = (g.L, g.\overline{\mathbf{x}}, \theta \circ g^{-1}).$$

We consider first the $\mathrm{SO}_3(\mathbb{Q})$-orbits in $\mathcal{P}_{(3,q)}$. From (5.1) any $\mathrm{SO}_3(\mathbb{Q})$-orbit contains a triple of the form $(L_0, \overline{\mathbf{x}}, \theta)$. Moreover, since reduction modulo 3 is an isomorphism $\mathrm{SO}_3(\mathbb{Z}) \simeq \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$, the map $\overline{\mathbf{x}} \mapsto (\overline{\mathbf{x}}, \mathrm{Id})$ yields a bijection between $\mathscr{H}_d(q)$ and the set of $\mathrm{SO}_3(\mathbb{Z})$-orbits of pairs $\{(\overline{\mathbf{x}}, \theta),\ \overline{\mathbf{x}} \in \mathscr{H}_d(q),\ \theta : (\mathbb{Z}/3\mathbb{Z})^3 \simeq (\mathbb{Z}/3\mathbb{Z})^3\}$. From this, we deduce that the map $\overline{\mathbf{x}} \in \mathscr{H}_d(q) \mapsto (L_0, \overline{\mathbf{x}}, \mathrm{Id}) \in \mathcal{P}_{(3,q)}$ induces a bijection:

$$\mathscr{H}_d(q) \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(3,q)}.$$

Returning to the action of the whole group $\mathrm{SO}_3(\mathbb{A}_f)$, we have

Lemma. *This action is transitive: fixing $\overline{\mathbf{x}}_q \in \mathscr{H}_d(q)$, we have*

$$\mathcal{P}_{(3,q)} = \mathrm{SO}_3(\mathbb{A}_f).(L_0, \overline{\mathbf{x}}_q, \mathrm{Id}).$$

Proof. The proof is exactly as above: any triple is in the orbit of a triple of the form $(L_0, \overline{\mathbf{x}}, \theta)$, $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$, $\theta : (\mathbb{Z}/3\mathbb{Z})^3 \simeq (\mathbb{Z}/3\mathbb{Z})^3$. This follows from the fact that $\prod_{p|q} \mathrm{SO}_3(\mathbb{Z}_p) \times \mathrm{SO}_3(\mathbb{Z}_3)$ acts transitively on the set of pairs $(\overline{\mathbf{x}}, \theta)$; recall (§2.1) that we are assuming that q is prime to 3. □

We have $\mathcal{P}_{(3,q)} \simeq \mathrm{SO}_3(\mathbb{A}_f)/K_f[3,q]$. (Recall the notation from §5.3). From this and the above discussion, it follows that the map

$$\overline{\mathbf{x}} \in \mathscr{H}_d(q) \mapsto (L_0, \overline{\mathbf{x}}, \mathrm{Id}) \in \mathcal{P}_{(3,q)}$$

induces a bijective map:

$$\mathscr{H}_d(q) \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(3,q)} \xleftarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K_f[3,q]. \tag{5.3}$$

5.5. **The sphere as an adelic quotient.** For completeness, we recall the interpretation of sphere S^2 as an adelic quotient; this will not be used there but it is the way to proceed in order to adapt the proof of Theorems 1.3 and 1.4 to obtain Theorems 1.1 and 1.2 or to obtain hybrid equidistribution theorems. Let $\mathbb{A} = \mathbb{R} \times \mathbb{A}_f$ denote the full ring of adèles, then $\mathrm{SO}_3(\mathbb{A}) = \mathrm{SO}_3(\mathbb{R}) \times \mathrm{SO}_3(\mathbb{A}_f)$; hence, by (5.1) we have the identification

$$\mathrm{SO}_3(\mathbb{Z})\backslash\mathrm{SO}_3(\mathbb{R}) \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A})/\mathrm{SO}_3(\widehat{\mathbb{Z}}).$$

Since $\mathrm{SO}_3(\mathbb{R})$ acts transitively on S^2, we obtain – choosing some point $x_\infty \in S^2$ and letting

$$K_\infty = \mathrm{SO}_{x_\infty}(\mathbb{R}) \simeq \mathrm{SO}_2(\mathbb{R})$$

– the identification

$$\mathrm{SO}_3(\mathbb{Z})\backslash S^2 \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A})/K_\infty.\mathrm{SO}_3(\widehat{\mathbb{Z}}). \tag{5.4}$$

As in the previous section, we may remove the quotient by $\mathrm{SO}_3(\mathbb{Z})$ by adding the principal 3-structure and obtain

$$\mathrm{SO}_3(\mathbb{R}) \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A})/K_f[3].$$

$$S^2 \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A})/K_\infty.K_f[3]. \tag{5.5}$$

6. Adelic interpretation of $\mathscr{H}_d$.

In this section, we describe the first line of the diagrams (4.6) and (4.8).

The significance of this in the proof is as follows: We previously described an action of $\mathrm{Pic}(\mathscr{O}_K)$ on $\widetilde{\mathscr{H}_d}$, and also a *lifting* of this action – at least in the case of a prime ideal above 5 – to $\mathscr{H}_d$. After the present section, we will understand this action (and also the lifted action, although this is only of cosmetic interest) in terms of adèles. This will be a key tool in the proof of Proposition 2.10 and Proposition 2.13.

The presentation of this section follows that in [EV08], but we emphasize that the material is in essence classical.

6.1. As before, let $\mathcal{P}$ be the set of pairs

$$\mathcal{P} = \{(L, \mathbf{x}),\ \mathbf{x}.\mathbf{x} = d,\ \mathbf{x} \in L,\ L \in \mathrm{genus}_{\mathrm{SO}_3}(L_0)\}.$$

By contrast with $\mathcal{P}_{(q)}$ or $\mathcal{P}_{(3,q)}$, the set $\mathcal{P}$ carries no natural action of $\mathrm{SO}_3(\mathbb{A}_f)$; on the other hand, the group $\mathrm{SO}_3(\mathbb{Q})$ acts on $\mathcal{P}$ diagonally. By (5.1), every $\mathrm{SO}_3(\mathbb{Q})$-orbit in $\mathcal{P}$ contains an element of the form $(L_0, \mathbf{x})$ (where $\mathbf{x}$ belongs to $\mathscr{H}_d$); two such pairs differ by the action of a unique element of $\mathrm{SO}_3(\mathbb{Q}) \cap \mathrm{SO}_3(\widehat{\mathbb{Z}}) = \mathrm{SO}_3(\mathbb{Z})$. From this is follows that the map $\mathbf{x} \in \mathscr{H}_d \mapsto (L_0, \mathbf{x}) \in \mathcal{P}$ induces a bijective map

$$\widetilde{\mathscr{H}_d} \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}. \tag{6.1}$$

We will now realize $\mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}$ as an adelic quotient by endowing it with a transitive action of a *subgroup* of $\mathrm{SO}_3(\mathbb{A}_f)$.

Choose an element $\mathbf{x}_0 \in \mathscr{H}_d$, and let $\mathrm{SO}_{\mathbf{x}_0}$ be its stabilizer in SO_3. Since the quadratic space is 3-dimensional, $\mathrm{SO}_{\mathbf{x}_0}$ is the special orthogonal group of a quadratic plane (namely $\mathbb{Q}\mathbf{x}_0^{\perp}$), and so is a 1-dimensional torus.

The group $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ acts (by multiplication on the first coordinate) on the subset of $\mathcal{P}$ consisting of pairs of the form $(L, \mathbf{x}_0)$. We can extend this action to the whole of $\mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}$ as follows. By Witt's theorem, every $\mathrm{SO}_3(\mathbb{Q})$-orbit in $\mathcal{P}$ is of the form $[L, \mathbf{x}_0]$, for some $L \in \mathrm{SO}_3(\mathbb{A}_f)L_0$. Then or any $t \in \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$, we define

$$t.[L, \mathbf{x}_0] := [t.L, \mathbf{x}_0].$$

This is well defined: if $[L, \mathbf{x}_0] = [L', \mathbf{x}_0]$ then L and L' differ by an element $\tau \in \mathrm{SO}_{x_0}(\mathbb{Q})$; because of the commutativity of $\mathrm{SO}_{\mathbf{x}_0}$,

$$[t\tau L, \mathbf{x}_0] = [\tau t L, \mathbf{x}_0] = [tL, \mathbf{x}_0].$$

This action is transitive: consider the orbit of $(L, \mathbf{x}_0)$ where $L = g.L_0$ and $g \in \mathrm{SO}_3(\mathbb{A}_f)$. Since L contains $\mathbf{x}_0$, $L_0 \otimes \mathbb{A}_f$ contains $g^{-1}\mathbf{x}_0$; by Lemma 5.4.1 there is an element k of $\mathrm{SO}_3(\widehat{\mathbb{Z}})$ which sends $\mathbf{x}_0$ to $g^{-1}\mathbf{x}_0$. In particular, $gk = t$ lies in $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ and $t.[L_0, \mathbf{x}_0] = [gkL_0, \mathbf{x}_0] = [L, \mathbf{x}_0]$.

Finally, from the definition of the action it is easy to see that the stabilizer of $[L_0, \mathbf{x}_0]$ is the subgroup $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$ where $\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}) = \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \cap \mathrm{SO}_3(\widehat{\mathbb{Z}})$. Hence (6.1) extends to bijections

$$\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P} \xleftarrow{\sim} \mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}). \tag{6.2}$$

Remark. This setup is very specific to the three dimensional situation, in particular the fact that SO_{x_0} is commutative. If one studies the question of the representations of an integer d by a higher rank quadratic form Q, the quotient $\mathrm{SO}_Q(\mathbb{Z})\backslash\mathscr{H}_d$, when non-empty, still has a description in terms of double cosets of an adelic group (see [EV08] for details) but will not carry an *action* of an adelic group as it does here.

6.2. We shall now identify $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$ with a quotient of the commutative group $\mathrm{Pic}(\mathscr{O}_K)$. Therefore, the identification (6.2) may be considered as giving an action of $\mathrm{Pic}(\mathscr{O}_K)$ on $\widetilde{\mathscr{H}_d}$.

For this purpose, it is again most convenient to phrase everything in terms of quaternions via the identifications $\mathbb{Q}^3 \simeq \mathrm{B}^{(0)}(\mathbb{Q})$ and $\mathrm{SO}_3 \simeq \mathrm{PB}^\times$. In particular we view $\mathbf{x}_0$ as a trace-free quaternion. Now the stabilizer $\mathrm{SO}_{\mathbf{x}_0} \simeq \mathrm{PB}^\times_{\mathbf{x}_0}$ is the multiplicative group generated by (conjugations by) invertible quaternions of the form $a + b\mathbf{x}_0$. In the previous section, we have defined a transitive action of $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \simeq \mathrm{PB}^\times_{\mathbf{x}_0}(\mathbb{A}_f)$ on $\widetilde{\mathscr{H}_d}$.

