J. Inst. Math. Jussieu (2023), **22**(5), 2075–2115 2075
doi:10.1017/S147474802100058X

QUANTUM ERGODICITY FOR COMPACT QUOTIENTS OF $\mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$ IN THE BENJAMINI–SCHRAMM LIMIT

FARRELL BRUMLEY[1] AND JASMIN MATZ[2]

[1] *Université Sorbonne Paris Nord, Laga – Institut Galilée, 99 Avenue Jean Baptiste Clément, Villetaneuse 93430, France*
(brumley@math.univ-paris13.fr)

[2] *Department of Mathematical Science, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark*
(matz@math.ku.dk)

(Received 20 October 2020; revised 2 November 2021; accepted 2 November 2021; first published online 13 December 2021)

Abstract We study the limiting behavior of Maass forms on sequences of large-volume compact quotients of $\mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$, $d \geq 3$, whose spectral parameter stays in a fixed window. We prove a form of quantum ergodicity in this level aspect which extends results of Le Masson and Sahlsten to the higher rank case.

Key words and phrases: quantum ergodicity; Benjamini–Schramm convergence; locally symmetric spaces

2020 Mathematics Subject Classification: Primary 58J51
Secondary 11F72; 37A46

1. Introduction

Let Y be a closed Riemann manifold, and let $\mathcal{B} = \{\psi_i\}$ be an orthonormal basis of $L^2(Y)$ consisting of Laplacian eigenfunctions $\Delta\psi_i = \lambda_i\psi_i$. A subsequence $\{\psi_{i_j}\}$ of $\mathcal{B}$ is called *quantum ergodic* if, for every degree 0 pseudo-differential operator A on Y with principal symbol $a \in C(S^*Y)$, we have

$$\langle A\psi_{i_j}, \psi_{i_j}\rangle_{L^2(Y)} \to \fint_{S^*Y} a\, d\mu_L \quad \text{as } j \to \infty.$$

Here μ_L is the Liouville measure on the cosphere bundle S^*Y and $\fint$ indicates normalization by the volume. The quantum ergodicity theorem of Šnirelman [32], Zelditch [35], and Colin de Verdière [10] states that if the geodesic flow on S^*Y is ergodic, one may extract from $\mathcal{B}$ a density 1 quantum ergodic subsequence. More precisely, they show

$$\frac{1}{N(\lambda)} \sum_{\lambda_i \leq \lambda} \left| \langle A\psi_i, \psi_i\rangle_{L^2(Y)} - \fint_{S^*Y} a\, d\mu_L \right|^2 \to 0 \tag{1}$$

as $\lambda \to \infty$, where $N(\lambda) = |\{i : \lambda_i \leq \lambda\}|$.

In this paper we are concerned with a version of quantum ergodicity where, in contrast to the foregoing semiclassical statement, in which the manifold is fixed and the eigenvalue goes to infinity (the large-frequency regime), we allow the manifold to vary while keeping the eigenvalue constrained to a fixed spectral window (the large-spacial regime).

The most natural setting in which one can formulate such variation is in the Gromov–Hausdorff space of pointed locally compact spaces – or, rather, its space of probability measures – where one has a notion of convergence due to Benjamini and Schramm [5]. Here one may consider a sequence of manifolds which converge, almost everywhere, to their common universal cover. Under certain auxiliary conditions, the question of quantum ergodicity for Benjamini–Schramm convergent sequences was recently settled for a large class of rank 1 symmetric spaces [2, 24]. The aim of the present paper is to address this question for higher-rank locally symmetric spaces; for the most part we restrict ourselves to compact quotients associated with $\mathrm{SL}_d(\mathbb{R})$.

1.1. The setting of locally symmetric spaces

As our setting will henceforth be that of locally symmetric spaces, we begin by commenting on some of their particular features in higher rank, and reviewing what is known for their quantum ergodic properties in the large-frequency regime.

We let S be a Riemannian globally symmetric space of noncompact type: nonpositively curved and having no Euclidean de Rham local factor. We may write $S = G/K$, where G is a connected semisimple Lie group with finite center and K is a maximal compact subgroup. The rank of S is defined to be the dimension of a maximal flat subspace in S; equivalently, it is the dimension of a maximal split torus A in G.

In rank 1, S is of strictly negative curvature, and A gives rise to the geodesic flow on the cosphere bundle of S via the identification of the latter with G/M, where M is the centralizer of A in K. It is important to note, however, that when the rank is strictly greater than 1, such as for $\mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$ when $d \geq 3$, the curvature of S, while nonpositive, can vanish, as indeed any geodesic triangle in a maximal flat will be Euclidean.

Now let $\Gamma \subset G$ be a uniform lattice in G and form $Y = \Gamma\backslash S$, a compact locally symmetric space with geometry S. In higher rank, due to the presence of maximal flats, the geodesic flow is not ergodic on the cosphere bundle of Y. For this reason, the quantum ergodic statement (1) cannot hold, as it is known to be equivalent with the ergodicity of the geodesic flow [36, Theorem 1].

Nevertheless, the higher-rank split torus A does act ergodically – in fact it is mixing – on $\Gamma\backslash G$, and this is enough for one to expect similar quantum ergodic phenomena to that of Šnirelman's theorem, at least if one refines the notion of eigenfunction as follows.

Recall that a Maass form on Y is a function $\psi \in L^2(Y)$ which is a joint eigenfunction of the algebra $\mathcal{D}(S)$ of left-G-invariant differential operators on S. Since $\Delta \in \mathcal{D}(S)$ a Maass form is again a Laplacian eigenfunction, but in higher rank, $L^2(Y)$ can be further diagonalized. A Maass form ψ gives rise to an algebra homomorphism $\chi \in \mathrm{Hom}_{\mathbb{C}-\mathrm{alg}}(\mathcal{D}(S),\mathbb{C})$ verifying $D\psi = \chi(D)\psi$ for all $D \in \mathcal{D}(S)$. The Harish-Chandra homomorphism allows one to realize χ as the Weyl group orbit of an element ν in $\mathfrak{a}_{\mathbb{C}}^*$, the complexification of the dual of the Lie algebra $\mathfrak{a}$ of A. We call ν the *spectral parameter* of ψ.

One can then formulate a natural extension of formula (1) for Maass forms on Y with growing spectral parameter. To the best of our knowledge, this form of quantum ergodicity has not yet been established, although recent work by Nelson and Venkatesh [28] on microlocal analysis and representation theory should shed light on this problem.

Remark 1.1. There are of course harder conjectures that one can formulate here, such as quantum *unique* ergodicity [25, Conjecture 1.2]. In the arithmetic setting, certain higher-rank congruence manifolds have been shown to satisfy the arithmetic quantum unique ergodicity property [30, 31], a generalization of the work (and techniques) of Lindenstrauss [26] (see also [33]). We will not make further comments on this active line of research, and refer the reader instead to [24, §1.3] for a discussion of a conjectural strengthening of quantum ergodicity for varying manifolds.

1.2. Our main result

We now pass to the large-spacial regime for higher-rank locally symmetric spaces $Y = \Gamma\backslash S$, and allow Γ to vary along a nonconjugate sequence of torsion-free uniform lattices with growing covolume.

We furthermore require that the sequence Γ be uniformly discrete, a property which we now recall. Let d be the Riemannian distance on S. Then the local injectivity radius about a point $x \in Y$ is the quantity

$$\mathrm{InjRad}_\Gamma(x) = \frac{1}{2}\min\{d(x,\gamma.x) : 1 \neq \gamma \in \Gamma\}.$$

The global injectivity radius $\mathrm{InjRad}(Y)$ is the infimum of $\mathrm{InjRad}_\Gamma(x)$ over all $x \in Y$; this is strictly positive, since Y is compact. We say that that a sequence of uniform torsion-free lattices Γ_n is *uniformly discrete* if $\mathrm{InjRad}(Y_n)$ is bounded away from zero. A conjecture of Margulis [27] states that this is automatic in higher rank; this is a weak form of the Lehmer conjecture on monic integral polynomials.

Our precise result is as follows. We refer to §3.1 for any unexplained notation.

