EFFECTIVE ESTIMATES ON INTEGRAL QUADRATIC FORMS: MASSER'S CONJECTURE, GENERATORS OF ORTHOGONAL GROUPS, AND BOUNDS IN REDUCTION THEORY

HAN LI AND GREGORY A. MARGULIS

Abstract. In this paper we prove a conjecture of David Masser on small height integral equivalence between integral quadratic forms. Using our resolution of Masser's conjecture we show that integral orthogonal groups are generated by small elements which is essentially an effective version of Siegel's theorem on the finite generation of these groups. We also obtain new estimates on reduction theory and representation theory of integral quadratic forms. Our line of attack is to make and exploit the connections between certain problems about quadratic forms and group actions, whence we may study the problem in the well-developed theory of homogeneous dynamics, arithmetic groups, and the spectral theory of automorphic forms.

1. Introduction

1.1. Effective search bounds and Masser's conjecture. A classical problem in the theory of quadratic forms is to decide whether two given integral quadratic forms are equivalent. By definition a quadratic form Q_A is integral if all the entries of A, the symmetric matrix of Q_A, are integers. Two integral quadratic forms Q_A and Q_B are said to be equivalent if there exists $\tau \in \mathbf{GL}_n(\mathbb{Z})$ such that $A = \tau^t B \tau$. Like many of the classical subjects in number theory, quadratic forms had a computational beginning. So for the early pioneers of the subject it was a natural question to ask if there exists a procedure, that can be preformed in a reasonable number of steps, to determine if two quadratic forms are equivalent. After all, one doesn't willingly begin a computation which they don't know will end.

Over the years people managed to prove that one can find such procedure for special classes of quadratic forms. However, no such procedure was known for all integral quadratic forms until the early 1970s when Siegel established an effective search bound [57]. Siegel showed that for any $n \geq 2$ there exists a non-negative function F_n such that if the non-singular, symmetric, $n \times n$ integral matrices A, B are equivalent, then there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ with

$$A = \gamma^t B \gamma, \qquad \|\gamma\| < F_n(A, B) \tag{1}$$

where $\|\cdot\|$, throughout the paper, is taken to be the maximum of the absolute values of all the entries. The function F_n was made explicit in the late 1990s by Straumann [58], who followed Siegel's argument [57] and showed that F_n can be taken as

$$F_n(A, B) = \exp(C_n |\det A|^{\frac{n^3+n^2}{2}}) \cdot \max(\|A\|, \|B\|)^{\frac{n^3-n^2}{2}}$$

where $C_n > 0$ is an absolute constant. The estimate (1) is an example of an effective search bound. Effective search bounds are useful because they give

(i) an effective procedure of deciding the equivalence of two given integral quadratic forms;
(ii) an effective procedure of finding an integral transformation between two given equivalent forms;
(iii) an estimate on the computational complexity of these procedure;
(iv) an estimate on the norm of an integral transformation between two given equivalent quadratic forms.

The procedure in (i) and (ii), which relies on search bounds, may not be efficient for practical use. But some good algorithms are known. For example, Conway-Sloane [10, Chapter 15.11] sketched an algorithm to

2010 *Mathematics Subject Classification.* Primary: 11E12; Secondary: 37A25.
Key words and phrases. integral quadratic forms; integral orthogonal groups; reduction theory of quadratic forms.
H. Li was supported in part by an AMS Simons Travel Grant.
G. A. Margulis was supported in part by NSF Grant #1265695.

decide whether two given integral quadratic forms are in the same spinor genus (see also [7, 3]). In particular, this applies to (i) for indefinite forms in at least three variables because in this case the spinor genera and the equivalence classes of integral quadratic forms coincide. They also mentioned there does not seem to be good algorithms for (ii). As of this writing, a list of the algorithms related to quadratic forms can be found on the website of Magma (http://magma.maths.usyd.edu.au/magma/overview).

Our main concern in this paper is to establish new and improved effective search bounds and to explore their applications to related problems. Since $|\det A|$ is in general comparable to $\|A\|$, Siegel's bound is in principle exponential in $\|A\|$ and $\|B\|$. It is natural to ask what is the optimal search bound? This question has an easy answer if the forms are definite. Indeed, if Q_B is definite and $A = \gamma^t B\gamma$, then every column of γ has to lie in the ellipsoid defined by $\{\boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{x}^t B\boldsymbol{x}| \le \|A\|\}$. By an eigenvalue argument (see e.g. [33, Lemma 10]) one gets that

$$\|\gamma\| < 2 \cdot (4n)^{\frac{n-1}{2}} \cdot \|A\|^{\frac{1}{2}} \cdot \|B\|^{\frac{n-1}{2}}. \tag{2}$$

Apparently, the method above does not apply to indefinite forms as the region $\{\boldsymbol{x} \in \mathbb{R}^n : |\boldsymbol{x}^t B\boldsymbol{x}| \le \|A\|\}$ is generally unbounded. Thus the difficulty of this problem lies in indefinite integral quadratic forms which we are going to focus on in the sequel. For $n = 2$, the exponential search bounds are sharp due to the large solutions of certain Pell's equations (see, for instance, [51, 34, 15]). For $n \ge 3$, in the late 1990s David Masser proposed the following conjecture on polynomial search bounds after Dietmann's proof for ternary forms (see also Fukshansky's note [21] for a discussion). As a notational remark, the set of $k \times l$ integral matrices will be denoted as $\mathbb{Z}^{k\times l}$, throughout the paper.

Conjecture 1. *(Masser* [41, p252]*) For any $n \ge 3$ there exist constants C_n, $\kappa_n > 0$ satisfying the following property. Let A, $B \in \mathbb{Z}^{n\times n}$ be symmetric, non-singular matrices, and suppose $A = \gamma_0^t B\gamma_0$ for some $\gamma_0 \in \mathbf{GL}_n(\mathbb{Z})$. Then there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ with $A = \gamma^t B\gamma$ and $\|\gamma\| < C_n \cdot (\|A\| + \|B\|)^{\kappa_n}$.*

Dietmann's work on ternary forms was sketched in Masser's survey [41], and his theorem appears in a later published paper [14, Theorem 4]. Unfortunately, Dietmann's method, which involves the circle method and also some ideas in arithmetic geometry, is not readily generalized to handle the forms in at least four variables. Nevertheless, he managed to prove Conjecture 1 assuming $\det B$ is cubic free, not divisible by four, and that not all coefficients on the diagonal of B are even [15, Theorem 3]. These assumptions let him use an induction argument based on his work for ternary forms. So for a large class of forms Masser's conjecture was already confirmed. Our first result is a complete resolution of Conjecture 1.

Theorem 1. *For any $n \ge 3$ there exists a constant $C_n > 0$ satisfying the following property. Let A, $B \in \mathbb{Z}^{n\times n}$ be symmetric, non-singular matrices, and suppose $A = \gamma_0^t B\gamma_0$ for some $\gamma_0 \in \mathbf{GL}_n(\mathbb{Z})$. Then there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ such that $\det\gamma = \det\gamma_0$, $A = \gamma^t B\gamma$, and that*

$$\|\gamma\| < C_n \cdot |\det B|^{-\frac{13n^3-39n^2+36n+52}{20n}} \cdot \|A\|^{\frac{13n^3-13n^2+52n+20}{40}} \cdot \|B\|^{\frac{13n^3-13n^2+72n-20}{40}}. \tag{3}$$

Moreover, if the integral quadratic form Q_B is indefinite and anisotropic (that is, does not represent the number 0 nontrivially), then there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ such that $\det\gamma = \det\gamma_0$, $A = \gamma^t B\gamma$, and that

$$\|\gamma\| < C_n \cdot |\det B|^{\frac{13n^2+8n-26}{10n}} \cdot \|A\|^{\frac{1}{2}} \cdot \|B\|^{\frac{n-1}{2}}. \tag{4}$$

The theorem implies that the κ_n in Conjecture 1 can be chosen to grow polynomially in n. On the other hand, it is not hard to show that κ_n should grow at least linearly in n (cf. [33, Lemma 10]).

We now explain how the proof of Theorem 1 is reduced to a problem on the quantitative behavior of the translates in G/Γ, where $G = \mathbf{SL}_n(\mathbb{R})$ and $\Gamma = \mathbf{SL}_n(\mathbb{Z})$. Let (p,q) be the signature of the integral quadratic forms Q_A and Q_B in question which we assume $\gamma_0 \in \Gamma$ to simplify the discussion. Let $X_0 = \mathrm{diag}(1,\cdots,1,-1,\cdots,-1)$ be of signature (p,q). It is easy to show that $A = |\det A|^{\frac{1}{n}} g_A^t X_0 g_A$ and $B =$

$|\det B|^{\frac{1}{n}} g_B^t X_0 g_B$ for some g_A, $g_B \in G$ whose norms are controlled by a polynomial of $\|A\|$ and $\|B\|$, respectively. Since A and B are equivalent, there exists $h \in H = \{h \in G : h^t X_0 h = X_0\}$ such that $g_B^{-1} h g_A \in \Gamma$. But for any $h \in H$ the matrix $g_B^{-1} h g_A$ transforms B into A. From this we see that to prove Theorem 1 it suffices to give a bound on the norm of an element $h \in H$ with $g_B^{-1} h g_A \in \Gamma$, in terms of a polynomial in $\|A\|$ and $\|B\|$. Another way of saying $g_B^{-1} h g_A \in \Gamma$ is that $h.g_A\Gamma = g_B\Gamma$ with $g_A\Gamma$ and $g_B\Gamma$ viewed as two elements in G/Γ. As we will see in Section 2, the orbit $Hg_A\Gamma$ is closed in G/Γ because the quadratic forms are integral. The proof of Theorem 1 is thus reduced to the following problem.

Problem 1. *In the above notation, let $H.x$ be a closed orbit in G/Γ. Suppose that the elements g_1, $g_2 \in G$ satisfy $g_1\Gamma$, $g_2\Gamma \in H.x$. Give a bound on the norm of an element $h \in H$ with $h.g_1\Gamma = g_2\Gamma$, in terms of a polynomial in $\|g_1\|$ and $\|g_2\|$.*

The study of Problem 1 involves the ideas and techniques from the theory of homogeneous dynamics, arithmetic groups and the spectral theory of automorphic representations. We will mention some of the ingredients in (iv) and (v) of Section 1.6 after more background is introduced.

1.2. Small generators of integral orthogonal groups. In a landmark paper [56], Siegel showed that for any integral quadratic form Q_B in $n \geq 3$ variables the integral orthogonal group $\mathbf{O}_{Q_B}(\mathbb{Z}) = \{\gamma \in \mathbf{GL}_n(\mathbb{Z}) : B = \gamma^t B \gamma\}$ is finitely generated. Our next result is essentially an effective generalization of Siegel's theorem.

Theorem 2. *For any $n \geq 3$ there exists $C_n > 0$ satisfying the following property. Let Q_B be an integral quadratic form in n-variables. Then $\mathbf{O}_{Q_B}(\mathbb{Z}) = \{\gamma \in \mathbf{GL}_n(\mathbb{Z}) : \gamma^t B \gamma = B\}$ is generated by the finite subset $\{\gamma \in \mathbf{O}_{Q_B}(\mathbb{Z}) : \|\gamma\| < C_n \cdot |\det B|^{n^6} \cdot \|B\|^{3n^4}\}$.*

In [22] Grunewald-Segal sketched an algorithm to compute a finite set of generators for arithmetic groups. As a byproduct of the estimate, Theorem 2 provides an alternative procedure of computing generators for integral orthogonal groups. Indeed, for given B, among all the integral matrices whose norms do not exceed $C_n |\det B|^{n^6} \|B\|^{3n^4}$ one eliminates those γ which does not satisfy the matrix equation $\gamma^t B \gamma = B$. Then the matrices remained will generate the group $\mathbf{O}_{Q_B}(\mathbb{Z})$, by the theorem.

To put this result in perspective, we note that if the non-singular, symmetric matrices A, $B \in \mathbb{Z}^{n\times n}$ and the unimodular matrix $\tau_1 \in \mathbf{GL}_n(\mathbb{Z})$ satisfy $A = \tau_1^t B \tau_1$, then $\{\tau \in \mathbf{GL}_n(\mathbb{Z}) : A = \tau^t B \tau\} = \mathbf{O}_{Q_B}(\mathbb{Z})\tau_1$. Hence the integral orthogonal group is closely related to the set of the transformations between two equivalent forms. Historically, the structure of these sets has been receiving a considerable attention since people started to consider the equivalence of integral quadratic forms.

The proof of Theorem 2 involves an argument of Borel-Harish-Chandra [4, Theorem 6.5] by which they proved the finite generation of arithmetic groups. We will give bounds on the norm of the exhibited generators. The main ingredients for achieving this goal are Theorem 1 together with some quantitative results in reduction theory (Corollary 2, Corollary 3). It is worth mentioning that the Grunewald-Segal algorithm is also largely based on making effective some of the ideas and arguments of Borel-Harish-Chandra.

1.3. Height bounds on reduction of integral quadratic forms. A general problem in reduction theory is to what extent can one simplify a given integral quadratic form by taking an equivalent form. To study this problem, one typically compares the largest coefficients of quadratic forms with the determinant, an invariant among equivalent quadratic forms. The next theorem is a result of this type.

Theorem 3. *Let $n \geq 3$. Let $A \in \mathbb{Z}^{n\times n}$ be symmetric, non-singular, and suppose that the quadratic form Q_A is indefinite. Then, there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ such that $\|\gamma^t A \gamma\| < C_n \cdot |\det A|^{\frac{1}{n}}$.*

Our approach is based on the theory of quantitative non-divergence of unipotent flows which is different from the classical methods in reduction theory. The theory of non-divergence of unipotent flows originates in the work of G. M. [37], and is further developed by Dani [12], Kleinbock-Margulis [30], and others.

We note that the estimate in Theorem 3 carries the best possible exponent of $|\det A|$, since $\|\gamma^t A\gamma\| \geq (n!)^{-\frac{1}{n}} |\det(\gamma^t A\gamma)|^{\frac{1}{n}} = (n!)^{-\frac{1}{n}} |\det A|^{\frac{1}{n}}$. Let us now summarize some bounds in reduction theory of integral quadratic forms. On one hand, as a consequence of Gauss' reduction theory, any indefinite binary integral quadratic form Q_A is equivalent to a form Q_B with $\|B\| \ll |\det A|^{\frac{1}{2}}$ (see, for instance, [59, Chapter 4]); On the other hand, any positive definite integral quadratic form Q_A is equivalent to a form Q_B with $\|B\| \ll_n |\det A|$. This follows from a theorem of Legendre for binary forms, and from the Hermite-Minkowski reduction theory ([7, p.287]) in general. For later use let us record the following corollary.

Corollary 1. *For any $n \geq 2$ there exists a constant $C_n > 0$ satisfying the following property. Let $A \in \mathbb{Z}^{n\times n}$ be symmetric and non-singular. Then, there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ such that $\|\gamma^t A\gamma\| < C_n \cdot |\det A|$.*

Once again, we note that the estimate in Corollary 1 carries the best possible exponent of $|\det A|$ because any integral quadratic form Q_B, that is equivalent to $Q_A = x_1^2 + \cdots + x_{n-1}^2 + dx_n^2$ $(d > 0)$, must satisfy $\|B\| \geq d = |\det A|$. For $n = 3$ the corollary follows from Gauss' first and second reduction procedure of ternary forms (see [59, p.146]). For $n \geq 4$ an approach of Cassels [7, p.135] would give any n-ary integral quadratic form Q_A an equivalent form Q_B with $\|B\| \ll_n |\det A|^{\alpha_n}$. The method is by induction and the exponent α_n is recursively defined as a function of n, and satisfies $\alpha_n > 1$. To our knowledge, Theorem 3 is new for any $n \geq 3$. We are not certain whether Corollary 1 was previously known for any $n \geq 4$.

1.4. Representations of integral quadratic forms. Let $A \in \mathbb{Z}^{m\times m}$, $B \in \mathbb{Z}^{n\times n}$ be symmetric integral matrices where $n \geq 2$, $n \geq m \geq 1$. By definition, the quadratic form Q_B represents Q_A if the matrix equation $A = \tau^t B\tau$ has a solution $\tau \in \mathbb{Z}^{n\times m}$. The representation is called *primitive* if the solution $\tau \in \mathbb{Z}^{n\times m}$ is primitive. That is, if the greatest common divisor of the maximal minors of τ is equal to 1. For $n \geq 3$, our next theorem addresses the polynomial search bounds for representations and primitive representations of integral quadratic forms. For special classes of quadratic forms some polynomial search bounds are known, and we will briefly remark later in Section 8. Let us note that the assumption $n \geq 3$ is necessary because when $n = 2$ the exponential bounds are sharp for either $m = 1$ ([51, 34]) or $m = 2$ ([14]).

