Algebra & Number Theory

Volume 8
2014
No. 4

Equidistribution of values of linear forms on quadratic surfaces

Oliver Sargent

msp

ALGEBRA AND NUMBER THEORY 8:4 (2014)
dx.doi.org/10.2140/ant.2014.8.895

Equidistribution of values of linear forms on quadratic surfaces

Oliver Sargent

In this paper, we investigate the distribution of the set of values of a linear map at integer points on a quadratic surface. In particular, it is shown that, subject to certain algebraic conditions, this set is equidistributed. This can be thought of as a quantitative version of the main result from a previous paper. The methods used are based on those developed by A. Eskin, S. Mozes and G. Margulis. Specifically, they rely on equidistribution properties of unipotent flows.

1. Introduction

Consider the following situation. Let X be a rational surface in $\mathbb{R}^d$, R be a fixed region in $\mathbb{R}^s$ and $F : X \to \mathbb{R}^s$ be a polynomial map. An interesting problem is to investigate the size of the set

$$Z = \{x \in X \cap \mathbb{Z}^d : F(x) \in R\}$$

consisting of integer points in X such that the corresponding values of F are in R. Suppose that the set of values of F at the integer points of X is dense in $\mathbb{R}^s$. In this case, the set Z will be infinite. However, the set

$$Z_T = \{x \in X \cap \mathbb{Z}^d : F(x) \in R,\ \|x\| \leq T\}$$

can be considered. This set will be finite, and its size will depend on T. Typically, the density assumption indicates that the set Z might be equidistributed within the set of all integer points in X. Namely, as T increases, the size of the set Z_T should be proportional to the appropriately defined volume of the set

$$\{x \in X : F(x) \in R,\ \|x\| \leq T\}$$

consisting of real points on X with values in R and bounded norm. Such a result, if it is obtained, can be seen as quantifying the denseness of the values of F at integral points.

MSC2010: primary 11E99; secondary 37A17, 37A45.
Keywords: quadratic forms, linear maps, integral values, unipotent flows.

The situation described above is too general, but it serves as motivation for what is to come. So far, what is proved is limited to special cases. For instance, when $M : \mathbb{R}^d \to \mathbb{R}^s$ is a linear map, classical methods can be used to establish necessary and sufficient conditions that ensure the values of M on $\mathbb{Z}^d$ are dense in $\mathbb{R}^s$. The equidistribution problem described above can also be considered in this case. It is straightforward to obtain an asymptotic estimate for the number of integer points with bounded norm whose values lie in some compact region of $\mathbb{R}^s$ [Cassels 1972].

When $Q : \mathbb{R}^d \to \mathbb{R}$ is a quadratic form, the situation is that of the Oppenheim conjecture. Margulis [1989] obtained necessary and sufficient conditions to ensure that the values of Q on $\mathbb{Z}^d$ are dense in $\mathbb{R}$. Considerable work has gone into the equidistribution problem in this case, first by Dani and Margulis [1993], who obtained an asymptotic lower bound for the number of integers with bounded height such that their images lie in a fixed interval. Later, Eskin, Margulis and Mozes [Eskin et al. 1998] gave the corresponding asymptotic upper bound for the same problem. The major ingredient, used in the proof of Oppenheim conjecture, is to relate the density of the values of a quadratic form at integers to the density of certain orbits inside a homogeneous space. This connection was first noted by M. S. Raghunathan in the late 1970s (appearing in print in [Dani 1981], for instance). It is, in this way, using tools from dynamical systems to study the orbit closures of subgroups corresponding to quadratic forms, that Margulis proved the Oppenheim conjecture. Similarly, the later refinement, due to Dani and Margulis [1990], who considered the values of quadratic forms at primitive integral points, and work on the equidistribution (quantitative) problem by Dani and Margulis and Eskin, Margulis and Mozes, was also obtained by studying the orbit closures of subgroups acting on homogeneous spaces.

Similar techniques were also used by Gorodnik [2004] to study the set of values of a pair, consisting of a quadratic and linear form, at integer points and in [Sargent 2013] to establish conditions sufficient to ensure that the values of a linear map at integers lying on a quadratic surface are dense in the range of the map. The main result of this paper deals with the corresponding equidistribution problem and is stated in the following:

Theorem 1.1. *Suppose Q is a quadratic form on $\mathbb{R}^d$ such that Q is nondegenerate and indefinite with rational coefficients. Let $M = (L_1, \dots, L_s) : \mathbb{R}^d \to \mathbb{R}^s$ be a linear map such that*:

(1) *The following relations hold*: $d > 2s$ *and* $\operatorname{rank}(Q|_{\ker(M)}) = d - s$.

(2) *The quadratic form $Q|_{\ker(M)}$ has signature (r_1, r_2), where $r_1 \geq 3$ and $r_2 \geq 1$.*

(3) *For all $\alpha \in \mathbb{R}^s \setminus \{0\}$, $\alpha_1 L_1 + \cdots + \alpha_s L_s$ is nonrational.*

Let $a \in \mathbb{Q}$ be such that the set $\{v \in \mathbb{Z}^d : Q(v) = a\}$ is nonempty. Then there exists $C_0 > 0$ such that, for every $\theta > 0$ and all compact $R \subset \mathbb{R}^s$ with piecewise smooth *boundary, there exists a* $T_0 > 0$ *such that, for all* $T > T_0$,

$$(1-\theta)C_0 \operatorname{Vol}(R)T^{d-s-2} \leq |\{v \in \mathbb{Z}^d : Q(v) = a,\ M(v) \in R,\ \|v\| \leq T\}| \\ \leq (1+\theta)C_0 \operatorname{Vol}(R)T^{d-s-2},$$

where $\operatorname{Vol}(R)$ *is the* s*-dimensional Lebesgue measure of* R.

Remark 1.2. The constant C_0 appearing in Theorem 1.1 is such that

$$C_0 \operatorname{Vol}(R)T^{d-s-2} \sim \operatorname{Vol}(\{v \in \mathbb{R}^d : Q(v) = a,\ M(v) \in R,\ \|v\| \leq T\}),$$

where the volume on the right is the Haar measure on the surface defined by $Q(v) = a$.

Remark 1.3. Theorem 1.1 should hold with the condition that $\operatorname{rank}(Q|_{\ker(M)}) = d - s$ replaced by the condition that $\operatorname{rank}(Q|_{\ker(M)}) > 3$. Dealing with the more general situation requires taking into account the nontrivial unipotent part of $\operatorname{Stab}_{\mathrm{SO}(Q)}(M)$; as such, lower bounds could probably be proved using methods of [Dani and Margulis 1993], but so far, no way has been found to obtain the statement that would be needed in order to obtain an upper bound.

Remark 1.4. As in [Eskin et al. 1998], it would be possible to obtain a version of Theorem 1.1 where the condition that $\|v\| < T$ was replaced by $v \in TK_0$, where K_0 is an arbitrary deformation of the unit ball by a continuous and positive function. It should also be possible to obtain a version of Theorem 1.1 where the parameters T_0 and C_0 remain valid for any pair (Q, M) coming from compact subsets of pairs satisfying the conditions of the theorem.

Remark 1.5. The cases when the quadratic form $Q|_{\ker(M)}$ has signature $(2, 2)$ or $(2, 1)$ can be considered exceptional. There are asymptotically more integers than expected (by a factor of $\log T$) lying on certain surfaces defined by quadratic forms of signature $(2, 2)$ or $(2, 1)$. This leads to counterexamples of Theorem 1.1 in the cases when the quadratic form $Q|_{\ker(M)}$ has signature $(2, 2)$ or $(2, 1)$. Details of these examples are found in Section 6.

Outline of the paper. The proof of Theorem 1.1 rests on statements about the distribution of orbits in certain homogeneous spaces. The philosophy is that equidistribution of the orbits corresponds to equidistribution of the points considered in Theorem 1.1. Consider the following:

Ratner's equidistribution theorem [Ratner 1994]. *Let* G *be a connected Lie group*, Γ *a lattice in* G *and* $U = \{u_t : t \in \mathbb{R}\}$ *a one-parameter unipotent subgroup of* G. *Then for all* $x \in G/\Gamma$, *the closure of the orbit* Ux *has an invariant measure* $\mu_{\overline{Ux}}$ *supported on it, and for all bounded continuous functions* f *on* G/Γ,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(u_t x) = \int_{\overline{Ux}} f \, d\mu_{\overline{Ux}}.$$

Recall that in the proof of the quantitative Oppenheim conjecture [Eskin et al. 1998] one needs to consider an unbounded function on the space of lattices. Similarly, in order to prove Theorem 1.1, one needs to consider an unbounded function F on a certain homogeneous space. The basic idea is to try to apply Ratner's equidistribution theorem to F in order to show that the average of the values of F evaluated along a certain orbit converges to the average of F on the entire space. This is the fact that corresponds to the fact that integral points on the quadratic surface with values in R are equidistributed. The main problem in doing this is that F is unbounded, and so one must obtain an ergodic theorem taking a similar form to Ratner's equidistribution theorem but valid for unbounded functions. In order to do this, one needs precise information about the behavior of the orbits near the cusp. This information is obtained in Section 3 and comes in the form of nondivergence estimates for certain dilated spherical averages. In order to obtain these estimates, we use a certain function defined by Benoist and Quint [2012]. The required ergodic theorem is then proved in Section 4. Finally in Section 5, the proof of Theorem 1.1 is completed using an approximation argument similar to that found in [Eskin et al. 1998]. Specifically, the averages of F over the space are related to the quantity $C_0 \operatorname{Vol}(R) T^{d-s-2}$ and the averages of F along an orbit are related to the number of integer points with bounded height, lying on the surface and with values in R. In Section 2, the basic notation is set up and the main results from Sections 3 and 4 are stated.

2. Set-up

2A. ***Main results.*** For the rest of the paper, the following convention is in place: s, d and p will be fixed natural numbers such that $2s < d$ and $0 < p < d$. Also, r_1 and r_2 will be varying, natural numbers such that $d - s = r_1 + r_2$. Let $\mathscr{L}$ denote the space of linear forms on $\mathbb{R}^d$, and let $\mathscr{C}_{\mathrm{Lin}}$ denote the subset of $\mathscr{L}^s$ such that for all $M \in \mathscr{C}_{\mathrm{Lin}}$ Condition (3) of Theorem 1.1 is satisfied. A quadratic form on $\mathbb{R}^d$ is said to be defined over $\mathbb{Q}$ if it has rational coefficients or is a scalar multiple of a form with rational coefficients. For a a rational number, let $\mathfrak{Q}(p, a)$ denote quadratic forms on $\mathbb{R}^d$ defined over $\mathbb{Q}$ with signature $(p, d - p)$ such that the set $\{v \in \mathbb{Z}^d : Q(v) = a\}$ is nonempty for all $Q \in \mathfrak{Q}(p, a)$. Define

$$\mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2) = \{(Q, M) : Q \in \mathfrak{Q}(p, a),\ M \in \mathscr{C}_{\mathrm{Lin}} \text{ and } Q|_{\ker(M)} \text{ has signature } (r_1, r_2)\}.$$

Note that for $r_1 \geq 3$ and $r_2 \geq 1$ the set $\mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)$ consists of pairs satisfying the conditions of Theorem 1.1. Although the set $\mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)$ and hence its subsets and sets derived from them depend on a, this dependence is not a crucial one, so from now on, most of the time, this dependence will be omitted from the notation. For $M \in \mathscr{L}^s$ and $R \subset \mathbb{R}^s$ a connected region with smooth boundary, let

$V_M(R) = \{v \in \mathbb{R}^d : M(v) \in R\}$. For $Q \in \mathcal{Q}(p, d-p)$, $a \in \mathbb{Q}$ and $\mathbb{K} = \mathbb{R}$ or $\mathbb{Z}$, let $X_Q^a(\mathbb{K}) = \{v \in \mathbb{K}^d : Q(v) = a\}$. Denote the annular region inside $\mathbb{R}^d$ by $A(T_1, T_2) = \{v \in \mathbb{R}^d : T_1 \leq \|v\| \leq T_2\}$. Using this notation, we state the following (equivalent) version of Theorem 1.1, which will be proved in Section 5:

Theorem 2.1. *Suppose that $r_1 \geq 3$, $r_2 \geq 1$ and $a \in \mathbb{Q}$. Then for all $(Q, M) \in \mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)$, there exists $C_0 > 0$ such that, for every $\theta > 0$ and all compact $R \subset \mathbb{R}^s$ with piecewise smooth boundary, there exists a $T_0 > 0$ such that, for all $T > T_0$,*

$$(1-\theta)C_0 \operatorname{Vol}(R)T^{d-s-2} \leq |X_Q^a(\mathbb{Z}) \cap V_M(R) \cap A(0,T)| \leq (1-\theta)C_0 \operatorname{Vol}(R)T^{d-s-2}.$$

Remark 2.2. As remarked previously, the cases when $r_1 = 2$ and $r_2 = 2$ or $r_1 = 2$ and $r_2 = 1$ are interesting. In dimensions 3 and 4, there can be more integer points than expected lying on some surfaces defined by quadratic forms of signature (2, 2) or (2, 1); this means that the statement of Theorem 2.1 fails for certain pairs. In Section 6, these counterexamples are explicitly constructed. Moreover, it is shown that this set of pairs is big in the sense that it is of second category. We note that as in [Eskin et al. 1998] one could also show that this set has measure 0 and one could prove the expected asymptotic formula as in Theorem 2.1 for almost all pairs.

Even though Theorem 2.1 fails when $r_1 = 2$ and $r_2 = 2$ or $r_1 = 2$ and $r_2 = 1$, we do have the following uniform upper bound, which will be proved in Section 5 and is analogous to Theorem 2.3 from [Eskin et al. 1998]:

Theorem 2.3. *Let $R \subset \mathbb{R}^s$ be a compact region with piecewise smooth boundary and $a \in \mathbb{Q}$.*

(I) *If $r_1 \geq 3$ and $r_2 \geq 1$, then for all $(Q, M) \in \mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)$ there exists a constant C depending only on (Q, M) and R such that, for all $T > 1$,*

$$|X_Q^a(\mathbb{Z}) \cap V_M(R) \cap A(0,T)| \leq CT^{d-s-2}.$$

(II) *If $r_1 = 2$ and $r_2 = 1$ or $r_1 = r_2 = 2$, then for all $(Q, M) \in \mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)$ there exists a constant C depending only on (Q, M) and R such that, for all $T > 2$,*

$$|X_Q^a(\mathbb{Z}) \cap V_M(R) \cap A(0,T)| \leq C(\log T)T^{d-s-2}.$$

2B. ***A canonical form.*** For $v_1, v_2 \in \mathbb{R}^d$, we will use the notation $\langle v_1, v_2 \rangle$ to denote the standard inner product in $\mathbb{R}^d$. For a set of vectors $v_1, \ldots, v_i \in \mathbb{R}^d$, we will also use the notation $\langle v_1, \ldots, v_i \rangle$ to denote the span of $v_1, \ldots, v_i$ in $\mathbb{R}^d$; although this could lead to some ambiguity, the meaning of the notation should be clear from the context.

For some computations, it will be convenient to know that our system is conjugate to a canonical form. Let $e_1, \ldots, e_d$ be the standard basis of $\mathbb{R}^d$. Let (Q_0, M_0) be the pair consisting of a quadratic form and a linear map defined by

$$Q_0(v) = Q_{1,\dots,s}(v) + 2v_{s+1}v_d + \sum_{i=s+2}^{s+r_1} v_i^2 - \sum_{i=s+r_1+1}^{d-1} v_i^2 \quad \text{and} \quad M_0(v) = (v_1, \dots, v_s),$$

where $v_i = \langle v, e_i \rangle$ and $Q_{1,\dots,s}(v)$ is a nondegenerate quadratic form in variables $v_1, \dots, v_s$. By Lemma 2.2 of [Sargent 2013], all pairs (Q, M) such that the signature of $Q|_{\ker(M)}$ is (r_1, r_2) and $\operatorname{rank}(Q|_{\ker(M)}) = d-s$ are equivalent to the pair (Q_0, M_0) in the sense that there exist $g_d \in \mathrm{GL}_d(\mathbb{R})$ and $g_s \in \mathrm{GL}_s(\mathbb{R})$ such that $(Q, M) = (Q_0^{g_d}, g_s M_0^{g_d})$, where for $g \in \mathrm{GL}_d(\mathbb{R})$ we write $Q = Q_0^g$ if and only if $Q_0(gv) = Q(v)$ for all $v \in \mathbb{R}^d$. Moreover, since $R \subset \mathbb{R}^s$ is arbitrary, up to rescaling and possibly replacing R by $g_s R$, we assume that $g_d \in \mathrm{SL}_d(\mathbb{R})$ and that g_s is the identity. Let

$$\mathscr{C}_{\mathrm{SL}}(a, r_1, r_2) = \{g \in \mathrm{SL}_d(\mathbb{R}) : (Q_0^g, M_0^g) \in \mathscr{C}_{\mathrm{Pairs}}(a, r_1, r_2)\}.$$

For $g \in \mathscr{C}_{\mathrm{SL}}(a, r_1, r_2)$, let G_g be the identity component of the group $\{x \in \mathrm{SL}_d(\mathbb{R}) : Q_0^g(xv) = Q_0^g(v)\}$, $\Gamma_g = G_g \cap \mathrm{SL}_d(\mathbb{Z})$, $H_g = \{x \in G_g : M_0^g(xv) = M_0^g(v)\}$ and $K_g = H_g \cap g^{-1} O_d(\mathbb{R}) g$. By examining the description of the subgroup H_g given in Section 2.3 of [Sargent 2013], it is clear that K_g is a maximal compact subgroup of H_g. It is a standard fact that G_g is a connected semisimple Lie group and hence has no nontrivial rational characters. Therefore, because Q_0^g is defined over $\mathbb{Q}$, the Borel–Harish-Chandra theorem [Platonov and Rapinchuk 1991, Theorem 4.13] implies Γ_g is a lattice in G_g. We will consider the dynamical system that arises from H_g acting on G_g/Γ_g. For $\mathbb{K} = \mathbb{R}$ or $\mathbb{Z}$, the shorthand $X^a_{Q_0^g}(\mathbb{K}) = X_g(\mathbb{K})$ will be used.