Theorem 1.1. *Set $d \geq 3$. Let $\Gamma_n \subset \mathrm{SL}_d(\mathbb{R})$ be a uniformly discrete sequence of torsion-free cocompact lattices such that $\mathrm{vol}(Y_n) \to \infty$ as $n \to \infty$, where $Y_n = \Gamma_n\backslash \mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$. Let a_n be a sequence of uniformly bounded, measurable functions on Y_n.*

There is $\varrho > 1$ such that for sufficiently regular $\nu \in i\mathfrak{a}^*$, we have

$$\frac{1}{N(B_0(\nu,\varrho),\Gamma_n)} \sum_{j:\nu_j^{(n)} \in B_0(\nu,\varrho)} \left| \left\langle a_n \psi_j^{(n)}, \psi_j^{(n)} \right\rangle_{L^2(Y_n)} - \fint_{Y_n} a_n d\,\mathrm{vol}_{Y_n} \right|^2 \to 0$$

as $n \to \infty$, where $\fint$ denotes the normalization of the integral by the volume of Y_n,

$$B_0(\nu,\varrho) = \{\lambda \in i\mathfrak{a}^* : \|\lambda - \nu\|_2 \leq \varrho\}$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

is the ball of radius ϱ *in the unramified tempered spectrum, and*

$$N(B_0(\nu,\varrho),\Gamma_n) = \left|\left\{j : \nu_j^{(n)} \in B_0(\nu,\varrho)\right\}\right|. \tag{2}$$

This extends to higher rank the rank 1 version of the same result in [24] (for hyperbolic surfaces) and [2] (for higher-dimensional hyperbolic manifolds), which themselves built upon the breakthrough results of Anantharaman and Le Masson [3] for large regular graphs. All of these works imposed the following additional hypotheses:

(i) The manifolds (or graphs) Y_n converge to S *in the sense of Benjamini and Schramm*: for every $R > 0$, we have

$$\frac{\mathrm{vol}\left((Y_n)_{\leq R}\right)}{\mathrm{vol}(Y_n)} \to 0$$

as $n \to \infty$, where the R-thin part of Y_n is defined as

$$(Y_n)_{\leq R} = \left\{x \in Y_n : \mathrm{InjRad}_{\Gamma_n}(x) \leq R\right\}. \tag{3}$$

(ii) The Y_n have a uniform spectral gap.

The proof of Theorem 1.1 also requires those properties, but they are automatic in higher rank (see [1, §4] and [20]).

We note that in [2] the large-frequency regime for Laplacian eigenfunctions was interpreted as the Benjamini–Schramm convergence to Euclidean space via a rescaling of the Riemannian metric.

Remark 1.2. We point out two differences between the quantum ergodicity theorem of [2] for rank 1 spaces and the higher-rank version we present in Theorem 1.1. In contrast to our result, the authors of [2] allow for two more general features:

(a) They allow the spectral window to shrink with n, thereby isolating in the large-n limit a fixed tempered eigenvalue for the universal cover. This added flexibility was also present in [3, Theorem 1.3].

(b) They take more general operators than scalar multiplication by functions a_n. This more advanced formulation was first put forward in [3, Theorem 1.7], using the pseudo-differential calculus for trees developed in [23].

Our Theorem 1.1 therefore more closely resembles the main results of [8, 24], in which neither of these two features is present. By taking more elaborate test functions, we believe we can incorporate (a) into our setup. By contrast, we have so far been unable to extend (b) to this higher-rank setting.

1.3. Comments on the proof

To prove the theorem we first reduce the assertion to several intermediate statements, as explained in §2. This reduction roughly follows along the lines of [24] and the subsequent [2]. To establish these intermediate results, however, we need several new ideas to deal with a number of issues that arise only in rank 2 or higher:

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

- The first main reduction step involves the use of a normalized averaging operator on $S = G/K$ (a kind of wave propagation) with expanding support C_t, $t \to \infty$. At a later critical point we need to estimate from above the volume of intersections $C_t \cap gC_t$ with $g \in G$. In [2, 24] they work with C_t being the metric ball B_t in S of radius t, and exploit the fact that S is a CAT(-1) space in rank 1. In higher rank, S is only a CAT(0) space, and working with intersection of metric balls becomes problematic. We therefore need to define new types of C_t that look more 'polytopal' and are easier to work with in higher rank (see equation (15) and §5.5).
- We need to establish a suitable lower bound for certain averages of spherical functions. In rank 1, this can be dealt with in a relatively straightforward manner, as the elementary spherical functions are basically linear combinations of trigonometric functions in one variable. In higher rank, we need to deal with linear combinations of exponential functions in several variables, which makes the analysis much more delicate (see §4). The techniques of §4 might also be of interest in other contexts.

Additionally, we need to establish a type of local Weyl law/limit multiplicity for the Y_n that gives a lower count for the number of eigenvalues locally around sufficiently regular points in the spectrum of Y_n. We only need the sharp lower bound in the level aspect stated in formula (7), but along the way we prove a stronger version that also yields the right order in the spectral parameter; for the exact statement, see Proposition 2.1. Such a result is also necessary for the rank 1 situation, but involves a much more careful analysis of the nontempered spectrum in higher rank.

2. Outline of proof and reduction steps

In this section, we shall reduce the proof of Theorem 1.1 to that of two auxiliary estimates: one spectral, one geometric. Let $G = \mathrm{SL}_d(\mathbb{R})$ and $K = \mathrm{SO}(d)$. As a preliminary step, we note that by replacing a_n with $a_n - \fint_{Y_n} a_n$, we may suppose that the measurable, right-K-invariant functions a_n on $\Gamma_n \backslash G$ satisfy

$$\int_{\Gamma_n \backslash G} a_n = 0 \quad \text{and} \quad \|a_n\|_\infty \leq 1. \tag{4}$$

Under that assumption it will then be enough to prove

$$\frac{1}{N(B_0(\nu,\varrho),\Gamma_n)} \sum_{j:\nu_j^{(n)} \in B_0(\nu,\varrho)} \left| \left\langle a_n \psi_j^{(n)}, \psi_j^{(n)} \right\rangle_{L^2(Y_n)} \right|^2 \to 0 \tag{5}$$

as $n \to \infty$. The rest of the paper is devoted to establishing formula (5).

2.1. Spectral estimate

We shall first need to control the spectral counting function defined in equation (2). This is provided in the following result, to be proved in §3 (see [2] for a more abstract approach, in rank 1, using the limit-multiplicity result of [1]):

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Proposition 2.1. *Let G be a connected noncompact simple Lie group with finite center, and K a maximal compact subgroup. Let $S = G/K$ be the associated irreducible Riemannian globally symmetric space of noncompact type. Let Γ_n be a sequence of uniformly discrete, torsion-free, cocompact lattices in G such that $Y_n = \Gamma_n \backslash S$ converges, in the sense of Benjamini and Schramm, to S. There exist $0 < \varrho_1 < 1 < \varrho_2$ such that the following holds: Fix $\epsilon > 0$, and let $\nu \in i\mathfrak{a}^*$ be ε-regular. Then*

$$N(B(\nu,\varrho_1),\Gamma_n) \ll \mathrm{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu) \ll_\varepsilon N(B(\nu,\varrho_2),\Gamma_n), \tag{6}$$

where $B(\nu,\varrho) = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} \mid \|\lambda - \nu\|_2 < \varrho\}$, and $N(B(\nu,\varrho),\Gamma_n) = \left|\left\{j \mid \nu_j^{(n)} \in B(\nu,\varrho)\right\}\right|$ is the counting function.

We shall only make use of Proposition 2.1 in the form

$$\mathrm{vol}(\Gamma_n\backslash G) \ll N(B(\nu,\varrho),\Gamma_n), \tag{7}$$

for a sufficiently regular spectral parameter ν. Indeed formula (7) follows from the second estimate in formula (6) using the fact that $\tilde{\beta}(\nu) \geq 1$ for all $\nu \in i\mathfrak{a}^*$. See §3.1 for our notational conventions on the dependence on the implied constants on the assumption of sufficient regularity.

We note that Proposition 2.1 is stated for general symmetric spaces S. By contrast, we have restricted the setting of the next two results, Theorems 2.2 and 2.3, to the symmetric space $S = \mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$.