Now, the map $a+b\sqrt{-d} \mapsto a+b\mathbf{x}_0$ defines via conjugation an isomorphism of $\mathbb{Q}$-algebraic groups

$$\mathrm{Res}_{K/\mathbb{Q}}\mathbf{G}_m/\mathbf{G}_m \simeq \mathrm{PB}^\times_{\mathbf{x}_0}; \tag{6.3}$$

in particular, for any field $F \supset \mathbb{Q}$, $\mathrm{PB}^\times_{\mathbf{x}_0}(F) \simeq (F \otimes_{\mathbb{Q}} K)^\times/F^\times$.

Thus, denoting by $\mathbb{A}^\times_{K,f} = (\mathbb{A}_f \otimes K)^\times$ the group of finite idèles of K, (6.3) defines an action of $\mathbb{A}^\times_{K,f}$ on $\widetilde{\mathscr{H}_d}$. The subgroups $\mathbb{A}^\times_f$ and $K^\times$ act trivially, as does $U \subset \mathbb{A}^\times_{K,f}$, the preimage of $\mathrm{PB}^\times_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$ under (6.3). The maximal compact subgroup $\widehat{\mathscr{O}}^\times_K = (\mathscr{O}_K \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}})^\times$ of $\mathbb{A}^\times_{K,f}$ is certainly contained in U; therefore, we have a well defined transitive action of

$$K^\times\backslash\mathbb{A}^\times_{K,f}/\widehat{\mathscr{O}}^\times_K = \mathrm{Pic}(\mathscr{O}_K)$$

on $\widetilde{\mathscr{H}_d}$. This action being transitive, $\widetilde{\mathscr{H}_d}$ is identified with a quotient of $\mathrm{Pic}(\mathscr{O}_K)$, and is in fact a torsor under the abelian group $K^\times\backslash\mathbb{A}^\times_{K,f}/U$. A careful study of the inclusion $\widehat{\mathscr{O}}^\times_K \subset U$ at 2 recovers Proposition 3.5.

A priori, the action we have defined depends on $\mathbf{x}_0$. We verify independence in §6.4.

6.3. **Lifting the action to $\mathscr{H}_d$.** As before, by introducing an extra level 3-structure, we may replace $\widetilde{\mathscr{H}_d}$ by its covering $\mathscr{H}_d$ and obtain a lift of the action of $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \simeq \mathrm{PB}^\times_{\mathbf{x}_0}(\mathbb{A}_f)$ and thus of $\mathbb{A}^\times_{K,f}$ on $\widetilde{\mathscr{H}_d}$ to an action on $\mathscr{H}_d$. Notice that this latter action is, in general, not necessarily transitive; nor does it, necessarily, factor through the class group, but only through a certain ray class group.

Consider the set of triples

$$\mathcal{Q} = \{(L, \mathbf{x}, \theta),\ L \in \mathrm{genus}_{\mathrm{SO}_3}(L_0),\ \mathbf{x} \in L,\ \mathbf{x}.\mathbf{x} = d, \theta : L/3L \simeq (\mathbb{Z}/3\mathbb{Z})^3\}.$$

Using that the reduction mod 3 map from $\mathrm{SO}_3(\mathbb{Z})$ to $\mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$ is an isomorphism, we may verify as above that the map

$$\mathscr{H}_d \to \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{Q}, \quad \mathbf{x} \mapsto [L_0, \mathbf{x}, \mathrm{Id}]$$

is bijective.

As above, the group $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ acts on $\mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{Q}$: by Witt's Theorem, every $\mathrm{SO}_3(\mathbb{Q})$-orbit in $\mathcal{Q}$ is of the form $[L, \mathbf{x}_0, \theta]$ and we set for $t \in \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$,

$$t.[L, \mathbf{x}_0, \theta] = [t.L, \mathbf{x}_0, \theta \circ t^{-1}];$$

from the previous discussion, this action is well defined. This action, however, is *not* transitive: the various $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$-orbits are parametrized by the orbits of $\mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$ under the action of $\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$. In other words, we have a bijection

$$\mathscr{H}_d \xrightarrow{\sim} \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{Q} \xleftarrow{\sim} \mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \times \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}), \tag{6.4}$$

where $\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$ acts diagonally on the product $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \times \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$. Under this identification, the action of $t \in \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ is the one induced by

$$t.[t', \kappa] = [tt', \kappa],\ (t', \kappa) \in \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f) \times \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z}).$$

Thus (6.4) gives us the desired way to lift the action of $\mathbb{A}_{K,f}^\times$ to $\mathscr{H}_d$.

6.4. **Independence w.r.t. $\mathbf{x}_0$.** Let us see that the above defined actions of $\mathbb{A}_{K,f}^\times$ on $\mathscr{H}_d$, $\widetilde{\mathscr{H}_d}$ do not depend on the choice of the base point $\mathbf{x}_0$.

Let $\mathbf{x}_0' \in \mathscr{H}_d$ be another point. By Witt's theorem $\mathbf{x}_0' = \rho.\mathbf{x}_0$, for some $\rho \in \mathrm{PB}^\times(\mathbb{Q})$. Then $\mathrm{PB}^\times_{\mathbf{x}_0'} = \rho \mathrm{PB}^\times_{\mathbf{x}_0} \rho^{-1}$. Let $u = a + b\sqrt{-d}$ be an element of $\mathbb{A}_K^\times$ ($a, b \in \mathbb{A}_f$), and let t_u (resp. t_u') denote the corresponding element in $\mathrm{PB}^\times_{\mathbf{x}_0}(\mathbb{A}_f)$ (resp. in $\mathrm{PB}^\times_{\mathbf{x}_0'}(\mathbb{A}_f)$). Then $t_u' = \rho t_u \rho^{-1}$. It will suffice to see that $t_u.\mathbf{x}_0 = t_u'.\mathbf{x}_0$, or equivalently

$$t_u[L_0, \mathbf{x}_0, \mathrm{Id}] = t_u'[L_0, \mathbf{x}_0, \mathrm{Id}];$$

the latter expression equals

$$\begin{aligned} t_u'[L_0, \rho^{-1}\mathbf{x}_0', \mathrm{Id}] &= t_u'[\rho L_0, \mathbf{x}_0', \rho^{-1}] = [t_u' \rho L_0, \mathbf{x}_0', \rho^{-1} t_u'^{-1}] \\ &= [\rho^{-1} t_u' \rho L_0, \rho^{-1}\mathbf{x}_0', \rho^{-1} t_u'^{-1} \rho] \\ &= [t_u L_0, \mathbf{x}_0, t_u^{-1}] = t_u[L_0, \mathbf{x}_0, \mathrm{Id}]. \end{aligned}$$

□

7. Adelic interpretation of $\mathrm{red}_q : \mathscr{H}_d \to \mathscr{H}_d(q)$.

In this section, we interpret the reduction maps, red_q and red_∞ in terms of the adelic quotients from the previous sections.

Recall that, in (5.4), (5.2), (6.2) (resp. (5.5), (5.3), (6.4)) we have given adelic identifications of $\widetilde{S}^2$, $\widetilde{\mathscr{H}_d(q)}$ and $\widetilde{\mathscr{H}_d}$, and of their finite coverings S^2, $\mathscr{H}_d(q)$ and $\mathscr{H}_d$.

7.1. First, let us recall the identifications

$$\widetilde{\mathscr{H}_d} \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P},\ \widetilde{\mathscr{H}_d}(q) \simeq \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)}$$

where

$$\mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P} = \{[L,\mathbf{x}], L \in \mathrm{SO}_3(\mathbb{A}_f).L_0,\ \mathbf{x} \in L,\ \mathbf{x}.\mathbf{x} = d\}$$

$$\mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{P}_{(q)} = \{[L,\overline{\mathbf{x}}],\ L \in \mathrm{SO}_3(\mathbb{A}_f).L_0,\ \overline{\mathbf{x}} \in L/qL,\ \overline{\mathbf{x}}.\overline{\mathbf{x}} \equiv d(q)\}.$$

The map $\mathrm{red}_q : \widetilde{\mathscr{H}_d} \to \widetilde{\mathscr{H}_d}(q)$ is induced by the natural map

$$\mathbf{x} \in L \mapsto \overline{\mathbf{x}} \in L/qL$$

which we also denote red_q.

We now explain how to write red_q in the adelic language. Let t be an element of $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ and $[t]$ its double coset in $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$. Recall that $\mathbf{x}_0$ and $\overline{\mathbf{x}}_q$ were basepoints in $\widetilde{\mathscr{H}_d}$ and $\widetilde{\mathscr{H}_d}(q)$ respectively. We will demonstrate that the reduction map red_q, thought of as a map

$$\mathrm{red}_q : \mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}) \to \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K_f[q]$$

is the one sending $[t]$ to $[t.k_{\overline{\mathbf{x}}_0}]$, where $k_{\overline{\mathbf{x}}_0} \in \mathrm{SO}_3(\widehat{\mathbb{Z}})$ is a fixed element satisfying $k_{\overline{\mathbf{x}}_0}.\overline{\mathbf{x}}_q \equiv \mathbf{x}_0 (\mathrm{mod}\, q)$.

To see this, let $t \in \mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)$ be a representative for one of the double cosets in $\mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}})$. We may choose t to have integral coordinates at all primes dividing q. Write $\beta = \gamma k$, with $\gamma \in \mathrm{SO}_3(\mathbb{Q})$ and $k \in \mathrm{SO}_3(\widehat{\mathbb{Z}})$. Then by the definitions of (6.2) and (5.3), one finds that the element of $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d$ corresponding to t is $\gamma^{-1}\mathbf{x}_0$, while the element of $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q)$ corresponding to $tk_{\overline{\mathbf{x}}_0}$ is $k\bar{\mathbf{x}}_0$. But γk fixes $\mathbf{x}_0$ (whence also $\bar{\mathbf{x}}_0$), so the reduction of $\gamma^{-1}\mathbf{x}_0$ is precisely $k\bar{\mathbf{x}}_0$.

Again, we can lift the situation to $\mathscr{H}_d$: there is a natural map

$$\mathrm{SO}_{\mathbf{x}_0}(\mathbb{Q})\backslash\mathrm{SO}_{\mathbf{x}_0}(\mathbb{A}_f)\times\mathrm{SO}(\mathbb{Z}/3\mathbb{Z})/\mathrm{SO}_{\mathbf{x}_0}(\widehat{\mathbb{Z}}) \mapsto \mathrm{SO}_3(\mathbb{Q})\backslash\mathrm{SO}_3(\mathbb{A}_f)/K_f[3,q]$$

which corresponds to the reduction map $\mathscr{H}_d \to \mathscr{H}_d(q)$. Explicitly, choosing for base points the triples $(L_0, \mathbf{x}_0, \mathrm{Id})$, $(L_0, \overline{\mathbf{x}}_q, \mathrm{Id})$ the map is given by

$$[t,\kappa] \mapsto [tk_{\overline{\mathbf{x}}_0}k_3] \tag{7.1}$$

where $k_3 \in \mathrm{SO}_3(\mathbb{Z}_3)$ is a lift of $\kappa \in \mathrm{SO}_3(\mathbb{Z}/3\mathbb{Z})$ and $k_{\overline{\mathbf{x}}_0}$ is as above but such that its component at the place 3 is trivial.