Theorem 4. *For any $n \geq 3$ there exists a constant $C_n > 0$ satisfying the following property. Let $A \in \mathbb{Z}^{m\times m}$, $B \in \mathbb{Z}^{n\times n}$ be symmetric, non-singular, and suppose $A = \tau_0^t B\tau_0$ for some $\tau_0 \in \mathbb{Z}^{n\times m}$. Then there exists $\tau \in \mathbb{Z}^{n\times m}$ such that*

$$A = \tau^t B\tau, \qquad \|\tau\| < C_n \cdot |\det A|^{n^4} \cdot |\det B|^{n^3} \cdot \|A\|^{n^3} \cdot \|B\|^{n^3}.$$

Moreover, if τ_0 is primitive, then τ can be chosen to be primitive as well.

The proof of the theorem uses Theorem 1 and Lemma 20 which is a quantitative version of a result Siegel in reduction theory.

1.5. An overview of the method. Our approach is based in a large part on homogeneous dynamics. Theorem 3, for instance, is proved almost exclusively via homogeneous dynamics. We must emphasize that from a dynamical point of view there is not much novelty in our work. The well-developed theory in homogeneous dynamics serves as a tool for solving the problems which arises in our investigation of integral quadratic forms rather than being the main subject studied in this paper.

In recent years there has been much work in quadratic forms which involve an approach of homogeneous dynamics. These include the works on the distribution of the values of indefinite irrational real quadratic

forms at integral points [38, 13, 19], and the proof of a local-global principle of representations of integral quadratic forms [18], to just mention a few.

We will also use a number of deep results in the automorphic representation theory related to orthogonal groups, such as the Jacquet-Langlands correspondence [27], the Burger-Sarnak restriction principle [5] and the Kim-Sarnak bound [28] on the Ramanujan conjecture of $\mathbf{SL}_2/\mathbb{Q}$. They will be used, in Section 3, to give an explicit spectral gap of the actions of $\mathrm{SO}(p,q)^\circ$ on its closed orbits in $\mathbf{SL}_n(\mathbb{R})/\Lambda$ where Λ is a congruence subgroup of $\mathbf{SL}_n(\mathbb{Z})$. Moreover, the spectral gap is uniform, that is to say, the gap is independent of the congruence subgroup Λ and also independent of the closed orbits in the quotient $\mathbf{SL}_n(\mathbb{R})/\Lambda$. As we will see, the uniform spectral gap crucially contributes to the polynomial rates in many effective estimates proved in this paper, including our main result Theorem 1.

The theory of arithmetic groups is also essential in our approach. It helps us utilize the underlying geometric structures of the arithmetic quotients such as the injectivity radius. Our work on Theorem 2 is also inspired by the ideas of Borel-Harish-Chandra in their proof of the finite generation of arithmetic groups.

Besides these tools, we will also use the ideas and techniques from the geometry of numbers and the classical reduction theory of quadratic forms at various stages of our investigation.

1.6. The structure of the paper and a summary of the main results.

(i) Section 2 and Section 3 are technical preparation. From Section 3, except Section 6, each section begins with a subsection called "statement of results" for the convenience of the readers.

(ii) In Section 2 we set up the dynamical background. The crucial fact is that each equivalence class of indefinite integral quadratic forms of signature (p,q) corresponds to a closed orbit of $H = \mathrm{SO}(p,q)^\circ$ in $X_n = \mathbf{SL}_n(\mathbb{R})/\mathbf{SL}_n(\mathbb{Z})$; and as a homogeneous space each closed orbit is a quotient of H by an arithmetic subgroup. In this setting, as we discussed, the proof of Conjecture 1 reduces to the study of Problem 1. The notation in Section 2 will be used throughout the paper, often without further reference.

(iii) In Section 3 we verify Lemma 2, a result on the uniform spectral gap which was mentioned before.

(iv) In Section 4 we make our first step to approach Problem 1. The main result in this section, Theorem 5, can be interpreted as for any two given points on a closed H-orbit in X_n, one point is translated to the other by an element $h \in H$ with $\|h\|$ controlled by the the injectivity radii of the two points together with the spectral gap and the volume of the orbit.

(v) In Section 5 we give two applications of quantitative non-divergence of unipotent flows. The first result to be settled in Section 5 is Theorem 3. The second result, Theorem 6, is our last step to approach Problem 1. Recall that the volume of a closed H-orbit is bounded by a polynomial in the determinant of the corresponding integral quadratic forms (see [36, §2.6]). In Theorem 6 we give an *explicit* exponent to this estimate. It seems this relation can also be studied via a volume formula of Prasad [49]. But certain intricate local consideration is likely to get involved. Instead, we present a proof of dynamical flavor.

(vi) In Section 6 we settle Theorem 1, or essentially Problem 1, based on Theorem 5 and Theorem 6.

(vii) In Section 7 we consider a strengthening of Conjecture 1 which involves certain congruence condition. This was suggested though not explicitly stated in Masser's survey [41]. To this end, let Q_B be an integral quadratic form, $\Lambda = \{\gamma \in \mathbf{SL}_n(\mathbb{Z}) : B = \gamma^t B \gamma\}$, and Λ_N be the principle congruence subgroup of Λ_N of level N. We show that each coset of Λ/Λ_N contains an element whose norm is controlled by N and $\|B\|$ polynomially. The proof is based on the geometric fact that Λ/Λ_N embeds into $\mathbf{SO}_{Q_B}(\mathbb{R})/\Lambda_N$, and the injectivity radii of the points in the image of the embedding are not too small. This allows us to use the spectral gap of the quotient $\mathbf{SO}_{Q_B}(\mathbb{R})/\Lambda_N$ to finish the proof.

(viii) In Section 8 we prove Theorem 4; and in Section 9 we settle Theorem 2.

1.7. The computability the constants in the statement of our results. We will explicitly evaluate every exponent in this paper. For the multiple constants we will use $x \ll_n y$ to represent an inequality $x < Cy$ in which the multiple constant C depends only on n. For example, $\|g^{-1}\| \ll_n |\det g|^{-1}\|g\|^{n-1}$ can stand for $\|g^{-1}\| \le (n-1)!|\det g|^{-1}\|g\|^{n-1}$ (by Cramer's rule) where $g \in \mathbf{GL}_n(\mathbb{R})$. This fact is subject to frequent use in the sequel. To evaluate the constant R_n in Lemma 14, one needs to exhibit a compact subset satisfying Lemma 13. This requires extra work and we will carry it out in the appendix. Other than that the reader should have no difficulty evaluating other multiple constants in terms of the dimension n directly from each step if she or he would so desire.

1.8. Acknowlegement. It was Wai Kiu Chan who suggested us study representations of quadratic forms besides the equivalence problem. This inspired us to work further and prove Theorem 4. We would like to express our gratitude to him for his helpful conversation and his insightful comments on the early drafts of the paper. We wish to thank David Masser for his question on the computability of the constants in the statement of our main results. We are also grateful to Daniel Allcock, Emmanuel Breuillard, Rainer Dietmann, Roger Howe, John Hsia, Alex Kontorovich, Jian-Shu Li, Amir Mohammadi, Rainer Schultz-Pillot, Alan Reid and Jiu-Kang Yu for stimulating discussions, and especially to Hee Oh for pointing out some inaccuracy which occurred in the preliminary draft and for her help on references. H. L. is deeply indebted to Michael Kelly for his detailed comments which significantly improved the exposition of this paper.

2. The dynamical setting and preliminary facts

2.1. An easy lemma. Throughout the paper, we will use standard terminology and notation. A quadratic form Q is defined by a non-singular symmetric matrix A through $Q(\boldsymbol{v}) = \boldsymbol{v}^t A\boldsymbol{v}$, in which case we will be free to use Q_A for Q, and $\det Q$ for $\det A$. The quadratic form Q_A is called integral, rational, or real, if all the entries of A are in $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, respectively. We shall start with an easy lemma in linear algebra.

Lemma 1. *Let the symmetric real matrices $X, Y \in \mathbb{R}^{n\times n}$ be such that $\det X = \det Y \neq 0$, and that the signature of the quadratic forms Q_X, Q_Y are the same. Then there exists $g \in \mathbf{SL}_n(\mathbb{R})$ satisfying*

$$X = g^t Y g, \qquad \|g\| \ll_n |\det Y|^{-\frac{1}{2}}\|X\|^{\frac{1}{2}}\|Y\|^{\frac{n-1}{2}}.$$

Proof. By Cramer's rule, $\|Y^{-1}\| \ll_n |\det Y|^{-1}\|Y\|^{n-1}$. The lemma is obvious when X, Y are diagonal matrices. For the general cases, we note that any real symmetric matrix can be diagonalized by an orthogonal matrix, and that the norm of any orthogonal matrix is less than or equal to 1. □

2.2. The dynamical setting for indefinite forms. In this section we set up the notation for further discussion. The congruence of integral matrices is always understood entrywise. We fix, once and for all, the positive integers n and N with $n \ge 3$. Set $G = \mathbf{SL}_n(\mathbb{R})$, $\Gamma = \mathbf{SL}_n(\mathbb{Z})$,

$$1_n = \text{the identity matrix of degree } n, \qquad \Gamma_N = \{\gamma \in \Gamma : \gamma \equiv 1_n \mod (N)\}.$$

The left coset $g\Gamma_N \in G/\Gamma_N$ will be denoted by $[g]_N$. For $\Gamma_1 = \Gamma$ we will abbreviate $[g]_1$ as $[g]$.

Let $Q = Q_X$ be a quadratic form in n-variables, and let $g \in \mathbf{GL}_n$. The quadratic form Q_{g^tXg} will be denoted by $Q \circ g$. The special orthogonal group of Q is the algebraic group $\mathbf{SO}_Q := \{h \in \mathbf{SL}_n : Q \circ h = Q\} \subset \mathbf{SL}_n$. The group $\mathbf{SO}_Q$ is realized as an algebraic subgroup of $\mathbf{SL}_n$ through the embedding, and is defined over the fields which all the coefficients of Q belong to. The readers are referred to [39, Chapter I] for the background materials on algebraic groups and their arithmetic subgroups.

Example 1. *(see e.g. [39, p.64]) Q be a rational quadratic form in n-variables. Then, $\mathbf{SO}_Q$ is a semisimple $\mathbb{Q}$-group; the subgroup $\mathbf{SO}_Q(\mathbb{R}) \cap \mathbf{SL}_n(\mathbb{Z})$ is a lattice in $\mathbf{SO}_Q(\mathbb{R})$; and the lattice is cocompact in $\mathbf{SO}_Q(\mathbb{R})$ if and only if Q does not represent the number 0 non-trivially over $\mathbb{Q}$.*

In the sequel, p and q are fixed integers with $p \geq 1$, $q \geq 2$ and $n = p + q$. The column vector $\boldsymbol{v} \in \mathbb{R}^n$ will be written as $\boldsymbol{v} = (v_1, \cdots, v_p, w_1, \cdots, w_q)^T$. We shall make use of the following integral quadratic form

$$Q_0(\boldsymbol{v}) = v_1^2 + \cdots + v_p^2 - w_1^2 - w_2^2 - \cdots - w_q^2 \quad (\boldsymbol{v} \in \mathbb{R}^n) \tag{5}$$

of signature (p, q). We let $H = \mathbf{SO}_{Q_0}(\mathbb{R})^\circ$ denote the identity component of $\mathbf{SO}_{Q_0}(\mathbb{R})$. The group H is also the identity component of the orthogonal group $\mathbf{O}_{Q_0}(\mathbb{R}) = \{g \in \mathbf{GL}_n(\mathbb{R}) : Q_0 \circ g = Q_0\}$ which has four connected components corresponding to the four components of its maximal compact subgroup $\mathrm{O}(p) \times \mathrm{O}(q)$ (see [26, p42-43]). Since the quotient group $\mathbf{O}_{Q_0}(\mathbb{R})/H$ is isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, we have if $u, s \in \mathbf{O}_{Q_0}(\mathbb{R})$ are in the same connected component, then their product us lies in the identity component H.

Let Q be a real quadratic form of signature (p, q). By Lemma 1 there exists $g \in G$ satisfying $Q = |\det Q|^{\frac{1}{n}}(Q_0 \circ g)$ and hence the identity component H_Q of $\mathbf{SO}_Q(\mathbb{R})$ satisfies $H_Q = g^{-1}Hg$. Clearly H_Q does not depend on the choice of g. As a closed subgroup of G, the group H acts on G/Γ_N by left translations. For any point $[g]_N \in G/\Gamma_N$ we have the natural bijections between the homogenous spaces

$$H.[g]_N \longleftrightarrow H/(H \cap g\Gamma_N g^{-1}) \longleftrightarrow H_Q/(H_Q \cap \Gamma_N) \tag{6}$$

which is explicitly given by $h.[g]_N \longleftrightarrow h.(H \cap g\Gamma_N g^{-1}) \longleftrightarrow g^{-1}hg.(H_Q \cap \Gamma_N)$.

2.3. Integral quadratic forms and closed orbits in homogeneous spaces. Let Q, g and H_Q be as in the preceding paragraph with the quadratic form Q being integral. As a finite index subgroup of $H_Q \cap \Gamma$, the discrete subgroup $H_Q \cap \Gamma_N$ is a lattice in $H_Q = g^{-1}Hg$ by Example 1. This, together with the relation (6), implies that the orbit $H.[g]_N \subseteq G/\Gamma_N$ supports an H-invariant probability measure and hence is closed in G/Γ_N. In fact, any closed H-orbit in G/Γ_N arises in this way. Indeed, let $H.[g]_N$ be a closed orbit in G/Γ. Since H is semisimple the discrete subgroup $H \cap g\Gamma_N g^{-1}$, which is the stabilizer of $[g]_N$, is a lattice of H ([40, p.420]). From this and the relation (6) we get that $H_Q \cap \Gamma_N$ is a lattice in H_Q. By Borel's density theorem the quadratic form $Q_0 \circ g$ is proportional to an integral quadratic form ([38, p.400]).

Recall that a rational quadratic form is called *isotropic* if it represents the number 0 nontrivially and *anisotropic* otherwise. In view of Example 1 and the relation (6), the orbit $H.[g]_N \subseteq G/\Gamma_N$ is compact if and only if $Q_0 \circ g$ is proportional to an indefinite anisotropic integral quadratic form, which only exists when $n \leq 4$ by Meyer's theorem. Hence there are no compact H-orbits in G/Γ_N for any $n \geq 5$.

2.4. Invariant measures. In this section we shall fix the Borel measures on G, H, and their homogeneous spaces. These measures are chosen in nice coordinates so that certain multiple constants can in principle be explicitly evaluated. For the group G, we let m_G stand for the Haar measure which satisfies

$$|\det(x_{ij})|^{-n} \bigwedge_{i,j} \mathrm{d}x_{ij} = \mathrm{d}m_G(g)\frac{\mathrm{d}t}{t}, \qquad \text{where } (x_{ij}) = (t^{\frac{1}{n}}1_n) \cdot g \in \mathbf{GL}_n(\mathbb{R}),\ g \in G. \tag{7}$$

We denote by $m_{G/\Gamma}$ the G-invariant measure on G/Γ induced by m_G. Recall that [16, Appendix]

$$m_{G/\Gamma}(G/\Gamma) = \zeta(2) \cdots \zeta(n). \tag{8}$$

For the group H, the Haar measure is chosen as follows. First we identify $\mathfrak{gl}_n(\mathbb{R}) = \mathrm{Lie}(\mathbf{GL}_n(\mathbb{R}))$ with the space $\mathbb{R}^{n^2}$ endowed by the Euclidean inner product. The inner product restricts to $\mathfrak{h} = \mathrm{Lie}(H)$ and induces a Lebesgue measure $\mathrm{d}X$ on $\mathfrak{h}$. We specify the measure m_H to be the Haar measure so that near the identity element it is given in exponential coordinates by (see [20, p.89])

$$\mathrm{d}m_H(h) = \det\Big(\frac{\mathrm{id}_{\mathfrak{h}} - \exp(-\mathrm{ad}_{\mathfrak{h}}X)}{\mathrm{ad}_{\mathfrak{h}}X}\Big)\mathrm{d}X, \qquad \text{where } h = \exp(X). \tag{9}$$

For any closed orbit $H.[g]_N$ in G/Γ_N, we will consider the two Borel measures which are supported on $H.[g]_N$. That is, the H-invariant probability measure $\mu_{H.[g]_N}$ (see Section 2.3) and the volume measure

which is defined as follows (cf. [17, §1.2]). Let $\mathcal{F} \subseteq H$ be a Borel fundamental domain (bijectively mapped onto $H.[g]_N$) of the lattice $H \cap g\Gamma_N g^{-1}$ in H. We define $(H.[g]_N, \mathrm{vol}_H)$ to be the pushforward measure of $(\mathcal{F}, m_H)$ with respect to the map $\mathcal{F} \to H.[g]_N$, $h \mapsto h.[g]_N$. It can be deduced from the definition that $\mathrm{vol}_H(H.[g]_N)$ equals the covolume of the lattice $H \cap g\Gamma_N g^{-1}$ in H with respect to m_H, and that for any Borel set $X \subseteq H.[g]_N$

$$\mathrm{vol}_H(X) = \mathrm{vol}_H(H.[g]_N) \cdot \mu_{H.[g]_N}(X). \tag{10}$$

3. A uniform spectral gap of the $\mathrm{SO}(p,q)^\circ$-actions on closed orbits

3.1. Statement of results. Let the notation be as in Section 2. Consider the linear subspace

$$V = \{(v_1, 0, \cdots, 0, w_1, w_2, 0, \cdots, 0) \in \mathbb{R}^n : v_1, w_1, w_2 \in \mathbb{R}\} \subseteq \mathbb{R}^n.$$

Let $V^\perp$ be the orthogonal complement of V for the symmetric bilinear form $\langle \boldsymbol{v}, \boldsymbol{w} \rangle_{Q_0} = \frac{1}{2}\{Q_0(\boldsymbol{v}+\boldsymbol{w}) - Q_0(\boldsymbol{v}) - Q_0(\boldsymbol{w})\}$ defined by Q_0. Let L be a copy of $\mathrm{SO}(1,2)^\circ$ in $H = \mathrm{SO}(p,q)^\circ$ given by

$$L := \{h \in H : h|_{V^\perp} = \mathrm{id}\}.$$

Our plan in this section is to prove the following lemma.