2C. ***Equidistribution of measures.*** Consider the function α as defined in [Eskin et al. 1998]. It is an unbounded function on the space of unimodular lattices in $\mathbb{R}^d$. It has the properties that it can be used to bound certain functions that we will consider and it is left-K_I-invariant. Similar functions have been considered in [Schnell 1995], where it is related to various quantities involving successive minima of a lattice. Let Δ be a lattice in $\mathbb{R}^d$. For any such Δ, we say that a subspace U of $\mathbb{R}^d$ is Δ-rational if $\operatorname{Vol}(U/U \cap \Delta) < \infty$. Let

$$\Psi_i(\Delta) = \{U : U \text{ is a } \Delta\text{-rational subspace of } \mathbb{R}^d \text{ with } \dim U = i\}.$$

For $U \in \Psi_i(\Delta)$, define $d_\Delta(U) = \operatorname{Vol}(U/U \cap \Delta)$. Note that $d_\Delta(U) = \|u_1 \wedge \cdots \wedge u_i\|$, where $u_1, \dots, u_i$ is a basis for $U \cap \Delta$ over $\mathbb{Z}$ and the norm on $\bigwedge^i(\mathbb{R}^d)$ is induced from the euclidean norm on $\mathbb{R}^d$. Now we recall the definition of the function α:

$$\alpha_i(\Delta) = \sup_{U \in \Psi_i(\Delta)} \frac{1}{d_\Delta(U)} \quad \text{and} \quad \alpha(\Delta) = \max_{0 \le i \le d} \alpha_i(\Delta).$$

Here we use the convention that, if U is the trivial subspace, then $d_\Delta(U) = 1$; hence,

$\alpha_0(\Delta) = 1$. Also note that, if Δ is a unimodular lattice, then $d_\Delta(\mathbb{R}^d) = 1$ and hence $\alpha_d(\Delta) = 1$.

In (2-2) and Theorem 2.5, we consider α as a function on G_g/Γ_g; this is done via the canonical embedding of G_g/Γ_g into the space of unimodular lattices in $\mathbb{R}^d$, given by $x\Gamma_g \to x\mathbb{Z}^d$. Specifically, every $x \in G_g/\Gamma_g$ can be identified with its image under this embedding before applying α to it. For $f \in C_c(\mathbb{R}^d)$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, we define the function $F_{f,g} : G_g/\Gamma_g \to \mathbb{R}$ by

$$F_{f,g}(x) = \sum_{v \in X_g(\mathbb{Z})} f(xv). \tag{2-1}$$

The function α has the property that there exists a constant $c(f)$ depending only on the support and maximum of f such that, for all x in G_g/Γ_g,

$$F_{f,g}(x) \le c(f)\,\alpha(x). \tag{2-2}$$

The last property is well known and follows from Minkowski's theorem on successive minima; see Lemma 2 of [Schmidt 1968] for example. Alternatively, see [Henk and Wills 2008] for an up-to-date review of many related results.

We will be carrying out integration on various measure spaces defined by the groups introduced at the beginning of the section. With this in mind, let us introduce the following notation for the corresponding measures. If v denotes some variable, the notation dv is used to denote integration with respect to Lebesgue measure and this variable. Let μ_g be the Haar measure on G_g/Γ_g; if $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, then since Γ_g is a lattice in G_g we can normalize so that $\mu_g(G_g/\Gamma_g) = 1$. In addition, ν_g will denote the measure on K_g normalized so that $\nu_g(K_g) = 1$. Let m_g^a denote the Haar measure on $X_g^a(\mathbb{R})$ defined by

$$\int_{\mathbb{R}^d} f(v)\,dv = \int_{-\infty}^{\infty} \int_{X_g^a(\mathbb{R})} f(v)\,dm_g^a(v)\,da. \tag{2-3}$$

The following provides us with our upper bounds and will be proved in Section 3:

Theorem 2.4. *Let $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ be arbitrary, and let $\Delta = g\mathbb{Z}^d$. Let $\{a_t : t \in \mathbb{R}\}$ denote a self-adjoint one-parameter subgroup of* $\mathrm{SO}(2,1)$ *embedded into H_I so that it fixes the subspace $\langle e_{s+2}, \dots, e_{d-1}\rangle$ and only has eigenvalues e^{-t},* 1 *and e^t.*

(I) *Suppose $r_1 \ge 3$, $r_2 \ge 1$ and $0 < \delta < 2$; then*

$$\sup_{t>0} \int_{K_I} \alpha(a_t k \Delta)^\delta \, d\nu_I(k) < \infty.$$

(II) *Suppose $r_1 = r_2 = 2$ or $r_1 = 2$ and $r_2 = 1$; then*

$$\sup_{t>1} \frac{1}{t} \int_{K_I} \alpha(a_t k \Delta)\, d\nu_I(k) < \infty.$$

In Section 4, we will modify the results from Section 4 of [Eskin et al. 1998] and combine them with Theorem 2.4 to prove the following, which will be a major ingredient of the proof of Theorem 2.1:

Theorem 2.5. *Suppose $r_1 \geq 3$ and $r_2 \geq 1$. Let $A = \{a_t : t \in \mathbb{R}\}$ be a one-parameter subgroup of H_g such that there exists a continuous homomorphism $\rho : \mathrm{SL}_2(\mathbb{R}) \to H_g$ with $\rho(D) = A$ and $\rho(\mathrm{SO}(2)) \subset K_g$, where $D = \left\{\left(\begin{smallmatrix} t & 0 \\ 0 & t^{-1}\end{smallmatrix}\right) : t > 0\right\}$. Let $\phi \in L^1(G_g/\Gamma_g)$ be a continuous function such that, for some $0 < \delta < 2$ and some $C > 0$,*

$$|\phi(\Delta)| < C\alpha(\triangle)^{\delta} \quad \text{for all } \Delta \in G_g/\Gamma_g. \tag{2-4}$$

Then for all $\epsilon > 0$ and all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, there exists $T_0 > 0$ such that, for all $t > T_0$,

$$\left| \int_{K_g} \phi(a_t k)\, d\nu_g(k) - \int_{G_g/\Gamma_g} \phi \, d\mu_g \right| \leq \epsilon.$$

3. The upper bounds

In this section, we prove Theorem 2.4. By definition, $H_I \cong \mathrm{SO}(r_1, r_2)$ and is embedded in $\mathrm{SL}_d(\mathbb{R})$ so that it fixes $\langle e_1, \ldots, e_s\rangle$. Let $\{a_t : t \in \mathbb{R}\}$ denote a self-adjoint one-parameter subgroup of $\mathrm{SO}(2, 1)$ embedded into H_I so that it fixes the subspace $\langle e_{s+2}, \ldots, e_{d-1}\rangle$. Moreover, suppose that the only eigenvalues of a_t are e^{-t}, 1 and e^t. For $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, let $\Delta = g\mathbb{Z}^d$.

3A. *Proof of Part (I) of Theorem 2.4.* The aim is to construct a function $f : H_I \to \mathbb{R}$ that is contracted by the operator

$$A_t f(h) = \int_{K_I} f(a_t k h)\, d\nu_I(k).$$

We say that f is contracted by the operator A_t if for any $c > 0$ there exists $t_0 > 0$ and $b > 0$ such that, for all $h \in H_I$,

$$A_{t_0} f(h) < c f(h) + b.$$

This fact will be used in conjunction with the following:

Proposition 3.1 [Eskin et al. 1998, Proposition 5.12]. *Let $f : H_I \to \mathbb{R}$ be a strictly positive function such that*:

(1) *For any $\epsilon > 0$, there exists a neighborhood $V(\epsilon)$ of 1 in H_I such that*

$$(1-\epsilon) f(h) \leq f(uh) \leq (1+\epsilon) f(h)$$

for all $h \in H_I$ and $u \in V(\epsilon)$.

(2) *The function f is left-K_I-invariant.*

(3) $f(1) < \infty$.

(4) *The function f is contracted by the operator A_t.*

Then $\sup_{t>0} A_t f(1) < \infty$.

It is clear that, if in addition to satisfying Properties (1)–(4) we have $\alpha(h\Delta)^\delta \leq f(h)$ for all $h \in H_I$, then the conclusion of Part (I) of Theorem 2.4 follows. We define the function in three stages. In the first stage, we define a function on the exterior algebra of $\mathbb{R}^d$; then this function is used to define a function on the space of lattices in $\mathbb{R}^d$. Finally we use that function to define a function with the required properties.

3A.1. *A function on the exterior algebra of* $\mathbb{R}^d$. Let $\bigwedge(\mathbb{R}^d) = \bigoplus_{i=1}^{d-1} \bigwedge^i(\mathbb{R}^d)$. We say that $v \in \bigwedge(\mathbb{R}^d)$ has degree i if $v \in \bigwedge^i(\mathbb{R}^d)$. Let $\Omega_i = \{v_1 \wedge \cdots \wedge v_i : v_1, \ldots, v_i \in \mathbb{R}^d\}$ be the set of monomial elements of $\bigwedge(\mathbb{R}^d)$ with degree i. Define $\Omega = \bigcup_{i=1}^{d-1} \Omega_i$. Consider the representation $\rho : H_I \to \mathrm{GL}(\bigwedge(\mathbb{R}^d))$. Since H_I is semisimple, this representation decomposes as a direct sum of irreducible subrepresentations. Associated to each of these subrepresentations is a unique highest weight. Let $\mathscr{P}$ denote the set of all these highest weights. For $\lambda \in \mathscr{P}$, denote by U^λ the sum of all of the subrepresentations with highest weight λ and let $\tau_\lambda : \bigwedge(\mathbb{R}^d) \to U^\lambda$ be the orthogonal projection.

Let $\epsilon > 0$. For $0 < i < d$ and $v \in \bigwedge^i(\mathbb{R}^d)$, the following function was defined by Benoist and Quint [2012]. Let

$$\varphi_\epsilon(v) = \begin{cases} \min_{\lambda \in \mathscr{P}\setminus\{0\}} \epsilon^{\gamma_i} \|\tau_\lambda(v\|)^{-1} & \text{if } \|\tau_0(v)\| \leq \epsilon^{\gamma_i}, \\ 0 & \text{else,} \end{cases}$$

where for $0 < i < d$ we define $\gamma_i = (d-i)i$. In fact, the definition of φ_ϵ given here is a special case of the definition given in [Benoist and Quint 2012]. In that definition of φ_ϵ, there is an extra set of exponents depending on $\lambda \in \mathscr{P}\setminus\{0\}$ appearing. However, we see that in our case we may choose all of these exponents to be equal to 1.

Let $\mathscr{F} = \{v \in \bigwedge(\mathbb{R}^d) : H_I v = v\}$ be the fixed vectors of H_I. Let $\mathscr{F}^c$ be the orthogonal complement of $\mathscr{F}$.

Remark 3.2. Since $\max_{\lambda \in \mathscr{P}\setminus\{0\}} \|\tau_\lambda(v)\|$ defines a norm on $\mathscr{F}^c$, there exist constants c_1 and c_2 depending on ϵ and the γ_i's such that

$$c_1\|v\|^{-1} \leq \varphi_\epsilon(v) \leq c_2\|v\|^{-1}$$

for all $v \in \mathscr{F}^c$.

Remark 3.3. For $0 < i < d$ and $v \in \bigwedge^i(\mathbb{R}^d) \setminus \{0\}$, we have $\varphi_\epsilon(v) = \infty$ if and only if v is H_I-invariant and $\|v\| \leq \epsilon^{\gamma_i}$.

We will need to refer to the constant defined as $b_1 = \sup\{\varphi_\epsilon(v) : v \in \bigwedge(\mathbb{R}^d), \|v\| \geq 1\}$. Benoist and Quint [2012, Lemma 4.2] showed that the function φ_ϵ satisfies the following convexity property:

Lemma 3.4. *There exists a positive constant C such that, for any $0 < \epsilon < C^{-1}$, $u \in \Omega_{i_1}$, $v \in \Omega_{i_2}$ and $w \in \Omega_{i_3}$ with $i_1 \geq 0$, $i_2 > 0$ and $i_3 > 0$ such that $\varphi_\epsilon(u \wedge v) \geq 1$ and $\varphi_\epsilon(u \wedge w) \geq 1$, one has*:

(1) *If $i_1 > 0$ and $i_1 + i_2 + i_3 < d$, then*

$$\min\{\varphi_\epsilon(u \wedge v), \varphi_\epsilon(u \wedge w)\} \leq (C\epsilon)^{1/2} \max\{\varphi_\epsilon(u), \varphi_\epsilon(u \wedge v \wedge w)\}.$$

(2) *If $i_1 = 0$ and $i_1 + i_2 + i_3 < d$, then*

$$\min\{\varphi_\epsilon(v), \varphi_\epsilon(w)\} \leq (C\epsilon)^{1/2} \varphi_\epsilon(v \wedge w).$$

(3) *If $i_1 > 0$, $i_1 + i_2 + i_3 = d$ and $\|u \wedge v \wedge w\| \geq 1$, then*

$$\min\{\varphi_\epsilon(u \wedge v), \varphi_\epsilon(u \wedge w)\} \leq (C\epsilon)^{1/2} \varphi_\epsilon(u).$$

(4) *If $i_1 = 0$, $i_1 + i_2 + i_3 = d$ and $\|v \wedge w\| \geq 1$, then*

$$\min\{\varphi_\epsilon(v), \varphi_\epsilon(w)\} \leq b_1.$$

We also need to obtain uniform bounds for the spherical averages of φ_ϵ. In order to do this, we use the following:

Lemma 3.5 [Eskin et al. 1998, Lemma 5.2]. *Let V be a finite-dimensional real inner-product space, A a self-adjoint linear transformation of V, K a closed connected subgroup of $O(V)$ and S a closed subset of the unit sphere in V. Assume the only eigenvalues of A are -1, 0 and 1, and denote by W^-, W^0 and W^+ the corresponding eigenspaces. Assume that $Kv \not\subset W^0$ for any $v \in S$ and that there exists a self-adjoint subgroup H_1 of $\mathrm{GL}(V)$ with the following properties*:

(1) *The Lie algebra of H_1 contains A.*

(2) *H_1 is locally isomorphic to $\mathrm{SO}(3, 1)$.*

(3) *$H_1 \cap K$ is a maximal compact subgroup of H_1.*

Then for any δ, $0 < \delta < 2$,

$$\lim_{t \to \infty} \sup_{v \in S} \int_K \|\exp(tA)kv\|^{-\delta} \, d\nu(k) = 0.$$

Using Lemma 3.5, we can obtain the following bound on the spherical averages:

Lemma 3.6. *Suppose $r_1 \geq 3$ and $r_2 \geq 1$. Then for all $\epsilon > 0$, $0 < \delta < 2$ and $c > 0$, there exists $t_0 > 0$ such that, for all $t > t_0$ and all $v \in \mathcal{F}^c \setminus \{0\}$,*

$$\int_{K_I} \varphi_\epsilon(a_t k v)^\delta \, d\nu_I(k) < c\varphi_\epsilon(v)^\delta.$$

Proof. The subset $S = \{v \in \bigwedge(\mathbb{R}^d) : \|v - \tau_0(v)\| = 1\}$ is a closed subset of the unit sphere in $\bigwedge(\mathbb{R}^d)$. We have $a_t = \exp(tA)$ for an appropriate choice of A satisfying the conditions of Lemma 3.5.