2.2. The averaging subset

Following [8, 24], the main idea behind formula (5) is the strategic use of a self-adjoint normalized averaging operator which on one hand acts nondecreasingly on the spectral average (Theorem 2.2) and on the other hand has small Hilbert–Schmidt norm (Theorem 2.3). This idea can be traced back to the work of Brooks, Le Masson, and Lindenstrauss [8], which presents an alternative proof of [3, Theorem 1.3], using discrete normalized averaging operators.

In the following we give a definition of an exhaustive sequence of sets $E_t \subset \mathrm{SL}_d(\mathbb{R})$, which we shall use in our averaging operators. The difficulty in higher rank is finding a set E_t which satisfies simultaneously the desired spectral and geometric properties (in rank 1, the obvious choice of a Riemannian metric ball is shown to work in [[2], 24]).

We begin by introducing some notation. Let $\mathfrak{g}$ be the trace 0 matrices in $M_d(\mathbb{R})$. Let

$$\mathfrak{a} = \{X = \mathrm{diag}(X_1,\ldots,X_d) \in \mathfrak{g}\}$$

be the standard Cartan subalgebra of diagonal matrices. Let $W = N_K(\mathfrak{a})/Z_K(\mathfrak{a})$ be the Weyl group of $(\mathfrak{a},\mathfrak{g})$; the action of $W \simeq S_d$ on $\mathfrak{a}$ is by permutation of the coordinates. Let $\mathfrak{a}^+$ be the standard positive Weyl chamber in $\mathfrak{a}$ given by $X_1 > X_2 > \cdots > X_d$. For a vector $X \in \mathfrak{a}$ we will write

$$\|X\|_\infty = \max\{|X_1|,\ldots,|X_d|\}.$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

The W-invariant norm $\|\cdot\|_\infty$ on $\mathfrak{a}$ induces a subadditive bi-K-invariant norm on $G = \mathrm{SL}_d(\mathbb{R})$ via the Cartan decomposition $G = K\exp\left(\overline{\mathfrak{a}^+}\right)K$. Namely, if $g \in G$ with $g \in Ke^XK$, then we write $|g| = \|X\|_\infty$. We have $|g| \geq 0$, $\left|g^{-1}\right| = |g|$, and $|g_1g_2| \leq |g_1| + |g_2|$.

With this notation, we put

$$E_t = \{g \in G : |g| \leq t\}. \tag{8}$$

This is the averaging set we shall use for our wave propagator. One may view E_t as a radially invariant subset of the symmetric space G/K. See §2.5 for more commentary on the nature of E_t.

2.3. The two main results

We now define

$$k_t = \frac{1}{\sqrt{m_G(E_t)}}\mathbf{1}_{E_t}.$$

Let $\rho_{\Gamma\backslash G}$ denote the right-regular representation of G on $L^2(\Gamma\backslash G)$, and consider the wave-propagation operator on $L^2(\Gamma\backslash G)$ given by

$$U_t = \rho_{\Gamma\backslash G}(k_t). \tag{9}$$

From $E_t^{-1} = E_t$ it follows that k_t is a self-adjoint operator. For a measurable right-K-invariant function a on $\Gamma\backslash G$ satisfying equation (4), and for $\tau > 0$, we consider the time average

$$\mathbf{A}(\tau) = \frac{1}{\tau}\int_0^\tau U_t a U_t dt. \tag{10}$$

In §4 we shall prove the following spectral estimate:

Theorem 2.2 (spectral estimate). *Let $S = \mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$. Let Γ a cocompact lattice in $\mathrm{SL}_d(\mathbb{R})$, and put $Y = \Gamma\backslash S$. Let $\Omega \subseteq i\mathfrak{a}^*$ be compact. There exist constants $c,\tau_0 > 0$, depending on Ω, such that for $\tau \geq \tau_0$, we have*

$$\sum_{j:\nu_j\in\Omega}\left|\langle a\psi_j,\psi_j\rangle_{L^2(Y)}\right|^2 \leq c\sum_{j:\nu_j\in\Omega}\left|\langle \mathbf{A}(\tau)\psi_j,\psi_j\rangle_{L^2(Y)}\right|^2.$$

In §5 we shall use the Nevo mean ergodic theorem [29] (see also [15, Theorem 4.1]) to prove the following geometric estimate. Recall that the Hilbert–Schmidt norm of a bounded operator $\mathbf{A}$ on a separable Hilbert space H is given by

$$\|\mathbf{A}\|_{\mathrm{HS}}^2 = \sum_i |\langle \mathbf{A}e_i, e_i\rangle|^2,$$

where $\{e_i\}$ is any orthonormal basis of H.

Theorem 2.3 (geometric estimate). *Let $S = \mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$. Let Γ be a cocompact torsion-free lattice in $\mathrm{SL}_d(\mathbb{R})$, and put $Y = \Gamma\backslash S$. Let a be a measurable function on*

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

$L^2(\Gamma\backslash S)$. *There are constants* $b, c_1, c_2 > 0$*, depending only on* d*, such that for* $\tau > 0$*,*

$$\|\boldsymbol{A}(\tau)\|_{\mathrm{HS}}^2 \ll_b \frac{\|a\|_2^2}{\tau} + \frac{e^{c_1\tau}}{\mathrm{InjRad}(Y)^{\dim S}} \mathrm{vol}\left((\Gamma\backslash G)_{\leq c_2(2\tau+b)}\right) \|a\|_\infty^2.$$

2.4. Reduction to two main results

We now deduce the main estimate (5) from the foregoing results.

We first note that if $\nu \in i\mathfrak{a}^*$ is sufficiently regular in the sense that $|\langle \nu, \alpha^\vee\rangle| \geq C_\varrho$ for all simple roots α, with $C_\varrho > 0$ a sufficiently large constant depending on ϱ, then $\left\{j \mid \nu_j^{(n)} \in B_0(\nu,\varrho)\right\} = \left\{j \mid \nu_j^{(n)} \in B(\nu,\varrho)\right\}$ because of formula (23). By the lower bound (7), which follows from Proposition 2.1, it suffices to prove an upper bound for

$$\frac{1}{\mathrm{vol}(\Gamma_n\backslash G)} \sum_{j:\nu_j^{(n)} \in B(\nu,\varrho)} \left|\left\langle a_n \psi_j^{(n)}, \psi_j^{(n)}\right\rangle_{L^2(Y_n)}\right|^2.$$

From Theorem 2.2, it follows that for $\tau \geq \tau_0$,

$$\frac{1}{\mathrm{vol}(\Gamma_n\backslash G)} \sum_{j:\nu_j^{(n)} \in B(\nu,\varrho)} \left|\left\langle a_n \psi_j^{(n)}, \psi_j^{(n)}\right\rangle_{L^2(Y_n)}\right|^2 \ll \frac{1}{\mathrm{vol}(\Gamma_n\backslash G)} \left\|\mathbf{A}^{(n)}(\tau)\right\|_{\mathrm{HS}}^2.$$

We now apply Theorem 2.3 to the right-hand side. To bound the first term, we use the estimate $\|a_n\|_2^2 \leq \|a_n\|_\infty^2 \mathrm{vol}(\Gamma_n\backslash G)$ along with the uniform boundedness of a_n. For the second term we insert the bound $\mathrm{InjRad}(Y_n) \gg 1$ coming from the uniformly discrete hypothesis in the statement of Theorem 1.1. We obtain

$$\frac{1}{\mathrm{vol}(\Gamma_n\backslash G)} \left\|\mathbf{A}^{(n)}(\tau_n)\right\|_{\mathrm{HS}}^2 \ll \frac{1}{\tau_n} + e^{c_1\tau_n} \frac{\mathrm{vol}\left((\Gamma_n\backslash G)_{\leq c_2(2\tau_n+b)}\right)}{\mathrm{vol}(\Gamma_n\backslash G)}. \tag{11}$$

We know from [1, §4] that for any sequence Γ_n for which $\mathrm{vol}(\Gamma_n\backslash G) \to \infty$, the space $\Gamma_n\backslash G/K$ converges in the sense of Benjamini and Schramm to $G/K = \mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$ (recall our assumption that $d \geq 3$). We deduce that there is a sequence $R_n \to \infty$ such that

$$\frac{\mathrm{vol}\left((\Gamma_n\backslash G)_{\leq R_n}\right)}{\mathrm{vol}(\Gamma_n\backslash G)} \to 0 \tag{12}$$

as $n \to \infty$. Given this sequence R_n, we let $r_n > 0$ be a sequence such that

(i) $r_n \to \infty$ as $n \to \infty$ and

(ii) $e^{c_1 r_n} \frac{\mathrm{vol}\left((\Gamma_n\backslash G)_{\leq c_2(2r_n+b)}\right)}{\mathrm{vol}(\Gamma_n\backslash G)} \to 0$ as $n \to \infty$.