Lemma 2. *Let L be as above. For any natural number N and any closed orbit $H.[g]_N$ in G/Γ_N the representation of L on $L^2_0(H.[g]_N) = \{f \in L^2(H.[g]_N) : \int f d\mu_{H.[g]_N} = 0\}$ is strongly $L^{\frac{64}{25}+\epsilon}$.*

Lemma 2 essentially says the following. There exists a neighborhood U of the trivial representation in the unitary dual of H (with the Fell topology), such that for *any* natural number N and *any* closed orbit $H.[g]_N \subseteq G/\Gamma_N$ the action of H on $L^2_0(H.[g]_N)$ does not weakly contain any irreducible unitary representation in U. As we will see the strong integrability exponent comes from deep results in the spectral theory of automorphic forms. Assuming the Ramanujan conjecture for $\mathbf{SL}_2/\mathbb{Q}$, the condition $L^{\frac{64}{25}+\epsilon}$ in the lemma can be improved to $L^{2+\epsilon}$.

3.2. Background of unitary representation theory. Let S be a semisimple Lie group with finite center, and $\widehat{S}$ be the unitary dual (irreducible representations) of S with the Fell topology. The 1-dimensional trivial representation of S will be denoted as id_S. The reader is referred to [39, Chapter I §5] or [17, §6] for the general background. By representations of the Lie groups we always mean unitary representations.

Let $(\pi, \mathcal{H}_\pi)$ be a representation of S, and k be a positive number. We say that $(\pi, \mathcal{H}_\pi)$ is strongly L^k if there exists a dense subset $W \subseteq \mathcal{H}_\pi$ such that for each $v, w \in W$ the function $s \mapsto \langle \pi(s)v, w \rangle$ belongs to $L^k(S)$ with respect to the Haar measure on S. The representation $(\pi, \mathcal{H}_\pi)$ is said to be strongly $L^{k+\epsilon}$ if it is strongly $L^m(S)$ for every $m > k$. If S is a connected simple Lie group, then $(\pi, \mathcal{H}_\pi)$ is strongly L^k for some $k > 0$ if and only if π does not weakly contain id_S ([11], [44]). In the modern literature this property is usually referred as π *has a spectral gap*. A vector $v \in \mathcal{H}_\pi$ is called smooth if the orbit map $S \to \mathcal{H}_\pi$, $s \mapsto s.v$ is smooth. The Lie algebra $\mathfrak{s}$ of S acts on the subspace of the smooth vectors in $\mathcal{H}_\pi$.

In what follows we shall review the unitary representation theory of $\mathbf{SL}_2(\mathbb{R})$ ([26]). Write

$$g_t = \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix} \in \mathbf{SL}_2(\mathbb{R}), \qquad \mathcal{D} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathfrak{sl}_2(\mathbb{R}). \tag{11}$$

Lemma 3. *Let $2 \le k < \infty$, and $(\pi, \mathcal{H}_\pi)$ be a representation of $\mathbf{SL}_2(\mathbb{R})$. Then*

(1) π is strongly $L^{k+\epsilon}$ if and only if any $\rho \in \widehat{\mathbf{SL}_2(\mathbb{R})}$ that is weakly contained in π is strongly $L^{k+\epsilon}$.

(2) If $(\pi, \mathcal{H}_\pi)$ is strongly $L^{k+\epsilon}$, then for any smooth vectors v, $w \in \mathcal{H}_\pi$ and $t \in \mathbb{R}$ one has

$$\left|\langle \pi(g_t).v, w \rangle\right| \ll \Xi(g_t)^{\frac{2}{k}} \cdot \|\mathcal{D}.v\| \cdot \|\mathcal{D}.w\|, \tag{12}$$

where Ξ is the Harish-Chandra function of $\mathbf{SL}_2(\mathbb{R})$ (see [26, Chapter V]).

The results in Lemma 3 are well known. In particular the assertion (1) follows from [55, Theorem 2.1] and (2) can be derived using the method of [17, §6.2.1] from the asymptotic behavior of K-finite vectors which can also be found in [55, Theorem 2.1]. The absolute multiple constant in (12) can be explicitly computed using the condition 2 of [55, Theorem 2.1] and following the proof of [32, Proposition 3.2].

By definition a *covering* of Lie groups is a continuous surjective homomorphism with finite kernel. It will be helpful to keep in mind the following

Lemma 4. *Let $\iota : \mathbf{SL}_2(\mathbb{R}) \to F$ be a covering of Lie groups, and ρ be a unitary representation of F. Then ρ is strongly L^k if and only if $\rho \circ \iota$ is strongly L^k as a unitary representation of $\mathbf{SL}_2(\mathbb{R})$.*

Proof. As the kernel of ι is finite, the pushforward measure of a Haar measure on $\mathbf{SL}_2(\mathbb{R})$ is also a Haar measure on F. The lemma is obvious. □

3.3. Automorphic spectrum. Let $\mathbf{G}$ be a semisimple algebraic $\mathbb{Q}$-subgroup of $\mathbf{GL}_d$. Set $\mathbf{G}(\mathbb{Z}) = \mathbf{G}(\mathbb{Q}) \cap \mathbf{GL}_d(\mathbb{Z})$. Then $\mathbf{G}(\mathbb{Z})$ is a lattice in $\mathbf{G}(\mathbb{R})$ ([4]). We say that a lattice Σ in $\mathbf{G}(\mathbb{R})$ is congruence if there exists a natural number N such that Σ contains $\Lambda_N = \{\gamma \in \mathbf{G}(\mathbb{Z}) : \gamma \equiv 1_n \mod (N)\} < \mathbf{G}(\mathbb{Z})$, the principle congruence subgroup of level N. By definition the automorphic spectrum $\widehat{\mathbf{G}}^{\mathrm{Aut}}$ is the closure of the subset

$$\{\rho \in \widehat{\mathbf{G}(\mathbb{R})} : \ \rho \text{ is weakly contained in } L^2(\mathbf{G}(\mathbb{R})/\Sigma) \text{ for some congruence lattice } \Sigma \in \mathbf{G}(\mathbb{R})\}$$

with respect to the Fell topology. We shall point out that the definition of $\widehat{\mathbf{G}}^{\mathrm{Aut}}$ in [5] only involves the spectra of the quotients of the principle congruence subgroups. Since for any subgroup $\Sigma' < \Sigma$ of finite index one has $L^2(\mathbf{G}(\mathbb{R})/\Sigma) \subseteq L^2(\mathbf{G}(\mathbb{R})/\Sigma')$, our definition agrees with the one in [5]. The notion of the congruence lattices and the automorphic spectrum do not depend on the $\mathbb{Q}$-embeddings of $\mathbf{G}$ into linear groups even though the groups $\mathbf{G}(\mathbb{Z})$ and Λ_N do. The readers are referred to Sarnak's survey [50] for a summary on the progress towards the understanding of the automorphic spectrum.

Let f be a rational quadratic form in at least three variables. The spinor group $\mathbf{Spin}_f$ is a simply connected semisimple $\mathbb{Q}$-group which covers the orthogonal group $\mathbf{SO}_f$ ([7], [48]). We shall make use of the automorphic representations of the spinor groups.

Lemma 5. *Let f be an indefinite ternary rational quadratic form and let $F = \mathbf{Spin}_f(\mathbb{R})$. Then as a Lie group F is isomorphic to $\mathbf{SL}_2(\mathbb{R})$, and the automorphic spectrum of $\mathbf{Spin}_f$ satisfies*

$$\widehat{\mathbf{Spin}_f}^{\mathrm{Aut}} \subseteq \{\mathrm{id}_F\} \cup \{\pi \in \widehat{F} : \pi \text{ is strongly } L^{\frac{64}{25}+\epsilon}\}$$

Lemma 5, as is well known, follows from the Kim-Sarnak's bound on the Ramanujan conjecture for $\mathbf{SL}_2/\mathbb{Q}$ [28] and the Jacquet-Langlands correspondence [27], two deep results in the theory of automorphic representations. The next result, which is known as the Burger-Sarnak restriction principle, also plays an important role in understanding the automorphic spectrum.

Lemma 6. *([5, Theorem 1.1]) Let $\mathbf{G}$ be a connected semisimple algebraic $\mathbb{Q}$-group and $\mathbf{L}$ a semisimple $\mathbb{Q}$-subgroup of $\mathbf{G}$. Let $\pi \in \widehat{\mathbf{G}}^{\mathrm{Aut}}$ and suppose that $\rho \in \widehat{\mathbf{L}(\mathbb{R})}$ is weakly contained in the restriction of π to the closed subgroup $\mathbf{L}(\mathbb{R})$. Then $\rho \in \widehat{\mathbf{L}}^{\mathrm{Aut}}$.*

3.4. Some preparation for the proof of Lemma 2. Next we shall deduce some consequences of Lemma 5 and Lemma 6. Let Q be a rational quadratic form of signature (p, q), and $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ be an orthogonal basis of the symmetric bilinear form $(Q, \mathbb{Q}^n)$. Suppose that

$$Q(\boldsymbol{e}_1) > 0, \qquad Q(\boldsymbol{e}_2) < 0, \qquad Q(\boldsymbol{e}_3) < 0.$$

Let f be the restriction of Q to the subspace spanned by $\boldsymbol{e}_1, \boldsymbol{e}_2$ and $\boldsymbol{e}_3$. Denote by τ_1 the natural $\mathbb{Q}$-embedding of $\mathbf{Spin}_f$ into $\mathbf{Spin}_Q$, and τ_2 the $\mathbb{Q}$-embedding of $\mathbf{SO}_f$ into $\mathbf{SO}_Q$. As is well known, there exist $\mathbb{Q}$-coverings $\iota_1 : \mathbf{Spin}_Q \to \mathbf{SO}_Q$ and $\iota_2 : \mathbf{Spin}_f \to \mathbf{SO}_f$ such that $\iota_1 \circ \tau_1 = \tau_2 \circ \iota_2$ (see e.g. [48, p.82]).

Lemma 7. *Let Q and f be as above and put $\Lambda = \iota^{-1}(\mathbf{SO}_Q(\mathbb{R})^\circ \cap \Gamma_N)$. Then, through the embeddings τ_1 and τ_2, we have*

(1) $\mathbf{Spin}_f(\mathbb{R})$ acts ergodically on $\mathbf{Spin}_Q(\mathbb{R})/\Lambda$ by left translations;
(2) the representation of $\mathbf{Spin}_f(\mathbb{R})$ on $L^2_0(\mathbf{Spin}_Q(\mathbb{R})/\Lambda)$ is strongly $L^{\frac{64}{25}+\epsilon}$;
(3) the representation of $\mathbf{SO}_f(\mathbb{R})^\circ$ on $L^2_0(\mathbf{SO}_Q(\mathbb{R})^\circ/(\mathbf{SO}_Q(\mathbb{R})^\circ \cap \Gamma_N))$ is strongly $L^{\frac{64}{25}+\epsilon}$.

Proof. (1) By Moore's ergodicity theorem [43, Theorem 1], it suffices to verify the assertion that the projection of $\mathbf{Spin}_f(\mathbb{R})$ to any simple factor of $\mathbf{Spin}_Q(\mathbb{R})$ is non-compact. If the signature $(p,q) \neq (2,2)$, then $\mathbf{Spin}_Q(\mathbb{R})$ is a simple Lie group and there is nothing to prove. If $(p,q) = (2,2)$ it is enough to verify $\mathbf{Spin}_f(\mathbb{R})$ is not a simple factor of $\mathbf{Spin}_Q(\mathbb{R})$. Suppose the contrary, then the image $\iota(\mathbf{Spin}_f(\mathbb{R})) = \mathbf{SO}_f(\mathbb{R})^\circ$ would also be a simple factor of $\mathbf{SO}_Q(\mathbb{R})^\circ$. This is impossible because both of the two simple factors of $\mathbf{SO}_Q(\mathbb{R})^\circ$ are isomorphic to $\mathbf{SL}_2(\mathbb{R})$, while $\mathbf{SO}_f(\mathbb{R})^\circ$ is isomorphic to $\mathbf{PSL}_2(\mathbb{R})$. The assertion is verified.

(2) Let ρ be the representation in question, and σ be any irreducible representation of $\mathbf{Spin}_f(\mathbb{R})$ which is weakly contained in ρ. Since Λ is a congruence lattice in $\mathbf{Spin}_Q(\mathbb{R})$ ([39, Chapter I 3.1.1]), σ lies in the automorphic spectrum of $\mathbf{Spin}_f$ by Lemma 6. It follows from Lemma 5 that σ is either trivial or strongly $L^{\frac{64}{25}+\epsilon}$. Because $\mathbf{Spin}_f(\mathbb{R})$ acts ergodically on $\mathbf{Spin}_Q(\mathbb{R})/\Lambda$ and id_F is isolated in the automorphic spectrum, σ cannot be trivial and hence is strongly $L^{\frac{64}{25}+\epsilon}$. By (1) of Lemma 3, ρ is strongly $L^{\frac{64}{25}+\epsilon}$.

(3) Identifying $\mathbf{Spin}_Q(\mathbb{R})/\Lambda$ with $\mathbf{SO}_Q(\mathbb{R})^\circ/(\mathbf{SO}_Q(\mathbb{R})^\circ \cap \Gamma_N)$, the assertion follows from Lemma 4. □

As we are about to deal with various group actions, let us introduce the following lemma.

Lemma 8. *Let S be a semisimple Lie group, Σ a lattice in S, and F a closed subgroup of S acting on S/Σ by left translations. Suppose that the representation of F on $L^2_0(S/\Sigma)$ is strongly L^k. Then,*

(1) for any $s \in S$, the representation of F on $L^2_0(S/s\Sigma s^{-1})$ is strongly L^k;
(2) for any Lie group isomorphism $\phi : S \to S'$, the action of $\phi(F)$ on $L^2_0(\phi(S)/\phi(\Sigma))$ is strongly L^k.

Proof. For (1) since the map $S/\Sigma \to S/s\Sigma s^{-1}$ given by $g.\Sigma \mapsto gs^{-1}.(s\Sigma s^{-1})$ is S-equivariant, the F actions on $L^2_0(S/\Sigma)$ and $L^2_0(S/s\Sigma s^{-1})$ are unitarily equivalent. The assertion (2) is clear. □

3.5. Proof of Lemma 2.

Proof. Since the orbit $H.[g]_N$ is closed, the quadratic form $Q_0 \circ g$ is proportional to an integral quadratic form Q (see Section 2.2). Let f be as in Lemma 7, and we naturally identify $\mathbf{SO}_f(\mathbb{R})$ with $\tau_2(\mathbf{SO}_f(\mathbb{R}))$. Put

$$L' = g\mathbf{SO}_f(\mathbb{R})^\circ g^{-1} \subset g\mathbf{SO}_Q(\mathbb{R})^\circ g^{-1} = H, \qquad \Sigma = H \cap g\Gamma_N g^{-1}.$$

Then, there is a linear subspace W of $\mathbb{R}^n$ such that the quadratic space $(Q_0|_W, W)$ has signature $(1,2)$ and that $L' = \{h \in H : h|_{W^\perp} = \mathrm{id}\}$. Here $W^\perp$ is taken with respect to the bilinear form defined by $(Q_0, \mathbb{R}^n)$.

Recall that in the lemma $L = \{h \in H : h|_{V^\perp} = \mathrm{id}\}$. By Witt's extension theorem ([7, p.21]), there exists $s \in \mathbf{GL}_n(\mathbb{R})$ which preserves Q_0 and maps W onto V. In fact, such an element s can be chosen to lie in H. To see this, notice that each connected component of the orthogonal group $\mathbf{O}_{Q_0}(\mathbb{R}) = \{g \in \mathbf{GL}_n(\mathbb{R}) : Q_0 \circ g = Q_0\}$ contains a connected component of $g\mathbf{O}_f(\mathbb{R})g^{-1}$ (see Section 2.2 or [26, p.42-43]). Let $u \in g\mathbf{O}_f(\mathbb{R})g^{-1}$ which lies on the same connected component in $\mathbf{O}_{Q_0}(\mathbb{R})$ as the element s. Then us maps W onto V, as u leaves W invariant; and us lies in H since u and s are in the same connected component (Section 2.2).