We claim that, for any $v \in S$, $Kv \not\subset W^0$. To see this, let

$$H_v = \{h \in H_I : hkv = kv \text{ for all } k \in K_I\}.$$

Note that K_I normalizes H_v. Let E_v be the subgroup generated by $K_I \cup H_v$. By its definition, E_v also normalizes H_v. Since K_I is a maximal proper subgroup of H_I, in the case that $H_v \not\subset K_I$, we must have $E_v = H_I$. Therefore, H_v is a normal subgroup of H_I. Since $r_1 \geq 3$ and $r_2 \geq 1$, H_I is simple and hence $H_v = H_I$ or H_v is trivial. Since $S \cap \mathscr{F} = 0$, the first case is impossible. Therefore, for all $v \in S$, $H_v \subset K_I$. In particular, this means that $\{a_t : t \in \mathbb{R}\}$ is not contained in H_v. This implies the claim.

Then if $r_1 \geq 3$ and $r_2 \geq 1$, the conditions of Lemma 3.5 are satisfied. Hence, for any δ with $0 < \delta < 2$,

$$\lim_{t\to\infty} \sup_{v\in S} \int_{K_I} \|a_t k v\|^{-\delta}\, d\nu_I(k) = 0.$$

This implies that for all $c > 0$ there exists $t_0 > 0$ such that, for all $t > t_0$ and all $v \in \mathscr{F}^c \setminus \{0\}$,

$$\int_{K_I} \|a_t k v\|^{-\delta}\, d\nu_I(k) < c\|v\|^{-\delta}.$$

Then the claim of the lemma follows from Remark 3.2. □

3A.2. *A function on the space of lattices.* For any lattice Λ, we say that $v \in \Omega$ is Λ-integral if one can write $v = v_1 \wedge \cdots \wedge v_i$ where $v_1, \ldots, v_i \in \Lambda$. Let $\Omega_i(\Lambda)$ and $\Omega(\Lambda)$ be the sets of Λ-integral elements of Ω_i and Ω, respectively. Define $f_\epsilon : \mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z}) \to \mathbb{R}$ by

$$f_\epsilon(\Lambda) = \max_{v\in\Omega(\Lambda)} \varphi_\epsilon(v).$$

Note that by Remark 3.2 for all $\epsilon > 0$ there exists some constant $c_\epsilon > 0$ such that, for any unimodular lattice Λ, we have

$$\begin{aligned} \max_{v\in\Omega(\Lambda)} \|v\|^{-1} &\leq \max_{0<i<d}\Big(\max_{v\in\Omega_i(\Lambda),\ \|\tau_0(v)\|\leq\epsilon^{\gamma_i}} \|v\|^{-1} + \max_{v\in\Omega_i(\Lambda),\ \|\tau_0(v)\|>\epsilon^{\gamma_i}} \|v\|^{-1}\Big) \\ &\leq c_\epsilon f_\epsilon(\Lambda) + \max_{0<i<d} \epsilon^{-\gamma_i}. \end{aligned} \tag{3-1}$$

Moreover, it follows from the definition of the α function that

$$\alpha(\Lambda) = \max\Big\{ \max_{v\in\Omega(\Lambda)} \|v\|^{-1}, 1\Big\}. \tag{3-2}$$

The following is necessary to ensure that the function $f_\epsilon(h\Delta)$ is finite for all $h \in H_I$:

Lemma 3.7. *For all $h \in H_I$, if $u \in \Omega(h\Delta)$, then $u \notin \mathcal{F}$.*

Proof. Suppose for a contradiction that $u \in \Omega(h\Delta) \cap \mathcal{F}$. Suppose that u has degree i for some $0 < i < d$, and let $u = u_1 \wedge \cdots \wedge u_i$ and $U = \langle u_1, \ldots, u_i \rangle$. Since $u \in \Omega(h\Delta)$, it follows that $U \cap h\Delta$ is a lattice in U. Moreover, because $u \in \mathcal{F}$, $U \cap \Delta$ is also a lattice in U or equivalently $g^{-1}U \cap \mathbb{Z}^d$ is a lattice in $g^{-1}U$. The subspace $g^{-1}U$ is H_g-invariant.

Conversely, it follows from Lemma 3.4 of [Sargent 2013] that, if V is any H_g-invariant subspace, then either

(1) $V \subseteq g^{-1}\langle e_1, \ldots, e_s \rangle$ or

(2) $V = g^{-1}\langle e_{s+1}, \ldots, e_d \rangle \oplus V'$ where $V' \subseteq g^{-1}\langle e_1, \ldots, e_s \rangle$.

Thus, either V or the orthogonal complement of V is contained in $g^{-1}\langle e_1, \ldots, e_s \rangle$. By Corollary 3.2 of [Sargent 2013], $g^{-1}\langle e_1, \ldots, e_s \rangle$ contains no subspaces defined over $\mathbb{Q}$. This implies that, if V is any H_g-invariant subspace, then V is not defined over $\mathbb{Q}$. In particular, $V \cap \mathbb{Z}^d$ cannot be a lattice in V. This gives a contradiction. □

3A.3. *A function on H_I.* Define $\tilde{f}_{\Delta,\epsilon} : H_I \to \mathbb{R}$ by

$$\tilde{f}_{\Delta,\epsilon}(h) = f_\epsilon(h\Delta).$$

In view of (3-1) and (3-2), the proof of Part (I) of Theorem 2.4 will be complete provided that Conditions (1)–(4) from Proposition 3.1 are satisfied by the function $\tilde{f}_{\Delta,\epsilon}$ for some $\epsilon > 0$. It is clear that $\tilde{f}_{\Delta,\epsilon}$ is left-K_I-invariant. Also since $\|\tau_\lambda(\rho(h^{-1})\|)^{-1} \le \|\tau_\lambda(hv)\|/\|v\| \le \|\tau_\lambda(\rho(h))\|$ for all $\lambda \in \mathcal{P}$, $v \in \Omega$ and $h \in H_I$, $\tilde{f}_{\Delta,\epsilon}$ also satisfies Condition (1) of Proposition 3.1. From Remark 3.3, we get that $\tilde{f}_{\Delta,\epsilon}(1) = \infty$ only if there exists $v \in \Omega(\Delta) \cap \mathcal{F}$, but by Lemma 3.7, we know that no such v exists and so $\tilde{f}_{\Delta,\epsilon}(1) < \infty$. It remains to show that $\tilde{f}_{\Delta,\epsilon}$ is contracted by the operator A_t. The proof is very similar to that of Proposition 5.3 in [Benoist and Quint 2012].

Lemma 3.8. *Suppose $r_1 \ge 3$ and $r_2 \ge 1$. There exists $\epsilon > 0$ such that, for all $0 < \delta < 2$, the function $\tilde{f}^{\delta}_{\Delta,\epsilon}$ is contracted by the operator A_t.*

Proof. Fix $c > 0$. By Lemma 3.6, there exists $t_0 > 0$ so that, for any $v \in \mathcal{F}^c \setminus \{0\}$,

$$\int_{K_I} \varphi_\epsilon(a_{t_0}kv)^\delta \, d\nu_I(k) < \frac{c}{d}\varphi_\epsilon(v)^\delta. \tag{3-3}$$

Let $m_0 = \|\rho(a_{t_0})\| = \|\rho(a_{t_0}^{-1})\|$. Then for all $v \in \bigwedge(\mathbb{R}^d)$,

$$m_0^{-1} \le \|a_{t_0}v\|/\|v\| \le m_0. \tag{3-4}$$

It follows from the definition of φ_ϵ and (3-4) that

$$m_0^{-1}\varphi_\epsilon(v) \le \varphi_\epsilon(a_{t_0}v) \le m_0\varphi_\epsilon(v). \tag{3-5}$$

Let

$$\Psi(h\Delta) = \{v \in \Omega(h\Delta) : f_\epsilon(h\Delta) \le m_0^2 \varphi_\epsilon(v)\}.$$

Note that

$$f_\epsilon(h\Delta) = \max_{\psi \in \Psi(h\Delta)} \varphi_\epsilon(\psi). \tag{3-6}$$

Let C be the constant from Lemma 3.4. Assume that ϵ is small enough so that

$$m_0^4 C\epsilon < 1. \tag{3-7}$$

There are now two cases.

Case 1: $f_\epsilon(h\Delta) \le \max\{b_1, m_0^2\}$. In this case, (3-5) and the fact that f_ϵ is left-K_I-invariant imply that $f_\epsilon(a_{t_0}kh\Delta) \le m_0 f_\epsilon(h\Delta)$. Hence,

$$\int_{K_I} f_\epsilon(a_{t_0}kh\Delta)^\delta \, d\nu_I(k) \le (m_0 \max\{b_1, m_0^2\})^\delta. \tag{3-8}$$

Case 2: $f_\epsilon(h\Delta) > \max\{b_1, m_0^2\}$. This implies:

Claim 3.9. *The set $\Psi(h\Delta)$ contains only one element up to sign change in each degree.*

Proof. Assume that, for some $0 < i < d$, $\Psi(h\Delta) \cap \Omega(h\Delta)$ contains two noncolinear elements, v_0 and w_0. Then because $f_\epsilon(h\Delta) > m_0^2$ and v_0 and w_0 are in $\Psi(h\Delta)$, we have $\varphi_\epsilon(v_0) \ge 1$ and $\varphi_\epsilon(w_0) \ge 1$. We can write $v_0 = u \wedge v$ and $w_0 = u \wedge w$, where $u \in \Omega_{i_1}(h\Delta)$, $v \in \Omega_{i_2}(h\Delta)$ and $w \in \Omega_{i_2}(h\Delta)$ with $i_1 \ge 0$ and $i_2 > 0$. There are four cases.

Case 2.1: $i_2 < i$ and $i_2 < d - i$. In this case,

$$f_\epsilon(h\Delta) \le m_0^2 \min\{\varphi_\epsilon(u \wedge v), \varphi_\epsilon(u \wedge w)\} \le (m_0^4 C\epsilon)^{1/2} \max\{\varphi_\epsilon(u), \varphi_\epsilon(u \wedge v \wedge w)\}$$

by Lemma 3.4(1). This implies that

$$f_\epsilon(h\Delta) \le (m_0^4 C\epsilon)^{1/2} f_\epsilon(h\Delta), \tag{3-9}$$

which contradicts (3-7).

Case 2.2: $i_2 = i < d - i$. In this case, $u = 1$. The same computation but using Lemma 3.4(2) still gives (3-9), which is still a contradiction.

Case 2.3: $i_2 = d - i < i$. In this case, $\|u \wedge v \wedge w\|$ is an integer. Therefore, the same computation but using Lemma 3.4(3) still gives (3-9).

Case 2.4: $i_2 = i = d - i$. The same computation, using Lemma 3.4(4), gives

$$f_\epsilon(h\Delta) \le b_1,$$

which is again a contradiction. ∎

Suppose $v \in \Omega$ is arbitrary. If $v \notin \Psi(h\Delta)$, then $f_\epsilon(h\Delta) > m_0^2 \varphi_\epsilon(v)$, and by left-$K_I$-invariance of φ_ϵ, (3-5) and (3-6), for all $k \in K_I$, we have

$$\begin{aligned}\varphi_\epsilon(a_{t_0}kv) \le m_0\varphi_\epsilon(v) \le m_0^{-1} f_\epsilon(h\Delta)&\\ \le m_0^{-1} \max_{\psi\in\Psi(h\Delta)} \varphi_\epsilon(\psi) \le \max_{\psi\in\Psi(h\Delta)} \varphi_\epsilon(a_{t_0}k\psi).&\end{aligned} \tag{3-10}$$

If $v \in \Psi(h\Delta)$, then (3-10) holds for obvious reasons. Therefore, (3-10) holds for all $v \in \Omega$. Thus, using the definition of f_ϵ and (3-10), we get

$$\begin{aligned}\int_{K_I} f_\epsilon(a_{t_0}kh\Delta)^\delta \, d\nu_I(k) &= \int_{K_I} \max_{v\in\Omega(h\Delta)} \varphi_\epsilon(a_{t_0}kv)^\delta \, d\nu_I(k)\\ &\le \sum_{\psi\in\Psi(h\Delta)} \int_{K_I} \varphi_\epsilon(a_{t_0}k\psi)^\delta \, d\nu_I(k).\end{aligned} \tag{3-11}$$

Using Lemma 3.7, we see that, for all $\psi \in \Psi(h\Delta)$, $\psi \notin \mathcal{F}$ and hence $\psi - \tau_0(\psi) \in \mathcal{F}^c \setminus \{0\}$. Moreover, if $\varphi_\epsilon(a_{t_0}k\psi) \neq 0$, then $\varphi_\epsilon(a_{t_0}k\psi) = \varphi_\epsilon(a_{t_0}k(\psi - \tau_0(\psi)))$ and we can apply (3-3) to get

$$\int_{K_I} \varphi_\epsilon(a_{t_0}k\psi)^\delta \, d\nu_I(k) \le \frac{c}{d} \varphi_\epsilon(\psi)^\delta \tag{3-12}$$

for each $\psi \in \Psi(h\Delta)$. If $\varphi_\epsilon(a_{t_0}k\psi) = 0$, then it is clear that (3-12) also holds. Using Claim 3.9, we obtain

$$\sum_{\psi\in\Psi(h\Delta)} \int_{K_I} \varphi_\epsilon(a_{t_0}k\psi)^\delta \, d\nu_I(k) \le d \max_{\psi\in\Psi(h\Delta)} \int_{K_I} \varphi_\epsilon(a_{t_0}k\psi)^\delta \, d\nu_I(k);$$

the claim of the lemma follows from (3-6), (3-8), (3-11) and (3-12). □

3B. ***Proof of Part (II) of Theorem 2.4.*** This time, the aim is to construct a function such that it satisfies the conditions of the following:

Lemma 3.10. *Suppose $r_1 = 2$ and $r_2 = 1$ or $r_1 = r_2 = 2$. Let $f : H_I \to \mathbb{R}$ be a strictly positive continuous function such that*:

(1) *For any $\epsilon > 0$, there exists a neighborhood $V(\epsilon)$ of 1 in H_I such that*

$$(1-\epsilon) f(h) \le f(uh) \le (1+\epsilon) f(h)$$

for all $h \in H_I$ and $u \in V(\epsilon)$.

(2) *The function f is left-K_I-invariant.*

(3) $f(1) < \infty$.

(4) *There exist $t_0 > 0$ and $b > 0$ such that, for all $h \in H_I$ and $0 \le t \le t_0$,*

$$A_t f(h) \le f(h) + b.$$

Then $\sup_{t>1} (1/t) A_t f(1) < \infty$.

Proof. Since SO(2, 1) is locally isomorphic to $\mathrm{SL}_2(\mathbb{R})$ and SO(2, 2) is locally isomorphic to $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$, this follows directly from Lemma 5.13 of [Eskin et al. 1998]. □

The general strategy of this subsection is broadly the same as in the last one. First we define a certain function on the exterior algebra of $\mathbb{R}^d$, and then we use this function to define a function that has the properties demanded by Lemma 3.10.

3B.1. *Functions on the exterior algebra of* $\mathbb{R}^d$. As before, we work with a function on the exterior algebra of $\mathbb{R}^d$. This time, the definition is simpler because in this case the vectors fixed by the action of H_I cause no extra problems. For $\epsilon > 0$, $0 < i < d$ and $v \in \bigwedge^i(\mathbb{R}^d)$, we define

$$\widetilde{\varphi}_\epsilon(v) = \epsilon^{\gamma_i} \|v\|^{-1}.$$

If $v \in \bigwedge^0(\mathbb{R}^d)$ or $v \in \bigwedge^d(\mathbb{R}^d)$, then we set $\widetilde{\varphi}_\epsilon(v) = 1$. The following is the analogue of Lemma 3.4:

Lemma 3.11. *Let $i_1 \geq 0$ and $i_2 > 0$ and Λ be a unimodular lattice. Then for all $u \in \Omega_{i_1}(\Lambda)$, $v \in \Omega_{i_2}(\Lambda)$ and $w \in \Omega_{i_2}(\Lambda)$,*

$$\widetilde{\varphi}_\epsilon(u \wedge v)\, \widetilde{\varphi}_\epsilon(u \wedge w) \leq \epsilon^{2i_2} \widetilde{\varphi}_\epsilon(u)\, \widetilde{\varphi}_\epsilon(u \wedge v \wedge w).$$

Proof. This is a direct consequence of Lemma 5.6 from [Eskin et al. 1998] and the fact that $2\gamma_{i_1+i_2} - \gamma_{i_1} - \gamma_{i_1+2i_2} = 2i_2$. □

The following lemma is used to bound the spherical averages. It is analogous to Lemma 3.6 (see also Lemma 5.5 of [Eskin et al. 1998]). It explains why in this case the fixed vectors do not cause problems.