Part 4. Graphs and Expanders

8. The graph structure on $\mathscr{H}_d(q)$

We shall now replace the role of the finite adèles $\mathbb{A}_f$ in the bijection (5.3) by the much smaller ring $\mathbb{Q}_5$. More precisely, we will show the existence of a bijection:

$$\mathscr{H}_d(q) \overset{\sim}{\leftarrow} \Gamma_{(3,q)}\backslash\mathrm{SO}_3(\mathbb{Q}_5)/K_5, \tag{8.1}$$

where $\Gamma_{(3,q)} \subset \mathrm{SO}_3(\mathbb{Q}_5)$ is a suitable lattice (i.e. a discrete cofinite subgroup) and K_5 is the maximal compact subgroup $\mathrm{SO}_3(\mathbb{Z}_5)$. We have

$$\mathrm{SO}_3(\mathbb{Q}_5) \simeq \mathrm{PB}^\times(\mathbb{Q}_5) \simeq \mathrm{PGL}_2(\mathbb{Q}_5)$$

(since $\mathrm{B}(\mathbb{Q}_5) \simeq M_2(\mathbb{Q}_5)$), therefore the quotient $\mathrm{SO}_3(\mathbb{Q}_5)/K_5$ is identified with

$$\mathrm{PGL}_2(\mathbb{Q}_5)/\mathrm{PGL}_2(\mathbb{Z}_5) =: \mathcal{T}_5$$

which has the structure of an infinite 6-valent tree (namely, the *Bruhat-Tits* tree of $\mathrm{PGL}_2(\mathbb{Q}_5)$, cf. [Ser03]).

The set $\mathscr{H}_d(q)$ has thus a structure of a finite quotient of $\mathcal{T}_5$ and we will see that this graph structure coincides with that described in §2.9. In particular, the latter is connected.

From this viewpoint, it will be possible to prove Proposition 2.10 and Proposition 2.13 (i.e. the graph $\mathscr{H}_d(q)$ is an expander). The latter relies on the theory of automorphic forms, especially the Jacquet-Langlands correspondence and the work of Eichler-Shimura.

Let us mention that a good part of this section is closely related to the book of Lubotzky [Lub94], especially Chapter 6 and the Appendix, in which the reader will find a motivated discussion of the passage between automorphic forms and expanders.

8.1. $\mathscr{H}_d(q)$ **as a quotient of a tree.** If w is an element of $\mathrm{PB}^\times(\mathbb{Q}_5)$, we denote by $[w]_5$ the element of $\mathrm{PB}^\times(\mathbb{A}_f)$ which projects to w at the 5-adic place, and to the identity everywhere else. When no confusion is likely (as in the statement of the following lemma) we will identify $\mathrm{PB}^\times(\mathbb{Q}_5)$ with the subgroup $[\mathrm{PB}^\times(\mathbb{Q}_5)]_5$ of $\mathrm{PB}^\times(\mathbb{A}_f)$.

Lemma. *One has*

$$\mathrm{PB}^\times(\mathbb{Q}).\mathrm{PB}^\times(\mathbb{Q}_5).K_f[3,q] = \mathrm{PB}^\times(\mathbb{A}_f).$$

Consequently, the map $g \in \mathrm{PB}^\times(\mathbb{Q}_5) \to [g]_5 \in \mathrm{PB}^\times(\mathbb{A}_f)$ *yields a bijective map*

$$\Gamma_{(3,q)}\backslash \mathrm{PB}^\times(\mathbb{Q}_5)/K_5 \xrightarrow{\sim} \mathrm{PB}^\times(\mathbb{Q})\backslash \mathrm{PB}^\times(\mathbb{A}_f)/K_f[3,q] \xleftarrow{\sim} \mathscr{H}_d(q) \tag{8.2}$$

where $\Gamma_{(3,q)}$ *is the lattice* $\mathrm{PB}^\times(\mathbb{Q}) \cap \mathrm{PB}^\times(\mathbb{Q}_5).K_f[3,q]$.

Proof. This main ingredient of the proof is the so-called *strong approximation property* (for simply connected semi-simple algebraic groups): we will not discuss this property in any generality and refer to [PR94, Chap. 7] for a complete treatment. Alternatively, the reader may also refer to [Vig80, Chap. III] for a discussion of the strong approximation property in the context of quaternion algebras. For now, let us merely say that, if $\mathrm{PB}^\times$ satisfied the strong approximation property the assertion would follow immediately; unfortunately, $\mathrm{PB}^\times$ is not simply connected.

One remedies this problem by passing from the $\mathbb{Q}$-algebraic group to its double cover B^1 (cf. (3.2)). The group B^1 is simply connected and

semisimple. Moreover, $\mathrm{B}^1(\mathbb{Q}_5)$ is noncompact. Thus, it satisfies a strong approximation property: for any open compact $\Omega \subset \mathrm{B}^1(\mathbb{A}_f)$, we have

$$\mathrm{B}^1(\mathbb{Q}).\mathrm{B}^1(\mathbb{Q}_5).\Omega = \mathrm{B}^1(\mathbb{A}_f).$$

It follows that $\mathrm{PB}^\times(\mathbb{Q}).\mathrm{PB}^\times(\mathbb{Q}_5).K_f[3,q]$ contains the image

$$\Theta = \mathrm{B}^1(\mathbb{A}_f)/\{\pm 1\}$$

of $\mathrm{B}^1(\mathbb{A}_f)$ in $\mathrm{PB}^\times(\mathbb{A}_f)$. It will suffice, then, to verify that

$$\mathrm{PB}^\times(\mathbb{Q}).\Theta.K_f[3,q] = \mathrm{PB}^\times(\mathbb{A}_f); \tag{8.3}$$

this will follow from (5.1) if $(\Theta \cap \mathrm{PB}^\times(\widehat{\mathbb{Z}})).K_f[3,q] = \mathrm{PB}^\times(\widehat{\mathbb{Z}})$; equivalently, if $\Theta \cap \mathrm{PB}^\times(\widehat{\mathbb{Z}})$ acts transitively on $\mathrm{PB}^\times(\widehat{\mathbb{Z}})/K_f[3,q] \simeq \mathscr{H}_d(q)$.

In turn, it is equivalent to show that $\Theta_p \cap \mathrm{PB}^\times(\mathbb{Z}_p)$ acts transitively on $\mathscr{H}_d(\mathbb{Z}_p)$ for each $p|3q$, where Θ_p is the image of $\mathrm{B}^1(\mathbb{Q}_p)$ in $\mathrm{PB}^\times(\mathbb{Q}_p)$.

Recall that $\mathscr{H}_d(\mathbb{Z}_p)$ is identified with $\mathrm{B}^{(0,d)}(\mathbb{Z}_p)$, the trace 0 quaternion of norm $d \in \mathbb{Z}_p^\times$; then $\Theta_p \cap \mathrm{PB}^\times(\mathbb{Z}_p)$ contains $\mathrm{B}^1(\mathbb{Z}_p)/\{\pm 1\}$ acting by conjugation. The transitivity of this action follows from the second part of Proposition 3.7, using the fact that $(p, 2d) = 1$. □

8.2. **The graph structure.** We can describe the (Bruhat-Tits) graph structure on $\mathrm{PB}^\times(\mathbb{Q}_5)/\mathrm{PB}^\times(\mathbb{Z}_5)$ in two equivalent ways. We write $K_5 = \mathrm{PB}^\times(\mathbb{Z}_5)$ in what follows. Recall that $\mathcal{A}_5$ consists of a set of six matrices (as in §2.6) or six quaternions (as in §3.12) depending on whether we are working in SO_3 or $\mathrm{PB}^\times$.

(1) Fix $\alpha_5 \in \mathrm{B}^\times(\mathbb{Q}_5)$ such that the 5-adic valuation of its norm is ± 1. Write $\overline{\alpha_5}$ for its image in $\mathrm{PB}^\times(\mathbb{Q}_5)$; then we have

$$K_5.\overline{\alpha_5}.K_5 = \bigsqcup_{w \in \mathcal{A}_5} [w]_5 K_5. \tag{8.4}$$

We join any coset gK_5 to the six cosets $g[w]_5K_5 : w \in \mathcal{A}_5$. The resulting structure is independent of α_5.

(2) More intrinsically, we may identify the quotient $\mathrm{SO}_3(\mathbb{Q}_5)/\mathrm{SO}_3(\mathbb{Z}_5)$ with the sublattices of $\mathbb{Q}_5^3$ in the orbit of $\mathrm{SO}_3(\mathbb{Q}_5).\mathbb{Z}_5^3$. Given any such sublattice L, the quadratic form takes values on $\mathbb{Z}_5$ on that lattice and the induced quadratic form on $L/5L \cong (\mathbb{Z}/5\mathbb{Z})^3$ takes values in $\mathbb{Z}/5\mathbb{Z}$; there are precisely six ($= |\mathbb{P}^1(\mathbb{Z}/5\mathbb{Z})|$) isotropic lines. Choosing any $\mathbf{v} \in L$ so that the image of $\mathbf{v}$ in $L/5L$ isotropic (but non-zero), we may construct the new lattice:

$$L' = \langle \mathbf{v}/5 \rangle + \{\mathbf{w} \in L : \mathbf{v}.\mathbf{w} \equiv 0 \bmod 5.\}$$

Then L' depends only on the line spanned by $\mathbf{v}$ in $L/5L$, and belongs to $\mathrm{SO}_3(\mathbb{Q}_5).\mathbb{Z}_5^3$ also. In particular, we construct six such L', which we declare to be the neighbours of L.

Then these give equivalent graph structures and moreover:

(8.5) The resulting graph is a tree, i.e., has no cycles.

Later on we shall use the following property the verification of which we leave to the reader: if L is a sublattice of $\mathbb{Q}_5^3$ representing an element of $\mathrm{SO}_3(\mathbb{Q}_5)/\mathrm{SO}_3(\mathbb{Z}_5)$, and $\mathbf{v} \in L$ is such that $\mathbf{v}.\mathbf{v}$ is not zero mod 5, then:

(8.6) The orbit $\mathrm{SO}_{\mathbf{v}}(\mathbb{Q}_5).L$ is an infinite geodesic in the tree,

(i.e. an isometric embedding of $\mathbb{Z}$ in the tree w.r.t. the obvious metrics); this is a simple consequence of the second definition.