Assume now that $L = sL's^{-1}$ for some $s \in H$. By (3) of Lemma 7 and (2) of Lemma 8, the action of L' on $L^2_0(H/\Sigma)$ is strongly $L^{\frac{64}{25}+\epsilon}$, and so is the action of L' on $L^2_0(H/s\Sigma s^{-1})$ by (1) of Lemma 8. From (2) of Lemma 8 we get that the action of L on $L^2_0(H/\Sigma)$ is also strongly $L^{\frac{64}{25}+\epsilon}$. By (6), the lemma is proved. □

3.6. Some additional remarks. One can also state the result on the spectral gap in terms of the representation of H instead of a copy of $\mathrm{SO}(1,2)^\circ$. We have avoided this to keep the background to a minimum. Let us now give a brief account on the other approach.

When $p \geq 2$ and $q \geq 3$ the group $\mathrm{SO}(p,q)$ has Kazhdan's property (T), and hence there exists a strong integrability exponent which applies to every infinite dimensional irreducible unitary representation. Following Howe's strategy these exponents for $\mathrm{SO}(p,q)$ were determined by Li [35] except $\mathrm{SO}(5,2)$, $\mathrm{SO}(4,3)$, $\mathrm{SO}(6,3)$. Li also showed that the irreducible representations with the slowest matrix coefficient decay do occur in the automorphic spectrum. Another way of quantifying Kazhdan's property (T) is to give uniform pointwise bounds of the matrix coefficients of the unitary representations without nonzero invariant vectors. This was the theme of Oh's work [46]. These results would give an explicit uniform spectral gap in terms of H. We are grateful to Oh for pointing out that her result [46, Theorem 1.4] (see also [35, Lemma 4.1]) implies when $p \geq 2$, $q \geq 3$ the condition $L^{\frac{64}{25}+\epsilon}$ in Lemma 2 can be replaced by $L^{2+\epsilon}$.

When $p = 1$, $q \geq 2$ we note that the unitary duals of $\mathrm{SO}(1,q)$ were classified by Hirai [23] (see also [42, §3] for an explicit description). The spectral gap for the congruence quotients were analyzed by Burger-Sarnak [5, Theorem 1.2]. For $q \geq 4$ the best possible spectral estimates were recently obtained by Bergeron-Clozel [2]. The case when $q = 3$ involves the Ramanujan bound of $\mathbf{SL}_2$ over imaginary quadratic number fields which we refer to [45] or [1] for the current record. For $q = 2$ the spectral estimate is equivalent to the progress towards the Ramanujan conjecture of $\mathbf{SL}_2/\mathbb{Q}$ and by far the best result is due to Kim-Sarnak [28] improving Selberg's 3/16-theorem [54] and the follow-up refinement.

For the spectral gap of congruence subgroups of general semisimple algebraic groups the reader is referred to Clozel's resolution of the property τ conjecture [9].

4. Quantitative behavior of the translates on homogenous spaces

4.1. Statement of results. Our goal in this section is to prove the following

Theorem 5. *Let $H.x$ be a closed orbit in G/Γ_N. Let $g_1, g_2 \in G$ be such that $[g_1]_N, [g_2]_N \in H.x$. Then, there exists $h \in H$ satisfying $h.[g_1]_N = [g_2]_N$ and*

$$\|h\| \ll_n \mathrm{vol}_H(H.x)^{\frac{13}{5}} \cdot (\|g_1\| \cdot \|g_2\|)^{\frac{13n}{5}\cdot(1+\frac{\dim H}{2})}. \tag{13}$$

Moreover, if the orbit $H.x$ is compact, then there exists an element $h \in H$ such that $h.[g_1]_N = [g_2]_N$ and $\|h\| \ll_n \mathrm{vol}_H(H.x)^{\frac{13}{5}}$.

The theorem may be compared to the isoperimetric inequalities in Riemannian geometry and spectral graph theory. An example of an isoperimetric inequality is that the diameter of a compact hyperbolic orbifold M is bounded by a function involving the volume, the injectivity radius, and the spectral gap of the Laplacian operator of M (see for instance [6]).

As we will see in Lemma 11 that for some $\delta \asymp \|g\|^{-n}$ the map $\{g_1 \in G : \|g_1 - 1_n\| < \delta\} \to G/\Gamma$ given by $g_1 \mapsto g_1.[g]$ is injective. Hence, roughly speaking, Theorem 5 says that the distance of the two points $[g_1]$, $[g_2]$ on the closed orbit $H.x$ is controlled by their injectivity radii and a polynomial of $\mathrm{vol}_H(H.x)$. Moreover, the degree of the polynomial (namely the exponent 13/5) comes from Lemma 2 which concerns a uniform spectral gap of the actions of H on the closed orbits. In this sense, Theorem 5 may be viewed as an explicit isoperimetric inequality for the closed orbit $H.x$ in G/Γ.

4.2. Injectivity radius. We fix for any $\eta > 0$ the neighborhood of the identity element of G and H

$$B_G(\eta) := \{g \in G : \|g - 1_n\| < \eta\}, \qquad B_H(\eta) := H \cap B_G(\eta).$$

Lemma 9. *For any $\eta < 1$ there exists $\eta_1 \gg_n \eta$ such that if $g_1, g_2 \in B_G(\eta_1)$, then $g_1^{-1}g_2 \in B_G(\eta)$.*

Proof. Since $\|g_1^{-1}g_2 - 1_n\| \ll_n \|g_1 - g_2\|$ for any $g_1, g_2 \in B_G(1)$, the assertion is clear. □

Lemma 10. *There exists a constant $\eta_0 > 0$ such that for any natural number N, any compact orbit $H.x$ in G/Γ_N, and any $y \in H.x$ the map $B_H(\eta_0) \to H.x$, $h \mapsto h.y$ is injective.*

Proof. Recall from Section 2.3 that compact H-orbits only exist when $n = 3$ or 4 by Meyer's theorem, and hence in this lemma we have $G = \mathbf{SL}_3(\mathbb{R})$ or $\mathbf{SL}_4(\mathbb{R})$.

Let g_1, g_2 be elements in $B_G(1)$. Let f_1 and f_2 be the characteristic polynomials of g_1 and g_2, respectively. It is clear that $\|f_1 - f_2\| \ll_n \|g_1 - g_2\|$ where $\|f_1 - f_2\|$ is the maximum of the absolute values of the coefficients of $f_1 - f_2$. Hence there exists $\eta > 0$ such that if $u \in B_G(\eta)$ and the coefficients of the characteristic polynomial of u are integers, then u is unipotent.

Let $H.x$ and y be as in the lemma assuming $y = [g]_N$, where $g \in G$. Recall that $H \cap g\Gamma g^{-1}$ is a cocompact lattice in H (see Section 2.3). We shall prove that if $h \in B_H(\eta) \cap g\Gamma g^{-1}$, then $h = 1_n$. Let h be such an element. As $g^{-1}hg \in \Gamma$ the coefficients of the characteristic polynomial of h are integers. By the choice of η, the element h is unipotent and hence so is $g^{-1}hg \in \mathbf{SO}_Q(\mathbb{R})$. On the other hand, since $H.[g]$ is a compact orbit in G/Γ, the algebraic group $\mathbf{SO}_Q$ has $\mathbb{Q}$-rank 0. By Godement's criterion $\mathbf{SO}_Q(\mathbb{R}) \cap \Gamma$ does not contain any non-trivial unipotent element (see e.g. [39, I.3.2.7]), and it follows that $h = 1_n$.

Let $\eta_0 \gg \eta$ satisfying Lemma 9. If $h_1, h_2 \in B_H(\eta_0)$ and $h_1.[g]_N = h_2.[g]_N$, then $h_2^{-1}h_1 \in B_H(\eta)$ and $h_2^{-1}h_1 \in g\Gamma_N g^{-1} \subset g\Gamma g^{-1}$. Hence $h_1 = h_2$ by what we have obtained in the preceding paragraph. □

In Theorem 5 the bound for compact orbits is related to the previous lemma. For non-compact closed H-orbits in G/Γ_N, the above argument does not apply because the corresponding integral quadratic forms are isotropic. Nevertheless, we have the following result on the injectivity radius.

Lemma 11. *For any $g \in G$ there exists $\delta \asymp_n \|g\|^{-n}$ such that for any natural number N the map $B_G(\delta) \to G/\Gamma_N$, $g_1 \mapsto g_1.[g]_N$ is injective. Here, and in the sequel, $\delta \asymp_n \|g\|^{-n}$ means $\|g\|^{-n} \ll_n \delta \ll_n \|g\|^{-n}$.*

Proof. Any $\gamma \in \Gamma$ can be written as $\gamma = 1_n + (\gamma - 1_n)$, a sum of integral matrices. Hence if $\gamma \neq 1_n$, then $\|\gamma - 1_n\| \geq 1$. By Cramer's rule for $\|g^{-1}\|$, we have

$$\|g\gamma g^{-1} - 1_n\| = \|g(\gamma - 1_n)g^{-1}\| \gg_n \|\gamma - 1_n\| \cdot \|g\|^{-1} \cdot \|g^{-1}\|^{-1} \gg_n \|g\|^{-n}.$$

Hence for some $\delta \asymp_n \|g\|^{-n}$, $B_G(\delta) \cap g\Gamma_N g^{-1} \subset B_G(\delta) \cap g\Gamma g^{-1} = \{1_n\}$. The lemma follows easily from Lemma 9 (cf. the last paragraph in the proof of Lemma 10). □

4.3. Smooth bump functions of small derivatives. Let g_t, $\mathcal{D}$ be as in (11) and L as in Lemma 2. It is easy to see that for $n = 3$ there exists a covering of $\mathbf{SL}_2(\mathbb{R})$ to H such that $g_t \mapsto \begin{pmatrix} \cosh t & 0 & \sinh t \\ 0 & 1 & 0 \\ \sinh t & 0 & \cosh t \end{pmatrix}$. Therefore, for any $n \geq 3$ we can fix a covering map $\iota : \mathbf{SL}_2(\mathbb{R}) \to L$ so that for every $t \in \mathbb{R}$

$$\|\iota(g_t)\| = \cosh t \tag{14}$$

Through the map ι the group $\mathbf{SL}_2(\mathbb{R})$ acts on $L^2(H)$ and $L^2(H.x)$ if the orbit $H.x$ is closed in G/Γ_N.

Lemma 12. *Let $H.x$ be a closed orbit in G/Γ_N and $0 < \delta < 1$. Suppose the map $B_H(\delta) \to H.x$, $h \mapsto h.x$ is injective. Then there exists a smooth function ψ on $H.x$, whose support is contained in $B_H(\delta).x$, satisfying*

$$\psi \geq 0, \qquad \mu_{H.x}(\psi) = \mathrm{vol}_H(H.x)^{-1}, \qquad \|\mathcal{D}.\psi\| \ll_n \mathrm{vol}_H(H.x)^{-\frac{1}{2}}\delta^{-(1+\frac{\dim H}{2})}. \tag{15}$$

Proof. We fix for every $0 < \eta < 1$ a non-negative smooth function φ_η on H supported in $B_H(\eta)$ satisfying

$$m_H(\varphi_\eta) = 1, \qquad \|\mathcal{D}.\varphi_\eta\| \ll_n \eta^{-1-\frac{\dim H}{2}}.$$

The functions can be constructed from the suitably chosen smooth functions on the Euclidean space $\mathfrak{h}$ (see e.g. [29, §2.4.7]) composed with the logarithm map from $B_H(\eta)$ to $\mathfrak{h}$.

Let us define a function ψ on $H.x$ by setting $\psi(h.x) = \varphi_\delta(h)$ for every $h \in B_H(\delta)$, and setting $\psi = 0$ outside the open subset $B_H(\delta).x$ of $H.x$. The function is well defined because the map $B_H(\delta) \to H.x$, $h \mapsto h.x$ is assumed to be injective in the lemma. Clearly the function ψ is smooth, and satisfies (15) because of the relation (10) and the bound on the norm of φ_δ. □

4.4. Proof of Theorem 5.

Proof. The plan is to show that the translate of a neighborhood of $[g_1]_N$ in $H.x$ by a large element g_t intersects a neighborhood of $[g_2]_N$.

Let $i = 1,2$. By Lemma 11 there exists $\delta_i \asymp_n \|g_i\|^{-n}$ such that the map $\pi_i : B_G(\delta_i) \to G/\Gamma_N$, $\pi_i(g) = g.[g_i]_N$ is injective and hence so is its restriction to $B_H(\delta_i)$. Let ψ_i be the non-negative smooth function on $H.x$ supported in $B_H(\delta_i).[g_i]_N$ satisfying the bounds (15) in Lemma 12.

By Lemma 2 and Lemma 4, the action of $\mathbf{SL}_2(\mathbb{R})$ on $L^2_0(H.x)$ through ι is strongly $L^{\frac{64}{25}+\epsilon}$. To use Lemma 3, let us recall the asymptotic behavior of the Harish-Chandra function [26, Chapter 5, §3.1]

$$|\Xi(g_t)|^{2\cdot\frac{25}{64}} \ll e^{-\frac{5}{13}|t|}, \qquad t \in \mathbb{R}. \tag{16}$$

We have used the number $13/5 > 64/25$ to make the multiple constant absolute.

Let $a_t = \iota(g_t)$ with ι given in Section 4.3. Recall that $\|a_t\| = \cosh(t)$ by (14). The functions $\psi_i - \mu_{H.x}(\psi_i)$ are smooth vectors of the action of $\mathbf{SL}_2(\mathbb{R})$ on $L^2_0(H.x)$. It follows from Lemma 3 and the estimate (16) that

$$\left|\langle a_t.\psi_1,\psi_2\rangle - \mu_{H.x}(\psi_1)\mu_{H.x}(\psi_2)\right| \ll e^{-\frac{5}{13}|t|}\cdot\|\mathcal{D}.(\psi_1)\|\cdot\|\mathcal{D}.(\psi_2)\|.$$

If $\mu_{H.x}(\psi_1)\mu_{H.x}(\psi_2)$ dominates the error term, then $\langle a_t.\psi_1,\psi_2\rangle \neq 0$. In particular, this will happen for some a_t with

$$\|a_t\| = \cosh(t) \le e^{|t|} \ll_n \mathrm{vol}_H(H.x)^{\frac{13}{5}}\cdot(\|g_1\|\cdot\|g_2\|)^{\frac{13n}{5}(1+\frac{\dim H}{2})}, \tag{17}$$

since $\delta_i \asymp_n \|g_i\|^{-n}$ and by (15). For later use in dealing with the compact orbits and in Section 7.5, we remark that it is essentially the lower bound of the injectivity radii of $[g_1]_N$ and $[g_2]_N$ that leads to (17).

Because $\langle a_t.\psi_1,\psi_2\rangle \neq 0$, we get

$$\Big(a_t.(B_H(\delta_1).[g_1]_N)\cap B_H(\delta_2).[g_2]_N\Big) \supseteq \Big(a_t.\mathrm{supp}(\psi_1)\cap\mathrm{supp}(\psi_2)\Big) \neq \emptyset.$$

As $\delta_i \ll_n 1$, it follows that $h.[g_1]_N = [g_2]_N$ for some $h \in B_H(\delta_2)^{-1}a_tB_H(\delta_1)$ and hence $\|h\| \ll_n$ (17). The theorem is now proved for closed orbits. For the compact H-orbits in G/Γ_N, the proof remains the same except we can use $\delta_i = \eta_0$ given by Lemma 10 in (17) above to get the desired bound. □

5. Quantitative non-divergence of unipotent flows and applications

5.1. Statement of results. In this section, we shall prove Theorem 3 and the following theorem.

Theorem 6. *Let Q be an integral quadratic form of signature (p,q), and suppose $Q = |\det Q|^{\frac{1}{n}}(Q_0\circ g)$ for some $g \in G$. Then, for any natural number N the closed orbit $H.[g]_N$ in G/Γ_N satisfies*

$$\mathrm{vol}_H(H.[g]_N) \ll_n N^{n^2}\cdot|\det Q|^{\frac{\dim G-\dim H}{n}}. \tag{18}$$

The value of $\mathrm{vol}_H(H.[g]_N)$ does not depend on the choice of g. Indeed, any $g_1 \in G$ with $Q_0\circ g_1 = Q_0\circ g$ must satisfy $g_1g^{-1} \in \mathbf{SO}_{Q_0}(\mathbb{R})$. As the measure m_H is invariant under the conjugation by any element in $\mathbf{SO}_{Q_0}(\mathbb{R})$, the covolumes of the lattices $H\cap g\Gamma_Ng^{-1}$ and $H\cap g_1\Gamma_Ng_1^{-1}$ in H are equal and hence $\mathrm{vol}_H(H.[g]_N) = \mathrm{vol}_H(H.[g_1]_N)$ (see Section 2.4).

5.2. Recurrence of closed H-orbits. Both theorems in Section 5.1 will be deduced from the following lemma [17, Lemma 3.2] (cf. [40, Remark p.421]) on the recurrence of the closed H-orbits in G/Γ.