Lemma 3.12. *Suppose $r_1 \geq 2$ and $r_2 \geq 1$. Then for all $t \geq 0$ and $v \in \bigwedge(\mathbb{R}^d) \setminus \{0\}$,*

$$\int_{K_I} \|a_t k v\|^{-1}\, d\nu_I(k) \leq \|v\|^{-1}.$$

Proof. Let $F_v(t) = \int_{K_I} \|a_t k v\|^{-1}\, d\nu_I(k)$. We will show that $\frac{d}{dt} F_v(t) \leq 0$ for all $t \geq 0$ and $v \in \bigwedge(\mathbb{R}^d) \setminus \{0\}$, from which it is clear that the claim of the lemma follows. Let π^- and π^+ be the projections from $\bigwedge(\mathbb{R}^d)$ onto the contracting and expanding eigenspaces of a_t, respectively. Note that

$$\begin{aligned} \frac{d}{dt} F_v(t) &= \int_{K_I} \frac{e^{-2t}\|\pi^-(kv)\|^2 - e^{2t}\|\pi^+(kv)\|^2}{\|a_t k v\|^3}\, d\nu_I(k) \\ &\leq \Big(\frac{\|a_t\|}{\|v\|}\Big)^3 \int_{K_I} (e^{-2t}\|\pi^-(kv)\|^2 - e^{2t}\|\pi^+(kv)\|^2)\, d\nu_I(k). \end{aligned} \tag{3-13}$$

Let Q_0 also denote the matrix that defines the quadratic form Q_0. Note that $\|\pi^-(Q_0 v)\| = \|\pi^+(v)\|$ and $\|\pi^+(Q_0 v)\| = \|\pi^-(v)\|$ for all $v \in \bigwedge(\mathbb{R}^d)$. Because

$Q_0^T = Q_0 = Q_0^{-1}$, if $\det(Q_0) = 1$, then $Q_0 \in K_I$, or if $\det(Q_0) = -1$, then $-Q_0 \in K_I$. This means that $Q_0 K_I(v - \tau_0(v)) = K_I \pm (v - \tau_0(v))$ and thus

$$\begin{aligned}\int_{K_I} \|\pi^-(kv)\|^2 \, d\nu_I(k) &= \int_{K_I} \|\pi^+(Q_0(kv))\|^2 \, d\nu_I(k) \\ &= \int_{K_I} \|\pi^+(kv)\|^2 \, d\nu_I(k). \end{aligned} \tag{3-14}$$

Therefore, using (3-13) and (3-14), we have

$$\frac{d}{dt} F_v(t) \le \left(\frac{\|a_t\|}{\|v\|}\right)^3 \int_{K_I} \|\pi^+(kv)\|^2 \, d\nu_I(k)(e^{-2t} - e^{2t}) \le 0$$

for all $t \ge 0$ and $v \in \bigwedge(\mathbb{R}^d) \setminus \{0\}$ as required. □

3B.2. *Functions on H_I.* Define $\tilde{f}_{\Delta,\epsilon} : H_I \to \mathbb{R}$ by

$$\tilde{f}_{\Delta,\epsilon}(h) = \sum_{i=1}^{d} \max_{v \in \Omega_i(h\Delta)} \widetilde{\varphi}_\epsilon(v).$$

Note that for all $\epsilon > 0$ there exists some constant $c_\epsilon > 0$ such that, for any unimodular lattice Λ,

$$\max_{v \in \Omega(\Lambda)} \|v\|^{-1} \le c_\epsilon \max_{v \in \Omega(\Lambda)} \widetilde{\varphi}_\epsilon(v) \le c_\epsilon \sum_{i=1}^{d} \max_{v \in \Omega_i(\Lambda)} \widetilde{\varphi}_\epsilon(v).$$

In view of this and (3-2), the proof of Part (II) of Theorem 2.4 will be complete provided that Conditions (1)–(4) from Lemma 3.10 are satisfied by the functions $\tilde{f}_{\Delta,\epsilon}$ for some $\epsilon > 0$. It is clear that $\tilde{f}_{\Delta,\epsilon}$ is left-K_I-invariant. Also since $\|\rho(h^{-1})\|^{-1} \le \|hv\|/\|v\| \le \|\rho(h)\|$ for all $v \in \Omega$ and $h \in H_I$, $\tilde{f}_{\Delta,\epsilon}$ also satisfies Condition (1) of Lemma 3.10. We also have that $\tilde{f}_{\Delta,\epsilon}(1) < \infty$. It remains to show that $\tilde{f}_{\Delta,\epsilon}$ satisfies Condition (4) of Lemma 3.10.

Lemma 3.13. *Suppose $r_1 = 2$ and $r_2 = 1$ or $r_1 = r_2 = 2$. Then there exist $\epsilon > 0$ and $t_0 > 0$ such that, for all $0 \le t < t_0$ and $h \in H_I$,*

$$\int_{K_I} \tilde{f}_{\Delta,\epsilon}(a_t k h) \, d\nu_I(k) \le \tilde{f}_{\Delta,\epsilon}(h).$$

Proof. Let $m_0 = \|\rho(a_{t_0})\|$. Then for all $v \in \bigwedge(\mathbb{R}^d)$ and $0 \le t < t_0$,

$$m_0^{-1} \le \|a_t v\|/\|v\| \le m_0. \tag{3-15}$$

It follows from the definition of $\widetilde{\varphi}_\epsilon$ and (3-15) that, for all $0 \le t < t_0$,

$$m_0^{-1} \widetilde{\varphi}_\epsilon(v) \le \widetilde{\varphi}_\epsilon(a_t v) \le m_0 \widetilde{\varphi}_\epsilon(v). \tag{3-16}$$

Let

$$\Psi(h\Delta) = \bigcup_{i=1}^{d} \Big\{ v \in \Omega_i(h\Delta) : \max_{v \in \Omega_i(h\Delta)} \widetilde{\varphi}_\epsilon(v) \le m_0^2 \widetilde{\varphi}_\epsilon(v) \Big\}.$$

Now we show that for ϵ small enough the set $\Psi(h\Delta)$ contains only one element up to sign change in each degree. To see this, assume that, for some $0 < i < d$, $\Psi(h\Delta) \cap \Omega(h\Delta)$ contains two noncolinear elements, v_0 and w_0. We can write $v_0 = u \wedge v$ and $w_0 = u \wedge w$ where $u \in \Omega_{i_1}(h\Delta)$, $v \in \Omega_{i_2}(h\Delta)$ and $w \in \Omega_{i_2}(h\Delta)$ with $i_1 \ge 0$ and $i_2 > 0$. In this case,

$$\tilde{f}_{\Delta,\epsilon}(h)^2 \le d^2 m_0^4 \widetilde{\varphi}_\epsilon(u \wedge v)\, \widetilde{\varphi}_\epsilon(u \wedge w) \le d^2 m_0^4 \epsilon^{2i_2} \tilde{f}_{\Delta,\epsilon}(h)^2$$

by Lemma 3.11. Hence, the claim is true since taking ϵ small enough gives a contradiction.

In view of this discussion, we can suppose that $\Psi(h\Delta) = \{\psi_i\}_{i=1}^d$, where ψ_i has degree i. Let $v \in \Omega_i(h\Delta)$ be arbitrary. If $v \notin \Psi(h\Delta)$, then $\max_{v \in \Omega_i(h\Delta)} \widetilde{\varphi}_\epsilon(v) > m_0^2 \widetilde{\varphi}_\epsilon(v)$, and by left-$K_I$-invariance of $\widetilde{\varphi}_\epsilon$ and (3-16), for all $k \in K_I$, we have

$$\widetilde{\varphi}_\epsilon(a_{t_0} k v) \le m_0 \widetilde{\varphi}_\epsilon(v) \le m_0^{-1} \max_{v \in \Omega_i(h\Delta)} \widetilde{\varphi}_\epsilon(v) = m_0^{-1} \widetilde{\varphi}_\epsilon(\psi_i) \le \widetilde{\varphi}_\epsilon(a_{t_0} k \psi_i). \tag{3-17}$$

If $v \in \Psi(h\Delta)$, then (3-17) holds for obvious reasons. Therefore, (3-17) holds for all $v \in \Omega$. Thus, using the definition of $\tilde{f}_{\Delta,\epsilon}$ and (3-17), we get

$$\begin{aligned} \int_{K_I} \tilde{f}_{\Delta,\epsilon}(a_{t_0} k h)\, d\nu_I(k) &= \sum_{i=1}^{d} \int_{K_I} \max_{v \in \Omega_i(h\Delta)} \widetilde{\varphi}_\epsilon(a_{t_0} k v)\, d\nu_I(k) \\ &\le \sum_{i=1}^{d} \int_{K_I} \widetilde{\varphi}_\epsilon(a_{t_0} k \psi_i)\, d\nu_I(k). \end{aligned} \tag{3-18}$$

By Lemma 3.12, there exists $t_0 > 0$ so that, for any $v \in \bigwedge(\mathbb{R}^d)$ and all $0 \le t < t_0$,

$$\int_{K_I} \widetilde{\varphi}_\epsilon(a_{t_0} k \psi_i)\, d\nu_I(k) \le \widetilde{\varphi}_\epsilon(\psi_i) \tag{3-19}$$

for each $\psi_i \in \Psi(h\Delta)$. The claim of the lemma follows from (3-18) and (3-19). □

4. Ergodic theorems

For subgroups W_1 and W_2 of G_g, let $X(W_1, W_2) = \{g \in G_g : W_2 g \subset g W_1\}$. As in [Eskin et al. 1998], the ergodic theory is based on Theorem 3 from [Dani and Margulis 1993], reproduced below in a form relevant to the current situation:

Theorem 4.1. *Suppose $r_1 \ge 2$ and $r_2 \ge 1$. Let $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ be arbitrary. Let $U = \{u_t : t \in \mathbb{R}\}$ be a unipotent one-parameter subgroup of G_g and ϕ be a bounded continuous function on G_g/Γ_g. Let $\mathfrak{D}$ be a compact subset of G_g/Γ_g, and let $\epsilon > 0$*

be given. Then there exist finitely many proper closed subgroups $H_1, \dots, H_k$ of G_g such that $H_i \cap \Gamma_g$ is a lattice in H_i for all i and compact subsets $C_1, \dots, C_k$ of $X(H_1, U), \dots, X(H_k, U)$, respectively, such that for all compact subsets F of $\mathfrak{D} - \bigcup_{1 \le i \le k} C_i \Gamma_g / \Gamma_g$ there exists a $T_0 > 0$ such that, for all $x \in F$ and $T > T_0$,

$$\left| \frac{1}{T} \int_0^T \phi(u_t x)\, dt - \int_{G_g/\Gamma_g} \phi \, d\mu_g \right| < \epsilon.$$

Remark 4.2. By construction, the subgroups H_i occurring are such that $H_i \cap \Gamma_g$ is Zariski-dense in H_i and hence H_i are defined over $\mathbb{Q}$. For a precise reference, see Theorem 3.6.2 and Remark 3.4.2 of [Kleinbock et al. 2002].

The next result is a reworking of Theorem 4.3 from [Eskin et al. 1998]. The difference is that in Lemma 4.3 the identity is fixed as the base point for the flow and the condition that H_g be maximal is dropped.

Lemma 4.3. *Suppose $r_1 \ge 2$ and $r_2 \ge 1$. Let $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ be arbitrary. Let $U = \{u_t : t \in \mathbb{R}\}$ be a one-parameter unipotent subgroup of H_g, not contained in any proper normal subgroup of H_g. Let ϕ be a bounded continuous function on G_g/Γ_g. Then for all $\epsilon > 0$ and $\eta > 0$, there exists a $T_0 > 0$ such that, for all $T > T_0$,*

$$\nu_g\left(\left\{ k \in K_g : \left| \frac{1}{T} \int_0^T \phi(u_t k)\, dt - \int_{G_g/\Gamma_g} \phi \, d\mu_g \right| > \epsilon \right\}\right) \le \eta. \tag{4-1}$$

Proof. Let $H_1, \dots, H_k$ and $C_1, \dots, C_k$ be as in Theorem 4.1. Let $\gamma \in \Gamma_g$; consider $Y_i(\gamma) = K_g \cap X(H_i, U)\gamma$. Suppose that $Y_i(\gamma) = K_g$; then $Uk\gamma^{-1} \subset k\gamma^{-1} H_i$ for all $k \in K_g$. In other words,

$$k^{-1} U k \subset \gamma^{-1} H_i \gamma \quad \text{for all } k \in K_g. \tag{4-2}$$

The subgroup $\langle k^{-1}Uk : k \in K_g \rangle$ is normalized by $U \cup K_g$ and clearly $\langle k^{-1}Uk : k \in K_g \rangle \subseteq \langle U \cup K_g \rangle \subseteq H_g$. If G is a simple Lie group with finite center, with maximal compact subgroup K, it follows from Exercise A.3, Chapter IV of [Helgason 2001] that K is also a maximal proper subgroup of G. This means that, because H_g is semisimple with finite center, any connected subgroup L of H_g containing K_g can be represented as $L = H' K_g$ where H' is a connected normal subgroup of H_g. Because U is not contained in any proper normal subgroup of H_g, this implies that $\langle U \cup K_g \rangle = H_g$. Therefore, $\langle k^{-1}Uk : k \in K_g \rangle$ is a normal subgroup of H_g, and because U is not contained in any proper normal subgroup of H_g, we have $\langle k^{-1}Uk : k \in K_g \rangle = H_g$. This and (4-2) imply that $H_g \subset \gamma^{-1} H_i \gamma$. Note that $\gamma \in \mathrm{SL}_d(\mathbb{Z})$ and, by Remark 4.2, H_i is defined over $\mathbb{Q}$. Therefore, $\gamma^{-1} H_i \gamma$ is defined over $\mathbb{Q}$; it follows from Theorem 7.7 of [Platonov and Rapinchuk 1991] that $\overline{\gamma^{-1} H_i \gamma \cap \mathrm{SL}_d(\mathbb{Q})} = \gamma^{-1} H_i \gamma$. Therefore, Lemma 3.7 and Proposition 4.1 of [Sargent 2013] imply that $\gamma^{-1} H_i \gamma = G_g$, which is a contradiction, and therefore,

$Y_i(\gamma) \subsetneq K_g$. This means, for all $1 \le i \le k$, $Y_i(\gamma)$ is a submanifold of strictly smaller dimension than K_g and hence

$$\nu_g(Y_i(\gamma)) = 0. \tag{4-3}$$

Note that, because $C_i \subseteq X(H_i, U)$,

$$K_g \cap \bigcup_{1\le i\le k} C_i\Gamma_g \subseteq K_g \cap \bigcup_{1\le i\le k} X(H_i, U)\Gamma_g = \bigcup_{1\le i\le k}\bigcup_{\gamma\in\Gamma_g} Y_i(\gamma),$$

and therefore, (4-3) implies

$$\nu_g\biggl(K_g \cap \bigcup_{1\le i\le k} C_i\Gamma_g\biggr) = 0. \tag{4-4}$$

Let $\mathfrak{D}$ be a compact subset of G_g such that $K_g \subset \mathfrak{D}$. Then from (4-4), it follows that, for all $\eta > 0$, there exists a compact subset F of $\mathfrak{D} - \bigcup_{1\le i\le k} C_i\Gamma_g$ such that

$$\nu_g(F \cap K_g) \ge 1 - \eta. \tag{4-5}$$

From Theorem 4.1, for all $\epsilon > 0$, there exists a $T_0 > 0$ such that, for all $x \in (F \cap K_g)/\Gamma_g$ and $T > T_0$,

$$\left|\frac{1}{T}\int_0^T \phi(u_t x)\,dt - \int_{G_g/\Gamma_g} \phi\,d\mu_g\right| < \epsilon.$$

Therefore, if $k \in K_g$, $T > T_0$ and

$$\left|\frac{1}{T}\int_0^T \phi(u_t k)\,dt - \int_{G_g/\Gamma_g} \phi\,d\mu_g\right| > \epsilon,$$

then $k \in K_g \setminus F$, but $\nu_g(K_g \setminus F) \le \eta$ by (4-5), and this implies (4-1). □

Lemma 4.4. *Suppose $r_1 \ge 2$ and $r_2 \ge 1$. Let $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ be arbitrary. Let $U = \{u_t : t \in \mathbb{R}\}$ be a one-parameter unipotent subgroup of H_g not contained in any proper normal subgroup of H_g. Let ϕ be a bounded continuous function on G_g/Γ_g. Then for all $\epsilon > 0$ and $\delta > 0$, there exists a $T_0 > 0$ such that, for all $T > T_0$,*

$$\left|\frac{1}{\delta T}\int_T^{(1+\delta)T}\int_{K_g} \phi(u_t k)\,d\nu_g(k)\,dt - \int_{G_g/\Gamma_g} \phi\,d\mu_g\right| < \epsilon.$$

Proof. Let ϕ be a bounded continuous function on G_g/Γ_g. Lemma 4.3 implies for all $\epsilon > 0$, $\eta > 0$ and $d > 0$ there exists a $T_0 > 0$ such that, for all $T > T_0$,

$$\nu_g\biggl(\biggl\{k \in K_g : \left|\frac{1}{dT}\int_0^{dT} \phi(u_t k)\,dt - \int_{G_g/\Gamma_g} \phi\,d\mu_g\right| > \epsilon\biggr\}\biggr) \le \eta. \tag{4-6}$$