Let us now check that the neighbors of $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ for the graph structure defined in §2.9 correspond to the neighbors under T_5 of the image of $\overline{\mathbf{x}}$ in (5.3). Thus, the graph structure on $\mathscr{H}_d(q)$ coincides with the graph structure on $\Gamma_{(3,q)}\backslash\mathcal{T}_5$.

In the notations of §5.4.2, the neighbors of $\overline{\mathbf{x}}$ i.e. $\{w.\overline{\mathbf{x}},\ w \in \mathcal{A}_5\}$ correspond to the orbits of the triples

$$\{[L_0, w.\overline{\mathbf{x}}, \mathrm{Id}],\ w \in \mathcal{A}_5\} = \{[w^{-1}L_0, \overline{\mathbf{x}}, \mathrm{Id} \circ w],\ w \in \mathcal{A}_5\};$$

now for any $p \neq 5$ and $w \in \mathcal{A}_5$, $[w]_p \in K_p$. Moreover, $[w]_3 \equiv \mathrm{Id} (\mathrm{mod}\, 3)$. Therefore the above set equals $\{[w^{-1}]_5.[L_0, \overline{\mathbf{x}}, \mathrm{Id}],\ w \in \mathcal{A}_5\}$ which manifestly agrees with the graph structure introduced above.

8.3. **Proof of Proposition 2.10.** We focus now on the action of a prime ideal $\mathfrak{p}$ above 5 on $\widetilde{\mathscr{H}_d}$. Fix π a uniformizer of $K_{\mathfrak{p}}$, which we write in the form $a + b\sqrt{-d} \in \mathbb{Z}_5[\sqrt{-d}]$. To any $\mathbf{x} \in \mathscr{H}_d$ we associate the quaternion $q_{\mathbf{x}} = a + b\mathbf{x}$ (the 5-adic valuation of $\mathrm{Nr}(q_{\mathbf{x}})$ equals 1) and the corresponding rotation $t_{\mathbf{x}} \in \mathrm{SO}_{\mathbf{x}}(\mathbb{Q}_5)$ induced by conjugation by $q_{\mathbf{x}}$; since $K_{\mathfrak{p}}^\times = \pi^{\mathbb{Z}}\mathscr{O}_{K_{\mathfrak{p}}}^\times$ we have $\mathrm{SO}_{\mathbf{x}}(\mathbb{Q}_5) = t_{\mathbf{x}}^{\mathbb{Z}}\mathrm{SO}_{\mathbf{x}}(\mathbb{Z}_5)$.

The action of $\mathfrak{p}$ can be interpreted in terms of our adelic viewpoint: We have seen in §6.3, that the group $\mathrm{SO}_{\mathbf{x}}(\mathbb{Q}_5)$ acts on $\mathscr{H}_d$ with $\mathrm{SO}_{\mathbf{x}}(\mathbb{Z}_5)$ acting trivially; by projection this also defines an action on $\widetilde{\mathscr{H}_d}$. Via the map $\pi \mapsto q_{\mathbf{x}} \mapsto t_{\mathbf{x}}$ one obtains an action of the group $\pi^{\mathbb{Z}}$ which in fact does not depend on the choice of $\mathbf{x}$ (§6.4); the action of π coincides with the action of the ideal class $[\mathfrak{p}]$ defined earlier.

Let us also recall that if $\mathbf{x} \in \mathscr{H}_d$ corresponds to the class $[L_0, \mathbf{x}, \mathrm{Id}] \in \mathrm{SO}_3(\mathbb{Q})\backslash\mathcal{Q}$, the element $\pi.\mathbf{x} \in \mathscr{H}_d$ corresponds to the class $t_{\mathbf{x}}[L_0, \mathbf{x}, \mathrm{Id}] = [t_{\mathbf{x}}L_0, \mathbf{x}, \mathrm{Id} \circ t_{\mathbf{x}}^{-1}] = [t_{\mathbf{x}}L_0, \mathbf{x}, \mathrm{Id}]$, the final equality holding because the 3-component of $t_{\mathbf{x}}$ is trivial. Therefore the trajectory $\pi^{\mathbb{Z}}.\mathbf{x}$ is described by the infinite sequence of lattices

$$\ldots, L_{-2}, L_{-1}, L_0, L_1, L_2, \ldots,\ L_i = t_{\mathbf{x}}^i L_0.$$

All the L_i contain $\mathbf{x}$. Write $L_{i,p}$ for $L_{i,p} = L_i \otimes_{\mathbb{Z}} \mathbb{Z}_p$; then $L_{i,p} = L_{0,p}$ for all i and all $p \neq 5$, the p-th component of $t_{\mathbf{x}}$ being trivial. So the sequence of lattices (L_i) is completely determined by the sequence $(L_{i,5})$ of its 5-adic components.

Since the rotation $t_{\mathbf{x}}$ comes from a quaternion whose norm has 5-adic valuation equal to 1, the sequence $(L_{i,5})_{i\in\mathbb{Z}}$ describes an infinite geodesic passing through $L_{0,5}$ in the tree (cf. 8.2 (1) or more generally (8.6)),

$$\mathrm{SO}_3(\mathbb{Q}_5).L_{0,5} \simeq \mathrm{SO}_3(\mathbb{Q}_5)/\mathrm{SO}_3(\mathbb{Z}_5).$$

Up to orientation, any such geodesic may be encoded by an infinite non-backtracking word in

$$\mathcal{A}_5 = \{A^{\pm 1}, B^{\pm 1}, C^{\pm 1}\},$$

where the i-th letter connects the $(i-1)$-st element of the geodesic along an edge of the tree to the i-th element. In the present case, the word associated with the sequence $(L_{i,5})_{i\in\mathbb{Z}}$ is the word $(w_i^{-1})_{i\in\mathbb{Z}}$ satifying $L_{i,5} = w_i^{-1} L_{i-1,5}$; equivalently, $(w_i)_{i\in\mathbb{Z}}$ is the word corresponding to the trajectory of $\mathbf{x}$ defined in §2.6.

Suppose that two points $\mathbf{x}, \mathbf{x}' \in \mathscr{H}_d$ give rise to the same truncated word of length 2ℓ:

$$W^{(\ell)} : [-\ell+1, \ell] \to \{A^{\pm 1}, B^{\pm 1}, C^{\pm 1}\} = \mathcal{A}_5.$$

This means exactly that the geodesics $\mathrm{SO}_{\mathbf{x}}(\mathbb{Q}_5).L_{0,5}$ and $\mathrm{SO}_{\mathbf{x}'}(\mathbb{Q}_5).L_{0,5}$ in the Bruhat-Tits tree coincide "from times $-\ell$ to ℓ " or in other terms

$$L_{i,5} = L'_{i,5},\ i = -\ell, \dots, \ell.$$

In particular $\mathbf{x}' \in L_{\ell,5} \cap L_{-\ell,5}$. The last intersection is a sublattice of $L_{0,5}$ of index $5^{2\ell}$; more precisely, it is the preimage in $L_{0,5}$ of the line generated by $\mathbf{x} (\mathrm{mod}\, 5^\ell)$ in $L_{0,5}/5^\ell L_{0,5}$. Thus $\mathbf{x}$ and $\mathbf{x}'$ are linearly dependent in $L_0/5^\ell L_0$, and since both vectors have norm d, we have $\mathbf{x} \equiv \pm\mathbf{x}'$ modulo 5^ℓ. This concludes the proof of Proposition 2.10. □

8.4. **Proof of Proposition 2.13.** We assume some familiarity with the theory of automorphic forms; in any case, we refer to Lubotzky's book [Lub94, Chap. 6 & Appendix].

The space $L^2(\mathrm{PB}^\times(\mathbb{Q})\backslash \mathrm{PB}^\times(\mathbb{A}_f)/K_f[3,q])$ admits an orthogonal decomposition into eigenspaces of the commutative algebra generated by Hecke operators. These eigenspaces are the set of $\mathrm{PB}^\times(\mathbb{R})K_f[3,q]$-invariant vectors of automorphic representations on $\mathrm{PB}^\times$. Such representations are of two types:

- one-dimensional representations;
- infinite dimensional representations.

The latter corresponds, via the Jacquet-Langlands correspondence [JL70] to automorphic representations of PGL_2 which are discrete series of weight 2 and unramified outside $2, 3$ and the primes dividing q (more precisely, one can check that the conductor divides $18.q^2$). From the work of Deligne (or rather Eichler/Igusa/Shimura since this is weight 2), the eigenvalue of the standard 5-Hecke operator for such spaces is bounded in absolute value by $2\sqrt{5}$.

As for the former: each such is the representation of $\mathrm{PB}^\times(\mathbb{A})$ on the one-dimensional subspace generated by the function

$$g \in \mathrm{PB}^\times(\mathbb{Q})\backslash\mathrm{PB}^\times(\mathbb{A}) \mapsto \chi(\mathrm{Nr}(g))$$

where $\chi : \mathbb{Q}^\times\backslash\mathbb{A}^\times \mapsto \{\pm 1\}$ is some quadratic character. The action of $\mathrm{PB}^\times(\mathbb{Q})$ on such a representation is trivial, as is the action of $K_f[3,q]$ (by definition); moreover since the elements of Θ come from quaternions of norm 1, the action of Θ is trivial as well; hence from (8.3), such representation has to be the trivial one. It follows from this enumeration that -6 does not occur as an eigenvalue of the 5-th Hecke operator so the graph, while connected (cf. above) is not bipartite.

Remark. The discussion above is also valid for $\widetilde{\mathscr{H}_d}(q)$: this follows immediately from the previous discussions by projection. In particular we have

$$\Gamma_{(q)}\backslash\mathrm{PB}^\times(\mathbb{Q}_5)/K_5 \xrightarrow{\sim} \mathrm{PB}^\times(\mathbb{Q})\backslash\mathrm{PB}^\times(\mathbb{A}_f)/K_f'[q] \xleftarrow{\sim} \widetilde{\mathscr{H}_d}(q)$$

with $\Gamma_{(q)} = \mathrm{PB}^\times(\mathbb{Q}) \cap K_f'[q]\mathrm{PB}^\times(\mathbb{Q}_5)$.

9. Expander graphs and random walks

The contents of this section follow lecture notes of Hoori, Linial and Wigderson [HLW06]. Our goal is to prove Proposition 2.14.

9.1. Let $\mathscr{G} = (V, E)$ be a (possibly directed) d-regular graph on $|V| = n$ vertices, i.e. the number of incoming edges to each vertex is d, and the number of outgoing edges is also d. We assume $d > 2$. The normalized adjacency matrix T of $\mathscr{G}$ acts on $L^2(V)$ by

$$Tf(x) = \frac{1}{d}\sum_{(x\mapsto y)\in E} f(y).$$

By an abuse of notation, we will use $L^2(\mathscr{G})$ and $L^2(V)$ interchangeably. More generally, $\mathscr{G}$ may be allowed to have multiple edges and loops, in which case we modify the definition of T in the evident way.