Lemma 13. *There exists a compact subset Ω_{cpt} of G/Γ such that for any closed orbit $H.x$ in G/Γ*

$$\text{vol}_H(H.x \cap \Omega_{\text{cpt}}) > 0.25\ \text{vol}_H(H.x). \tag{19}$$

This lemma is essentially a further development of the works [37, 12, 30]. From now on we shall fix for each $n \geq 3$ a compact subset Ω_{cpt} of G/Γ which satisfies (19). In particular, any constant which depends on the choice of Ω_{cpt} will be treated as dependent only on n.

Lemma 14. *For any $n \geq 3$ there exists a constant $R_n > 0$ such that any point $[g] \in \Omega_{\text{cpt}}$ can be written as $[g] = [g_1]$ with $g_1 \in G$ and $\|g_1\| < R_n$.*

Proof. It follows from the compactness of Ω_{cpt}. □

5.3. Proof of Theorem 3.

Proof. Let (p, q) be the signature of Q_A. Denote by X_0 the symmetric matrix of the quadratic form Q_0 defined in (5). Then $A = |\det A|^{\frac{1}{n}}(g_A^t X_0 g_A)$ for some $g_A \in G$ (Lemma 1), and $H.[g_A]$ is a closed orbit in G/Γ (Section 2.3). By Lemma 13 and Lemma 14, the subset $Hg_A\Gamma$ of G contains an element $g_1 = hg_A\gamma$ ($h \in H$, $\gamma \in \Gamma$) with $\|g_1\| < R_n$. Clearly $A_1 = |\det A|^{\frac{1}{n}}(g_1^t X_0 g_1)$ is equivalent to A with $\|A_1\| < nR_n^2 |\det A|^{\frac{1}{n}}$.

In the appendix, we will exhibit a compact subset Ω_{cpt} satisfying (19). Once this is done, there will be no obstruction evaluating the constant R_n in Lemma 14. □

5.4. Recurrence of closed H-orbits in transversal directions.

Lemma 15. *For any $n \geq 3$ there exists a constant C_n satisfying the following property. For any closed orbit $H.[g]$ in G/Γ there exist a point $[g_1] \in H.[g] \cap \Omega_{\text{cpt}}$ and an element $u \in G$, such that $u \notin \mathbf{SO}_{Q_0}(\mathbb{R})$ and that*

$$u.[g_1] \in H.[g], \qquad \|u - 1_n\| < C_n \cdot \text{vol}_H(H.[g])^{-\frac{1}{\dim G - \dim H}}. \tag{20}$$

Proof. Let $\mathfrak{g}$ be the Lie algebra of G endowed by the Euclidean norm coming from $\mathfrak{gl}_n(\mathbb{R})$. Suppose $\mathfrak{g} = \mathfrak{h} + \mathfrak{h}'$ as a direct sum of $\mathfrak{h}$-modules with respect to the adjoint action. Put $B_{\mathfrak{h}'}(\delta) := \{\boldsymbol{w} \in \mathfrak{h}' : \|\boldsymbol{w}\| < \delta\}$ for any $\delta > 0$. We fix a constant $0 < \eta_0 < 1$ such that $\exp(-\boldsymbol{w}_2) \cdot \exp(\boldsymbol{w}_1) \notin \mathbf{SO}_{Q_0}(\mathbb{R})$ for any $\boldsymbol{w}_1, \boldsymbol{w}_2 \in B_{\mathfrak{h}'}(\eta_0)$ with $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$. Let $\eta < \eta_0$ be such that the map

$$B_{\mathfrak{h}'}(\eta) \times (H.[g] \cap \Omega_{\text{cpt}}) \to G/\Gamma, \qquad (\boldsymbol{w}, [g_1]) \mapsto \exp(\boldsymbol{w}).[g_1] \tag{21}$$

is injective. We shall prove $\eta \ll_n \text{vol}_H(H.[g])^{-\frac{1}{\dim \mathfrak{h}'}}$. Since Ω_{cpt} is compact, there exists an absolute constant[1] $b_n > 0$ such that the $m_{G/\Gamma}$-measure of the image of the map in (21) (a tubular neighborhood of $H.[g] \cap \Omega_{\text{cpt}}$) is at least $b_n \cdot \eta^{\dim \mathfrak{h}'} \cdot \text{vol}_H(H.[g] \cap \Omega_{\text{cpt}})$ (cf. [17, Prop. 14.2]). This, together with (8) and (19), gives

$$0.25 \cdot b_n \cdot \eta^{\dim \mathfrak{h}'} \cdot \text{vol}_H(H.[g]) < b_n \cdot \eta^{\dim \mathfrak{h}'} \cdot \text{vol}_H(H.[g] \cap \Omega_{\text{cpt}}) < m_{G/\Gamma}(G/\Gamma) = \zeta(2) \cdots \zeta(n). \tag{22}$$

From this we get $\eta < c_n \cdot \text{vol}_H(H.[g])^{-\frac{1}{\dim \mathfrak{h}'}}$ as we wanted.

It is enough to prove the lemma for those orbits with $c_n \cdot \text{vol}_H(H.[g])^{-\frac{1}{\dim \mathfrak{h}'}} < \eta_0$, because for any given number M there are only finitely many closed H-orbits in G/Γ whose volumes are less than M. Now let $H.[g]$ be a closed orbit with $\eta_1 = c_n \cdot \text{vol}_H(H.[g])^{-\frac{1}{\dim \mathfrak{h}'}} < \eta_0$. By what we obtained in the proceeding paragraph, the map (21) fails to be injective for $\eta = \eta_1$, and hence there exist $\boldsymbol{w}_1, \boldsymbol{w}_2 \in B_{\mathfrak{h}'}(\eta_1)$ and $[g_1], [g_2] \in H.[g] \cap \Omega_{\text{cpt}}$ such that $(\boldsymbol{w}_1, [g_1]) \neq (\boldsymbol{w}_2, [g_2])$ but $\exp(\boldsymbol{w}_1).[g_1] = \exp(\boldsymbol{w}_2).[g_2]$. Notice that $\boldsymbol{w}_1 \neq \boldsymbol{w}_2$, since otherwise $[g_1] = [g_2]$ which leads to a contradiction $(\boldsymbol{w}_1, [g_1]) = (\boldsymbol{w}_2, [g_2])$. Therefore, $u = \exp(-\boldsymbol{w}_2) \cdot \exp(\boldsymbol{w}_1) \notin \mathbf{SO}_{Q_0}(\mathbb{R})$ by the choice of η_0. In conclusion, $[g_1]$ and u satisfy the lemma. □

[1] The constant b_n can explicitly evaluated in terms of n as the measures introduced in Section 2.4 are defined using coordinates.

5.5. Isolation of closed H-orbits in transversal directions.

Lemma 16. *Let Q_1 be an integral quadratic form of signature (p,q) and suppose $Q_1 = |\det Q_1|^{\frac{1}{n}}(Q_0 \circ g_1)$ where $g_1 \in G$ and $\|g_1\| < R_n$ with R_n given in Lemma 14. Assume that the element $u \in G$ satisfies $u \notin \mathbf{SO}_{Q_0}(\mathbb{R})$ and $u.[g_1] \in H.[g_1]$. Then*

$$\|u - 1_n\| > (3n^2R_n^2)^{-1}|\det Q_1|^{-\frac{1}{n}}. \tag{23}$$

Proof. Let $X_0 \in \mathbb{Z}^{n\times n}$ be the symmetric matrix of Q_0. As $u.[g_1] \in H.[g_1]$, one has that $ug_1 = hg_1\gamma$ for some $h \in H$ and $\gamma \in \Gamma$. Since $u \notin \mathbf{SO}_{Q_0}(\mathbb{R})$ by assumption, we get $\gamma \notin \mathbf{SO}_{Q_1}(\mathbb{R}) = g_1^{-1}\mathbf{SO}_{Q_0}(\mathbb{R})g_1$. Therefore, $|\det Q_1|^{\frac{1}{n}} g_1^t u^t X_0 u g_1 = |\det Q_1|^{\frac{1}{n}} \gamma^t g_1^t X_0 g_1 \gamma$ is a symmetric integral matrix that is equivalent to but different from $|\det Q_1|^{\frac{1}{n}} g_1^t X_0 g_1$, and we have $|\det Q_1|^{\frac{1}{n}}\|g_1^t u^t X_0 u g_1 - g_1^t X_0 g_1\| \geq 1$ for two different integral matrices. From this (23) follows, otherwise we would get $\|g_1^t u^t X_0 u g_1 - g_1^t X_0 g_1\| < |\det Q_1|^{-\frac{1}{n}}$. □

5.6. Proof of Theorem 6.

Proof. As $\mathrm{vol}_H(H.[g]_N) \leq [\Gamma : \Gamma_N] \cdot \mathrm{vol}_H(H.[g])$ and the index $[\Gamma : \Gamma_N] < N^{n^2}$, it suffices to prove (18) for $N = 1$. Let $g_1 \in G$ be such that the point $[g_1] \in H.[g] \cap \Omega_{\mathrm{cpt}}$ satisfies Lemma 15 and $\|g_1\| < R_n$ as in Lemma 14. Because $[g_1] \in H.[g]$, $|\det Q|^{\frac{1}{n}} g_1^t X_0 g_1$ defines an integral quadratic form Q_1 which is equivalent to Q. In particular, $H.[g_1] = H.[g]$ and $\det Q_1 = \det Q$. The theorem follows from (20), (23). □

6. Proof of Theorem 1 for indefinite integral quadratic forms

Proof. Let (p,q) be the signature of Q_A (and hence Q_B). Let Q_0 be the integral quadratic form of signature (p,q) defined in Section 2.2, and let $X_0 \in \mathbb{Z}^{n\times n}$ be the symmetric matrix of Q_0. By Lemma 1 there exist g_A, $g_B \in G$ such that $A = |\det A|^{\frac{1}{n}}(g_A^t X_0 g_A)$, $B = |\det B|^{\frac{1}{n}}(g_B^t X_0 g_B)$, and moreover

$$\|g_A\| \ll_n |\det A|^{-\frac{1}{2n}}\|A\|^{\frac{1}{2}}, \qquad \|g_B\| \ll_n |\det B|^{-\frac{1}{2n}}\|B\|^{\frac{1}{2}}. \tag{24}$$

Since A and B are assumed to be equivalent ($A = \gamma_0^t B\gamma_0$) in the theorem, we get that $\det A = \det B$ and hence

$$g_A^t X_0 g_A = \gamma_0^t g_B^t X_0 g_B \gamma_0.$$

We first assume that $\det\gamma_0 = 1$, that is, $\gamma_0 \in \Gamma$. By Section 2.3, the orbits $H.[g_A]$, $H.[g_B]$ are closed in G/Γ. Let s_0 be a fixed element in $\mathbf{SO}_{Q_0}(\mathbb{R})$ so that $\|s_0\| = 1$, and $s \notin H$. Replacing g_B by $s_0 g_B$ if necessary ($[\mathbf{SO}_{Q_0}(\mathbb{R}) : H] = 2$), we may assume that $g_A \in Hg_B\gamma_0$ and thus $H.[g_A] = H.[g_B] \subseteq G/\Gamma$. By Theorem 5 there exists $h \in H$ which satisfies $h.[g_A] = [g_B]$ and (13). We note that $\gamma = g_B^{-1}hg_A \in \Gamma$ and $A = \gamma^t B\gamma$. Since $\|\gamma\| \ll_n \|g_B\|^{-1} \cdot \|h\| \cdot \|g_A\|$, the theorem follows from (13), (18), (24) and the fact that $\dim G = n^2 - 1$ and $\dim H = (n^2 - n)/2$.

Suppose now $\det\gamma_0 = -1$. Let ξ be a fixed element in $\mathbf{GL}_n(\mathbb{Z})$ such that $\det\xi = -1$ and $\|\xi\| = 1$. Then, for $\xi^t A\xi$ and B, by what we have just obtained, there exists $\gamma_1 \in \Gamma$ such that $\xi^t A\xi = \gamma_1^t B\gamma_1$, and γ_1 satisfies the norm bound in (3). Clearly the matrix $\gamma_1\xi^{-1}$ transforms B into A, and satisfies $\det(\gamma_1\xi^{-1}) = -1$.

For indefinite anisotropic forms we apply the bound in Theorem 5 for compact H-orbits (Section 2.3). □

7. Masser's conjecture with congruence conditions

7.1. Statement of results.

We now turn to a strengthening of Theorem 1.

Theorem 7. *For any $n \geq 3$ there exists a constant $C_n > 0$ satisfying the following property. Let A, $B \in \mathbb{Z}^{n\times n}$ be symmetric, non-singular matrices, and suppose $A = \gamma_0^t B\gamma_0$ for some $\gamma_0 \in \mathbf{GL}_n(\mathbb{Z})$. Then, for any natural number N there exists $\gamma_1 \in \mathbf{GL}_n(\mathbb{Z})$ with $A = \gamma_1^t B\gamma_1$, $\gamma_1 \equiv \gamma_0 \mod (N)$, and*

$$\|\gamma_1\| \ll C_n \cdot N^{\frac{13n^2}{5}} \cdot |\det B|^{-\frac{39n^3-117n^2+98n+156}{20n}} \cdot \|A\|^{\frac{13n^3-13n^2+52n+20}{40}} \cdot \|B\|^{\frac{13n^3-13n^2+68n-4}{8}}. \tag{25}$$

The study of search bounds with congruence condition was suggested in Masser's survey [41]. For ternary forms the problem was settled by Dietmann [14, Theorem 4]. Our theorem resolves the general cases.

An important step of proving the theorem is to analyze small coset representatives of congruence subgroups. Let Σ_1 be a group, and Σ_2 a subgroup of Σ_1 of index k. We say that $\xi_1, \cdots, \xi_k \in \Sigma_1$ is a *system of coset representatives* of Σ_1/Σ_2 if Σ_1 can be written as the disjoint union $\Sigma_1 = \xi_1\Sigma_2 \sqcup \cdots \sqcup \xi_k\Sigma_2$.

Theorem 8. *Let $Q = Q_B$ be an integral quadratic form in n-variables. Write $\Sigma = \mathbf{SO}_Q(\mathbb{R}) \cap \Gamma$ and $\Sigma_N = \Sigma \cap \Gamma_N$. Then, for any N there exists a system of coset representatives $\xi_1, \cdots, \xi_k$ of Σ/Σ_N with*

$$\|\xi_i\| \ll_n N^{\frac{13n^2}{5}} \cdot |\det B|^{-\frac{13n^3-39n^2+31n+52}{10n}} \cdot \|B\|^{\frac{13n^3-13n^2+67n}{10}}, \qquad i = 1, \cdots, k. \tag{26}$$

We first finish the proof of Theorem 7 and then we proceed to prove Theorem 8.

7.2. Proof of Theorem 7 for indefinite integral quadratic forms.

Proof. Let $Q = Q_B$. By Theorem 1 there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$, such that $A = \gamma^t B\gamma$, $\det\gamma = \det\gamma_0$, and satisfies the norm bound (3). Hence $\gamma_0\gamma^{-1} \in \Sigma = \mathbf{SO}_Q(\mathbb{R}) \cap \Gamma$. Let the level N be given. By Theorem 8 there exists a coset $\xi_i\Sigma_N \subseteq \Sigma$ such that ξ_i satisfies (26) and $\gamma_0\gamma^{-1} \in \xi_i\Sigma_N = \Sigma_N\xi_i$ (the principle congruence subgroups are normal). Let $\gamma_1 = \xi_i\gamma$. Then, $\gamma_0 \equiv \gamma_1 \mod (N)$ and $A = \gamma_1^t B\gamma_1$. The norm bound in the theorem follows from $\|\gamma_1\| \ll_n \|\xi_i\| \cdot \|\gamma\|$ along with the bounds in (3) and (26). □

7.3. The integral points in the identity component of $\mathbf{O}_Q(\mathbb{R})$.

Lemma 17. *Let $Q = Q_B$ be an integral quadratic form of signature (p, q), and H_Q the identity component of $\mathbf{O}_Q(\mathbb{R}) = \{g \in \mathbf{GL}_n(\mathbb{R}) : Q = Q \circ g\}$ (or equivalently $\mathbf{SO}_Q(\mathbb{R})$). Write $\Lambda = H_Q \cap \Gamma$ and $\Lambda_N = H_Q \cap \Gamma_N$. Then, for any N there exists a system of coset representatives $\xi_1, \cdots, \xi_k$ of Λ/Λ_N with*

$$\|\xi_i\| \ll_n N^{\frac{13n^2}{5}} \cdot |\det B|^{-\frac{13n^3-39n^2+26n+52}{20n}} \cdot \|B\|^{\frac{13n^3-13n^2+62n}{20}}, \qquad i = 1, \cdots, k. \tag{27}$$

Proof. The idea of the proof is to use the geometric fact (in a slightly modified setting) that Λ/Λ_N embeds into H_Q/Λ_N. Moreover, the points in the image of the embedding share the same injectivity radius.