Using (4-6) with $d=1$ and $d=1+\delta$, we get that for all $\epsilon>0$ and $\eta>0$ there exists a subset $\mathscr{C}\subseteq K_g$ with $\nu_g(\mathscr{C})\geq 1-\eta$ such that for all $k\in\mathscr{C}$ the following hold:

$$\left|\int_0^T \phi(u_t k)\,dt - T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| < \epsilon T,$$

$$\left|\int_0^{(1+\delta)T} \phi(u_t k)\,dt - (1+\delta)T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| < (1+\delta)T\epsilon.$$

Hence, for all $k\in\mathscr{C}$, we have

$$\begin{aligned}
&\left|\int_T^{(1+\delta)T} \phi(u_t k)\,dt - \delta T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| \\
&\quad= \left|\int_0^{(1+\delta)T} \phi(u_t k)\,dt - (1+\delta)T\int_{G_g/\Gamma_g}\phi\,d\mu_g \right.\\
&\qquad\qquad\qquad\qquad \left. - \int_0^T \phi(u_t k)\,dt + T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| \\
&\quad\leq \left|\int_0^T \phi(u_t k)\,dt - T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| \\
&\qquad\qquad\qquad\qquad + \left|\int_0^{(1+\delta)T} \phi(u_t k)\,dt - (1+\delta)T\int_{G_g/\Gamma_g}\phi\,d\mu_g\right| \\
&\quad\leq (2+\delta)T\epsilon.
\end{aligned}$$

This means that, for all $\delta>0$, $\eta>0$ and $\epsilon>0$,

$$\nu_g\left(\left\{k\in K_g : \left|\frac{1}{\delta T}\int_T^{(1+\delta)T}\phi(u_t k)\,dt - \int_{G_g/\Gamma_g}\phi\,d\mu_g\right| < \frac{(2+\delta)\epsilon}{\delta}\right\}\right)\geq 1-\eta.$$

Since we can make ϵ and η as small as we wish, this implies the claim. □

Lemma 4.5. *Suppose $r_1\geq 2$ and $r_2\geq 1$. Let $A=\{a_t : t\in\mathbb{R}\}$ be a one-parameter subgroup of H_g, not contained in any proper normal subgroup of H_g, such that there exists a continuous homomorphism $\rho:\mathrm{SL}_2(\mathbb{R})\to H_g$ with $\rho(D)=A$ and $\rho(\mathrm{SO}(2))\subset K_g$, where $D=\left\{\begin{pmatrix} t & 0\\ 0 & t^{-1}\end{pmatrix} : t>0\right\}$. Let ϕ be a continuous function on G_g/Γ_g vanishing outside of a compact set. Then for all $g\in\mathscr{C}_{\mathrm{SL}}(r_1,r_2)$ and $\epsilon>0$ there exists $T_0>0$ such that, for all $t>T_0$,*

$$\left|\int_{K_g}\phi(a_t k)\,d\nu_g(k) - \int_{G_g/\Gamma_g}\phi\,d\mu_g\right|\leq\epsilon.$$

Proof. This is very similar to the proof of Theorem 4.4 from [Eskin et al. 1998], and some details will be omitted. Fix $\epsilon>0$. Assume that ϕ is uniformly continuous. Let $u_t=\begin{pmatrix} 1 & t\\ 0 & 1\end{pmatrix}$ and $w=\begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix}$. Then it is clear that $d_t=\begin{pmatrix} t & 0\\ 0 & t^{-1}\end{pmatrix}=b_t u_t k_t w$, where $b_t=(1+t^{-2})^{-1/2}\begin{pmatrix} 1 & 0\\ -t^{-1} & 1+t^{-2}\end{pmatrix}$ and $k_t=(1+t^{-2})^{-1/2}\begin{pmatrix} 1 & t^{-1}\\ -t^{-1} & 1\end{pmatrix}$. By our assumptions on A, there exists a continuous homomorphism $\rho:\mathrm{SL}_2(\mathbb{R})\to H_g$ such that $\rho(D) = A$ and $\rho(\mathrm{SO}(2)) \subset K_g$. Let $\rho(d_t) = d'_t$, $\rho(b_t) = b'_t$, $\rho(k_t) = k'_t$ and $\rho(w) = w'$. Then for all $t > 0$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$,

$$\int_{K_g} \phi(d'_t k)\, d\nu_g(k) = \int_{K_g} \phi(b'_t u'_t k'_t w' k)\, d\nu_g(k)$$

$$= \int_{K_g} \phi(b'_t u'_t k)\, d\nu_g(k) \tag{4-7}$$

since $k'_t, w' \in K_g$. It follows from (4-7) that, for $r, t > 0$,

$$\left|\int_{K_g} \phi(d'_t k)\, d\nu_g(k) - \int_{K_g} \phi(u'_{rt} k)\, d\nu_g(k)\right|$$

$$\leq \left|\int_{K_g} (\phi(d'_{rt} k) - \phi(d'_t k))\, d\nu_g(k)\right| + \left|\int_{K_g} (\phi(d'_{rt} k) - \phi(u'_{rt} k)) d\nu_g(k)\right|$$

$$= \left|\int_{K_g} (\phi(d'_r d'_t k) - \phi(d'_t k))\, d\nu_g(k)\right| + \left|\int_{K_g} (\phi(b'_{rt} u'_{rt} k) - \phi(u'_{rt} k))\, d\nu_g(k)\right|. \tag{4-8}$$

By uniform continuity, the fact that $\lim_{t\to\infty} b_t = I$ and (4-8) imply there exist $T_1 > 0$ and $\delta > 0$ such that for $t > T_1$ and $|r - 1| < \delta$ we have

$$\left|\int_{K_g} \phi(d'_t k)\, d\nu_g(k) - \int_{K_g} \phi(u'_{rt} k)\, d\nu_g(k)\right| \leq \epsilon.$$

Thus, if $T > T_1$, then

$$\left|\int_{K_g} \phi(d'_t k)\, d\nu_g(k) - \frac{1}{\delta T} \int_T^{(1+\delta)T} \int_{K_g} \phi(u'_t k)\, d\nu_g(k)\, dt\right| \leq \epsilon. \tag{4-9}$$

Combining (4-9) with Lemma 4.5 via the triangle inequality finishes the proof of the lemma. □

The section is completed by the proof of the main ergodic result, whose proof follows that of Theorem 3.5 in [Eskin et al. 1998].

Proof of Theorem 2.5. Assume that ϕ is nonnegative. Let $A(r) = \{x \in G_g/\Gamma_g : \alpha(x) > r\}$. Choose a continuous nonnegative function g_r on G_g/Γ_g such that $g_r(x) = 1$ if $x \in A(r+1)$, $g_r(x) = 0$ if $x \notin A(r)$ and $0 \leq g_r(x) \leq 1$ if $x \in A(r) \setminus A(r+1)$. Then

$$\int_{K_g} \phi(a_t k)\, d\nu_g(k)$$

$$= \int_{K_g} \phi(a_t k)\, g_r(a_t k)\, d\nu_g(k) + \int_{K_g} (\phi(a_t k) - \phi(a_t k)\, g_r(a_t k))\, d\nu_g(k). \tag{4-10}$$

Let $\beta = 2-\delta$; then for $x \in G_g/\Gamma_g$,

$$\begin{aligned}\phi(x)\, g_r(x) &\le C\alpha(x)^{2-\beta} g_r(x)\\ &= C\alpha(x)^{2-\beta/2} g_r(x)\,\alpha(x)^{-\beta/2} \le Cr^{-\beta/2}\alpha(x)^{2-\beta/2}.\end{aligned}$$

The last inequality is true because $g_r(x) = 0$ if $\alpha(x) \le r$. Therefore,

$$\int_{K_g} \phi(a_t k)\, g_r(a_t k)\, d\nu_g(k) \le Cr^{-\beta/2} \int_{K_g} \alpha(a_t k)^{2-\beta/2}\, d\nu_g(k). \tag{4-11}$$

Since $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, $r_1 \ge 3$ and $r_2 \ge 1$, Theorem 2.4(I) implies there exists B such that

$$\begin{aligned}\int_{K_g} \alpha(a_t k)^{2-\beta/2}\, d\nu_g(k) &= \int_{K_I} \alpha \circ g^{-1}(a_t k g)^{2-\beta/2}\, d\nu_I(k)\\ &\le c(g) \int_{K_I} \alpha(a_t k g)^{2-\beta/2}\, d\nu_I(k) < B\end{aligned}$$

for all $t \ge 0$. Then (4-11) implies that

$$\int_{K_g} \phi(a_t k)\, g_r(a_t k)\, d\nu_g(k) \le BCr^{-\beta/2}. \tag{4-12}$$

For all $\epsilon > 0$, there exists a compact subset, $\mathscr{C}$ of G_g/Γ_g, such that $\mu_g(\mathscr{C}) \ge 1-\epsilon$. The function α is bounded on $\mathscr{C}$, and hence, for all $\epsilon > 0$,

$$\begin{aligned}&\lim_{r\to\infty} \mu_g(A(r))\\ &\quad= \lim_{r\to\infty} \bigl(\mu_g(\{x \in \mathscr{C} : \alpha(x) > r\}) + \mu_g(\{x \in (G_g/\Gamma_g) \setminus \mathscr{C} : \alpha(x) > r\})\bigr) \le \epsilon.\end{aligned}$$

This means that

$$\lim_{r\to\infty} \mu_g(A(r)) = 0. \tag{4-13}$$

Note that

$$\int_{G_g/\Gamma_g} \phi(x)\, g_r(x)\, d\mu_g(x) \le \int_{A(r)} \phi(x)\, d\mu_g(x). \tag{4-14}$$

Since $\phi \in L^1(G_g/\Gamma_g)$, (4-13) and (4-14) imply that

$$\lim_{r\to\infty} \int_{G_g/\Gamma_g} \phi(x)\, g_r(x)\, d\mu_g(x) = 0. \tag{4-15}$$

Since the function $\phi(x) - \phi(x)g_r(x)$ is continuous and has compact support, Lemma 4.5 implies for all $\epsilon > 0$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $T_0 > 0$ such that, for all $t > T_0$,

$$\left| \int_{K_g} (\phi(a_t k) - \phi(a_t k)\, g_r(a_t k))\, d\nu_g(k) \right.$$
$$\left. - \int_{G_g/\Gamma_g} (\phi(x) - \phi(x)\, g_r(x))\, d\mu_g(x) \right| < \frac{\epsilon}{2}. \quad (4\text{-}16)$$

It is straightforward to check that (4-10), (4-12), (4-15) and (4-16) imply the conclusion of the theorem if r is sufficiently large. □

5. Proof of Theorem 2.1

The proof of Theorem 2.1 follows the same route as that of Sections 3.4–3.5 of [Eskin et al. 1998]. The main modification we make in order to handle the present situation is that we work inside the surface $X_g(\mathbb{R})$ rather than in the whole of $\mathbb{R}^d$. For $t \in \mathbb{R}$ and $v \in \mathbb{R}^d$, define a linear map a_t by

$$a_t v = (v_1, \dots, v_s, e^{-t} v_{s+1}, v_{s+2}, \dots, e^t v_d).$$

Note that the one-parameter group $\{\hat{a}_t : t \in \mathbb{R}\} = g^{-1}\{a_t : t \in \mathbb{R}\}g \subset H_g$ and that there exists a continuous homomorphism $\rho : \mathrm{SL}_2(\mathbb{R}) \to H_g$ with $\rho(D) = \{\hat{a}_t : t \in \mathbb{R}\}$ and $\rho(\mathrm{SO}(2)) \subset K_g$ where $D = \left\{\left(\begin{smallmatrix} t & 0 \\ 0 & t^{-1}\end{smallmatrix}\right) : t > 0\right\}$. Moreover, note that $\{a_t : t \in \mathbb{R}\}$ is self-adjoint and not contained in any normal subgroup of H_g and the only eigenvalues of a_t are e^{-t}, 1 and e^t. In other words, $\{\hat{a}_t : t \in \mathbb{R}\}$ satisfies the conditions of Theorems 2.5 and 2.4. For any natural number n, let S^{n-1} denote the unit sphere in an n-dimensional Euclidean space and let $\gamma_n = \mathrm{Vol}(S^n)$ and $c_{r_1,r_2} = \gamma_{r_1-1}\gamma_{r_2-1}$; then define

$$C_1 = c_{r_1,r_2} 2^{(2-r_1-r_2)/2} = c_{r_1,r_2} 2^{(2-d+s)/2}. \quad (5\text{-}1)$$

5A. ***Proof of Theorem 2.3.*** In Lemma 5.1, it is shown that it is possible to approximate certain integrals over K_g by integrals over $\mathbb{R}^{d-s-2}$. The integral over $\mathbb{R}^{d-s-2}$ can be used like the characteristic function of $R \times A(T/2, T)$; in particular, Theorem 2.3 is proved as an application of Lemma 5.1. It should be noted that Lemma 5.1 is analogous to Lemma 3.6 from [Eskin et al. 1998] and its proof is similar.

Lemma 5.1. *Let f be a continuous function of compact support on $\mathbb{R}^d_+ = \{v \in \mathbb{R}^d : \langle v, e_{s+1}\rangle > 0\}$, and for $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, let*

$$J_{f,g}(\ell_1, \dots, \ell_s, r) = \frac{1}{r^{d-s-2}} \int_{\mathbb{R}^{d-s-2}} f(\ell_1, \dots, \ell_s, r, v_{s+2}, \dots, v_{d-1}, v_d)\, dv_{s+2} \cdots dv_{d-1},$$

where $v_d = (a - Q_0^g(\ell_1, \dots, \ell_s, 0, v_{s+2}, \dots, v_{d-1}, 0))/2r$ so that $Q_0^g(\ell_1, \dots, \ell_s, r, v_{s+2}, \dots, v_{d-1}, v_d) = a$. Then for every $\epsilon > 0$, there exists $T_0 > 0$ such that, for *every t with $e^t > T_0$ and every $v \in \mathbb{R}^d_+$ with $\|v\| > T_0$,*

$$\left| C_1 e^{(d-s-2)t} \int_{K_g} f(\hat{a}_t k v)\, d\nu_g(k) - J_{f,g}(M_0^g(v), \|v\| e^{-t}) \right| < \epsilon.$$

Proof. By Lemma 2.2 of [Sargent 2013], for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, there exists a basis of $\mathbb{R}^d$, denoted by $b_1, \dots, b_d$, such that

$$Q_0^g(v) = Q_{1,\dots,s}(v) + 2v_{s+1}v_d + \sum_{i=s+2}^{s+r_1} v_i^2 - \sum_{i=s+r_1+1}^{d-1} v_i^2 \quad \text{and} \quad M_0^g(v) = (v_1, \dots, v_s)$$

and

$$\hat{a}_t(v) = (v_1, \dots v_s, e^{-t} v_{s+1}, v_{s+2}, \dots v_{d-1}, e^t v_d),$$

where $v_i = \langle v, b_i \rangle$ for $1 \le i \le d$ and $Q_{1,\dots,s}(v)$ is a nondegenerate quadratic form in variables $v_1, \dots, v_s$. Let E denote the support of f. Let $c_1 = \inf_{v \in E} \langle v, b_{s+1} \rangle$ and $c_2 = \sup_{v \in E} \langle v, b_{s+1} \rangle$. From the definition of $\hat{a}_t$, it follows that $f(\hat{a}_t w) = 0$ unless

$$|\langle w, b_{s+1} \rangle \langle w, b_d \rangle| \le \beta, \tag{5-2}$$

$$c_1 \le \langle w, b_{s+1} \rangle e^{-t} \le c_2, \tag{5-3}$$

$$\pi'(w) \in \pi'(E), \tag{5-4}$$

where β depends only on E and π' denotes the projection onto the span of $b_1, \dots, b_s$, $b_{s+2}, \dots, b_{d-1}$. For w satisfying (5-2) and (5-3), we have $\langle w, b_d \rangle = O(e^{-t})$. This, together with (5-4) and (5-3), implies that, if $f(\hat{a}_t w) \ne 0$ and t is large, then

$$\|w\| = \langle w, b_{s+1} \rangle + O(e^{-t}). \tag{5-5}$$

Note that by (5-5),

$$\langle \hat{a}_t w, b_{s+1} \rangle = \langle w, b_{s+1} \rangle e^{-t} = e^{-t} \|w\| + O(e^{-2t}) \tag{5-6}$$

and

$$\langle \hat{a}_t w, b_i \rangle = \langle w, b_i \rangle \quad \text{for } 1 \le i \le s \text{ or } s+2 \le i \le d-1. \tag{5-7}$$

Finally,

$$\begin{aligned} \langle \hat{a}_t w, b_d \rangle &= \big(Q_0^g(w) - Q_0^g\big(\langle w, b_1 \rangle, \dots, \langle w, b_s \rangle, 0, \\ &\qquad\qquad \langle w, b_{s+1} \rangle, \dots, \langle w, b_{d-1} \rangle, 0 \big)\big) / 2\langle \hat{a}_t w, b_{s+1} \rangle \\ &= \big(Q_0^g(w) - Q_0^g\big(\langle w, b_1 \rangle, \dots, \langle w, b_s \rangle, 0, \\ &\qquad\qquad \langle w, b_{s+1} \rangle, \dots, \langle w, b_{d-1} \rangle, 0 \big)\big) / 2e^{-t} \|w\| + O(e^{-t}). \end{aligned} \tag{5-8}$$

Hence, using (5-6), (5-7) and (5-8) together with the uniform continuity of f, applied with $w = kv$ for $v \in \mathbb{R}^d_+$ and $k \in K_g$, we see that for all $\delta > 0$ there exists a

$t_0 > 0$ so that if $t > t_0$ then

$$\left| f(\hat{a}_t k v) - f(v_1, \dots, v_s, \|v\| e^{-t}, \langle kv, b_{s+1} \rangle, \dots, \langle kv, b_{d-1} \rangle, v_d) \right| < \delta, \tag{5-9}$$

where v_d is determined by

$$Q_0^g(v_1, \dots, v_s, \|v\| e^{-t}, \langle kv, b_{s+1} \rangle, \dots, \langle kv, b_{d-1} \rangle, v_d) = Q_0^g(v) = a.$$

Change basis by letting $f_{s+1} = (b_{s+1} + b_d)/\sqrt{2}$, $f_d = (b_{s+1} - b_d)/\sqrt{2}$ and $f_i = b_i$ for $1 \le i \le s$ or $s+2 \le i \le d-1$. In this basis, $K_g \cong \mathrm{SO}(r_1) \times \mathrm{SO}(r_2)$ consists of orthogonal matrices preserving the subspaces $L_1 = \langle f_1, \dots, f_s \rangle$, $L_2 = \langle f_{s+1}, \dots, f_{s+r_1} \rangle$ and $L_3 = \langle f_{s+r_1+1}, \dots, f_d \rangle$. For $i = 1$, 2 or 3, let π_i denote the orthogonal projection onto L_i. Write $\rho_i = \|\pi_i(v)\|$; then the orbit $K_g v$ is the product of a point and two spheres $\{v_1, \dots, v_s\} \times \rho_2 S^{r_1 - 1} \times \rho_3 S^{r_2 - 1}$, where $S^{r_1 - 1}$ denotes the unit sphere in L_2 and $S^{r_2 - 1}$ the unit sphere in L_3.