Taking $\tau_n = r_n$, both terms in formula (11) go to 0 with n, establishing formula (5).

Remark 2.1. One could deduce an effective rate of convergence in Theorem 1.1 for congruence subgroups Γ_n of $\mathrm{SL}_d(\mathbb{Z})$ by inserting, in place of the $o(1)$ result in formula (12), the explicit bounds on the R-thin part for such sequences proved in [1, Theorem 5.2].

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

2.5. Remarks on the averaging subset

We now make several remarks about E_t:

(i) In rank 1, the set E_t recovers the metric ball centered at $i \in \mathbb{H}$ of radius $2t$ for the hyperbolic metric $\left(dx^2 + dy^2\right)/y^2$ on the upper half-plane $\mathbb{H} = \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2)$.

(ii) In higher rank, the set E_t differs substantially from a metric ball centered at the identity for the metric induced by the trace form $\mathrm{Tr}\left(X^\top Y\right)$ (a constant multiple of the Killing form) on $\mathfrak{g}$.

To see this, let us first compare a geodesic ball and the set E_t by means of the Cartan decomposition of $G = \mathrm{SL}_d(\mathbb{R})$. Let $\mathcal{B} = \{X \in \mathfrak{a} : \|X\|_2 \leq 1\}$, where $\|X\|_2^2 = \mathrm{tr}\left(X^2\right) = X_1^2 + \cdots + X_d^2$. The geodesic ball B_t of radius $t > 0$ centered at the origin is given by

$$B_t = K \exp\left(\mathcal{B}_t^+\right) K, \tag{13}$$

where $\mathcal{B}_t^+ = t\mathcal{B} \cap \overline{\mathfrak{a}^+}$ (see [4, Proposition 4.2]). On the other hand, let

$$\mathcal{P} = \left\{X \in \mathfrak{a} \mid \|X\|_\infty \leq 1\right\}. \tag{14}$$

Note that $\mathcal{P}$ is a bounded convex polytope in $\mathfrak{a}$, being the intersection of the half-planes $\pm X_i \leq 1$. It is clear that

$$E_t = K \exp\left(\mathcal{P}_t^+\right) K, \tag{15}$$

where $\mathcal{P}_t^+ = t\mathcal{P} \cap \overline{\mathfrak{a}^+}$. One sees from equations (13) and (15) that an element in B_t or E_t has Cartan radial component contained to, respectively, an expanding ball or a dilated W-invariant polytope in $\mathfrak{a}$.

(iii) Related to the foregoing description is the difference in the volume asymptotics between B_t and E_t. Both are expressed, naturally, with respect to the half-sum of the positive roots $\rho \in \mathfrak{a}^*$, as that quantity governs the Jacobian factor in the Cartan decomposition of the Haar measure on G (see §4.1).

For the geodesic ball, a result of Knieper [21], valid more generally for irreducible Riemannian symmetric spaces $S = G/K$ of noncompact type, states that

$$m_G(B_t) \asymp t^{(r-1)/2} e^{2t\|\rho\|_2}, \tag{16}$$

where r is the rank of S. On the other hand, we show in Corollary A.4 that, for $G = \mathrm{SL}_d(\mathbb{R})$,

$$m_G(E_t) \sim c_2 e^{2t\left\langle \rho, X^0\right\rangle}, \quad t \to \infty,$$

where $X^0 \in \mathcal{P}^+$ is the unique point satisfying

$$\left\langle \rho, X^0 \right\rangle = \max_{X \in \mathcal{P}^+} \langle \rho, X \rangle. \tag{17}$$

One can compute that

$$\|\rho\|_2 = \left\langle \rho, \frac{\rho}{\|\rho\|_2} \right\rangle = \sqrt{d\left(d^2 - 1\right)/12}, \tag{18}$$

whereas using Lemma A.1, we have

$$\left\langle \rho, \frac{X^0}{\|X^0\|_2} \right\rangle = \begin{cases} d^{3/2}/4, & d \text{ even}, \\ (d+1)(d-1)^{1/2}/4, & d \text{ odd}. \end{cases}$$

We have normalized by $\|X^0\|_2$ so that B_t is the smallest geodesic ball containing $E_{t/\|X^0\|_2}$. We see that the exponential volume growth of B_t in $\mathrm{SL}_d(\mathbb{R})/\mathrm{SO}(d)$ is of order $2t\sqrt{d^3/12}$, whereas that of $E_{t/\|X^0\|_2}$ is of order $2t\sqrt{d^3/16}$, which is smaller by a factor of $\sqrt{3}/2$.

(iv) Another norm that one often considers in the context of $G = \mathrm{SL}_d(\mathbb{R})$ is the restriction to $\mathrm{SL}_d(\mathbb{R}) \subset M_d(\mathbb{R})$ of the *Frobenius norm*, defined on $M_d(\mathbb{R})$ as $\|g\|^2 = \operatorname{tr}\left(g^\top g\right)$, where $g^\top$ is the transpose of g. Note that $\|g^{-1}\| \neq \|g\|$ in general. Since invariance under inversion is important for the self-adjointness of the propagator U_t from equation (9), we define the *Frobenius ball* as

$$\boldsymbol{E}_t = \left\{ g \in G : \max\left\{ \|g\|, \|g^{-1}\| \right\} \le e^t \right\}. \tag{19}$$

From the point of view of their large-scale geometry, the Frobenius balls are similar to the sets E_t from equation (8). Indeed, in the proof of Proposition 5.8 we show that $m_G(E_t) \asymp m_G(\boldsymbol{E}_t)$. Moreover, the very statement of Proposition 5.8, which is a key component of Theorem 2.3, holds equally well for $\boldsymbol{E}_t$. On the other hand, because they are not defined by their radial Cartan component, Frobenius balls are not amenable to our proof of Theorem 2.2.

3. Weyl-type law

Our aim in this section is to prove Proposition 2.1. Throughout this section only, we let G denote a connected noncompact simple Lie group with finite center and K a maximal compact subgroup.

3.1. Notation

The notation we introduce here will be consistent with that already introduced for the particular case of $G = \mathrm{SL}_d(\mathbb{R})$ and $K = \mathrm{SO}(d)$.

Let θ be the Cartan involution on G for which $K = G^\theta$. Let Θ be its differential and let $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ be the Cartan decomposition of the Lie algebra $\mathfrak{g}$ of G into the ± 1 eigenspaces of Θ. We may identify $\mathfrak{k}$ with the Lie algebra of K.

Let $\kappa(X,Y) = \operatorname{tr}(\operatorname{ad} X \operatorname{ad} Y)$ denote the Killing form on $\mathfrak{g}$. Then $-\kappa(X, \Theta Y)$ defines an Ad_K-invariant inner product on $\mathfrak{g}$, which induces a left-invariant Haar measure on G, denoted dg or m_G.

The Killing form defines an Ad_K-invariant inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{p}$, and in particular on $\mathfrak{a}$, the maximal abelian subalgebra of $\mathfrak{p}$. Let $W = N_K(\mathfrak{a})/Z_K(\mathfrak{a})$ be the Weyl group. Let $\|\cdot\|_2$ denote the W-invariant norm[1] on $\mathfrak{a}$ induced by $\langle \cdot, \cdot \rangle$.