Let $\|T\|$ be the operator norm of T acting on the orthogonal complement of the constants in $L^2(V)$. The graph $\mathscr{G}$ is said to be an α-expander, for some $\alpha < 1$, if $\|T\| \leqslant 1 - \alpha$. In rough terms, the smaller $\|T\|$ is, the more "strongly connected" the graph $\mathscr{G}$.

When we speak of a "random walk on $\mathscr{G}$," we mean that we select a vertex uniformly and randomly from V, and then proceed to walk along directed edges, at each stage choosing one of the adjacent edges one with each choice assigned probability $1/d$.

Lemma 9.1.1. *Let $Q_1, \dots, Q_\ell$ be subsets of V, with densities $\mu_i := |Q_i|/n$. The probability that a random walk on $\mathscr{G}$ is in Q_j at step j, for all $1 \leqslant j \leqslant \ell$, is at most*

$$\prod_{i=1}^{\ell-1}\left(\sqrt{\mu_i\mu_{i+1}} + \|T\|\right).$$

Proof. Let χ_{Q_i} be the characteristic function of Q_i, and A_i be the endomorphism of $L^2(V)$ defined by $f \mapsto \chi_{Q_i} f$. Let Π denote the projection onto the constants, so that $T\Pi = \Pi$. The endomorphism $A_i.T.A_{i+1}$ may be decomposed:

$$A_i T A_{i+1} = A_i.\Pi.A_{i+1} + A_i T(1-\Pi)A_{i+1}.$$

The endomorphism $A_i.\Pi.A_{i+1}$ may be written as $f \mapsto \frac{\chi_{Q_i}}{|V|}\langle f, \chi_{Q_{i+1}}\rangle$, and thus has operator norm $\mu_i^{1/2}\mu_{i+1}^{1/2}$. Since the operator norm of $A_i.T(1-\Pi).A_{i+1}$ is at most $\|T\|$, we conclude that the operator norm of $A_i T A_{i+1}$ is at most $\|T\| + (\mu_i\mu_{i+1})^{1/2}$. The probability that a random walk visits Q_i at step i for every $i \in \{1,\dots,\ell\}$ is given by

$$|V|^{-1}\langle 1_V, (A_1 T A_2)(A_2 T A_3)\dots(A_{\ell-1} T A_\ell)1_V\rangle$$

and the result follows. □

Lemma 9.1.2. *Let Q be a subset of V, with $\mu := |Q|/n$, and let γ be a random walk of length ℓ. Then the probability that $|\gamma \cap Q| \geqslant (\mu+\epsilon)\ell$ is at most*

$$c_1 \exp(-c_2\ell)$$

for positive constants c_1, c_2 depending only on $d, \|T\|, \mu, \epsilon$.

In other words, the number of "bad" walks of length ℓ, with respect to some fixed notion of "bad", decays exponentially with ℓ.

Proof. The constants $C_1, C_2, \dots$ appearing in the proof are all understood to be positive constants depending only on $d, \|T\|, \mu, \epsilon$.

Let S be a subset of $\{1..\ell\}$ of size k. It follows from Lemma 9.1.1 that the probability that $\gamma_i \in Q$ for $i \in S$ and $\gamma_i \notin Q$ for $i \notin S$ is at most

$$C_1 \mu^k (1-\mu)^{\ell-k}\big(1 + \frac{\|T\|}{\min(\mu, 1-\mu)}\big)^\ell.$$

Summing over all choices of S we have that the probability that $|\gamma \cap Q| = k$ is at most

$$C_1 \binom{\ell}{k} \mu^k (1-\mu)^{\ell-k}\big(1 + \frac{\|T\|}{\min(\mu, 1-\mu)}\big)^\ell$$

On the other hand, the sum

$$\sum_{k \geqslant (\mu+\epsilon)\ell} \binom{\ell}{k} \mu^k (1-\mu)^{\ell-k}$$

is at most $\exp(-C_2\ell)$. Thus, we are done if $\|T\|$ is small enough that $(1 + \frac{\|T\|}{\min(\mu,1-\mu)}) < e^{C_2}$.

If this is not the case, we fix an integer $C_3 \geqslant 1$, and replace the graph $\mathscr{G}$ by the graph $\mathscr{G}^{(C_3)}$ for which a directed edge from x to y corresponds to a directed path of length C_3 in the graph $\mathscr{G}$. (Note that $\mathscr{G}^{(C_3)}$ may have multiple edges and loops.) This improves the spectral gap: if $T^{(C_3)}$ is

the normalized adjacency matrix of $\mathscr{G}^{(C_3)}$, we have $T^{(C_3)} = T^{C_3}$. Accordingly, $\|T^{(C_3)}\| = \|T\|^{C_3}$. Choosing C_3 large enough, the argument above shows that the probability that a random walk of length ℓ on the graph $\mathscr{G}^{(C_3)}$ of remains within Q for at least $(\mu+\epsilon)\ell$ steps is bounded above by $C_4 \exp(-C_5\ell)$.

It follows immediately that the probability that a random walk on $\mathscr{G}$ of length $C_3\ell$ spends time $\geqslant (\mu+\epsilon)C_3\ell$ inside Q is at most $C_4\exp(-C_5\ell)$.

This proves the desired claim. □

9.2. **The arc graph.** Lemma 9.1.2, which tells us that an exponentially small proportion of random walks are poorly distributed in $\mathscr{G}$, will not quite suffice for our purposes; what we need to know is that an exponentially small proportion of *non-backtracking* walks are poorly distributed in $\mathscr{G}$. In this section we explain how to derive such a statement from Lemma 9.1.2.

We now assume $\mathscr{G}$ to be *symmetric*; that is, E is closed under reversal. For each edge $a \in E$, write a^+ for the target of a and a^- for the source of a. We denote the reversal of a by $\bar{a}$. With these notations, define the arc graph $\mathscr{G}'$ to be a directed graph whose vertices are the directed edges, or arcs, of $\mathscr{G}$. There is an edge from the arc a to the arc b exactly when $a^+ = b^-$ and $a \neq \bar{b}$.

Thus, $\mathscr{G}'$ is regular of degree $d-1$. We denote by T' the normalized adjacency matrix of $\mathscr{G}'$:

$$T'F(a) = \frac{1}{d-1} \sum_{\substack{b^- = a^+ \\ b \neq \bar{a}}} F(b)$$

The key feature of the arc graph $\mathscr{G}'$, for us, is that we have a natural bijection between non-backtracking paths of length ℓ on $\mathscr{G}$, and paths of length $\ell - 1$ on $\mathscr{G}'$.

9.3. Our goal will be to deduce a spectral gap for T' from that for T. This is a simple analogue of Atkin-Lehner theory in the subject of modular forms: when $d = p+1$ for some prime p, we can think of $\mathscr{G}$ as a quotient of the Bruhat-Tits tree attached to $\mathrm{PGL}_2(\mathbb{Q}_p)/\mathrm{PGL}_2(\mathbb{Z}_p)$; then the arc graph of the Bruhat-Tits tree is obtained by replacing $\mathrm{PGL}_2(\mathbb{Z}_p)$ with an Iwahori subgroup, so that the passage from graph to arc graph is much the same as the passage from the congruence subgroup $\Gamma_0(N)$ to the smaller subgroup $\Gamma_0(Np)$.

There are natural maps $B, E : L^2(\mathscr{G}) \to L^2(\mathscr{G}')$ ("beginning" and "end") defined via

$$Bf(a) = f(a^-),\ Ef(a) = f(a^+).$$

Moreover, the orthogonal complement to $\mathrm{Im}(B) \oplus \mathrm{Im}(E)$ consists of those functions $F \in L^2(\mathscr{G}')$ with the property that

$$\sum_{a^- = v} F(a) = \sum_{a^+ = v} F(a) = 0,$$

for all $v \in V\mathscr{G}$. On this orthogonal complement (the "new space") , the operator T' acts via

$$F \mapsto -\frac{1}{d-1}\bar{F}, \tag{9.1}$$

where $\bar{F}(a) = F(\bar{a})$. Moreover, one checks that

$$\langle Bf_1, Ef_2\rangle = d\langle Tf_1, f_2\rangle, \quad T' \circ B = E.$$

Thus, if $w \in L^2(\mathscr{G})$ is an eigenfunction for T with eigenvalue λ, then the "old space" $\mathbb{C}(Bw)+\mathbb{C}(Ew)$ is stable under T'. From this we see that every eigenvalue of T' on this space is also an eigenvalue of the matrix :

$$\begin{pmatrix} 0 & 1 \\ \frac{-1}{(d-1)} & \frac{d\lambda}{(d-1)} \end{pmatrix}$$

It is easily computed that the eigenvalues are bounded away from 1 if λ is. By (9.1), the eigenvalues of T' on the new space are bounded in absolute value by $1/(d-1) < 1$. We conclude that $\|T'\|$ is bounded away from 1 if $\|T\|$ is.

Proposition 9.4. *Let $\mathscr{G} = (V, E)$ be an undirected graph with $\|T\| < 1$, and let Q be a subset of V, with $\mu := |Q|/n$. The probability that a random walk* without backtracking *of length ℓ spends more than $(\mu+\epsilon)\ell$ time in Q is at most*

$$c_1 \exp(-c_2\ell)$$

for constants $c_1, c_2 > 0$ depending only on $d, \|T\|, \mu, \epsilon$.

Proof. Let $Q' \subset V\mathscr{G}'$ be the subset of arcs whose initial vertex lies inside Q. Noting that $\frac{|Q'|}{|\mathscr{G}'|} = \frac{|Q|}{|V|}$, we apply Lemma 9.1.2 to $(\mathscr{G}', Q')$ taking into account that $\|T'\|$ is bounded away from 1 in terms of $\|T\|$. □

Part 5. Related topics

10. Miscellanea and extensions of the ergodic method.

In this section we comment on various modifications that can be made in the assumptions or the proofs.