Suppose $w \in K_g v$ is such that $f(\hat{a}_t w) \neq 0$. Then from (5-2) and (5-3), it follows that $\langle w, b_d \rangle = O(e^{-t})$. Now, set $w_i = \langle w, f_i \rangle$; then $w_{s+1} = 2^{-1/2} \langle w, b_{s+1} \rangle + O(e^{-t})$, $w_d = 2^{-1/2} \langle w, b_{s+1} \rangle + O(e^{-t})$ and, for $1 \le i \le s$ or $s+2 \le i \le d-1$, $w_i = O(1)$. Hence, for $i = 2$ or 3,

$$\rho_i = \|\pi_i(w)\| = 2^{-1/2} \langle w, b_{s+1} \rangle + O(e^{-t}) = 2^{-1/2} \|w\| + O(e^{-t}), \tag{5-10}$$

where the last estimate follows from (5-5).

By integrating (5-9) with respect to K_g, we see that for all $\epsilon > 0$ there exists a $t_0 > 0$ so that if $t > t_0$ then

$$\begin{aligned} \Bigg| &\int_{K_g} f(\hat{a}_t k v) \, d\nu_g(k) \\ &- \int_{K_g} f(v_1, \dots, v_s, \|v\| e^{-t}, \langle kv, b_{s+1} \rangle, \dots, \langle kv, b_{d-1} \rangle, v_d) \, d\nu_g(k) \Bigg| < \epsilon. \end{aligned} \tag{5-11}$$

Equation (5-4) implies that, if $f(\hat{a}_t k v) \neq 0$, then kv is within a bounded distance from $\rho_2 f_{s+1} + \rho_3 f_d$. As $\|v\|$ increases, so do the ρ_i and the normalized Haar measure on $\rho_2 S^{r_1 - 1}$ near $\rho_2 f_{s+1}$ tends to $(1/\operatorname{Vol}(\rho_2 S^{r_1 - 1})) \, dv_{s+2} \cdots dv_{s+r_1}$ and similarly the Haar measure on $\rho_3 S^{r_2 - 1}$ near $\rho_3 f_d$ tends to $(1/\operatorname{Vol}(\rho_3 S^{r_2 - 1})) \, dv_{s+r_1+1} \cdots dv_{d-1}$. This means that as $\|v\|$ tends to infinity the second integral in (5-11) tends to

$$\begin{aligned} \frac{\rho_2^{1-r_1} \rho_3^{1-r_2}}{c_{r_1, r_2}} &\int_{\mathbb{R}^{d-s-2}} f(v_1, \dots, v_s, \|v\| e^{-t}, v_{s+1}, \dots, v_d) \, dv_{s+2} \cdots dv_{d-1} \\ &= \frac{(\|v\| e^{-t})^{d-s-2}}{\rho_2^{r_1 - 1} \rho_3^{r_2 - 1} c_{r_1, r_2}} J_{f,g}(M_0^g(v), \|v\| e^{-t}). \end{aligned} \tag{5-12}$$

Because (5-10) implies that $\rho_2^{r_1 - 1} \rho_3^{r_2 - 1} = 2^{(s+2-d)/2} \|v\|^{d-s-2} + O(e^{-t})$, we can use (5-11) and (5-12) to get that for all $\epsilon > 0$ there exists a $t_0 > 0$ so that if $t > t_0$

and $\|v\| > t_0$ then

$$\left|\int_{K_g} f(\hat{a}_t k v)\, d\nu_g(k) - \frac{e^{t(s+2-d)}}{C_1} J_{f,g}(M_0^g(v), \|v\|e^{-t})\right| < \epsilon.$$

By dividing through by the factor $\frac{e^{t(s+2-d)}}{C_1}$, we obtain the desired conclusion. □

For f_1 and f_2 continuous functions of compact support on $\mathbb{R}^d_+ = \{v \in \mathbb{R}^d : \langle v, e_{s+1}\rangle > 0\}$, define $J_{f_1,g} + J_{f_2,g} = J_{f_1+f_2,g}$ and $J_{f_1,g}J_{f_2,g} = J_{f_1 f_2,g}$. These operations make the collection of functions of the form $J_{f,g}$ into an algebra of real-valued functions on the set $\mathbb{R}^s \times \{v \in \mathbb{R} : v > 0\}$. Denote this algebra by $\mathscr{A}$. The following will be used in the proofs of Theorems 2.3 and 2.1:

Lemma 5.2. *$\mathscr{A}$ is dense in $C_c(\mathbb{R}^s \times \{v \in \mathbb{R} : v > 0\})$.*

Proof. Let B be a compact subset of $\mathbb{R}^s \times \{v \in \mathbb{R} : v > 0\}$. Let $\mathscr{A}_B$ denote the subalgebra of $\mathscr{A}$ of functions with support B. It is straightforward to check that the algebra $\mathscr{A}_B$ separates points in B and does not vanish at any point in B. Therefore, by the Stone–Weierstrass theorem [Rudin 1976, Theorem 7.32], $\mathscr{A}_B$ is dense in the space of continuous functions on B. Since B is arbitrary, this implies the claim. □

Proof of Theorem 2.3. Let $\epsilon > 0$ be arbitrary and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$. By Lemma 5.2, there exists a continuous nonnegative function f on $\mathbb{R}^d_+$ of compact support so that $J_{f,g} \geq 1+\epsilon$ on $R \times [1, 2]$. Then if $v \in \mathbb{R}^d$ satisfies $e^t \leq \|v\| \leq 2e^t$, $M_0^g(v) \in R$ and $Q_0^g(v) = a$, then $J_{f,g}(M_0^g(v), \|v\|e^{-t}) \geq 1+\epsilon$. Then by Lemma 5.1, for sufficiently large t,

$$C_1 e^{(d-s-2)t} \int_{K_g} f(\hat{a}_t k v)\, d\nu_g(k) \geq 1$$

if $e^t \leq \|v\| \leq 2e^t$, $M_0^g(v) \in R$ and $Q_0^g(v) = a$. Then summing over $v \in X_g(\mathbb{Z})$, we get

$$\begin{aligned} |X_g(\mathbb{Z}) \cap V_M([a, b]) \cap A(e^t, 2e^t)| &\leq \sum_{v \in X_g(\mathbb{Z})} C_1 e^{(d-s-2)t} \int_{K_g} f(\hat{a}_t k v)\, d\nu_g(k) \\ &= C_1 e^{(d-s-2)t} \int_{K_g} F_{f,g}(\hat{a}_t k)\, d\nu_g(k). \end{aligned} \tag{5-13}$$

Note that

$$\int_{K_g} F_{f,g}(\hat{a}_t k)\, d\nu_g(k) = \int_{K_I} F_{f,g}(g^{-1} a_t k g)\, d\nu_I(k). \tag{5-14}$$

By (2-2), we have the bound $F_{f,g}(x) \leq c(f)\, \alpha(x)$ for all $x \in G_g/\Gamma_g$, where $c(f)$ is a constant depending only on f. Since $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, Part (I) of Theorem 2.4

implies that if $r_1 \geq 3$ and $r_2 \geq 1$ then

$$\int_{K_I} F_{f,g}(g^{-1}a_t kg)\, d\nu_I(k) < c(f \circ g^{-1}) \int_{K_I} \alpha(a_t kg)\, d\nu_I(k) < \infty. \tag{5-15}$$

In the case when $r_1 = 2$ and $r_2 = 1$ or $r_1 = r_2 = 2$, Part (II) of Theorem 2.4 implies that for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists a constant C so that

$$\int_{K_I} F_{f,g}(g^{-1}a_t kg)\, d\nu_I(k) < c(f \circ g^{-1}) \int_{K_I} \alpha(a_t kg)\, d\nu_I(k) < Ct. \tag{5-16}$$

Hence, (5-13), (5-14) and (5-15) imply that as long as $r_1 \geq 3$ and $r_2 \geq 1$ there exists a constant C_2 such that

$$|X_g(\mathbb{Z}) \cap V_M(R) \cap A(e^t, 2e^t)| \leq C_2 e^{(d-s-2)t}.$$

Similarly, (5-13), (5-14) and (5-16) imply that, if $r_1 = 2$ and $r_2 = 1$ or $r_1 = r_2 = 2$, then

$$|X_g(\mathbb{Z}) \cap V_M(R) \cap A(e^t, 2e^t)| \leq C_2 t e^{(d-s-2)t}.$$

Since we can write $T = e^t$ and

$$A(0, T) = \lim_{n \to \infty} \left(A(0, T/2^n) \bigcup_{i=1}^{n} A(T/2^i, T/2^{i-1}) \right),$$

the theorem follows by summing a geometric series. □

Theorem 2.3 has the following corollary, which is comparable with Proposition 3.7 from [Eskin et al. 1998] and will be used in the proof of Theorem 2.1.

Corollary 5.3. *Let f be a continuous function of compact support on $\mathbb{R}^d_+$. Then for every $\epsilon > 0$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, there exists $t_0 > 0$ so that, for $t > t_0$,*

$$\left| e^{-(d-s-2)t} \sum_{v \in X_g(\mathbb{Z})} J_{f,g}(M_0^g(v), \|v\| e^{-t}) - C_1 \int_{K_g} F_{f,g}(\hat{a}_t k)\, d\nu_g(k) \right| < \epsilon. \tag{5-17}$$

Proof. Since $J_{f,g}$ has compact support, the number of nonzero terms in the sum on the left-hand side of (5-17) is bounded by $ce^{(d-s-2)t}$ because of Theorem 2.3. Hence, summing the result of Lemma 5.1 over $v \in X_g(\mathbb{Z})$ proves (5-17). □

5B. ***Volume estimates.*** For a compactly supported function h on $\mathbb{R}^s \times \mathbb{R}^d \setminus \{0\}$, we define

$$\Theta(h, T) = \int_{X_g(\mathbb{R})} h(M_0^g(v), v/T)\, dm_g(v).$$

For $\mathscr{X} \subseteq \mathbb{R}^d$, we will use the notation $\mathrm{Vol}_{X_g}(\mathscr{X}) = \int_{X_g(\mathbb{R})} \mathbb{1}_{\mathscr{X} \cap X_g(\mathbb{R})}\, dm_g$ to mean the volume of $\mathscr{X}$ with respect to the volume measure on $X_g(\mathbb{R})$.

The following lemma and its corollary are analogous to Lemma 3.8 from [Eskin et al. 1998], and the proofs share some similarities although it is here that the fact we are integrating over $X_g(\mathbb{R})$ rather than the whole of $\mathbb{R}^d$ becomes an important distinction. In Lemma 5.4, we compute $\lim_{T\to\infty}(1/T^{d-s-2})\Theta(h,T)$. Here it is crucial that h is not defined on $\mathbb{R}^s\times\{0\}$; if it was, using the fact that h can be bounded by an integrable function, one could directly pass the limit inside the integral and the limit would be 0. The basic strategy of Lemma 5.4 is that we evaluate the integral $\int_{X_g(\mathbb{R})} dm_g$ by switching to polar coordinates. This has the effect that the integral changes into an integral over two spheres; then we approximate the spheres by an orbit of K_g and an integral over the coordinates fixed by K_g.

Lemma 5.4. *Suppose that h is a continuous function of compact support in $\mathbb{R}^s\times\mathbb{R}^d\setminus\{0\}$. Then*

$$\lim_{T\to\infty}\frac{1}{T^{d-s-2}}\Theta(h,T)=C_1\int_{K_g}\int_0^\infty\int_{\mathbb{R}^s} h(z,rke_0)r^{d-s-2}\,dz\,\frac{dr}{2r}\,d\nu_g(k),$$

where e_0 is a unit vector in $\mathbb{R}^d$ and C_1 is the constant defined by (5-1).

Proof. By Lemma 2.2 of [Sargent 2013], for all $g\in\mathscr{C}_{\mathrm{SL}}(r_1,r_2)$, there exists a basis of $\mathbb{R}^d$, denoted by $f_1,\dots,f_d$, such that

$$Q_0^g(v)=\sum_{i=1}^{s_1}v_i^2-\sum_{i=s_1+1}^{s}v_i^2+\sum_{i=s+1}^{s+r_1}v_i^2-\sum_{i=s+r_1+1}^{d}v_i^2 \quad\text{and}\quad M_0^g(v)=J(v_1,\dots,v_s),$$

where $v_i=\langle v,f_i\rangle$ for $1\le i\le d$, $J\in\mathrm{GL}_s(\mathbb{R})$, s_1 is a nonnegative integer such that $r_1+s_1=p$ and s_2 is a nonnegative integer such that $r_2+s_2=d-p$. Let $L_1=\langle v_1,\dots,v_{s_1},v_{s+1},\dots,v_{s+r_1}\rangle$, $L_2=\langle v_{s_1+1},\dots,v_s,v_{s+r_1+1},\dots,v_d\rangle$, S^{p-1} be the unit sphere inside L_1 and S^{d-p-1} be the unit sphere inside L_2. Let $\alpha\in S^{p-1}$ and $\beta\in S^{d-p-1}$. Using polar coordinates, we can parametrize $v\in X_g(\mathbb{R})$ so that

$$v_i=\begin{cases}\sqrt{a}\alpha_i\cosh t & \text{for } 1\le i\le s_1,\\ \sqrt{a}\beta_{i-s_1}\sinh t & \text{for } s_1+1\le i\le s,\\ \sqrt{a}\alpha_{i-s+s_1}\cosh t & \text{for } s+1\le i\le s+r_1,\\ \sqrt{a}\beta_{i-s_1-r_1}\sinh t & \text{for } s+r_1+1\le i\le d.\end{cases}\tag{5-18}$$

In these coordinates, we may write

$$dm_g(v)=\frac{a^{(d-2)/2}}{2}\cosh^{p-1}t\,\sinh^{q-1}t\,dt\,d\xi(\alpha,\beta)=P(e^t)\,dt\,d\xi(\alpha,\beta),$$

where $P(x)=(a^{(d-2)/2}/2^{d-1})x^{d-2}+O(x^{d-3})$ and ξ is the Haar measure on $S^{p-1}\times S^{q-1}$. Making the change of variables, $r=\sqrt{a}e^t/2T$, gives

$$\sqrt{a}\cosh t=Tr+a/4Tr \quad\text{and}\quad \sqrt{a}\sinh t=Tr-a/4Tr.\tag{5-19}$$