[1] When $G = \mathrm{SL}_d(\mathbb{R})$, we introduced in §2.5(ii) the norm induced by the trace form on $\mathfrak{g}$, denoted there by the same symbol $\|\cdot\|_2$. The Killing form for $\mathrm{SL}_d(\mathbb{R})$, in fact, differs from the trace form by a constant factor of $2d$. When we return to the specific situation of $\mathrm{SL}_d(\mathbb{R})$ in later sections, we shall always take the norm $\|\cdot\|_2$ on $\mathfrak{a}$ to mean $\|X\|_2^2 = \operatorname{tr}\left(X^2\right)$.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Let $\Phi \subset \mathfrak{a}^*$ denote the system of roots for the adjoint action of $\mathfrak{a}$ on $\mathfrak{g}$. For $\alpha \in \Phi$, let $\mathfrak{g}_\alpha$ denote the corresponding root space. Let $\mathfrak{m} \subset \mathfrak{g}$ be the centralizer of $\mathfrak{a}$ in $\mathfrak{g}$. Then we have the root-space decomposition

$$\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{a} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha.$$

Choose a positive system of roots Φ^+ in Φ. Let $\Delta \subseteq \Phi^+$ be the set of simple roots.

Let $\mathfrak{a}^* = \mathrm{Hom}(\mathfrak{a},\mathbb{R})$ be the dual vector space of $\mathfrak{a}$. We may identify $\mathfrak{a}$ with $\mathfrak{a}^*$ via the Killing form. We again denote by $\|\cdot\|_2$ the induced norm on $\mathfrak{a}^*$ and extend it to a complex bilinear form on the complexification $\mathfrak{a}^*_{\mathbb{C}} = \mathfrak{a}^* \otimes_{\mathbb{R}} \mathbb{C}$. We call $\nu \in i\mathfrak{a}^*$ *regular* if $\langle \alpha, \nu \rangle \neq 0$ for all $\alpha \in \Delta$, and for $c > 0$ we call $\nu \in i\mathfrak{a}^*$ *c-regular* if $|\langle \alpha, \nu \rangle| \geq c$ for all $\alpha \in \Phi^+$. An element $\nu \in i\mathfrak{a}^*$ is *sufficiently regular* if there exists a sufficiently regular $c > 0$ (which may vary at each occurrence) such that ν is c-regular. Unless otherwise noted, implied constants for statements valid for sufficiently regular ν can depend on the value of c.

Let $\mathfrak{n} = \sum_{\alpha \in \Phi^+} \mathfrak{g}_\alpha$ and write $N = \exp \mathfrak{n}$. Then we have the Iwasawa decomposition $G = NAK$, where $A = \exp \mathfrak{a}$, which gives rise to the Iwasawa projection

$$H_0 : G \to \mathfrak{a}, \qquad g = nak \mapsto \log a,$$

along the A component.

We may decompose the Riemannian Haar measure dg on G according to the Iwasawa decomposition. We let da denote the Haar measure on A obtained by pushing forward the Lebesgue measure on $\mathfrak{a}$ by means of the exponential. We let dk denote the probability Haar measure dk on K. We normalize the left-invariant Haar measure du on N as in [12, §3.1]. As usual, we put

$$\rho = \frac{1}{2} \mathrm{tr}\left(\mathrm{ad}(\mathfrak{a})|_{\mathfrak{n}}\right) = \frac{1}{2} \sum_{\alpha \in \Phi^+} (\dim \mathfrak{g}_\alpha) \alpha.$$

Then we have [14, Proposition 2.4.10]

$$dg = c_I e^{-2\rho(H_0(g))} du da dk, \quad c_I = 2^{-(1/2)\dim N}.$$

Since A normalizes N and $\det \mathrm{Ad}(a)|_{\mathfrak{n}} = e^{2\rho(H_0(a))}$,

$$dg = c_I da du dk \tag{20}$$

in the decomposition $G = ANK$.

Recall the Cartan decomposition $G = K \exp\left(\overline{\mathfrak{a}^+}\right) K$, where $\overline{\mathfrak{a}^+}$ denotes the closure of the (open) positive Weyl chamber $\mathfrak{a}^+ = \{X \in \mathfrak{a} : \alpha(X) > 0 \, \forall \alpha \in \Phi^+\}$. We have the following decomposition of the Riemannian Haar measure dg into the Cartan decomposition:

$$dg = c_C J(X) dk_1 da dk_2, \quad c_C = 2^{-\dim N} \mathrm{vol}(K) \mathrm{vol}(K/M). \tag{21}$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Here, $\mathrm{vol}(K)$ is the Riemannian volume on K for the measure induced by the inner product $-\kappa(X,Y)$ on $\mathfrak{k}$, and similarly with $\mathrm{vol}(K/M)$. Finally, $J(X)$ is the Jacobian factor, given by

$$J(X) = \prod_{\alpha\in\Phi^+} \sinh(\alpha(X))^{\dim \mathfrak{g}_\alpha}, \quad X \in \overline{\mathfrak{a}^+} \tag{22}$$

(see [14, Proposition 2.4.11]).

Let P_0 be the normalizer of $\mathfrak{m}\oplus\mathfrak{a}\oplus\mathfrak{n}$ in G. Then P_0 has Langlands decomposition $P_0 = MAN$, where M is the centralizer of A in G. For $\lambda \in \mathfrak{a}^*_{\mathbb{C}}$, let π_λ denote the unique unramified irreducible subquotient of $\mathrm{Ind}_{P_0}^G\left(1\otimes e^\lambda \otimes 1\right)$. Let

$$\mathfrak{a}^*_{\mathrm{un}} = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} : \pi_\lambda \text{ unitarizable}\}$$

be the spherical unitary dual of G. Furthermore, let

$$\mathfrak{a}^*_{\mathrm{hm}} = \bigcup_{w\in W} \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} : w\lambda = -\bar{\lambda}\}$$

denote the spherical Hermitian dual of G [22, §3.3]. Let $i\mathfrak{a}$ be the subspace of $\mathfrak{a}^*_{\mathbb{C}}$ consisting of λ taking on purely imaginary values. We may write $\lambda \in \mathfrak{a}^*_{\mathbb{C}}$ uniquely as $\lambda = \mathrm{Re}\,\lambda + \mathrm{Im}\,\lambda \in \mathfrak{a}^* \oplus i\mathfrak{a}^*$. We have

$$i\mathfrak{a}^* \subset \mathfrak{a}^*_{\mathrm{un}} \subset \mathfrak{a}^*_{\mathrm{hm}} \cap \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} : \|\mathrm{Re}\,\lambda\|_2 \le \|\rho\|_2\}. \tag{23}$$

For $\mu \in i\mathfrak{a}^*$ and $r > 0$, we let

$$B_0(\mu,r) = \{\lambda \in i\mathfrak{a}^* : \|\lambda-\mu\|_2 \le r\}$$

denote the ball of radius r about μ in the tempered spectrum $i\mathfrak{a}^*$. When $r = 1$ we write $B_0(\mu)$ for $B_0(\mu,1)$. We also write

$$B(\mu,r) = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} : \|\lambda-\mu\|_2 \le r\}$$

for the ball of radius r around μ in $\mathfrak{a}^*_{\mathbb{C}}$.

3.2. Plancherel density and the c-function

We denote by β the density function for the Plancherel measure on the spherical unitary dual, which can be identified with $\mathfrak{a}^*_{\mathrm{un}}/W$. Then β is a W-invariant function supported on $i\mathfrak{a}^*$ that can be described as a product of Γ-functions [17, Ch. IV, §6]. Following [22], we put

$$\tilde{\beta}(t,\lambda) = \prod_{\alpha\in\Phi^+} \left(t + |\langle\lambda,\alpha^\vee\rangle|\right)^{\dim \mathfrak{g}_\alpha}, \quad \lambda \in i\mathfrak{a}^*,\ t \ge 1,$$

and $\tilde{\beta}(\lambda) = \tilde{\beta}(1,\lambda)$. From [17, Ch. IV, Theorem 6.14] and standard estimates for the Γ-function, it follows that $\beta(\lambda) \ll \tilde{\beta}(\lambda)$ for all $\lambda \in i\mathfrak{a}^*$. Since

$$\begin{aligned} 1 + |\langle t\lambda+\mu,\alpha^\vee\rangle| &\le 1 + t|\langle\lambda,\alpha^\vee\rangle| + |\langle\mu,\alpha^\vee\rangle| \\ &\le t(1+\|\lambda\|_2) + |\langle\mu,\alpha^\vee\rangle| \le (1+\|\lambda\|_2)\left(t + |\langle\mu,\alpha^\vee\rangle|\right) \end{aligned}$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

for all $\lambda, \mu \in i\mathfrak{a}^*$ and $t \geq 1$, we get

$$\tilde{\beta}(t\lambda + \mu) \ll (1 + \|\lambda\|_2)^{\dim \mathfrak{n}} \tilde{\beta}(t, \mu) \tag{24}$$

for all such λ, μ, and t.