10.1. **The squarefreeness assumption.** Through this paper, we have assumed that d is *squarefree*. This is mainly for simplifying the exposition. It is true in general that some quotient of $\mathscr{H}_d$ by a subgroup of $\mathrm{SO}_3(\mathbb{Z})$ is acted on by the ideal class group of some quadratic order containing $\sqrt{-d}$. However, when d is not squarefree this action is not necessarily homogeneous: there may be *several orbits*. This has to do with the existence of *non-primitive elements in $\mathscr{H}_d$*. An element $\mathbf{x} = (a, b, c) \in \mathscr{H}_d$ is primitive iff the integers a, b, c are coprime; when d is squarefree, every element of $\mathscr{H}_d$ is primitive. On the other hand, if we choose d of the form $d_0 f^2$, the subset $f^2\mathscr{H}_{d_0} \subset \mathscr{H}_d$ forms a separate orbit of the class group action. The appropriate repair is to consider just the set of primitive elements $\mathscr{H}_d^{\mathrm{p}} \subset \mathscr{H}_d$; these

elements again form a homogeneous space for a certain class group. By the methods of this paper one can then show the equidistribution of $\mathrm{red}_q(\mathscr{H}_d^{\mathrm{P}})$ or $\mathrm{red}_\infty(\mathscr{H}_d^{\mathrm{P}})$ as $d \to \infty$. Using the decomposition

$$\mathscr{H}_d = \bigsqcup_{f^2|d} f.\mathscr{H}_{d/f^2}^{\mathrm{P}}$$

one can conclude that $\mathrm{red}_q(\mathscr{H}_d)$ and $\mathrm{red}_\infty(\mathscr{H}_d)$ are indeed equidistributed.

10.2. **Distribution of trajectories.** One of the important ingredients for the proof of Theorem 1.4 has been to consider the *trajectories* associated to the various points in $\mathscr{H}_d$ and to show that these trajectories are, in a suitable sense, "well-spaced".

The principle of lifting a point to its trajectory and then of studying the distribution of the trajectories occurs in several recent applications of ergodic methods to number theory. For instance, a (much less evident) version of this principle occurs in the proof of the arithmetic quantum unique ergodicity conjecture, in which the role of the lifted trajectories is played by the "microlocal lifts" [Lin06].

In fact it is possible to extrapolate the above principle to showing that *segments of trajectories of bounded length* become equidistributed. For instance, in the context of Theorem 1.3, one can prove the following result:

Theorem. *Under the assumptions of Theorem 1.3, fix $\ell_0 > 0$, and let $\gamma_0^{(\ell_0)}$ be a fixed non-backtracking marked path on $\mathscr{H}_d(q)$ of length $2\ell_0+1$; then the cardinality of the set of $\mathbf{x} \in \mathscr{H}_d$ for which*

$$\gamma_{\mathbf{x}}^{(\ell_0)} = \gamma_0^{(\ell_0)}$$

is equal to

$$\frac{1}{3 \cdot 5^{2\ell_0-1}|\mathscr{H}_d(q)|}|\mathscr{H}_d|(1+o(1)),\ d \to \infty.$$

(Note that $3 \cdot 5^{2\ell_0-1}|\mathscr{H}_d(q)|$ is the total number of non-backtracking paths on $\mathscr{H}_d(q)$ of length $2\ell_0+1$.)

In combination with Proposition 2.10 this implies that $\mathscr{H}_d$ is well-distributed 5-adically, as well as mod q. This may be seen as the analogue of Duke's theorem on the equidistribution of closed geodesics on the *unit tangent bundle of* the modular surface $X_0(1)$ (cf. [Duk88][Theorem 1]). See [ELMV09b] for a purely dynamical proof of the latter, following principles similar to the ones of the present paper (although in a different language). This theorem also admits a variant in the context of Theorem 1.1 and uniform versions analogous to Theorems 1.2 and 1.4.

10.3. **Application to mixing.** As an application of the uniformity reached in Theorems 1.1 and 1.4, we have the following mixing type results:

Corollary. *Fix $\delta > 0$. Let q be a fixed integer coprime with 30. For each squarefree $d \equiv \pm 1 (5)$ and coprime with q, let $\mathfrak{p}_d$ be a primitive*[7] *ideal of $\mathcal{O}_K$ such that $\mathrm{N}(\mathfrak{p}_d) < d^{1/2-\delta}$; we suppose moreover that as $d \to +\infty$, $\mathrm{N}(\mathfrak{p}_d) \to +\infty$. Then, for $d \to \infty$ the set*

$$\{\mathrm{red}_\infty(\tilde{\mathbf{x}}, [\mathfrak{p}_d].\tilde{\mathbf{x}}),\ \tilde{\mathbf{x}} \in \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d\} \subset \mathrm{SO}_3(\mathbb{Z})\backslash S^2 \times \mathrm{SO}_3(\mathbb{Z})\backslash S^2 \tag{10.1}$$

becomes equidistributed on $\mathrm{SO}_3(\mathbb{Z})\backslash S^2 \times \mathrm{SO}_3(\mathbb{Z})\backslash S^2$ with respect to the product of Lebesgue measures on each factor. Similarly, the multiset

$$\{\mathrm{red}_q(\tilde{\mathbf{x}}, [\mathfrak{p}_d].\tilde{\mathbf{x}}),\ \tilde{\mathbf{x}} \in \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d\} \subset \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q) \times \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q) \tag{10.2}$$

becomes equidistributed on $\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q) \times \mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q)$ with respect to the counting measure.

Notice that if the norm $\mathrm{N}(\mathfrak{p}_d)$ remains constant for d varying (possibly over a suitable subsequence), a simple variant of Theorem 1.2 above shows that the set (10.1) (resp. (10.2)) becomes equidistributed along some codimension two subvariety of $(\mathrm{SO}_3(\mathbb{Z})\backslash S^2)^2$ (resp. a multiset of order $\approx \mathrm{N}(\mathfrak{p}_d)q^2$ in $(\mathrm{SO}_3(\mathbb{Z})\backslash\mathscr{H}_d(q))^2$), namely the graph of the $\mathrm{N}(\mathfrak{p}_d)$-th Hecke correspondance. This explains the condition $\mathrm{N}(\mathfrak{p}_d) \to \infty$.

In [MV07], the second and third authors. proposed a "mixing conjecture": namely the quoted statement should remain true so long as $\mathrm{N}(\mathfrak{p}_d)$ tends to ∞ as $d \to \infty$. Not surprisingly, the range *not* covered by the corollary (i.e. $d^{1/2-\varepsilon(d)} \leqslant \mathrm{N}(\mathfrak{p}_d) \ll d^{1/2}\log d$ for $0 < \varepsilon(d) \to 0$) is by far the most interesting range of the conjecture; for instance, the applications sketched in [MV07] depend crucially on this range.

10.4. **Extensions to other quadratic forms.** For the *indefinite* discriminant quadratic form $x^2 - yz$, results analogous to Linnik's theorems have been obtained by Skubenko using the ergodic method (see [Lin68, Chap VI] and the references inside). In [ELMV09b], the second and third authors together with M. Einsiedler and E. Lindenstrauss gave an alternative exposition of the ergodic approach which in fact allows for stronger results.

More generally, Linnik's ergodic method carries over when one replaces the quadratic form $x^2+y^2+z^2$ defined over $\mathbb{Q}$ by any non-degenerate ternary quadratic form Q defined over a fixed number field F. For instance, suitable generalizations of Theorems 1.1 and 1.3 over number fields have been obtained by Teterin [Tet83]. All the main theorems of the present text remain valid when suitably adapted to this more general context. In this setting, the condition $d \equiv \pm 1$ modulo 5 is to be replaced by the existence of a fixed prime ideal in F which *splits* in $F(\sqrt{-d\,\mathrm{disc}(Q)})$.

10.5. **The discriminant aspect.** Theorems 1.4 and 1.2 give results that have nearly optimal uniformity in the modulus q or in the radius of the ball ρ. However, the ternary form Q whose representations of integers we study is fixed as $Q(x, y, z) = x^2 + y^2 + z^2$. It would be interesting to investigate

[7] A primitive ideal is an integral ideal of minimal norm in its ideal class.

the option of proving analogues of these results where Q is allowed to vary, with the aim of obtaining optimal uniformity in the discrimant $\mathrm{disc}(Q)$. In particular, if the form is definite, its genus grows on the order of $\mathrm{disc}(Q)$. One would envisage an analogue of Theorem 1.4 that establishes that *almost every* $Q' \in \mathrm{genus}(Q)$ represents d (if everywhere locally representable), so long as $|\mathrm{genus}(Q)| \ll d^{1/2-\delta}$.

One example where this can be carried out completely is when Q is the reduced norm on the space of trace zero elements in the quaternion algebra which is ramified at ∞ and at some prime $q > 2$: then $|\mathrm{genus}(Q)| \sim \frac{q}{12}$. In that case the equidistribution results may be interpreted in terms of supersingular reduction of CM elliptic curves (cf. the paper of D. Gross[Gro87]):

Let $\mathcal{E}_q^{ss}$ denote the set of isomorphism classes of supersingular elliptic curves over $\overline{\mathbb{F}}_q$; these are in fact defined over $\mathbb{F}_{q^2}$. Then $\mathcal{E}_q^{ss}$ is canonically identified with $\mathrm{genus}(Q)$ and so has size $\sim q/12$. For K an imaginary quadratic field in which q is inert, let $\mathcal{E}_{\mathscr{O}_K}$ denote the set of elliptic curves in characteristic 0 with complex multiplication by $\mathscr{O}_K$; these curves are defined over the Hilbert class field of K, and

$$|\mathcal{E}_{\mathscr{O}_K}| = |\mathrm{disc}(K)|^{1/2+o(1)}.$$

For $\mathfrak{q}$, a place in $\overline{\mathbb{Q}}$ above p, we have a "reduction mod $\mathfrak{q}$"-map

$$\mathrm{red}_{\mathfrak{q}} : \mathcal{E}_{\mathscr{O}_K} \mapsto \mathcal{E}_q^{ss}.$$

For any $\overline{E} \in \mathcal{E}_q^{ss}$ define

$$\mathrm{dev}_d(\overline{E}) = \frac{1}{\mu(\overline{E})} \frac{|\mathrm{red}_{\mathfrak{q}}^{-1}(\overline{E})|}{|\mathcal{E}_{\mathscr{O}_K}|} - 1$$

where

$$\mu(\overline{E}) = \frac{1/|\mathrm{Aut}(\overline{E})|}{\sum_{\overline{E}' \in \mathcal{E}_q^{ss}} 1/|\mathrm{Aut}(\overline{E}')|};$$

μ defines a probability measure on $\mathcal{E}_q^{ss}$.

An analog of Theorem 1.4 in this context is the following:

Theorem 10.6 (Equidistribution of "Grauss" points)**.** *Fix $\delta, \eta > 0$. Let q be as above, and let p, be a fixed auxiliary prime with $(p, 2q) = 1$. Let K range over the imaginary quadratic fields such that*

1. *q is inert in K,*
2. *p is split in K.*

Set $d = |\mathrm{disc}(K)|$. As long as $q \leqslant d^{1/2-\delta}$, the μ-measure[8] of $\overline{E} \in \mathcal{E}_q^{ss}$ such that

$$|\mathrm{dev}_d(\overline{E})| > \eta$$

tends to 0 as $d \to +\infty$.