We now work out this idea. By Lemma 1, there exists $g \in G$ such that

$$\|g\| \ll_n |\det B|^{-\frac{1}{2n}} \|B\|^{\frac{1}{2}}, \qquad Q = |\det B|^{\frac{1}{n}}(Q_0 \circ g). \tag{28}$$

The orbit $H.[g]_N$ is closed in G/Γ_N (see Section 2.3). By Lemma 11, there is $\delta \asymp_n \|g\|^{-n}$ such that for every level N the map $B_H(\delta) \to H.[g]_N$, $h \mapsto h.[g]_N$ is injective. Let $\zeta_1, \cdots, \zeta_k$ be an arbitrary system of the coset representatives for Λ/Λ_N, and let $i \in \{1, \cdots, k\}$ with $k = [\Lambda : \Lambda_N]$.

Since $\zeta_i \in \Lambda \subset H_Q = g^{-1}Hg$, one has $g\zeta_i \in Hg$ and hence $[g\zeta_i]_N$ lies on the same orbit $H.[g]_N$. The map $B_H(\delta) \to H.[g]_N$, $h \mapsto h.[g\zeta_i]_N$ is injective. Indeed, if $h_1.[g\zeta_i]_N = h_2.[g\zeta_i]_N$ with $h_1, h_2 \in B_H(\delta)$; then $h_1.[g]_N = h_2.[g]_N$ since $\zeta_i\Gamma_N\zeta_i^{-1} = \Gamma_N$, and hence $h_1 = h_2$ by the choice of δ.

Using same argument in the proof of Theorem 5 (cf. the remark below (17)) we get that there exists an element $h_i \in H$ satisfying $h_i.[g]_N = [g\zeta_i]_N$ and

$$\|h_i\| \ll_n \mathrm{vol}_H(H.[g]_N)^{\frac{13}{5}} \cdot (\|g\|^{n(1+\frac{\dim H}{2})} \cdot \|g\|^{n(1+\frac{\dim H}{2})})^{\frac{13}{5}}. \tag{29}$$

Let $\xi_i = g^{-1}h_ig$. Then $\xi_i \in (\zeta_i\Gamma_N) \cap H_Q = \zeta_i\Lambda_N$ and $\|\xi_i\| \ll_n \|g\|^n \cdot \|h_i\|$. The norm bound in the lemma follows easily from (18), (28) and (29). □

7.4. The integral points in non-identity components of $\mathbf{O}_Q(\mathbb{R})$.

Lemma 18. *Let $Q = Q_B$ be an integral quadratic form of signature (p, q), H_Q be the identity component of $\mathbf{O}_Q(\mathbb{R})$, and u be an element in $\mathbf{O}_Q(\mathbb{R})$. Suppose that the connected component $uH_Q = H_Qu$ of $\mathbf{O}_Q(\mathbb{R})$*

intersects with $\mathbf{GL}_n(\mathbb{Z})$. *Then there exists* $\gamma \in H_Q u \cap \mathbf{GL}_n(\mathbb{Z})$ *with*

$$\|\gamma\| \ll_n |\det B|^{-\frac{13n^3-39n^2+26n+52}{20n}} \cdot \|B\|^{\frac{13n^3-13n^2+62n}{20}}. \tag{30}$$

Proof. Clearly we only have to deal with the case when $u \notin H_Q$. Let $g \in G$ which satisfies (28) in Lemma 17. We assume as we may that $u = g^{-1}sg$ with $s \in \mathbf{O}_{Q_0}(\mathbb{R})$, $s \notin H$ and $\|s\| = 1$. By assumption of the lemma there exists $s_1 \in sH = Hs$ such that the element $g^{-1}s_1 g$ belongs to $\mathbf{GL}_n(\mathbb{Z})$.

Recall from Section 2.2 that $ss_1 \in H$. Hence for the action of $\mathbf{O}_{Q_0}(\mathbb{R})$ on $\mathbf{GL}_n(\mathbb{R})/\mathbf{GL}_n(\mathbb{Z})$, we have $Hsg\mathbf{GL}_n(\mathbb{Z}) = Hsg(g^{-1}s_1 g)\mathbf{GL}_n(\mathbb{Z}) = Hg\mathbf{GL}_n(\mathbb{Z})$. In other words, the cosets $sg\mathbf{GL}_n(\mathbb{Z})$ and $g\mathbf{GL}_n(\mathbb{Z})$ lie on the same H-orbit. If $hsg\mathbf{GL}_n(\mathbb{Z}) = g\mathbf{GL}_n(\mathbb{Z})$ with $h \in H$, then $g^{-1}hsg \in H_Q u \cap \mathbf{GL}_n(\mathbb{Z})$. Hence we only need to give a bound on the norm of such an element $h \in H$.

The stabilizer of the coset $g\mathbf{GL}_n(\mathbb{Z})$ in H is the arithmetic lattice $H \cap g\mathbf{GL}_n(\mathbb{Z})g^{-1} = H \cap g\Gamma g^{-1}$. Recall that the homogeneous spaces $H/(H \cap g\Gamma g^{-1})$ and $H.[g]$ are naturally identified by (6). The injectivity radii at $sg\mathbf{GL}_n(\mathbb{Z})$ and $g\mathbf{GL}_n(\mathbb{Z})$ are both $\gg_n \|g\|^{-n}$ by the same proof of Lemma 11. Using the same argument in the proof of Theorem 5 (cf. Lemma 17), we get that there exists $h \in H$ such that

$$hsg\mathbf{GL}_n(\mathbb{Z}) = g\mathbf{GL}_n(\mathbb{Z}), \qquad \|h\| \ll_n \mathrm{vol}_H(H/(H \cap g\Gamma g^{-1}))^{\frac{13}{5}} \cdot (\|g\|^{n(1+\frac{\dim H}{2})} \cdot \|g\|^{n(1+\frac{\dim H}{2})})^{\frac{13}{5}}.$$

This, together with (18) and (28), gives the norm bound in (30). □

7.5. Proof of Theorem 8.

Proof. By (2) we only need to work with indefinite forms. Let (p,q) be the signature of $Q = Q_B$. Let $\xi_1, \cdots, \xi_k$ be the system of the coset representatives for Λ/Λ_N satisfying Lemma 17. If $\Sigma \neq \Lambda$ (otherwise Lemma 17 already proves the theorem), then $[\Sigma : \Lambda] = 2$. By Lemma 18, there exists $\xi \in \Sigma$ such that $\xi \notin \Lambda$ and that ξ satisfies (30). The elements $\xi_1, \cdots, \xi_k, \xi\xi_1, \cdots, \xi\xi_k$ already form a system of coset representatives of Σ/Λ_N. As $\Lambda_N \subset \Sigma_N$, a subset of $\{\xi_1, \cdots, \xi_k, \xi\xi_1, \cdots, \xi\xi_k\}$ will form a system of coset representatives of Σ/Σ_N. The bound in (26) follows from $\|\xi\xi_i\| \ll_n \|\xi\| \cdot \|\xi_i\|$ and the bounds in (27) and (30). □

8. Representations of integral quadratic forms

8.1. Statement of results. Our goal in this section is to prove Theorem 4.

8.2. Remarks on some related works. Let us say a few words on the existing work. The readers are referred to the papers of Hsia [24] and Schulze-Pillot [53] for informative surveys on this topic.

(i) In the 1950s, Cassels [7, p.87] proved that if $B = (b_{ij})$ is a non-singular symmetric integral matrix with Q_B isotropic, then $\tau^t B\tau = 0$ has a solution $\tau \in \mathbb{Z}^n$ with $0 < \|\tau\| \leq (3\sum |b_{ij}|)^{\frac{n-1}{2}}$. His proof involves an ingenious use of Diophantine approximation.

(ii) Starting from this example, we shall assume that Q_A (resp. Q_B) is an m-ary (resp. n-ary) non-singular integral quadratic form. For $m = 1$ Watson [60] showed that if B represents A, then $\tau^t B\tau = A$ has a solution $\tau \in \mathbb{Z}^n$ with $\|\tau\| < c|A|^{\frac{1}{2}}$ provided that Q_B is anisotropic or isotropic with $n \geq 4$. For isotropic forms his proof involves the study of congruence relations. For anisotropic forms he uses reduction theory of positive definite quadratic forms. The constant c is effective and depends only on B, but its relation to B was not explicit in [60]. For $n \geq 5$ the constant was described by Kornhouser [33]. For the polynomial search bounds for the integral solutions of quadratic Diophantine equations, see the work of Kornhouser [33] for $n \geq 5$ and the work of Dietmann [14] for $n \geq 3$.

(iii) On the arithmetic side, O'Meara [47] has given necessary and sufficient conditions to decide if Q_B represents Q_A over the p-adic integers for any prime p (in fact, his result applies to any non-dyadic local field). Suppose that Q_B is indefinite, and that there is no local obstruction or equivalently the genus of Q_B represents Q_A. Then Kneser's strong approximation theorem [31] for the spinor groups implies that

Q_B represents Q_A provided that $n-m \geq 3$. When $n-m \leq 2$, an effective procedure of deciding which classes in the genus of Q_B represent Q_A is given in the paper of Hsia-Shao-Xu [25]. For more related works, the readers shall consult the surveys [24, 53] and the references therein.

8.3. Hermite normal form. We now recall a result on the Hermite normal form [52] of an integral matrix.

Lemma 19. *Let $\tau \in \mathbb{Z}^{n\times m}$ be of full column rank ($n \geq m$). Then there exists $\sigma \in \mathbf{GL}_n(\mathbb{Z})$ such that (a) the first m rows of $\sigma\tau$ form an upper triangular square matrix whose entries are non-negative; (b) the entries of the last $n-m$ rows of $\sigma\tau$ are all zero; (c) $\|\sigma\tau\| \leq d$ where d is the g.c.d. of the maximal minors of τ. Moreover, if $n = m$ then there exists $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ such that $\tau\gamma$ satisfies (a), (b) and (c).*

Proof. Let d_1 be the g.c.d. of the first column. Then by Euclid's algorithm, there exists $\sigma_1 \in \mathbf{GL}_n(\mathbb{Z})$ so that the first column of $\sigma_1\tau$ equals $(d_1, 0, \cdots, 0)^t$ (this amounts to perform elementary row operations). For the second column $(x_{12}, \cdots, x_{n2})$ of $\sigma_1\tau$, let d_2 be the g.c.d. of $(x_{22}, \cdots, x_{n2})$. Again by Euclid's algorithm, there exists $\sigma_2 \in \mathbf{GL}_n(\mathbb{Z})$ keeping the first column of $\sigma_2\sigma_1\tau$ unchanged while transforming the second column into $(x_{12}, d_2, 0\cdots, 0)^t$ which also satisfies $0 \leq x_{12} < d_2$. One can continue the process to find $\sigma \in \mathbf{GL}_n(\mathbb{Z})$ with $\sigma\tau$ satisfying (a), (b), and moreover the largest entry in each column of $\sigma\tau$ is the one on the main diagonal. As elementary row operations do not change the g.c.d. of the maximal minors ([52]), the product of the diagonal elements (positive integers) of $\sigma\tau$ equals d and hence (c) follows. When $n = m$, the desired γ can be obtained using the same procedure as the above (starting from the last row). □

8.4. An argument of Siegel. The following lemma originates from Siegel's work [57, Satz 3, Satz 4].

Lemma 20. *Let $n > m$, and $A \in \mathbb{Z}^{m\times m}$, $B \in \mathbb{Z}^{n\times n}$ be non-singular and symmetric. Suppose that $A = \tau_0^t B\tau_0$ for some $\tau_0 \in \mathbb{Z}^{n\times m}$, and let d be the g.c.d. of the maximal minors of τ_0. Then, $0 < d \leq |\det A|$; and there exists $\tau_1 \in \mathbb{Z}^{n\times(n-m)}$ completing τ_0 into a square matrix $\gamma = (\tau_0|\tau_1) \in \mathbb{Z}^{n\times n}$ which satisfies*

$$\det(\gamma) = d, \qquad \|\gamma^t B\gamma\| \ll_n |\det A|^{n-m}|\det B| + \|A\|. \tag{31}$$

Proof. Since $A = \tau_0^t B\tau_0$, the Cauchy-Binet formula (which computes $\det A$ in terms of the $m\times m$ minors of $\tau_0^t B$ and τ_0) implies that $d \neq 0$ and that $\det A$ is divisible by d. Hence $0 < d \leq |\det A|$.

Put $k = n-m$. There exists $\tau' \in \mathbb{Z}^{n\times k}$ completing τ_0 into $\gamma_0 = (\tau_0|\tau')$ with $\det\gamma_0 = d$. To see this, we note that the matrix $\sigma\tau_0$ satisfying (a), (b), (c) in Lemma 19 can be extended into a square integral matrix by setting the last k diagonal elements equal to 1.

To complete the proof of the lemma, we reproduce the argument in [14, Lemma 24] which may be seen as a quantitative version of [57, Satz 3, Satz 4]. With $\gamma_0 = (\tau_0|\tau')$ in the preceding paragraph, one has that

$$B_0 = \gamma_0^t B\gamma_0 = \begin{pmatrix} A & \boldsymbol{c}^t \\ \boldsymbol{c} & D \end{pmatrix}, \qquad \text{with } \boldsymbol{c} \in \mathbb{Z}^{k\times m},\ D \in \mathbb{Z}^{k\times k}.$$

Let $\boldsymbol{r} \in \mathbb{Z}^{k\times m}$, $T \in \mathbf{GL}_k(\mathbb{Z})$, and put $\boldsymbol{s} = \boldsymbol{r} + T^t\boldsymbol{c}A^{-1}$. Direct computation gives us

$$B_0\gamma_1 = \begin{pmatrix} A & A\boldsymbol{s}^t \\ \boldsymbol{c} & \boldsymbol{c}\boldsymbol{r}^t + DT \end{pmatrix} \qquad \text{where } \gamma_1 = \begin{pmatrix} I_m & \boldsymbol{r}^t \\ \boldsymbol{0} & T \end{pmatrix} \in \mathbf{GL}_n(\mathbb{Z}),$$

and

$$\gamma_1^t\gamma_0^t B\gamma_0\gamma_1 = \gamma_1^t B_0\gamma_1 = \begin{pmatrix} A & A\boldsymbol{s}^t \\ \boldsymbol{s}A & T^t\boldsymbol{c}\boldsymbol{r}^t + T^t DT + \boldsymbol{r}A\boldsymbol{s}^t \end{pmatrix}.$$

Let $\gamma = \gamma_0\gamma_1$. Clearly the $n\times m$ integral matrix formed by the first m columns of γ is τ_0, that is, $\gamma = (\tau_0|\tau_1) \in \mathbb{Z}^{n\times n}$. To finish the proof, we need to choose T and $\boldsymbol{r}$ so as to ensure γ satisfying (31). Let us first consider the choice of T. With $E = D - \boldsymbol{c}A^{-1}\boldsymbol{c}^t$, one has that

$$T^t\boldsymbol{c}\boldsymbol{r}^t + T^t DT + \boldsymbol{r}A\boldsymbol{s}^t = (T^t\boldsymbol{c}\boldsymbol{r}^t + T^t DT - T^t\boldsymbol{c}\boldsymbol{s}^t) + (T^t\boldsymbol{c}\boldsymbol{s}^t + \boldsymbol{r}A\boldsymbol{s}^t) = T^t ET + \boldsymbol{s}A\boldsymbol{s}^t.$$

First we would like to control the norm of T^tET. Observe that

$$S^tB_0S = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & E \end{pmatrix} \qquad \text{where} \quad S = \begin{pmatrix} I_m & -A^{-1}\boldsymbol{c}^t \\ \mathbf{0} & I_k \end{pmatrix} \in \mathbf{SL}_n(\mathbb{R}).$$

Write $\lambda = \det A$. Recall that we already showed $0 < d \le |\lambda|$. Therefore,

$$|\lambda \det E| = |\det B_0| = |\det(\gamma_0^t B\gamma_0)| = d^2|\det B| \le \lambda^2|\det B|.$$

Notice that λE is a symmetric, non-singular *integral* matrix. We choose, by Corollary 1, an element $T \in \mathbf{GL}_k(\mathbb{Z})$ such that $\|T^t(\lambda E)T\| \ll_k |\det \lambda E| = |\lambda^{k+1}\det B|$. It follows that $\|T^tET\| \ll_k |\lambda^k \det B|$. Finally we complete the proof of the lemma by choosing an $\boldsymbol{r}$ so that $\|\boldsymbol{s}\| \le 1$. □

8.5. Proof of Theorem 4.

Proof. Suppose that $n = m$. Let $d = \det\tau_0$. The assumption $A = \tau_0^tB\tau_0$ implies that $d^2\det B = \det A$. By Lemma 19, τ_0 can be written as the product of $\tau_0 = \eta_1\eta_2$ where $\eta_1 \in \mathbb{Z}^{n\times n}$ satisfies $\|\eta_1\| \le |d|$ and $\eta_2 \in \mathbf{GL}_n(\mathbb{Z})$. Then, $B_1 = \eta_1^tB\eta_1$ is equivalent to A, and $\|B_1\| \ll_n d^2\|B\| \ll_n |\det A|^2 \cdot \|B\|$. By Theorem 1 there exists an $\eta \in \mathbf{GL}_n(\mathbb{Z})$ with $A = \eta^tB_1\eta$ and

$$\|\eta\| \ll_n |\det A|^{-\frac{13n^3-39n^2+36n+52}{20n}} \cdot \|A\|^{\frac{13n^3-13n^2+52n+20}{40}} \cdot \|B_1\|^{\frac{13n^3-13n^2+72n-20}{40}}.$$

The matrix $\tau = \eta_1\eta$, by which B represents A, satisfies $\|\tau\| \ll |d| \cdot \|\eta\|$.