The Harish-Chandra $\mathbf{c}$-function $\mathbf{c} : \mathfrak{a}_{\mathbb{C}}^* \longrightarrow \mathbb{C}$ for G asymptotically describes the behavior of the elementary spherical functions $\phi_\lambda\left(e^H\right)$ of G as the group parameter H grows. The quantity $\mathbf{c}(\lambda)$ depends only on the root system of G, and can be explicitly computed as described in [17, Ch. IV, Theorem 6.14]. Up to normalization, the Plancherel density $\beta(\lambda)$ equals $|\mathbf{c}(\lambda)|^{-2}$ for $\lambda \in i\mathfrak{a}^*$.

3.3. Test functions

For $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$, let

$$\varphi_\lambda(g) = \int_K e^{\langle \lambda + \rho, H_0(kg) \rangle} dk$$

denote the spherical function on G with spectral parameter λ. Here $H_0 : G \longrightarrow \mathfrak{a}$ is the Iwasawa projection. Let $C_c^\infty(G//K)$ be the space of complex-valued bi-K-invariant functions on G. The Harish-Chandra transform of a function $k \in C_c^\infty(G//K)$ is defined to be

$$\hat{k}(\lambda) = \int_G \varphi_\lambda(g) k(g) dg.$$

For $t > 0$, let $\mathcal{B}_t$ denote the ball of radius t centered at 0 in $\mathfrak{a}$ with respect to the usual Euclidean norm $\|\cdot\|_2$. Let $G_{\leq t} = K \exp \mathcal{B}_t K$. Let $C_c^\infty(G//K)_{\leq t}$ denote the space of smooth compactly supported bi-K-invariant functions on G, supported on $G_{\leq t}$. Let $\mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)_t$ denote the class of Paley–Wiener functions of exponential type t. The Paley–Wiener theorem with supports (see [13, Theorem 3.5]) states that the Harish-Chandra transform is a topological isomorphism $C_c^\infty(G//K)_{\leq t} \xrightarrow{\sim} \mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)_t^W$ of Fréchet spaces. The inverse map sends $h \in \mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)_t^W$ to

$$\frac{1}{|W|} \int_{i\mathfrak{a}^*} h(\mu) \varphi_\mu(g) \beta(\mu) d\mu.$$

In particular, for $k \in C_c^\infty(G//K)$ we have

$$k(e) = \frac{1}{|W|} \int_{i\mathfrak{a}^*} \hat{k}(\mu) \beta(\mu) d\mu. \tag{25}$$

We shall need a Paley–Wiener function concentrating about ν and verifying certain positivity properties.

Lemma 3.1 ([9, §4]). *For $t \geq 1$ and $\nu \in i\mathfrak{a}^*$, there is $k_{\nu,t} \in C_c^\infty(G/\!/K)_{\leq 1/t}$ whose Harish-Chandra transform $h_{\nu,t} \in \mathcal{PW}(\mathfrak{a}_{\mathbb{C}}^*)_t^W$ satisfies the following:*

1. *$h_{\nu,t}$ is real and nonnegative on $\mathfrak{a}_{\mathrm{hm}}^*$.*
2. *For all $\lambda \in \mathfrak{a}_{\mathrm{un}}^*$ we have*

$$h_{\nu,t}(\lambda) \ll_A (1 + \|(\mathrm{Im}\,\lambda - \nu)/t\|_2)^{-A}.$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

(3) *There are constants* $0 < c_1, c_2 < 1$ *such that for* $\lambda \in \mathfrak{a}_{\mathrm{un}}^*$ *with* $\|\mathrm{Im}\,\lambda - \nu\|_2 \le c_1 t$, *we have* $c_2 \le h_{\nu,t}(\lambda) \le 2$.

(4) $k_{\nu,t}(g) \ll \tilde{\beta}(\nu) t^r (1 + \|\nu\|_2 d(g,K))^{-1/2}$, *where* $r = \dim \mathfrak{a}$.

3.4. Proof of Proposition 2.1.

Recall that for $\nu \in i\mathfrak{a}^*$ and $c > 0$, we denote by $B(\nu, c)$ the ball $\{\lambda \in \mathfrak{a}_{\mathbb{C}}^* \mid \|\lambda - \nu\|_2 \le c\}$ in $\mathfrak{a}_{\mathbb{C}}^*$. Moreover, for any set $\Omega \subseteq \mathfrak{a}_{\mathbb{C}}^*$, we write $N(\Omega, \Gamma_n) = \left|\left\{ j \mid \nu_j^{(n)} \in \Omega \right\}\right|$.

In the notation of Lemma 3.1, we write $h_\nu = h_{\nu,1}$ and $k_\nu = k_{\nu,1}$. Then

$$\sum_{\lambda \in \Lambda_n} h_\nu(\lambda) = \mathrm{vol}(\Gamma_n \backslash G) k_\nu(1) + \int_{\Gamma_n \backslash G} \sum_{\substack{\gamma \in \Gamma_n \\ \gamma \ne 1}} k_\nu \left(g^{-1} \gamma g\right) dg,$$

where Λ_n denotes the spectrum of $\Gamma_n \backslash G$ (each λ appearing with its respective multiplicity).

To deal with the second term, we use Lemma 3.1(4), as well as the support condition on k_ν, to deduce

$$\int_{\Gamma_n \backslash G} \sum_{\substack{\gamma \in \Gamma_n \\ \gamma \ne 1}} k_\nu \left(g^{-1} \gamma g\right) dg \ll \tilde{\beta}(\nu) \int_{\Gamma_n \backslash G} |\{\gamma \in \Gamma_n \setminus \{1\} : d(g, \gamma g) \le 1\}| dg.$$

As Γ_n is torsion free, the inner sum is empty for all $g \in (\Gamma_n \backslash G)_{>1}$, so that it suffices to bound

$$\int_{(\Gamma_n \backslash G)_{\le 1}} N_R(g) dg,$$

where

$$N_R(g) = |\{\gamma \in \Gamma_n : d(g, \gamma g) \le 1\}|. \tag{26}$$

For $g \in (\Gamma_n \backslash G)_{\le 1}$ we apply [1, Lemma 6.18], which provides a constant $C > 0$, depending only on G, such that $N_R(g) \le C\, \mathrm{InjRad}_{\Gamma_n}(g)^{-\dim S}$. The uniform discreteness of the (torsion-free) Γ_n implies that $\mathrm{InjRad}_{\Gamma_n}(g)^{-1} \ll 1$, uniformly in n and g. Taking these estimates together, we get

$$\sum_{\lambda \in \Lambda_n} h_\nu(\lambda) = \mathrm{vol}(\Gamma_n \backslash G) k_\nu(1) + O\left(\tilde{\beta}(\nu)\, \mathrm{vol}(\Gamma_n \backslash G)_{\le 1}\right). \tag{27}$$

Furthermore, we have

$$k_\nu(1) \asymp_\epsilon \tilde{\beta}(\nu), \tag{28}$$

where only the lower bound depends on ε. Indeed, the upper bound results from Lemma 3.1(4). To obtain the lower bound, one applies the Plancherel inversion formula (25) and Lemma 3.1(1) and (3) to get

$$k_\nu(1) \ge \int_{\substack{\lambda \in i\mathfrak{a}^* \\ \|\lambda - \nu\|_2 \le \delta}} h_\nu(\lambda) \beta(\lambda) d\lambda \ge c_2 \int_{\substack{\lambda \in i\mathfrak{a}^* \\ \|\lambda - \nu\|_2 \le \delta}} \beta(\lambda) d\lambda,$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

for any $0 < \delta < c_1$. Recall that $\nu \in i\mathfrak{a}^*$ is ε-regular. Taking δ small enough, we may assume that the $\lambda \in i\mathfrak{a}^*$ such that $\|\lambda - \nu\|_2 \le \delta$ are $\varepsilon/2$-regular. Then the lower bound follows from the inequality $\beta(\lambda) \gg_\varepsilon \tilde{\beta}(\lambda)$ for such λ (see [12, (3.44a)]).