[8] one could replace "the μ-measure of" by "the fraction of" since $1 \leqslant \mathrm{Aut}(\overline{E}) \leqslant 12$

The proof is similar to that of Theorem 1.4; what is needed is the analog of Corollary 2.12. It is proven by a suitable generalization of the geometric part of the Gross formulas [Gro87, Prop. 10.8]. As noted in [Mic04], using the formulas of Gross and subconvex bounds for L-functions, one can prove a version of this result, with μ-measure 0 and without the condition (2), as long as q is less than a small positive power of d; under GRH this may be further extended to $q \leqslant d^{1/4-\delta}$.

10.7. **Representation of quadratic forms by quadratic forms.** As we briefly discussed in §4 the problem of studying the representation of an integer by a quadratic form admits another sort of generalization. Namely, one can ask about representations of a varying integral quadratic form of rank m by another, fixed, quadratic form of higher rank $n \geqslant m$. When $m = 1$, $n = 3$, and the rank-3 form is $x^2 + y^2 + z^2$, we are in the situation of the present paper. The general problem basically splits into two questions:

- the existence of such representations;
- how these are distributed (when they exist) in real or p-adic topologies.

As is explained in [EV08], both the existence and the distribution problem can be understood in terms of the distribution properties of adelic orbits of some algebraic subgroup H inside an adelic homogeneous space associated to some ambient algebraic group $G \supset H$. More precisely, we can take G to be a form of SO_n and H a form of SO_{n-m}.

The work [EV08] considered cases for which $n-m \geqslant 3$. In that case, since noncompact forms of SO_{n-m} have plenty of unipotents, one can apply the results of M. Ratner to show that (under suitable technical hypotheses) the adelic orbit of the smaller group equidistributes in the adelic quotient of the larger group; this suffices to prove the existence of such representations. The second question, regarding equidistribution, was not addressed in [EV08] but likely follows from similar methods.

This is in sharp contrast with the present situation where $n = 3$, $m = 1$. In this case, the smaller group is a torus (a form of SO_2); in particular, Ratner's theorems do not apply, and one needs to use the special methods described in the present paper. It should be noted, moreover, that we are helped by the fact that $\mathrm{SO}_2 \subset \mathrm{SO}_3$ is a *maximal* torus; it seems that the problem of studying representations of, say, forms of rank 2 by a given form of rank 4, would be significantly harder.

11. Harmonic analysis and Linnik's method.

11.1. **Duke's approach via harmonic analysis.** Although our main concern is with extending Linnik's method, we briefly remark on Duke's method and the relevance of modular L-functions to our problem.

As we have remarked, the condition "$d \equiv \pm 1$ modulo 5" has the effect of ensuring that 5 splits in $\mathbb{Q}(\sqrt{-d})$; one could replace the role of 5 by any

other fixed prime, as did Linnik. The problem of removing such a constraint, however, proved difficult. This question, i.e.:

(11.1) Can we guarantee the existence of a small p with $\left(\frac{-d}{p}\right) = 1$?

is a very difficult and deep one, for it is intimately connected to the issue of the Landau-Siegel "exceptional" zero, i.e. to effective lower bounds for the class number of $\mathbb{Q}(\sqrt{-d})$. For instance, observe that proving that there exists a prime p as in (11.1) with $p \leqslant d^\epsilon$ would immediately show that the class number of $\mathbb{Q}(\sqrt{-d})$ tends to ∞ as $d \to \infty$; itself, far from a trivial result.

It was therefore a considerable breakthrough when W. Duke[9] [Duk88, DSP90] established Theorem 1.3 and Theorem 1.1 without the auxiliary assumption. Duke's original approach used the Maass-Shimura theta correspondence to express the Weyl sums attached to these equidistribution problems, i.e.

$$W(\varphi; d) = \frac{1}{|\mathscr{H}_d|} \sum_{\mathbf{x} \in \mathscr{H}_d} \varphi(\overline{\mathbf{x}})$$

for φ is a (smooth) function on S^2 (resp. $\mathscr{H}_d(q)$) with zero mean with respect to the Lebesgue (resp. the counting) measure, in terms of Fourier coefficients of some half-integral weight holomorphic form $\theta(\varphi)$ [Wal81]; indeed, $\theta(\varphi)$ is a classical θ-series. The decay of the Weyl sums then followed from non-trivial bounds for these Fourier coefficients, which were obtained using in a key way ideas and results of H. Iwaniec [Iwa87].

11.2. **Central values of automorphic L-function: Waldspurger's formula.** It was quickly realized that these problems are also related to L-functions and to the problem of bounding them non-trivially at the central point: if φ is an eigenfunction of (almost all) the Hecke operators[10], acting on S^2 (resp. $\mathscr{H}_d(q)$), then by a formula of Waldspurger [Wal85] one has

$$\frac{|W(\varphi; d)|^2}{\langle \varphi, \varphi \rangle} = c \frac{d^{1/2}}{|\mathscr{H}_d|^2} \frac{L(\pi, 1/2) L(\pi \otimes \chi_d, 1/2)}{L(\pi, \mathrm{Ad}, 1/2)} I_\infty(\varphi; d) \prod_{p | 6q} I_p(\varphi; d) \tag{11.2}$$

where

- $\langle\ ,\ \rangle$ is the inner product associated to the Lebesgue (resp. counting) probability measure on S^2 (resp. $\mathscr{H}_d(q)$) and $c > 0$ is an absolute explicit constant;
- π is the automorphic representation of PGL_2 in Jacquet-Langlands correspondence with the automorphic representation of $\mathrm{SO}_3 \cong \mathrm{PB}^\times$ generated by φ; π is a discrete series representation (i.e. corresponding to an holomorphic modular form) of conductor dividing $18q^2$.
- χ_d is the Kronecker symbol associated with the quadratic field $\mathbb{Q}(\sqrt{-d})$;

[9] partly in collaboration with Schulze-Pillot

[10] This is not a restriction, since the space of functions on S^2 or $\mathscr{H}_d(q)$ is generated by such Hecke eigenforms.

- the factors $I_\infty(\varphi;d)$, $I_p(\varphi;d)$, $p|6q$ are some local integrals (real or p-adic) which can be bounded explicitly in terms of φ but *independently* of d.

From this formula, one sees that the decay of $W(\varphi;d)$ as $d\to\infty$ is a consequence of the so-called *subconvex* bound: there is some absolute $\delta>0$

$$L(\pi\otimes\chi_d,1/2)\ll_\pi d^{1/2-\delta}. \tag{11.3}$$

Such a bound was proven, for the first time, in [Iwa87]. A few years later, a more general bound (with $1/2$ replaced by $1/2+it$) was obtained by Duke, Friedlander, Iwaniec by an entirely different method [DFI93] (again without any condition on d).

Remark. The method of Duke has been generalized by Duke and Schulze-Pillot to general ternary quadratic forms over $\mathbb{Q}$ [DSP90]. More recently, the alternative approach involving Waldspurger's formula (11.2) has been extended to treat ternary forms over general number fields, in particular, for totally definite ternary forms over totally real number fields: this has been carried out by Cogdell, Piatetsky-Shapiro and Sarnak [Cog03]; in this generality the required subconvex bound may be found in [Ven05]. For more comments about this we refer to [MV06] and to [BH09] where a complete exposition of these arguments is given along with sharp form of (11.3).

11.3. **Ergodic Theory vs. Harmonic Analysis.** Now let us compare what the ergodic and harmonic-analysis techniques yield, as regards to Theorem 1.4 and 1.2. It is possible to show that the product of the local integrals $I_\infty(\varphi;d)\prod_{p|6q}I_p(\varphi;d)$ is bounded in terms of some Sobolev norm on φ. Moreover, the dependence on π in the bound (11.3) is polynomial in $Q(\pi)$, the analytic conductor of π (as defined by Iwaniec and Sarnak [IS00]). From this one can deduce, without any condition on d that, for *every* $\overline{\mathbf{x}}$ (resp. every $x\in S^2$), $|\mathrm{dev}_d(\overline{\mathbf{x}})|$ (resp. $|\mathrm{dev}_d(\Omega(x,\rho))|$) is bounded by a small negative power of d as long as q (resp. ρ^{-1}) is less than a (small) positive power of d [Blo04, Duk05, GF87]. As we explain below, assuming the *Generalized Riemann Hypothesis* (GRH), this result holds when q (resp. ρ^{-1}) is less than $\leqslant d^{1/8-\nu}$ for any fixed $\nu>0$.

We consider for simplicity the case of $\mathscr{H}_d(q)$. Let $\mathcal{B}$ denote an orthogonal basis in the space of functions on $\mathscr{H}_d(q)$. The average

$$\sum_{\varphi\in\mathcal{B}}\frac{|\sum_{\mathbf{x}\in\mathscr{H}_d}\varphi(\overline{\mathbf{x}})|^2}{\langle\varphi,\varphi\rangle}$$

is independent of $\mathcal{B}$. Therefore, (picking as a basis $\{\delta_{\overline{\mathbf{x}}},\ \overline{\mathbf{x}}\in\mathscr{H}_d(q)\}$) such a sum equals

$$|\mathscr{H}_d(q)|\sum_{\overline{\mathbf{x}}\in\mathscr{H}_d(q)}\left|\mathrm{red}_q^{-1}(\overline{\mathbf{x}})\right|^2.$$

Taking $\mathcal{B}$ of the form $\{1\} \cup \mathcal{B}_0$ where $\mathcal{B}_0$ is constituted of Hecke eigenforms, we obtain

$$\sum_{\varphi \in \mathcal{B}_0} \frac{|\sum_{\mathbf{x} \in \mathscr{H}_d} \varphi(\overline{\mathbf{x}})|^2}{\langle \varphi, \varphi \rangle} = |\mathscr{H}_d(q)| \sum_{\overline{\mathbf{x}} \in \mathscr{H}_d(q)} \big(\mathrm{red}_q^{-1}(\overline{\mathbf{x}}) - \frac{|\mathscr{H}_d|}{|\mathscr{H}_d(q)|}\big)^2$$
$$= \frac{|\mathscr{H}_d|^2}{|\mathscr{H}_d(q)|} \sum_{\overline{\mathbf{x}}} \mathrm{dev}_d(\overline{\mathbf{x}})^2.$$

Hence by (11.2),

$$\frac{|\mathscr{H}_d|^2}{|\mathscr{H}_d(q)|} \sum_{\overline{\mathbf{x}}} \mathrm{dev}_d(\overline{\mathbf{x}})^2 = c d^{1/2} \sum_{\varphi \in \mathcal{B}_0} \frac{L(\pi, 1/2) L(\pi \otimes \chi_d, 1/2)}{L(\pi, \mathrm{Ad}, 1/2)} I_\infty(\varphi; d) \prod_{p | 6q} I_p(\varphi; d).$$

The Generalized Riemann Hypothesis and a local computation[11], shows that for any $\varepsilon > 0$

$$\sum_{\overline{\mathbf{x}}} \mathrm{dev}_d(\overline{\mathbf{x}})^2 \ll_\varepsilon (qd)^\varepsilon q^4 / d^{1/2}. \tag{11.4}$$

Linnik's basic Lemma alone shows the same result with an additional factor $(1 + d/q^4)$ on the right-hand side; in particular, (11.4) is known unconditionally for $q^4 > d$. Now (11.4) shows, for $\nu > 0$, that:

- $\mathrm{dev}_d(\overline{\mathbf{x}}) \ll d^{-\nu}$ for any $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ as long as $q \leqslant d^{1/8-\nu}$ (thus proving the surjectivity of the reduction map red_q in this range);
- given $\delta > 0$, the fraction of $\overline{\mathbf{x}} \in \mathscr{H}_d(q)$ for which $|\mathrm{dev}_d(\overline{\mathbf{x}})| > \delta$ is bounded by $\ll_\delta d^{-4\nu}$ as long as $q \leqslant d^{1/4-\nu-\epsilon}$.