Let $L_1' = \langle v_{s+1}, \dots, v_{s+r_1}\rangle$, $L_2' = \langle v_{s+r_1+1}, \dots, v_d\rangle$, S^{r_1-1} be the unit sphere inside L_1', S^{r_2-1} be the unit sphere inside L_2', $\alpha' \in S^{r_1-1}$ and $\beta' \in S^{r_2-1}$. We may write

$$d\xi(\alpha, \beta) = \delta(\alpha, \beta)\, d\alpha_1 \cdots d\alpha_{s_1}\, d\beta_1 \cdots d\beta_{s_2}\, d\xi'(\alpha', \beta'),$$

where $\delta(\alpha, \beta)$ is the appropriate density function and $d\xi'$ is the Haar measure on $S^{r_1-1} \times S^{r_2-1}$. This gives

$$dm_g(v) = P\Big(\frac{2Tr}{\sqrt{a}}\Big)\, \delta(\alpha, \beta)\, \frac{dr}{r}\, d\alpha_1 \cdots d\alpha_{s_1}\, d\beta_1 \cdots d\beta_{s_2}\, d\xi'(\alpha', \beta'). \tag{5-20}$$

Let $z \in \mathbb{R}^s$. Make the further change of variables

$$(\alpha_1, \dots, \alpha_{s_1}, \beta_1, \dots, \beta_{s-s_1}) = \frac{1}{Tr} J^{-1} z; \tag{5-21}$$

this means that

$$d\alpha_1 \cdots d\alpha_{s_1}\, d\beta_1 \cdots d\beta_{s_2} = \frac{1}{\det(J)(Tr)^s}\, dz. \tag{5-22}$$

Moreover, using (5-18), (5-19) and (5-21) gives

$$M_0^g(v) = z + O(1/T) \quad \text{and} \quad v/T = r(\alpha' + \beta') + O(1/T). \tag{5-23}$$

Since h is continuous and compactly supported, it may bounded by an integrable function and hence

$$\begin{aligned}
\lim_{T\to\infty} &\frac{1}{T^{d-s-2}} \Theta(h, T) \\
&= \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} h(M_0^g(v), v/T)\, dm_g(v) \\
&= \int_{X_g(\mathbb{R})} \lim_{T\to\infty} \frac{1}{T^{d-s-2}} h(M_0^g(v), v/T)\, dm_g(v) \\
&= \int_{S^{r_1-1}\times S^{r_2-1}} \int_0^\infty \int_{\mathbb{R}^s} h(z, r(\alpha'+\beta')) r^{d-s-2} \delta(\alpha', \beta')\, dz\, \frac{dr}{2r}\, d\xi'(\alpha', \beta'),
\end{aligned}$$

where the last step follows from (5-20), the definition of $P(x)$, (5-22) and (5-23). Note that from the definition of δ it is clear that $\delta(\alpha', \beta') = 1$. Finally, let $e_0 = \frac{1}{\sqrt{2}}(f_1 + f_{p+1})$ and $\frac{1}{\sqrt{2}}(\alpha' + \beta') = ke_0$ and $r' = \sqrt{2}r$ to get that

$$\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h, T) = C_1 \int_{K_g} \int_0^\infty \int_{\mathbb{R}^s} h(z, r'ke_0) r'^{d-s-2}\, dz\, \frac{dr'}{2r'}\, d\nu_g(k). \quad \square$$

Corollary 5.5. *For all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, there exists a constant $C_3 > 0$ such that for all compact regions $R \subset \mathbb{R}^s$ with piecewise smooth boundary*

$$\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \mathrm{Vol}_{X_g}(V_{M_0^g}(R) \cap A(T/2, T)) = C_3\, \mathrm{Vol}(R).$$

Proof. Let $\mathbb{1}$ denote the characteristic function of $R \times A(1/2, 1)$; then it is clear that

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \operatorname{Vol}_{X_g}(V_{M_0^g}(R) \cap A(T/2, T)) \\ &= \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} \mathbb{1}(M_0(gv), v/T)\, dm_g(v) \\ &= \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(\mathbb{1}, T).\end{aligned}$$

Since R has piecewise smooth boundary, there exist regions $R_\delta^- \subseteq R \times A(1/2, 1) \subseteq R_\delta^+$ such that $\lim_{\delta\to 0} R_\delta^+ = \lim_{\delta\to 0} R_\delta^- = R$, and for all $\delta > 0$, we can choose continuous compactly supported functions h_δ^- and h_δ^+ on $\mathbb{R}^s \times \mathbb{R}^d \setminus 0$ such that $0 \le h_\delta^- \le \mathbb{1} \le h_\delta^+ \le 1$, $h_\delta^-(v) = \mathbb{1}(v)$ if $v \in R_\delta^-$ and $h_\delta^+(v) = 0$ if $v \notin R_\delta^+$. By Lemma 5.4,

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h_\delta^-, T) &\le \liminf_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} \mathbb{1}(M_0(gv), v/T)\, dm_g(v) \\ &\le \limsup_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} \mathbb{1}(M_0(gv), v/T)\, dm_g(v) \\ &\le \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h_\delta^+, T).\end{aligned}$$

It is clear that

$$\begin{aligned}\lim_{\delta\to\infty} \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h_\delta^-, T) &= \lim_{\delta\to\infty} \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h_\delta^+, T) \\ &= \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(\mathbb{1}, T);\end{aligned}$$

hence, we can apply Lemma 5.4 to get that

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(\mathbb{1}, T) \\ &= C_1 \int_{K_g} \int_0^\infty \int_{\mathbb{R}^s} \mathbb{1}(z, rk^{-1}e_0) r^{d-s-2}\, dz \frac{dr}{2r}\, d\nu_g(k) \\ &= C_1 \int_{\mathbb{R}^s} \mathbb{1}_R(z)\, dz \int_{K_g} \int_0^\infty \mathbb{1}_{A(1/2,1)}(rk^{-1}e_0) r^{d-s-2} \frac{dr}{2r}\, d\nu_g(k) = C_3 \operatorname{Vol}(R).\end{aligned}$$

The last equality holds because

$$\int_{K_g} \int_0^\infty \mathbb{1}_{A(1/2,1)}(rk^{-1}e_0) r^{d-s-2} \frac{dr}{2r}\, d\nu_g(k) < \infty$$

as $\mathbb{1}_{A(1/2,1)}$ has compact support and K_g is compact. $\square$

5C. ***Proof of Theorem 2.1.*** By Theorem 4.9 of [Platonov and Rapinchuk 1991], there exist $v_1, \dots, v_j \in X_g(\mathbb{Z})$ such that $X_g(\mathbb{Z}) = \bigsqcup_{i=1}^{j} \Gamma_g v_i$. Let $P_i(g) = \{x \in G_g : x v_i = v_i\}$ and $\Lambda_i(g) = P_i(g) \cap \Gamma_g$. By Proposition 1.13 of [Helgason 2000], there exist Haar measures ϱ_{Λ_i}, p_{Λ_i} and γ_{Λ_i} on $G_g/\Lambda_i(g)$, $P_i(g)/\Lambda_i(g)$ and $\Gamma_g/\Lambda_i(g)$, respectively, such that, for $f \in C_c(G_g/\Lambda_i(g))$ and hence for integrable functions on $G_g/\Lambda_i(g)$,

$$\int_{G_g/\Lambda_i(g)} f \, d\varrho_{\Lambda_i} = \int_{X_g(\mathbb{R})} \int_{P_i(g)/\Lambda_i(g)} f(xp)\, dp_{\Lambda_i}(p)\, dm_g(x), \tag{5-24}$$

$$\int_{G_g/\Lambda_i(g)} f \, d\varrho_{\Lambda_i} = \int_{G_g/\Gamma_g} \int_{\Gamma_g/\Lambda_i(g)} f(x\gamma)\, d\gamma_{\Lambda_i}(\gamma)\, d\mu_g(x). \tag{5-25}$$

Note that $\Gamma_g/\Lambda_i(g) = \Gamma_g v_i$ is discrete and its Haar measure $d\gamma_{\Lambda_i}$ is just the counting measure and so

$$\int_{\Gamma_g/\Lambda_i(g)} f(x\gamma)\, d\gamma_{\Lambda_i}(\gamma) = \sum_{v \in \Gamma_g v_i} f(xv). \tag{5-26}$$

Therefore, the normalizations already present on m_g and μ_g induce a normalization on p_{Λ_i}. Moreover, it follows from the Borel–Harish-Chandra theorem [Platonov and Rapinchuk 1991, Theorem 4.13] that the measure of $p_{\Lambda_i}(P_i(g)/\Lambda_i(g)) < \infty$ for each $1 \le i \le j$. As in [Eskin et al. 1998; Dani and Margulis 1993], where the proofs rely on Siegel's integral formula, here the proof relies on the following result:

Lemma 5.6. *For all $f \in C_c(X_g(\mathbb{R}))$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, there exists a constant*

$$C(g) = \sum_{i=1}^{j} p_{\Lambda_i}(P_i(g)/\Lambda_i(g))$$

such that

$$C(g) \int_{X_g(\mathbb{R})} f \, dm_g = \int_{G_g/\Gamma_g} F_{f,g}\, d\mu_g. \tag{5-27}$$

Proof. Note that, for $1 \le i \le j$, $G_g/P_i(g) \cong X_g(\mathbb{R})$. If $f \in C_c(X_g(\mathbb{R}))$, then f is $\Lambda_i(g)$-invariant and therefore can be considered as an integrable function on $G_g/\Lambda_i(g)$ and so

$$\int_{X_g(\mathbb{R})} \int_{P_i(g)/\Lambda_i(g)} f(xp)\, dp_{\Lambda_i}(p)\, dm_g(x) = \int_{P_i(g)/\Lambda_i(g)} dp_{\Lambda_i} \int_{X_g(\mathbb{R})} f\, dm_g. \tag{5-28}$$

Now it follows from the definition of $F_{f,g}$ (i.e., (2-1)), (5-24), (5-25), (5-26) and (5-28) that

$$\int_{G_g/\Gamma_g} F_{f,g}\, d\mu_g = \sum_{i=1}^{j} \int_{G_g/\Gamma_g} \sum_{v\in\Gamma_g v_i} f(xv)\, d\mu_g(x)$$
$$= \sum_{i=1}^{j} \int_{P_i(g)/\Lambda_i(g)} dp_{\Lambda_i} \int_{X_g(\mathbb{R})} f\, dm_g,$$

which is the desired result. □

The final lemma of this section is the counterpart of Lemma 3.9 from [Eskin et al. 1998], and again the proof there is mimicked.

Lemma 5.7. *Let f be a continuous function of compact support on $\mathbb{R}^d_+$. Then for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$,*

$$\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} J_{f,g}(M_0^g(v), \|v\|/T)\, dm_g(v) = C_1 C(g) \int_{G_g/\Gamma_g} F_{f,g}\, d\mu_g,$$

where C_1 is defined by (5-1) *and $C(g)$ is defined in Lemma 5.6.*

Proof. Let v_i be the components of v when written in the basis $b_1, \dots, b_d$ from Lemma 5.1. Using the change of variables $(v_1, \dots, v_d) \to (z_1, \dots, z_s, r, v_{s+2}, \dots, a)$ where $Q_0^g(v_1, \dots, v_d) = a$, we see that

$$\int_{\mathbb{R}^d} f(v)\, dv = \int_{-\infty}^{\infty} \int_0^{\infty} \int_{\mathbb{R}^s} J_{f,g}(z, r) r^{d-s-2}\, dz\, \frac{dr}{2r}\, da.$$

Hence, it follows from how m_g is defined (i.e., (2-3)) that

$$\int_{X_g(\mathbb{R})} f(v)\, dm_g(v) = \int_0^{\infty} \int_{\mathbb{R}^s} J_{f,g}(z, r) r^{d-s-2}\, dz\, \frac{dr}{2r}. \tag{5-29}$$

Lemma 5.4 and (5-29) imply that

$$\lim_{T\to\infty} \frac{1}{T^{d-s-2}} \int_{X_g(\mathbb{R})} J_{f,g}(M_0^g(v), \|v\|/T)\, dm_g(v)$$
$$= C_1 \int_{K_g} \left(\int_{X_g(\mathbb{R})} f(v)\, dm_g \right) d\nu_g(k).$$

Now the conclusion follows from Lemma 5.6. □

The purpose of Lemma 5.7 is to relate the integral over G_g/Γ_g to an integral over $X_g(\mathbb{R})$ in order that the integral over $X_g(\mathbb{R})$ can be approximated by an integral over K_g via Theorem 2.5. Then the integral over K_g can be approximated by the appropriate counting function via Corollary 5.3. We now proceed to put this into action in the proof of our main theorem, which is just a modification of the proof in [Eskin et al. 1998].

Proof of Theorem 2.1. By Lemma 5.4, the functional Ψ on $C_c(\mathbb{R}^s \times \mathbb{R}^d \setminus \{0\})$ given by

$$\Psi(h) = \lim_{T\to\infty} \frac{1}{T^{d-s-2}} \Theta(h, T)$$

is continuous. For all connected regions $R \subset \mathbb{R}^s$ with smooth boundary, if $\mathbb{1}$ denotes the characteristic function of $R \times A(1/2, 1)$, then for every $\epsilon > 0$ there exist continuous functions h_+ and h_- on $\mathbb{R}^s \times \mathbb{R}^d \setminus \{0\}$ such that, for all $(r, v) \in \mathbb{R}^s \times \mathbb{R}^d \setminus \{0\}$,

$$h_-(r, v) \le \mathbb{1}(r, v) \le h_+(r, v) \tag{5-30}$$

and

$$|\Psi(h_+) - \Psi(h_-)| < \epsilon. \tag{5-31}$$

Let $\mathcal{J}$ denote the space of linear combinations of functions on $\mathbb{R}^s \times \mathbb{R}^d$ of the form $J_{f,g}(r, \|v\|)$, where f is a continuous function of compact support on $\mathbb{R}^d_+$. Let $\mathcal{H}$ denote the collection of functions in $C_c(\mathbb{R}^s \times \mathbb{R}^d \setminus \{0\})$ such that if $h \in \mathcal{H}$ then h takes an argument of the form $(r, \|v\|)$. By Lemma 5.2, $\mathcal{J}$ is dense in $\mathcal{H}$, and since h_+ and h_- belong to $\mathcal{H}$, we may suppose that h_+ and h_- may be written as a finite linear combination of functions from $\mathcal{J}$. The function $F_{f,g}$ defined by (2-1) obeys the bound (2-4) with $\delta = 1$ by (2-2). Moreover, Lemma 3.10 of [Eskin et al. 1998] implies that $F_{f,g} \in L_1(G_g/\Gamma_g)$. Therefore, if $h' \in \{h_+, h_-\}$, then for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$, we can apply Theorem 2.5 with the function $F_{f,g}$ followed by Corollary 5.3 and Lemma 5.7 to get that there exists $t_0 > 0$ so that, for all $\epsilon > 0$ and $t > t_0$,

$$\left| \frac{C(g)}{e^{(d-s-2)t}} \sum_{v\in X_g(\mathbb{Z})} h'(M_0^g(v), ve^{-t}) - \Psi(h') \right| < \epsilon. \tag{5-32}$$

From the definition of $\Psi(h)$, we see that for all $h \in C_c(\mathbb{R}^s \times \mathbb{R}^d \setminus \{0\})$ and $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $t_0 > 0$ so that, for all $\epsilon > 0$ and $t > t_0$,

$$\left| \frac{1}{e^{(d-s-2)t}} \int_{X_g(\mathbb{R})} h(M_0^g(v), ve^{-t})\, dm_g(v) - \Psi(h) \right| < \epsilon. \tag{5-33}$$

Clearly (5-30) implies

$$\begin{aligned} \frac{C(g)}{e^{(d-s-2)t}} \sum_{v\in X_g(\mathbb{Z})} h_-(M_0^g(v), ve^{-t}) - \Psi(h_+) \\ &\le \frac{C(g)}{e^{(d-s-2)t}} \sum_{v\in X_g(\mathbb{Z})} \mathbb{1}(M_0^g(v), ve^{-t}) - \Psi(h_+) \\ &\le \frac{C(g)}{e^{(d-s-2)t}} \sum_{v\in X_g(\mathbb{Z})} h_+(M_0^g(v), ve^{-t}) - \Psi(h_+). \end{aligned} \tag{5-34}$$

Apply (5-31) to the left-hand side of (5-34), and then apply (5-32) with suitable choices of ϵ's to get that for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $t_0 > 0$ so that, for all $\theta > 0$ and $t > t_0$,

$$\left| \frac{C(g)}{e^{(d-s-2)t}} \sum_{v \in X_g(\mathbb{Z})} \mathbb{1}(M_0^g(v), ve^{-t}) - \Psi(h_+) \right| \leq \frac{\theta}{2}. \tag{5-35}$$

Similarly using (5-30), (5-31) and (5-33), we see that for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $t_0 > 0$ so that, for all $\theta > 0$ and $t > t_0$,

$$\left| \frac{1}{e^{(d-s-2)t}} \int_{X_g(\mathbb{R})} \mathbb{1}(M_0^g(v), ve^{-t})\, dm_g(v) - \Psi(h_+) \right| \leq \frac{\theta}{2}. \tag{5-36}$$