From formulas (27) and (28), it follows that

$$\sum_{\lambda \in \Lambda_n} h_\nu(\lambda) \asymp_\varepsilon \operatorname{vol}(\Gamma_n \backslash G)\tilde{\beta}(\nu)\left(1 + O\left(\frac{\operatorname{vol}(\Gamma_n\backslash G)_{\le 1}}{\operatorname{vol}(\Gamma_n\backslash G)}\right)\right),$$

where only the lower bound depends on ε. By the Benjamini–Schramm assumption, we have $\operatorname{vol}(\Gamma_n\backslash G)_{\le 1} = o(\operatorname{vol}(\Gamma_n\backslash G))$ as $n \to \infty$. Thus for n large enough we have

$$\sum_{\lambda \in \Lambda_n} h_\nu(\lambda) \asymp_\varepsilon \operatorname{vol}(\Gamma_n \backslash G)\tilde{\beta}(\nu). \tag{29}$$

For the first bound in formula (6), it suffices at this point to take $\varrho_1 = c_1$, apply Lemma 3.1(1) and (3), and then quote the upper bound in formula (29) to get

$$N(B(\nu,c_1),\Gamma_n) \le c_2 \sum_{\substack{\lambda \in \Lambda_n \\ \|\lambda-\nu\|_2 \le c_1}} h_\nu(\lambda) \le c_2 \sum_{\lambda \in \Lambda_n} h_\nu(\lambda) \ll \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu).$$

For the second bound in formula (6), we must show that the left-hand side in formula (29) approximates $N(B(\nu,\varrho),\Gamma_n)$, for some $\varrho > 0$.

A crucial ingredient for passing from a smooth count as before to a sharp count will be a good upper bound on $N(B(\mu,t),\Gamma_n)$, for any center $\mu \in i\mathfrak{a}^*$ and any $t \ge 1$. This is proved similarly to the preceding analysis. Indeed, from Lemma 3.1 and the preceding geometric argument, we obtain

$$\sum_{\substack{\lambda \in \Lambda_n \\ \|\operatorname{Im}\lambda-\mu\|_2 \le t}} 1 \ll \sum_{\lambda \in \Lambda_n} h_{\mu,c_1^{-1}t}(\lambda) = \operatorname{vol}(\Gamma_n\backslash G) k_{\mu,c_1^{-1}t}(1) + o\left(\operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\mu)t^r\right).$$

Now $k_{\mu,t}(1) \ll \tilde{\beta}(\mu)t^r$ from Lemma 3.1(4). We conclude that (for n large enough)

$$N(B'(\mu,t),\Gamma_n) \ll \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\mu)t^r, \tag{30}$$

where $B'(\mu,t) = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}} \mid \|\operatorname{Im}\lambda - \mu\|_2 \le t\}$. We now return to the left-hand side of formula (29). We first claim the following:

Lemma 3.2. *For every $\varrho > 1$,*

$$\sum_{\substack{\lambda \in \Lambda_n \\ \|\operatorname{Im}\lambda-\nu\|_2 > \varrho}} h_\nu(\lambda) \ll_A \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu)\varrho^{-A}. \tag{31}$$

Proof. To prove formula (31) we cover $\{\lambda \in i\mathfrak{a}^* : \|\lambda - \nu\|_2 > \varrho\}$ by balls of unit radius with centers based at an affine lattice $\Lambda \subset i\mathfrak{a}^*$ containing ν. Let $A(\nu,\varrho)$ denote the set of all $\lambda \in \Lambda$ such that the ball of radius 1 in $i\mathfrak{a}^*$ around λ is entirely contained in the ball of radius ϱ in $i\mathfrak{a}^*$ around ν.

For any fixed $\mu \in \Lambda - A(\nu,\varrho)$, we bound the contribution of $\lambda \in \Lambda_n$ for which $\operatorname{Im}\lambda \in B(\mu)$ by using Lemma 3.1(2) to get

$$\sum_{\substack{\lambda\in\Lambda_n\\ \|\operatorname{Im}\lambda-\mu\|_2\le 1}} h_\nu(\lambda) \ll_A (1+\|\mu-\nu\|_2)^{-A} N(B'(\mu,1),\Gamma_n).$$

Using formula (30) with $t=1$ and summing over $\mu \in \Lambda - A(\nu,\varrho)$ we get

$$\sum_{\substack{\lambda\in\Lambda_n\\ \|\operatorname{Im}\lambda-\nu\|_2>\varrho}} h_\nu(\lambda) \ll_A \operatorname{vol}(\Gamma_n\backslash G) \sum_{\mu\in\Lambda-A(\nu,\varrho)} (1+\|\mu-\nu\|_2)^{-A}\tilde{\beta}(\mu).$$

Now by formula (24) we have $\tilde{\beta}(\mu) \ll (1+\|\mu-\nu\|_2)^{\dim\mathfrak{n}}\tilde{\beta}(\nu)$, so that

$$\sum_{\|\operatorname{Im}\lambda-\nu\|_2>\varrho} h_\nu(\lambda) \ll_A \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu) \sum_{\mu\in\Lambda-A(\nu,\varrho)} (1+\|\mu-\nu\|_2)^{-A+\dim\mathfrak{n}}.$$

For ϱ sufficiently large, this last sum is at most $O_A\left(\varrho^{-A+\dim\mathfrak{n}}\right)$, as desired. □

From this lemma, we want to deduce the following estimate:

$$\sum_{\substack{\lambda\in\Lambda_n\\ \|\lambda-\nu\|_2>\varrho}} h_\nu(\lambda) \ll_A \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu)\varrho^{-A}. \tag{32}$$

For this we split the sum on the left-hand side into three parts:

$$\sum_{\substack{\lambda\in\Lambda_n\\ \|\operatorname{Im}\lambda-\nu\|_2>\varrho}} h_\nu(\lambda) + \sum_{\substack{\lambda\in\Lambda_n\\ \varrho^2-\|\rho\|_2^2<\|\operatorname{Im}\lambda-\nu\|_2^2\le\varrho^2\\ \|\operatorname{Re}\lambda\|_2^2>\varrho^2-\|\operatorname{Im}\lambda-\nu\|_2^2}} h_\nu(\lambda) + \sum_{\substack{\lambda\in\Lambda_n\\ \|\operatorname{Im}\lambda-\nu\|_2^2\le\varrho^2-\|\rho\|_2^2\\ \|\operatorname{Re}\lambda\|_2^2>\varrho^2-\|\operatorname{Im}\lambda-\nu\|_2^2}} h_\nu(\lambda).$$

Using Lemma 3.2, the first sum can be bounded by $\ll_A \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu)\varrho^{-A}$. The third sum is in fact empty: The conditions on the real and complex parts of λ imply that $\|\operatorname{Re}\lambda\|_2 > \|\rho\|_2$, which is not possible for λ in the spectrum of $L^2(\Gamma_n\backslash G)$, by formula (23). Finally, for the second term, we extend the sum to all $\lambda \in \Lambda_n$ with $\|\operatorname{Im}\lambda-\nu\|_2^2 \ge \varrho^2 - \|\rho\|_2^2$. This is possible because $h_\nu(\lambda)$ is nonzero on the unitary spectrum, by Lemma 3.1. Hence we can use Lemma 3.2 to bound this term by $\ll_A \left(\varrho^2-\|\rho\|_2^2\right)^{-A/2}$.

We deduce from the lower bound in formula (29), as well as from formula (32) that with $A>1$ fixed and ϱ large enough with respect to the implied constant,

$$\sum_{\substack{\lambda\in\Lambda_n\\ \|\lambda-\nu\|_2\le\varrho}} h_\nu(\lambda) \gg_\varepsilon \operatorname{vol}(\Gamma_n\backslash G)\tilde{\beta}(\nu).$$

On the other hand, by applying Lemma 3.1(3) we have

$$N(B(\nu,\varrho),\Gamma_n) \gg \sum_{\substack{\lambda\in\Lambda_n\\ \|\lambda-\nu\|_2\le\varrho}} h_\nu(\lambda).$$

Combing these yields the second bound in formula (6), with $\varrho_2 = \varrho$. □

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

4. Spectral side

We now return to the setting of $G = \mathrm{SL}_d(\mathbb{R})$ and prove Theorem 2.2. In the course of the proof, we will make use of both Appendices A and B.