Suppose that $n > m$. Let $\gamma = (\tau_0|\tau_1)$ satisfy Lemma 20 and, in particular, $\gamma^tB\gamma \in \mathbb{Z}^{n\times n}$ satisfy (31). Applying the argument in the preceding paragraph to $B_2 = \gamma^tB\gamma \in \mathbb{Z}^{n\times n}$ and B, we conclude that there exists a non-singular integral matrix $\gamma_2 \in \mathbb{Z}^{n\times n}$ satisfying $B_2 = \gamma_2^tB\gamma_2$ and

$$\|\gamma_2\| \ll_n |\det A|^{n^4} \cdot |\det B|^{n^3} \cdot \|A\|^{n^3} \cdot \|B\|^{n^3}.$$

Let $\tau \in \mathbb{Z}^{n\times m}$ be formed by the first m columns of γ_2. Then $\tau^tB\tau = \tau_0^tB\tau_0 = A$. Apparently, we have kept a cruder but cleaner exponents for Theorem 4 instead of using a more precise bound.

For primitive representations we refer to Theorem 1 when $n = m$. If $n > m$ and τ_0 is primitive, then so is the $\gamma = (\tau_0|\tau_1)$ in the preceding paragraph and the argument remains the same. □

9. Small generators of integral orthogonal groups

9.1. Statement of results. Throughout Section 9, we let $n = p + q$, Q_0 be as fixed in Section 2, and X_0 be the diagonal matrix of the quadratic form Q_0. We say that two real quadratic forms Q_1, Q_2 are integrally equivalent if $Q_1 = Q_2 \circ \gamma$ for some $\gamma \in \mathbf{GL}_n(\mathbb{Z})$. Our goal in this section is to prove Theorem 2.

9.2. The Siegel domains and reduced real quadratic forms. For any $\alpha, \beta > 0$ we consider

$$D_\alpha = \{\mathrm{diag}(a_1, \cdots, a_n) : 0 < a_i \le \alpha a_{i+1} \text{ for any } 1 \le i \le n-1\} \subset \mathbf{GL}_n(\mathbb{R});$$

$$N_\beta = \{(x_{ij}) : x_{ij} = 1 \text{ if } i = j;\ x_{ij} = 0 \text{ if } i > j;\ |x_{ij}| \le \beta \text{ if } i < j\} \subset \mathbf{SL}_n(\mathbb{R}).$$

Definition 1. *For any $\alpha, \beta > 0$ the set $\mathcal{S}_{\alpha,\beta} = \{kau : k \in O(n), a \in D_\alpha, u \in N_\beta\}$ is called a Siegel domain of $\mathbf{GL}_n(\mathbb{R})$, where $O(n) = \{g \in \mathbf{GL}_n(\mathbb{R}) : g^tg = 1_n\}$ is the orthogonal group of degree n.*

Lemma 21. *(see* [48, Chap. 4, Theorem 4.4]*) If $\alpha \ge \frac{2}{\sqrt{3}}$ and $\beta \ge \frac{1}{2}$, then $\mathbf{GL}_n(\mathbb{R}) = \mathcal{S}_{\alpha,\beta}\mathbf{GL}_n(\mathbb{Z})$.*

The notion of Hermite majorant plays an essential role in reduction theory of indefinite quadratic forms.

Definition 2. *Let Q_B be a real quadratic form of signature (p,q) and suppose $B = g^tX_0g$ where $g \in \mathbf{GL}_n(\mathbb{R})$. The positive definite real quadratic form Q_A with $A = g^tg$ is called an Hermite majorant of B.*

The following lemma can be easily verified using the definition.

Lemma 22. [7, Chap. 13, Lemma 11.1] *If a positive definite quadratic form Q_A is an Hermite majorant of an indefinite real quadratic form Q_B, then $A = BA^{-1}B$.*

Definition 3. *A positive definite real quadratic form Q_A is called (α,β)-reduced if $A = g^t g$ for some $g \in \mathcal{S}_{\alpha,\beta}$. An indefinite real quadratic form is called (α,β)-reduced if it possesses an (α,β)-reduced Hermite majorant.*

Remark 1. *(i) Lemma 21 says that any positive definite real quadratic form is integrally equivalent to a $(\frac{2}{\sqrt{3}},\frac{1}{2})$-reduced quadratic form. This fact is also a consequence of Minkowski's reduction theory.*
(ii) If $A = g^t g = s^t s$ and $g \in \mathcal{S}_{\alpha,\beta}$, then $sg^{-1} \in O(n)$ and hence $s \in \mathcal{S}_{\alpha,\beta}$.
(iii) If Q_A is an (α,β)-reduced positive definite quadratic form, then A can be written as $A = (au)^t(au)$ for some $a \in D_\alpha$ and $u \in N_\beta$.

Lemma 23. *Let $J = (x_{ij})$ be such that $x_{ij} = 1$ if $i + j = n + 1$ and $x_{ij} = 0$ otherwise. Then, for any positive definite (α,β)-reduced quadratic form Q_A with $\beta \le 1$, the quadratic form $Q_{JA^{-1}J}$ is $(\alpha,(n-1)!)$-reduced.*

Proof. (cf. [7, p322]) By (iii) of Remark 1 we can write $A = (au)^t(au)$ with $a \in D_\alpha$, $u \in N_\beta$. Plainly $A^{-1} = u^{-1}a^{-1}a^{-1}(u^{-1})^t$. It is straightforward to check that $Ja^{-1}J \in D_\alpha$. Notice that $\|u\| = 1$ as $\beta \le 1$. So $\|u^{-1}\| \le (n-1)!$ by Cramer's rule, and $J(u^{-1})^t J \in N_{(n-1)!}$. □

9.3. Successive minima of positive definite real quadratic forms.

Definition 4. *Let $1 \le i \le n$, and Q be a positive definite real quadratic form. The i-th minimum of Q is the minimum among all the numbers M such that $\{\boldsymbol{v} \in \mathbb{Z}^n : Q(\boldsymbol{v}) \le M\}$ spans a subspace of dimension $\ge i$.*

Lemma 24. *Let $\alpha > 1$, Q be an (α,β)-reduced positive definite quadratic form, and M be the n-th minimum of Q. Then, any $\boldsymbol{w} \in \mathbb{Z}^n$ with $Q(\boldsymbol{w}) \le M$ satisfies $\|\boldsymbol{w}\| \ll_n \alpha^{2n-2}\beta^n$.*

Proof. The proof is given in [7, p266-269]. Write $Q = (au)^t(au)$ with $a \in D_\alpha$ and $u \in N_\beta$. In the notation of [7, p266-269], the quadratic form Q is in the Siegel domain $\mathscr{S}_n(\alpha^2,\beta)$ (cf. [7, Definition, p259]). As $\alpha > 1$ and $Q(\boldsymbol{w}) \le M$, we have $\|u\boldsymbol{w}\|^2 \ll_n (\alpha^2)^{n-1}$ for case (i) in [7, p268], and $\|u\boldsymbol{w}\|^2 \ll_n (\alpha^2)^{n-1}(\alpha^2)^{n-1}\beta^2$ for case (ii). Hence $\|\boldsymbol{w}\| \ll_n \|u^{-1}\| \cdot \|u\boldsymbol{w}\| \ll_n \alpha^{2n-2}\beta^n$ (cf. [7, Corollary, p268]). □

The following result can be seen as a quantitative version of [7, Chap. 12, Theorem 1.4].

Corollary 2. *Let $\alpha > 1$ and $\beta > 0$. Assume that $\gamma \in (\mathcal{S}^{-1}_{\alpha,\beta} \cdot \mathcal{S}_{\alpha,\beta}) \cap \mathbf{GL}_n(\mathbb{Z})$. Then $\|\gamma\| \ll_n \alpha^{2n^2-2n}\beta^{n^2}$.*

Proof. By assumption $g_2\gamma = g_1$ for some $g_1, g_2 \in \mathcal{S}_{\alpha,\beta}$. Let Q_1 and Q_2 be the positive definite quadratic forms defined by $g_1^t g_1$ and $g_2^t g_2$, respectively. Then, Q_1 and Q_2 are integrally equivalent and hence they share the same n-th minimum M. Let $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n \in \{\boldsymbol{v} \in \mathbb{Z}^n : Q_1(\boldsymbol{v}) \le M\}$ be a linearly independent subset. As Q_1 is (α,β)-reduced, the integral matrix $\tau_1 = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n) \in \mathbb{Z}^{n\times n}$ (the i-th column of τ_1 is given by the column vector $\boldsymbol{w}_i \in \mathbb{Z}^n$) satisfies $\|\tau_1\| \ll_n \alpha^{2n-2}\beta^n$ by Lemma 24. We note that $Q_2(\gamma\boldsymbol{w}_i) = Q_1(\boldsymbol{w}_i) \le M$ and recall that Q_2 is also (α,β)-reduced. Hence by Lemma 24 the integral matrix $\tau_2 = (\gamma\boldsymbol{w}_1, \cdots, \gamma\boldsymbol{w}_n) \in \mathbb{Z}^{n\times n}$ also satisfies $\|\tau_2\| \ll_n \alpha^{2n-2}\beta^n$. As $\gamma = \tau_2\tau_1^{-1}$, we have $\|\gamma\| \ll_n \|\tau_2\| \cdot \|\tau_1^{-1}\|$. The corollary follows. □

9.4. The size of the coefficients of reduced integral quadratic forms. The following lemma is a quantitative version of [7, Chap. 13, Lemma 11.3].

Lemma 25. *Let $T \in \mathbb{Z}^{n\times n}$ be non-singular, and $Q = Q_C$ be an (α,β)-reduced positive definite real quadratic form, where $\alpha > 1$ and $\beta > 0$. Denote by Q_1 the quadratic form of $T^t CT$, and M the n-th minimum of Q_1. Then, any $\boldsymbol{w} \in \mathbb{Z}^n$ with $Q_1(\boldsymbol{w}) \le M$ satisfies $\|T\boldsymbol{w}\| \ll_n |\det T|^{3n-1}\alpha^{2n-2}\beta^n$.*

To explain this result, we note that for a given positive definite quadratic form Q and a lattice Λ of $\mathbb{R}^n$, one can similarly define the successive minima of Q with respect to Λ. In Lemma 25 the n-th minimum of $Q = Q_C$ with respect to the lattice $T\mathbb{Z}^n$ is equal to the n-th minimum of $Q_1 = Q_{T^tCT}$ with respect to $\mathbb{Z}^n$. Essentially, the lemma says that for any lattice point $\boldsymbol{v} \in T\mathbb{Z}^n$ with $Q(\boldsymbol{v}) \le M$, the norm $\|\boldsymbol{v}\|$ is bounded by a polynomial in the parameters α, β of the Siegel domain where Q_C lies, and the index $[\mathbb{Z}^n : T\mathbb{Z}^n] = |\det T|$ (see [7, p.323] for a similar discussion). Clearly, the lemma generalizes Lemma 24.

Proof. Let $d = |\det T|$. By Lemma 19, we can write $T = T_0\gamma$ with $\gamma \in \mathbf{GL}_n(\mathbb{Z})$ and $T_0 = a_1u_1$, where a_1 is diagonal with $\|a_1\| \le d$ and u_1 is a unipotent upper triangular real matrix with $\|u_1\| = 1$ (as in proof of Lemma 19 we assume the largest number in each row of T_0 to be the diagonal entry). Let $Q = (au)^t(au)$ with $a \in D_\alpha$, and $u \in N_\beta$. It is easy to check that $aua_1u_1 = a'u'$ with $a' \in D_{d\alpha}$ and $u' \in N_{nd\beta}$. Let Q_2 be the quadratic form defined by $(a'u')^t a'u'$. Clearly Q_2 is $(d\alpha, nd\beta)$-reduced, and $Q_2 \circ \gamma = Q_1$. Hence the n-th minimum of Q_2 is M; and if $Q_1(\boldsymbol{w}) \le M$, then $Q_2(\gamma\boldsymbol{w}) \le M$. By Lemma 24, $\|\gamma\boldsymbol{w}\| \ll_n (d\alpha)^{2n-2}(nd\beta)^n$. The lemma follows immediately from the fact that $\|T\boldsymbol{w}\| \ll_n \|T_0\| \cdot \|\gamma\boldsymbol{w}\|$. □

The next result is a quantitative version of [7, Chap. 13, Theorem 11.1].

Corollary 3. *Let Q_B be a $(2,1)$-reduced indefinite integral quadratic form. Then $\|B\| \ll_n |\det B|^{3n-1}$.*

Proof. Let Q_A be an Hermite majorant of Q_B which is $(2,1)$-reduced, M be the n-th minimum of A, and $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n \in \{\boldsymbol{v} \in \mathbb{Z}^n : Q_A(\boldsymbol{v}) \le M\}$ be a linearly independent subset. As Q_A is $(2,1)$-reduced, one has $\|\boldsymbol{w}_i\| \ll_n 1$ by Lemma 24. Write $JB = T$ and $C = JA^{-1}J$. By Lemma 22 and Lemma 23, we get that $A = T^tCT$ and that Q_C is $(2,(n-1)!)$-reduced. By Lemma 25, $\|T\boldsymbol{w}_i\| \ll_n |\det T|^{3n-1}$. The same argument in the proof of Corollary 2 will give that $\|T\| \ll_n |\det T|^{3n-1}$. The corollary follows since $B = JT$. □

9.5. Proof of Theorem 2 for indefinite integral quadratic forms.

Proof. Assume that the signature of Q_B is (p,q). Write $Q = Q_B$ and H_Q the identity component of $\mathbf{O}_Q(\mathbb{R})$. Let $g \in \mathbf{GL}_n(\mathbb{R})$ be such that $B = g^tX_0g$ and hence $H_Q = g^{-1}Hg$. We use $\alpha = 2 > \frac{2}{\sqrt{3}}$ and $\beta = 1 > \frac{1}{2}$ for the parameters of the Siegel domain. As a remark, we will make no effort to optimize the bounds.

In this paragraph, we make a reduction of the proof and also indicate the plan. Let s_1H_Q, s_2H_Q, s_3H_Q, s_4H_Q be an enumeration of the four connected components of $\mathbf{O}_Q(\mathbb{R})$ with $s_1 = 1_n$. Put $\sigma_1 = 1_n$. For $j > 1$, put $\sigma_j = 1_n$ if s_jH_Q does not contain any integral point; otherwise put $\sigma_j \in \mathbf{O}_Q(\mathbb{Z}) \cap s_jH_Q$ with $\|\sigma_j\| \ll_n \|B\|^{n^3}$, by the norm bound (30) in Lemma 18. It remains to give a bound on the norm of a finite set of generators of $\Lambda = H_Q \cap \mathbf{GL}_n(\mathbb{Z})$. To approach this problem, we will construct an open subset $U \subset H_Q$ with $U\Lambda = H_Q$ following an argument of Borel-Harish-Chandra [4, Theorem 6.5]. Then, as a well known result, $\{\gamma \in \Lambda : U \cdot \gamma \cap U \ne \emptyset\}$ generates Λ (see e.g. [4, Lemma 6.6]).

In this paragraph, we make some preparation. By Corollary 3, there are only finitely many $(2,1)$-reduced quadratic forms which are integrally equivalent to Q. Let $Q_{A_1}, \cdots, Q_{A_i}, \cdots Q_{A_k}$ be an enumeration of these forms, and we have $\|A_i\| \ll_n |\det B|^{3n-1}$ by Corollary 3. We fix, by Theorem 1, for each $1 \le i \le k$ an element $\tau_i \in \mathbf{GL}_n(\mathbb{Z})$ such that $\tau_i^t A_i \tau = B$, and $\|\tau_i\| \ll_n |\det B|^{n^5}\|B\|^{n^3}$ according to the estimate (3).

Let $\{\gamma_1, \cdots, \gamma_l, \cdots, \gamma_{4k}\} = \{\tau_i\sigma_j : 1 \le i \le k, 1 \le j \le 4\}$. Our task in this paragraph is to justify

$$H_Q \subset U \cdot \Lambda, \qquad \text{where} \quad U = \bigcup_{1 \le l \le 4k} \left((g^{-1} \cdot \mathcal{S}^\circ_{2,1} \cdot \gamma_l) \cap H_Q\right).$$

That is, let h' be any element in H_Q, and we want to show $h' = u\sigma$ with $u \in U$ and $\sigma \in \Lambda$. Since $H_Q = g^{-1}Hg$ and $h' \in H_Q$, we can write $h' = g^{-1}hg$ for some $h \in H$. By Lemma 21, $hg = s\gamma$ for some $s \in \mathcal{S}_{\frac{2}{\sqrt{3}},\frac{1}{2}} \subset \mathcal{S}^\circ_{2,1}$ and $\gamma \in \mathbf{GL}_n(\mathbb{Z})$. As $(g^{-1}s\gamma)^tB(g^{-1}s\gamma) = B$, the symmetric matrix $(g^{-1}s)^tB(g^{-1}s) = s^tX_0s$ defines a $(\frac{2}{\sqrt{3}},\frac{1}{2})$-reduced quadratic form of signature (p,q) which is equivalent to Q through γ. Hence

$s^t X_0 s = A_i$ for some $1 \le i \le k$. There exists $1 \le l \le 4k$, by our choice of γ_i and σ_j, such that $\gamma_l^t A_i \gamma_l = B$ and that $\gamma_l^{-1}\gamma \in \Lambda$. Therefore, $h' = g^{-1}s\gamma_l\gamma_l^{-1}\gamma$ with $u = g^{-1}s\gamma_l \in U$ and $\gamma_l^{-1}\gamma \in \Lambda$. We are done.