In particular we recover Theorem 1.4. On the other hand we don't see at the moment, even under GRH, how to prove[12] that there are *no* $\overline{\mathbf{x}}$ for which $\mathrm{red}_q^{-1}(\overline{\mathbf{x}}) = \emptyset$ in the range $d^{1/8} \leqslant q \leqslant d^{1/4-\nu}$.

Remark 11.1. It should be noted that (11.4), obtained by pointwise application of GRH, is essentially a "sharp on average" bound for the family $\mathcal{F}$ of L-values

$$\{L(\pi, 1/2) L(\pi \otimes \chi_d, 1/2),\ \pi \text{ of conductor } 18q^2 \text{ and trivial central character}\}.$$

The size $|\mathcal{F}|$ of that family is $\approx q^2$ while the conductor C of any of these L-functions is of size $\approx q^4 d^2$. In particular, in the range $q^4 \approx d$, we have $|\mathcal{F}| \approx |C|^{1/6}$. Starting with the original work of Weyl, there are, by now, several examples of families of L-functions satisfying a similar relation between the size of the family and the size of the conductor and for which the central values could be bounded sharply on average without any hypothesis [Wey21, CI00, Li08].

[11] to bound the local integrals $I_p(\varphi; d)$

[12] This type of situation is not uncommon; for instance, it is known that the GRH implies that the smallest prime in an arithmetic progression with modulus q is $\leqslant c(\varepsilon) q^{2+\varepsilon}$; while we surely expect the result with $2 + \varepsilon$ replaced by $1 + \varepsilon$, no proof of this is known even under GRH.

References

[Blo04] V. Blomer, *Uniform bounds for Fourier coefficients of theta-series with arithmetic applications*, Acta Arith. **114** (2004), no. 1, 1–21.

[BH09] V. Blomer and G. Harcos, *The spectral decomposition of shifted convolution sums*, GAFA (to appear) (2009). `http://arxiv.org/abs/0904.2429`.

[BM66] A. Borel and G. Mostow Edts., *Proceedings of Symposia in Pure Mathematics. Vol. IX: Algebraic groups and discontinuous subgroups*, Proceedings of the Symposium in Pure Mathematics of the American Mathematical Society held at the University of Colorado, Boulder, Colorado (July 5-August 6, vol. 1965, American Mathematical Society, Providence, R.I., 1966.

[Cog03] J. W. Cogdell, *On sums of three squares*, J. Théor. Nombres Bordeaux **15** (2003), no. 1, 33–44 (English, with English and French summaries). Les XXIIèmes Journées Arithmetiques (Lille, 2001).

[CI00] J. B. Conrey and H. Iwaniec, *The cubic moment of central values of automorphic L-functions*, Ann. of Math. (2) **151** (2000), no. 3, 1175–1216.

[Duk88] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. **92** (1988), no. 1, 73–90.

[Duk05] W. Duke, *On ternary quadratic forms*, J. Number Theory **110** (2005), no. 1, 37–43.

[DFI93] W. Duke, J. Friedlander, and H. Iwaniec, *Bounds for automorphic L-functions*, Invent. Math. **112** (1993), no. 1, 1–8.

[DSP90] W. Duke and R. Schulze-Pillot, *Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids*, Invent. Math. **99** (1990), no. 1, 49–57.

[ELMV07] M. Einsiedler, E. Lindenstrauss, Ph. Michel, and A. Venkatesh, *The distribution of periodic torus orbits on homogeneous spaces : Duke's theorem for cubic fields*, (to appear) (2007). `http://arxiv.org/abs/0708.1113`.

[ELMV09a] M. Einsiedler, E. Lindenstrauss, Ph. Michel, and A. Venkatesh, *The distribution of periodic torus orbits on homogeneous spaces*, Duke Math. Journal **148** (2009), no. 1, 119-174. `http://arxiv.org/abs/math/0607815`.

[ELMV09b] M. Einsiedler, E. Lindenstrauss, Ph. Michel, and A. Venkatesh, *The distribution of periodic torus orbits on homogeneous spaces : Duke's theorem for quadratic fields*, Preprint (2009).

[EV08] J. S. Ellenberg and A. Venkatesh, *Local-global principles for representations of quadratic forms*, Invent. Math. **171** (2008), no. 2, 257–279.

[GF87] E. P. Golubeva and O. M. Fomenko, *Asymptotic distribution of lattice points on the three-dimensional sphere*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **160** (1987), no. Anal. Teor. Chisel i Teor. Funktsii. 8, 54–71, 297 (Russian); English transl., J. Soviet Math. **52** (1990), no. 3, 3036–3048.

[Gau01] C. F. Gauss, *Disquisitiones arithmeticae*, Springer-Verlag, New York, 1801. Translated and with a preface by Arthur A. Clarke; Revised by William C. Waterhouse, Cornelius Greither and A. W. Grootendorst and with a preface by Waterhouse.

[Gro87] B. H. Gross, *Heights and the special values of L-series*, Number theory (Montreal, Que., 1985), CMS Conf. Proc., vol. 7, Amer. Math. Soc., Providence, RI, 1987, pp. 115–187.

[HLW06] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561 (electronic).

[Iwa87] H. Iwaniec, *Fourier coefficients of modular forms of half-integral weight*, Invent. Math. **87** (1987), no. 2, 385–401.

[IS00] H. Iwaniec and P. Sarnak, *Perspectives on the analytic theory of L-functions*, Geom. Funct. Anal. **Special Volume** (2000), 705–741. GAFA 2000 (Tel Aviv, 1999).

[JL70] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Springer-Verlag, Berlin, 1970. Lecture Notes in Mathematics, Vol. 114.

[Jon50] B. W. Jones, *The Arithmetic Theory of Quadratic Forms.*, Carus Monograph Series, vol. no. 10., The Mathematical Association of America, Buffalo, N. Y., 1950.

[Kna97] A. W. Knapp, *Introduction to the Langlands program*, Representation theory and automorphic forms (Edinburgh, 1996), Proc. Sympos. Pure Math., vol. 61, Amer. Math. Soc., Providence, RI, 1997, pp. 245–302.

[Li08] X. Li, *Bounds for $GL(3) \times GL(2)$ L-functions and $GL(3)$ L-functions*, Ann. of Math. (2008). (to appear).

[Lin06] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) **163** (2006), no. 1, 165–219.

[Lin68] Yu. V. Linnik, *Ergodic properties of algebraic fields*, Translated from the Russian by M. S. Keane. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 45, Springer-Verlag New York Inc., New York, 1968.

[Lub94] A. Lubotzky, *Discrete groups, expanding graphs and invariant measures*, Progress in Mathematics, vol. 125, Birkhäuser Verlag, Basel, 1994. With an appendix by Jonathan D. Rogawski.

[LPS86] A. Lubotzky, R. Phillips, and P. Sarnak, *Hecke operators and distributing points on the sphere. I*, Comm. Pure Appl. Math. **39** (1986), no. S, suppl., S149–S186. Frontiers of the mathematical sciences: 1985 (New York, 1985).

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), no. 3, 261–277.

[Mal84] A. V. Malyshev, *Discrete ergodic method and its applications to the arithmetic of ternary quadratic forms*, Topics in classical number theory, Vol. I, II (Budapest, 1981), Colloq. Math. Soc. János Bolyai, vol. 34, North-Holland, Amsterdam, 1984, pp. 1023–1049.

[Mic04] Ph. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. (2) **160** (2004), no. 1, 185–236.

[MV06] Ph. Michel and A. Venkatesh, *Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik*, International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, pp. 421–457.

[MV07] Ph. Michel and A. Venkatesh, *Heegner points and non-vanishing of Rankin/Selberg L-functions*, Analytic number theory, Clay Math. Proc., vol. 7, Amer. Math. Soc., Providence, RI, 2007, pp. 169–183.

[Pal49] G. Pall, *Representation by quadratic forms*, Canadian J. Math. **1** (1949), 344–364.

[PR94] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., Boston, MA, 1994. Translated from the 1991 Russian original by Rachel Rowen.

[Sam70] P. Samuel, *Algebraic theory of numbers*, Translated from the French by Allan J. Silberger, Houghton Mifflin Co., Boston, Mass., 1970.

[Ser73] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York, 1973. Translated from the French; Graduate Texts in Mathematics, No. 7.

[Ser03] J.-P. Serre, *Trees*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003. Translated from the French original by John Stillwell; Corrected 2nd printing of the 1980 English translation.

[She86] T. R. Shemanske, *Representations of ternary quadratic forms and the class number of imaginary quadratic fields*, Pacific J. Math. **122** (1986), no. 1, 223–250.

[Tet83] Yu. G. Teterin, *Representation of algebraic integers by ternary quadratic forms*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **121** (1983), 157–168 (Russian). Studies in number theory, 8.

[Ven05] A. Venkatesh, *Sparse equidistribution problems, period bounds, and subconvexity.*, Ann. of Math. (to appear) (2005). `http://arxiv.org/abs/math/0506224`.

[Ven22] B. A. Venkov, *On the arithmetic of quaternion algebras.*, Izv. Akad. Nauk (1922), 205–241.

[Ven29] B. A. Venkov, *On the arithmetic of quaternion algebras.*, Izv. Akad. Nauk (1929), 489–509, 532–562, 607–622.

[Vig80] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics, vol. 800, Springer, Berlin, 1980 (French).

[Wal81] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier*, J. Math. Pures Appl. (9) **60** (1981), no. 4, 375–484 (French).

[Wal85] J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie*, Compositio Math. **54** (1985), no. 2, 173–242 (French).

[Wey21] H. Weyl, *Zur Abschätzung von* $\zeta(1+it)$, Math. Zeit. **10** (1921), 88-101.