Hence, using (5-35) and (5-36), we see that for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $t_0 > 0$ so that, for all $\theta > 0$ and $t > t_0$,

$$\left| C(g) \sum_{v \in X_g(\mathbb{Z})} \mathbb{1}(M_0^g(v), ve^{-t}) - \int_{X_g} \mathbb{1}(M_0^g(v), ve^{-t})\, dm_g(v) \right| \leq \theta. \tag{5-37}$$

This means that for all $g \in \mathscr{C}_{\mathrm{SL}}(r_1, r_2)$ there exists $t_0 > 0$ so that, for all $\theta > 0$ and $t > t_0$,

$$\begin{aligned}(1-\theta) \int_{X_g(\mathbb{R})} \mathbb{1}(M_0^g(v), ve^{-t})\, dm_g(v) &\leq C(g) \sum_{v \in X_g(\mathbb{Z})} \mathbb{1}(M_0^g(v), ve^{-t}) \\ &\leq (1+\theta) \int_{X_g(\mathbb{R})} \mathbb{1}(M_0^g(v), ve^{-t})\, dm_g(v). \end{aligned} \tag{5-38}$$

Hence, for all $(Q, M) \in \mathscr{C}_{\mathrm{Pairs}}(r_1, r_2)$, there exists $t_0 > 0$ so that, for all $\theta > 0$ and $t > t_0$,

$$\begin{aligned}(1-\theta)\, \mathrm{Vol}_{X_Q}(V_M(R) \cap A(T/2, T)) &\leq C(g) |X_Q(\mathbb{Z}) \cap V_M(R) \cap A(T/2, T)| \\ &\leq (1+\theta)\, \mathrm{Vol}_{X_Q}(V_M(R) \cap A(T/2, T)).\end{aligned}$$

The conclusion of the theorem follows by applying Corollary 5.5 and summing a geometric series. □

6. Counterexamples

In small dimensions, there are slightly more integer points than expected on the quadratic surfaces defined by forms with signature $(1, 2)$ and $(2, 2)$. This fact was exploited in [Eskin et al. 1998] to show that the expected asymptotic formula for the situation they consider is not valid for these special cases. In a similar manner, it is possible to construct examples that show that Theorem 1.1 is not valid in the cases that the signature of H_g is $(1, 2)$ or $(2, 2)$. In this section, for the sake of brevity, we restrict our attention to the case when $s = 1$, but we note that similar arguments would hold in the case when $s > 1$. To start with, make the following definitions:

$$Q_1(x) = -x_1x_2 + x_3^2 + x_4^2,$$
$$Q_2(x) = x_1x_2 + x_3^2 - x_4^2,$$
$$Q_3(x) = -x_1x_2 + x_3^2 + x_4^2 - \alpha x_5^2,$$
$$L_\alpha(x) = x_1 - \alpha x_2.$$

We can now prove:

Lemma 6.1. *Let $\epsilon > 0$; suppose $[a, b] = [1/2 - \epsilon, 1]$ or $[-1, -1/2 + \epsilon]$. Let $a > 0$; then for every $T_0 > 0$, the set of $\beta \in \mathbb{R}$ for which there exists a $T > T_0$ such that*

$$\left|X^a_{Q_1}(\mathbb{Z}) \cap V_{L_\beta}([a, b]) \cap A(0, T)\right| > T(\log T)^{1-\epsilon}$$
$$\text{or} \quad \left|X^a_{Q_2}(\mathbb{Z}) \cap V_{L_\beta}([a, b]) \cap A(0, T)\right| > T(\log T)^{1-\epsilon}$$

is dense. Similarly if $a = 0$, then for every $T_0 > 0$, the set of $\beta \in \mathbb{R}$ for which there exists a $T > T_0$ such that

$$\left|X^a_{Q_3}(\mathbb{Z}) \cap V_{L_\beta}([a, b]) \cap A(0, T)\right| > T^2(\log T)^{1-\epsilon}$$

is dense.

Proof. Let $S_i(\alpha, T, a) = \{x \in \mathbb{Z}^{d_i} : L_\alpha(x) = 0,\ Q_i(x) = a,\ \|x\| \le T\}$, where $d_i = 4$ if $i = 1$ or 2 and $d_i = 5$ if $i = 3$. Lemma 3.14 of [Eskin et al. 1998] implies that

$$|S_i(\alpha, T, a)| \sim c_{i,\alpha} T \log T \qquad \text{for } i = 1, 2 \text{ and } \sqrt{\alpha} \in \mathbb{Q} \text{ and } a > 0, \tag{6-1}$$
$$|S_3(\alpha, T, 0)| \sim c_{3,\alpha} T^2 \log T \qquad \text{for } \sqrt{\alpha} \in \mathbb{Q}, \tag{6-2}$$

where $c_{i,\alpha}$ are constants that depend on α. Note that if $i = 1, 2$ and $x \in S_i(\alpha, T, a) \setminus S_i(\alpha, T/2, a)$, then

$$\frac{T^2}{4} - (\alpha^2 + 1)x_2^2 \le x_3^2 + x_4^2 \le T^2 - (\alpha^2 + 1)x_2^2 \tag{6-3}$$

and

$$x_3^2 + x_4^2 = \alpha x_2^2 + a. \tag{6-4}$$

Similarly if $x \in S_3(\alpha, T, 0) \setminus S_3(\alpha, T/2, 0)$,

$$\frac{T^2}{4} - (\alpha^2 + 1)x_2^2 \le x_3^2 + x_4^2 + x_5^2 \le T^2 - (\alpha^2 + 1)x_2^2 \tag{6-5}$$

and

$$x_3^2 + x_4^2 = \alpha(x_2^2 + x_5^2). \tag{6-6}$$

Combining (6-3) and (6-4) gives

$$\frac{T^2 - 4a}{4(\alpha^2 + \alpha + 1)} \le x_2^2 \le \frac{T^2 - a}{\alpha^2 + \alpha + 1}. \tag{6-7}$$

Respectively, combining (6-5) and (6-6) gives

$$\frac{T^2-(\alpha+1)x_5^2}{4(\alpha^2+\alpha+1)} \le x_2^2 \le \frac{T^2-(\alpha+1)x_5^2}{\alpha^2+\alpha+1}, \tag{6-8}$$

which upon noting that $-T \le x_5 \le T$ offers

$$\frac{T^2-(\alpha+1)T}{4(\alpha^2+\alpha+1)} \le x_2^2 \le \frac{T^2+(\alpha+1)T}{\alpha^2+\alpha+1}. \tag{6-9}$$

Take

$$\beta_\pm = \alpha \pm \sqrt{\frac{\alpha^2+\alpha+1}{T^2}}. \tag{6-10}$$

It is clear that $L_{\beta_\pm}(x) = L_\alpha(x) \pm \sqrt{(\alpha^2+\alpha+1)/T^2}\,x_2$, and hence, if $i = 1, 2$ and $x \in S_i(\alpha, T, a) \setminus S_i(\alpha, T/2, a)$, then (6-7) implies

$$\begin{aligned} \sqrt{\frac{1}{4}-\frac{a}{T^2}} \le L_{\beta_+}(x) &\le \sqrt{1-\frac{a}{T^2}}, \\ -\sqrt{1-\frac{a}{T^2}} \le L_{\beta_-}(x) &\le -\sqrt{\frac{1}{4}-\frac{a}{T^2}}. \end{aligned} \tag{6-11}$$

Similarly if $x \in S_3(\alpha, T, 0) \setminus S_3(\alpha, T/2, 0)$, then (6-9) implies

$$\begin{aligned} \sqrt{\frac{1}{4}-\frac{(\alpha+1)}{T}} \le L_{\beta_+}(x) &\le \sqrt{1-\frac{(\alpha+1)}{T}}, \\ -\sqrt{1-\frac{(\alpha+1)}{T}} \le L_{\beta_-}(x) &\le -\sqrt{\frac{1}{4}-\frac{(\alpha+1)}{T}}. \end{aligned} \tag{6-12}$$

This means for all $\epsilon > 0$ there exists $T_+ > 0$ such that if $T > T_+$ then $S_i(\alpha, T, a) \subset X^a_{Q_i}(\mathbb{Z}) \cap V_{L_{\beta_+}}([1/2-\epsilon, 1]) \cap A(0, T)$; respectively, there also exists $T_- > 0$ such that if $T > T_-$ then $S_i(\alpha, T, a) \subset X^a_{Q_i}(\mathbb{Z}) \cap V_{L_{\beta_-}}([-1, -1/2+\epsilon]) \cap A(0, T)$. By (6-1) and (6-2), for $i = 1, 2$ and large enough T (depending on α), $|S_i(\alpha, T, a)| > T(\log T)^{1-\epsilon}$ and $|S_i(\alpha, T, a)| > CT^2(\log T)^{1-\epsilon}$. The set of β satisfying (6-10) for rational α and large T is clearly dense, and this proves the lemma. □

Theorem 6.2. *Let $j = 1, 2$. For every $\epsilon > 0$ and every interval $[a, b]$, there exists a rational quadratic form Q and an irrational linear form L such that* $\mathrm{Stab}_{\mathrm{SO}(Q)}(L) \cong \mathrm{SO}(j, 2)$ *such that, for an infinite sequence $T_k \to \infty$,*

$$\left|X^{a_j}_Q(\mathbb{Z}) \cap V_L([a, b]) \cap A(0, T_k)\right| > T_k^j (\log T_k)^{1-\epsilon},$$

where $a_1 > 0$ and $a_2 = 0$.

Proof. Since the interval $[a, b]$ must intersect either the positive or negative reals, there is no loss of generality in assuming, after passing to a subset and rescaling, that $[a, b] = [1/4, 5/4]$ or $[-5/4, -1/4]$. For a given $S > 0$ and $i = 1, 2$, let $\mathcal{U}_S$ be the set of $\gamma \in \mathbb{R}$ for which there exist $\beta \in \mathbb{R}$ and $T > S$ with

$$\left|X^{a_1}_{Q_i}(\mathbb{Z}) \cap V_{L_\beta}([1/2, 1]) \cap A(0, T)\right| > CT \log T \tag{6-13}$$

and

$$|\beta - \gamma| < T^{-2}. \tag{6-14}$$

Then $\mathcal{U}_S$ is open and dense by Lemma 6.1. By the Baire category theorem [Rudin 1987, Theorem 5.6], $\bigcap_{k=1}^{\infty} \mathcal{U}_{2^{k+1}}$ is dense in $\mathbb{R}$ and is in fact of second category and hence uncountable. Let $\gamma \in \bigcap_{k=1}^{\infty} \mathcal{U}_{2^{k+1}} \setminus \mathbb{Q}$; then there exist infinite sequences β_k and T_k such that (6-13) and (6-14) hold with β replaced by β_k and T by T_k. Note that (6-14) implies that, for $\|x\| < T_k$,

$$|L_{\beta_k}(x) - L_\gamma(x)| < \frac{1}{T_k} < \frac{1}{4}$$

so that

$$X^{a_1}_{Q_i}(\mathbb{Z}) \cap V_{L_{\beta_k}}([1/2, 1]) \cap A(0, T_k) \subseteq X^{a_1}_{Q_i}(\mathbb{Z}) \cap V_{L_\gamma}([1/4, 5/4]) \cap A(0, T_k)$$

and hence $|X^{a_1}_{Q_i}(\mathbb{Z}) \cap V_{L_\gamma}([1/4, 5/4]) \cap A(0, T_k)| > CT_k \log T_k$ by (6-13). If $i = 3$, then we can carry out the same process, but we replace $\mathcal{U}_S$ by the set $\mathcal{W}_S$ of $\gamma \in \mathbb{R}$ for which there exist $\beta \in \mathbb{R}$ and $T > S$ with

$$\left|X^{0}_{Q_3}(\mathbb{Z}) \cap V_{L_\beta}([1/2, 1]) \cap A(0, T)\right| > CT^2 \log T$$

and

$$|\beta - \gamma| < T^{-2}. \qquad \square$$

Acknowledgment

The author would like to thank Alex Gorodnik for many helpful discussions and remarks about earlier versions of this paper.

References

[Benoist and Quint 2012] Y. Benoist and J.-F. Quint, "Random walks on finite volume homogeneous spaces", *Invent. Math.* **187**:1 (2012), 37–59. MR 2874934 Zbl 1244.60009

[Cassels 1972] J. W. S. Cassels, *An introduction to Diophantine approximation*, Cambridge Tracts in Mathematics and Mathematical Physics **45**, Hafner, New York, 1972. MR 50 #2084 Zbl 0077.04801

[Dani 1981] S. G. Dani, "Invariant measures and minimal sets of horospherical flows", *Invent. Math.* **64**:2 (1981), 357–385. MR 83c:22009 Zbl 0498.58013

[Dani and Margulis 1990] S. G. Dani and G. A. Margulis, "Orbit closures of generic unipotent flows on homogeneous spaces of SL(3, $\mathbb{R}$)", *Math. Ann.* **286**:1-3 (1990), 101–128. MR 91k:22026 Zbl 0679.22007

[Dani and Margulis 1993] S. G. Dani and G. A. Margulis, "Limit distributions of orbits of unipotent flows and values of quadratic forms", pp. 91–137 in *I. M. Gelfand Seminar* (Moscow, 1993), vol. 1, edited by S. Gelfand and S. Gindikin, Adv. Soviet Math. **16**, American Mathematical Society, Providence, RI, 1993. MR 95b:22024 Zbl 0814.22003

[Eskin et al. 1998] A. Eskin, G. A. Margulis, and S. Mozes, "Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture", *Ann. of Math.* (2) **147**:1 (1998), 93–141. MR 99a:11043 Zbl 0906.11035

[Gorodnik 2004] A. Gorodnik, "Oppenheim conjecture for pairs consisting of a linear form and a quadratic form", *Trans. Amer. Math. Soc.* **356**:11 (2004), 4447–4463. MR 2005h:11146 Zbl 1051.11034

[Helgason 2000] S. Helgason, *Groups and geometric analysis: integral geometry, invariant differential operators, and spherical functions*, Mathematical Surveys and Monographs **83**, American Mathematical Society, Providence, RI, 2000. MR 2001h:22001 Zbl 0965.43007

[Helgason 2001] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Graduate Studies in Mathematics **34**, American Mathematical Society, Providence, RI, 2001. MR 2002b:53081 Zbl 0993.53002

[Henk and Wills 2008] M. Henk and J. M. Wills, "Minkowski's successive minima", pp. 129–142 in *Number theory and discrete geometry* (Chandigarh, 2005), edited by R. Balasubramanian et al., Ramanujan Math. Soc. Lect. Notes Ser. **6**, Ramanujan Math. Soc., Mysore, 2008. MR 2010f:11114 Zbl 1234.11085

[Kleinbock et al. 2002] D. Kleinbock, N. Shah, and A. Starkov, "Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory", pp. 813–930 in *Handbook of dynamical systems*, vol. 1A, edited by B. Hasselblatt and A. Katok, North-Holland, Amsterdam, 2002. MR 2004b:22021 Zbl 1050.22026

[Margulis 1989] G. A. Margulis, "Discrete subgroups and ergodic theory", pp. 377–398 in *Number theory, trace formulas and discrete groups* (Oslo, 1987), edited by K. E. Aubert et al., Academic Press, Boston, 1989. MR 90k:22013a Zbl 0675.10010

[Platonov and Rapinchuk 1991] V. P. Platonov and A. S. Rapinchuk, *Алгебраические группы и теория чисел*, Nauka, Moscow, 1991. Translated as *Algebraic groups and number theory*, Pure and Applied Mathematics **139**, Academic Press, Boston, 1994. MR 93j:11023 Zbl 0732.20027

[Ratner 1994] M. Ratner, "Invariant measures and orbit closures for unipotent actions on homogeneous spaces", *Geom. Funct. Anal.* **4**:2 (1994), 236–257. MR 95c:22018 Zbl 0801.22008

[Rudin 1976] W. Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill, New York, 1976. MR 52 #5893 Zbl 0346.26002

[Rudin 1987] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill, New York, 1987. MR 88k:00002 Zbl 0925.00005

[Sargent 2013] O. Sargent, "Density of values of linear maps on quadratic surfaces", preprint, 2013. To appear in *J. Number Theor.* arXiv 1111.4428v3

[Schmidt 1968] W. M. Schmidt, "Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height", *Duke Math. J.* **35** (1968), 327–339. MR 37 #161 Zbl 0172.06304

[Schnell 1995] U. Schnell, "Successive minima, intrinsic volumes, and lattice determinants", *Discrete Comput. Geom.* **13**:2 (1995), 233–239. MR 95k:52029 Zbl 0818.52003

Communicated by Peter Sarnak
Received 2013-03-05 Revised 2013-12-10 Accepted 2014-01-22

oliver.sargent@bris.ac.uk *Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, United Kingdom*

mathematical sciences publishers msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2014 is US $225/year for the electronic version, and $400/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebra & Number Theory

Volume 8 No. 4 2014

1937-0652(2014)8:4;1-7