Recall that the group Λ is generated by the subset $\{\gamma \in \Lambda : U\cdot\gamma \cap U \neq \emptyset\}$. Notice that

$$\{\gamma \in \Lambda : U\cdot\gamma \cap U \neq \emptyset\} \subset (U^{-1}U) \cap \mathbf{GL}_n(\mathbb{Z}) \subset \bigcup_{1\le l,l'\le 4k} \left(\gamma_{l'}^{-1}\cdot\left((\mathcal{S}_{2,1}^{-1}\cdot\mathcal{S}_{2,1})\cap \mathbf{GL}_n(\mathbb{Z})\right)\cdot\gamma_l\right).$$

By Corollary 2, any $\tau \in (\mathcal{S}_{2,1}^{-1}\cdot\mathcal{S}_{2,1})\cap\mathbf{GL}_n(\mathbb{Z})$ satisfies $\|\tau\| \ll_n 1$. For any $1 \le l \le 4k$ we have $\|\gamma_l\| \ll_n \|\tau_i\sigma_j\| \ll_n |\det B|^{n^5}\|B\|^{2n^3}$. Hence $\|\xi\| \ll_n |\det B|^{n^6}\|B\|^{2n^4}$ for any element $\xi \in \{\gamma \in \Lambda : U\cdot\gamma\cap U \neq \emptyset\}$.

The generators that we exhibit for the group $\mathbf{O}_Q(\mathbb{Z})$ are $\{\sigma_j\xi : 1 \le j \le 4,\ \xi \in (U^{-1}U)\cap\Lambda\}$. Recall that $\|\sigma_j\| \ll_n \|B\|^{n^3}$. Therefore, each element in this generating set satisfies the norm bound in the theorem. □

Appendix: On the compact subset in Lemma 13

This is a special case of [17, Appendix]. As in the approach of [17], our plan is also to make connection with [30, Theorem 5.2, Theorem 5.3].

In this paragraph, we set up notation and review some related facts. Let L be a lattice in $\mathbb{R}^n$. We denote the successive minima (with respect to the Euclidean norm in $\mathbb{R}^n$, see [8]) of L by $\lambda_1(L) \le \cdots \le \lambda_n(L)$. Recall that G/Γ is naturally identified with the space of unimodular lattices in $\mathbb{R}^n$ via the map $[g] \mapsto g\mathbb{Z}^n$. Thanks to the Mahler's criterion $\mathfrak{S}(\epsilon) := \{[g] \in G/\Gamma : \lambda_1(g\mathbb{Z}^n) \ge \epsilon\}$ is a compact subset in G/Γ for every $\epsilon > 0$. Our goal is to prove that $\Omega_{\mathrm{cpt}} = \mathfrak{S}(\epsilon_n)$ satisfies (19), where $\epsilon_n = (4n^3 6^n (n^3+n)^{1/n^2})^{-n^2}$.

In this paragraph, we prove a lemma. Let us fix a one dimensional unipotent subgroup $U = \{u_t : t \in \mathbb{R}\}$ inside H such that the normal subgroup generated by U coincides with H.

Lemma 26. *Let $g \in G$ and $\Sigma_1, \cdots, \Sigma_k$ be an enumeration of the primitive sub-lattices of $g\mathbb{Z}^n$ of covolume less than $1/n$ and dimension strictly less than n (there are only finitely many of them, see [30]). Then*

$$\dim\left(\Psi_{[g]} = \{h \in H : h^{-1}Uh.\Sigma_i = \Sigma_i \text{ for some } i = 1, \cdots, k\}\right) < \dim H.$$

Proof. (cf. [17, p.207-208]) For each i let $Y_i := \{h \in H : h^{-1}Uh.\Sigma_i = \Sigma_i\}$. Suppose $\dim Y_i = \dim H$. Then $Y_i = H$ since it is Zariski closed in H. Recall it is assumed that the normal subgroup generated by U is H. From this we get H preserves Σ_i. But this cannot happen, because for the standard representation of $\mathbf{SL}_n(\mathbb{R})$ on $\mathbb{R}^n$ the restriction to H is irreducible. Hence $\dim Y_i < \dim H$ and the lemma follows. □

Now let $H.[g]$ be a closed orbit in G/Γ. By Mautner's phenomenon, U acts ergodically on the probability space $(G/\Gamma, \mu_{H.[g]})$ by left translations. Let χ_m be the indicator function of the set $\mathfrak{S}(1/m)$ where m is any positive integer. By Birkhoff ergodic theorem, at $\mu_{H.[g]}$-almost all $x \in H.[g]$ the ergodic average for the function χ_m along U converges to $\int \chi_m d\mu$ for every positive integer m. Let $h.[g]$ be such a point. For later use let us record the fact that, since $\Psi_{[g]}$ is nowhere dense in H, the element h can be chosen in $H - \Psi_{[g]}$, and arbitrarily close to the identity element 1_n.

Notice that $h^{-1}Uh$ does not preserve any of the primitive sub-lattices $\Sigma_1, \cdots, \Sigma_k$ of $g\mathbb{Z}^n$ in Lemma 26. By [30, Theorem 5.2, Theorem 5.3], there exists $T_0 = T_0(U, h, g)$ such that for any $T \ge T_0$ and $0 < \epsilon < 1/n$

$$|\{0 < t < T : h^{-1}u_t hg\mathbb{Z}^n \notin \mathfrak{S}(\epsilon)\}| \le a_n \epsilon^{1/n^2} T, \qquad a_n = 2n^3 6^n (n^3+n)^{1/n^2}. \tag{32}$$

Let $\sigma_n = (4a_n/3)^{-n^2}$, and $\epsilon_n = (2a_n)^{-n^2}$. We assume, as we mentioned, that h is so close to the identity element 1_n that $h\mathfrak{S}(\sigma_n) \subset \mathfrak{S}(\epsilon_n)$. We shall now justify that $\Omega_{\mathrm{cpt}} = \mathfrak{S}(\epsilon_n)$ satisfies (19). Since $1/\epsilon_n$ is a positive integer, by the choice of $h.[g]$ we have

$$\mu(\mathfrak{S}(\epsilon_n)) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \chi_{1/\epsilon_n}(u_t hg\mathbb{Z}^n)dt \ge 1 - \liminf_{T\to\infty}\frac{|\{0<t<T : u_t hg\mathbb{Z}^n \notin h\mathfrak{S}(\sigma_n)\}|}{T} \overset{(32)}{\ge} 0.25.$$

References

[1] V. Blomer, F. Brumley: *On the Ramanujan conjecture over number fields.* Ann. of Math. (2) 174 (2011), 581-605.

[2] N. Bergeron, L. Clozel: *Quelques conséquences des travaux d'Arthur pour le spectre et la topologie des variétés hyperboliques.* Invent. Math. 192 (2013), no. 3, 505-532.

[3] J. W. Benham, J. S. Hsia: *Spinor equivalence of quadratic forms.* J. Number Theory 17 (1983), no. 3, 337-342.

[4] A. Borel, Harish-Chandra: *Arithmetic subgroups of algebraic groups.* Ann. of Math. (2) 75 (1962) 485-535.

[5] M. Burger, P. Sarnak: *Ramanujan duals II.* Invent. Math. 106 (1991), no. 1, 1-11.

[6] M. Burger, V. Schroeder: *Volume, diameter and the first eigenvalue of locally symmetric spaces of rank one.* J. Differential Geom. 26 (1987), no. 2, 273-284.

[7] J. W. S. Cassels: *Rational quadratic forms.* Academic Press, London. (1978)

[8] J. W. S. Cassels: *An introduction to the geometry of numbers.* (second printing) Springer-Verlag, 1971.

[9] L. Clozel: *Démonstration de la conjecture τ.* Invent. Math. 151 (2003), no. 2, 297-328.

[10] J. H. Conway, N. J. A. Sloane: Sphere packings, lattices and groups. Springer-Verlag, New York, 1999.

[11] M. Cowling: *Sur les coefficients des representations unitaires des groupes de Lie simples,* In: Analyse Harmonique sur les Groupes de Lie, II. Lecture Notes in Math. (739), 132-178. Springer, Berlin (1979)

[12] S. G. Dani: *On orbits of unipotent flows on homogeneous spaces. II.* Ergodic Theory Dynam. Systems 6 (1986), 67-182.

[13] S. G. Dani, G. A. Margulis: *Values of quadratic forms at primitive integral points.* Invent. Math. 98 (1989), no. 2, 405-424.

[14] R. Dietmann: *Small solutions of quadratic Diophantine equations.* Proc. London Math. Soc. (3) 86 (2003), 545-582.

[15] R. Dietmann: *Polynomial bounds for equivalence of quadratic forms with cube-free determinant,* Math. Proc. Cambridge Philos. Soc. 143 (2007), no. 3, 521-532.

[16] W. Duke, Z. Rudnick, P. Sarnak: *Density of integer points on affine homogeneous varieties.* Duke Math. J. 71 (1993), no. 1, 143-179.

[17] M. Einsiedler, G. A. Margulis, A. Venkatesh: *Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces.* Invent. Math. 177 (2009), no. 1, 137-212.

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

[19] A. Eskin, G. A. Margulis, S. Mozes: *Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture.* Ann. of Math. (2) 147 (1998), no. 1, 93-141.

[20] Faraut, J: *Analysis on Lie groups. An introduction.* Cambridge University Press, Cambridge, 2008.

[21] L. Fukshansky: *Heights and quadratic forms: on Cassels' theorem and its generalizations,* Contemp. Math. (2013), 77-94.

[22] F. Grunewald, D. Segal: Some general algorithms. I. Arithmetic groups. Ann. of Math. (2) 112 (1980), no. 3, 531-583.

[23] T. Hirai: *On irreducible representations of the Lorentz group of n-th order.* Proc. Japan. Acad, 38, 258-262, 1962.

[24] J. S. Hsia: *Arithmetic of indefinite quadratic forms.* Contemp. Math., 249, Amer. Math. Soc., Providence, RI, 1999.

[25] J. S. Hsia, Y. Y. Shao, F. Xu: *Representations of indefinite quadratic forms.* J. Reine Angew. Math. 494 (1998), 129-140.

[26] R. Howe, E-C. Tan: *Nonabelian harmonic analysis. Applications of* $\mathbf{SL}(2,\mathbb{R})$. Springer-Verlag, New York, 1992.

[27] H. Jacquet, R. Langlands: *Automorphic forms on $GL(2)$.* Springer-Verlag, Berlin, New York, 1970.

[28] H. H. Kim, P. Sarnak: *Refined estimates towards the Ramanujan and Selberg conjectures,* Appendix to *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2, J. Amer. Math. Soc. **16** (2003), no. 1, 139-183.

[29] D. Kleinbock, G. A. Margulis: *Bounded orbits of nonquasiunipotent flows on homogeneous spaces.* Amer. Math. Soc. Transl. (2) 171 (1996), 141-172.

[30] D. Kleinbock, G. A. Margulis: *Flows on homogeneous spaces and Diophantine approximation on manifolds.* Ann. Math. 148, 339-360 (1998)

[31] M. Kneser: Klassenzahlen indefiniter quadratischer Formen in drei oder mehr Veränderlichen. Arch. Math. 7 (1956), 323-332.

[32] A, Kontorovich, H. Oh: *Almost prime Pythagorean triples in thin orbits.* J. Reine Angew. Math. 667 (2012), 89-131.

[33] D. M. Kornhauser: *On small solutions of the general nonsingular quadratic Diophantine equation in five and more unknowns.* Math. Proc. Cam. Philos. Soc., 107, 197-211 (1990)

[34] J. Lagarias: *On the computational complexity of determining the solvability or unsolvability of the equation $x^2 - dy^2 = -1$.* Trans. Amer. Math. Soc. 260, 485-508 (1980)

[35] J-S. Li: *The minimal decay of matrix coefficients for classical groups.* Harmonic analysis in China, 146-169, Math. Appl., 327, Kluwer Acad. Publ., Dordrecht, 1995.

[36] E. Lindenstrauss, G. A. Margulis: *Effective estimates on indefinite ternary forms,* Israel J. of Math., to appear.

[37] G. A. Margulis: *The action of unipotent groups in a lattice space,* Mat. Sb. (N.S.) 86(128), 552-556 (1971)

[38] G. A. Margulis: *Indefinite quadratic forms and unipotent flows on homogeneous spaces.* Proceeding of "Semester on dynamical systems and ergodic theory" (Warsaw 1986) 399-409, Banach Center Publ., **23** (1989).

[39] G. A. Margulis: *Discrete subgroups of semisimple Lie groups,* Springer-Verlag, (1991)

[40] G. A. Margulis: *Random walks on the space of lattices and the finiteness of covolumes of arithmetic subgroups.* Algebraic groups and arithmetic, 409-425, Tata Inst. Fund. Res., Mumbai, 2004.

[41] D. W. Masser: *Search bounds for Diophantine equations,* A Panorama of Number Theory or the View from Baker's Garden (Zurich 1999), 247-259.

[42] A. Mohammadi, H. Oh: *Matrix coefficients, Counting and Primes for orbits of geometrically finite groups.* To appear in Journal of European Math. Society. arXiv:1208.4139.

[43] C. C. Moore: *Ergodicity of flows on homogeneous spaces,* Amer. J. Math. 88, 154-178 (1966).

[44] C. C. Moore: *Exponential decay of correlation coefficients for geodesic flows.* Group representations, ergodic theory, operator algebras, and mathematical physics (Berkeley, Calif., 1984), 163-181, Springer, New York, 1987.

[45] M. Nakasuji: *Generalized Ramanujan conjecture over general imaginary quadratic fields.* Forum Math. 24 (2012), no. 1, 85-98.

[46] H. Oh: *Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants.* Duke Math. J. 113 (2002), no. 1, 133-192.

[47] O. T. O'Meara: *The integral representations of quadratic forms over local fields.* Amer. J. Math. 80, 1958, 843-878.

[48] V. Platonov, A. Rapinchuk: *Algebraic groups and number theory.* Academic Press, INC, 1994

[49] G. Prasad: *Volumes of S-arithmetic quotients of semi-simple groups.* I.H.E.S. Publ. Math. No. 69 (1989), 91-117.

[50] P. Sarnak: *Notes on the generalized Ramanujan conjectures.* Harmonic analysis, the trace formula, and Shimura varieties, 659-685, Clay Math. Proc., 4, Amer. Math. Soc., Providence, RI, 2005.

[51] A. Schinzel: *Integer points on conics,* Ann. Soc. Math. Polon., Ser I: Comment. Math. 16, 133-135.

[52] A. Schrijver: *Theory of linear and integer programming.* John Wiley & Sons, Ltd., Chichester, 1986.

[53] R. Schulze-Pillot: *Representation of quadratic forms by integral quadratic forms.* Quadratic and higher degree forms, 233-253, Dev. Math., 31, Springer, New York, 2013.

[54] A. Selberg: *On the estimation of Fourier coefficients of modular forms,* Proceedings of Symposia in Pure Mathematics VIII, Providence, R.I. AMS, 1-15 (1965)

[55] Y. Shalom: *Rigidity, unitary representations of semisimple groups, and fundamental groups of manifolds with rank one transformation group.* Ann. of Math. (2), **152**(1), 113-182, (2000).

[56] C. L. Siegel: *Einheiten quadratischer Formen,* Abh. Math. Sem. Hansischen Univ. 13, (1940). 209-239.

[57] C. L. Siegel: *Zur Theorie der quadratische Formen,* Nachr. Akad. Wiss. Göttingen MTH. Phys. Kl. II, 21-46 (1972)

[58] S. Straumann: *Das Äquivalenzproblem ganzer quadratischer Formen: Einige explizite Resultate,* Diplomarbeit. Universität Basel (1999)

[59] B. A. Venkov: *Elementary number theory,* Translated from the Russian and edited by Helen Alderson Wolters-Noordhoff Publishing, Groningen 1970

[60] G. L. Watson: *Bounded representations of integers by quadratic forms.* Mathematika 4, 1957, 17-24.

(H. L.) Department of Mathematics and Computer Science, Wesleyan University, Middletown, CT 06459, USA.
E-mail address: hli03@wesleyan.edu

(G. M.) Department of Mathematics, Yale University, New Haven, CT 06511, USA.
E-mail address: margulis@math.yale.edu