We first examine how k_t acts on eigenfunctions ψ_λ. The Harish-Chandra transform of k_t is given by

$$h_t(\lambda) = \frac{1}{\sqrt{m_G(E_t)}} \int_{E_t} \varphi_\lambda(g)\, dg. \tag{33}$$

Then, recalling the definition (9), we have

$$U_t \psi_\lambda = h_t(\lambda)\psi_\lambda,$$

which follows from the doubling formula of elementary spherical functions.

Let $\mathbf{A}(\tau)$ be the time-averaging operator of equation (10). Since k_t is self-adjoint, we have

$$\left|\langle \mathbf{A}(\tau)\psi_\lambda, \psi_\lambda\rangle_{L^2(Y)}\right|^2 = \left(\frac{1}{\tau}\int_0^\tau |h_t(\lambda)|^2 dt\right)^2 \left|\langle a\psi_\lambda, \psi_\lambda\rangle_{L^2(Y)}\right|^2.$$

To prove Theorem 2.2, it will therefore be enough to show that $|h_t(\lambda)|$ is bounded away from 0 on average over t. More precisely, we prove the following higher-rank generalization of [2, §8.1]:

Proposition 4.1. *Let $G = \mathrm{SL}_d(\mathbb{R})$. Given a compact nonempty set $\Omega \subseteq i\mathfrak{a}^*$, there exist constants $C, \tau_0 > 0$, depending on Ω, such that for every $\tau \geq \tau_0$ and $\lambda \in \Omega$,*

$$\frac{1}{\tau}\int_0^\tau |h_t(\lambda)|^2 dt \geq C.$$

4.1. Main idea of proof

We begin by writing the integral defining h_t in polar coordinates, according to the Cartan decomposition. Recall the averaging set E_t from §2.2 and its polar decomposition (15). Letting $f_\lambda(X) = e^{\langle \rho, X\rangle}\varphi_\lambda\left(e^X\right)$, and using the Cartan measure decomposition (21), we find that equation (33) becomes

$$h_t(\lambda) = \frac{c_C}{\sqrt{m_G(E_t)}} \int_{\mathcal{P}_t^+} f_\lambda(X) J(X) e^{-\langle \rho, X\rangle} dX. \tag{34}$$

The idea of the proof of Proposition 4.1 is to replace f_λ by the main term Φ_λ, discussed later, of its Harish-Chandra asymptotic expansion relative to the Levi determined by X^0, defined in equation (17). We do this by showing in Appendix A (see Lemma A.2) that $\mathcal{P}^+$ is the intersection with $\overline{\mathfrak{a}^+}$ of a translated cone $\mathcal{C} = X^0 + \mathcal{C}^0$ in $\mathfrak{a}$. We then show that when λ is taken to be rational, Φ_λ behaves like a sum of characters in the direction of X^0. An argument using linear independence of characters then yields the result. We carry out this argument in this section, using some combinatorial results which we establish in Appendix B.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

4.2. The main term and periodicity

We wish to replace f_λ by the main term in its Harish-Chandra expansion. This is possible thanks to Proposition 4.2. To state it, we must introduce some notation beyond that of §3.1. For simplicity, we continue to take $G = \mathrm{SL}_d(\mathbb{R})$, although Proposition 4.2 applies more generally.

A Levi subgroup of G will be called *semistandard* if it contains A. Let $\Delta_0 \subseteq \Delta$ be a (possibly empty) subset of the set of simple roots, and let Φ_0^+ be the subset of Φ^+ generated by Δ_0. The sets Δ_0 correspond to semistandard Levi subgroups $M \subseteq G$ such that Φ_0^+ is the set of positive roots of A on M. With this identification, $\Delta_0 = \emptyset$ corresponds to $M = A$ and $\Delta_0 = \Delta$ to $M = G$.

For fixed Δ_0 and corresponding M, we further introduce the following notation:

- ρ^M is the half-sum of all $\alpha \in \Phi_0^+$;
- W^M is the Weyl group of A in M;
- $\mathbf{c}^M$ denotes the Harish-Chandra $\mathbf{c}$-function for Φ_0^+;
- φ_λ^M is the spherical function on M for $\lambda \in \mathfrak{a}_\mathbb{C}^*$ – that is,

$$\varphi_\lambda^M(m) = \int_{K\cap M} e^{\langle \lambda+\rho^M, H_0(km)\rangle} dk;$$

- for $X \in \mathfrak{a}$, $\lambda \in \mathfrak{a}_\mathbb{C}^*$, and $t > 0$, let $f_\lambda^M(t,X) = e^{\langle \rho^M, tX\rangle}\varphi_\lambda^M\left(e^{tX}\right)$ and (setting $t = 1$) let

$$f_\lambda^M(X) = e^{\langle \rho^M, X\rangle}\varphi_\lambda^M\left(e^{X}\right);$$

- $\mathfrak{a}_M$ is the common null space of all $\alpha \in \Phi_0^+$, and $\mathfrak{a}^M$ is its orthocomplement in $\mathfrak{a}$ relative to the inner product $\langle X,Y\rangle = \operatorname{tr} XY$.

We follow the standard practice of omitting the superscript M from the notation whenever $M = G$. In this way we recover the $f_\lambda(X)$ introduced earlier. Other than G, the only Levi which will be of importance to us throughout this section is the centralizer of X^0 in G. *Henceforth, we shall take M to be the centralizer of X^0 in G, and Δ_0 such that it corresponds to M.* In this case, $X^0 \in \mathfrak{a}_M$.

For $\lambda \in i\mathfrak{a}^*$ regular, we define

$$\Phi_\lambda(X) = \sum_{w\in W^M\backslash W} \frac{\mathbf{c}(w\lambda)}{\mathbf{c}^M(w\lambda)} f_{w\lambda}^M(X). \tag{35}$$

In fact, we can define $\Phi_\lambda(X)$ for any regular $\lambda \in \mathfrak{a}_\mathbb{C}^*$ for which the $\mathbf{c}(w\lambda)$ do not have a pole. We can make this more precise, as follows. For a real number $\eta > 0$, let $A_\eta \subseteq \mathfrak{a}^*$ denote the convex hull of the Weyl group orbit of $\eta\rho$. Define $\mathcal{T}_\eta \subseteq \mathfrak{a}_\mathbb{C}^*$ to be the set of all $\lambda \in \mathfrak{a}_\mathbb{C}^*$ with $\mathrm{Re}\,\lambda \in A_\eta$. Let η be so small that for every positive root α and all $\lambda \in \mathcal{T}_\eta$, we have $\langle \mathrm{Re}\,\lambda, \alpha^\vee\rangle \notin \mathbb{Z}\setminus\{0\}$. We fix such a $0 < \eta < 1$ from now on (we will make our choice more precise later). Note that $\mathcal{T}_\eta$ is W-invariant by construction. We may extend the definition of $\Phi_\lambda(X)$ in equation (35) to $\lambda \in \mathcal{T}_\eta$ regular and $X \in \mathfrak{a}$. Indeed, the $\mathbf{c}$-function is nonzero on $\mathfrak{a}_\mathbb{C}^*$ and does not have any poles for regular $\lambda \in \mathcal{T}_\eta$, by our choice of η.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

The relevance of $\Phi_\lambda(X)$ is apparent from the next result, due to Trombi and Varadarajan [34, Theorem 2.11.2], providing the full asymptotic expansion of the spherical function. We have taken the exact statement from [14, Theorem 5.9.4]. This extends to all $X \in \mathfrak{a}$ the asymptotic expansion proved by Harish-Chandra and Gangolli for conical subsets of regular elements.

Proposition 4.2. *Let $\Omega \subseteq i\mathfrak{a}^*$ be a compact set and $c > 0$. Then there exist $C > 0$ and $m \geq 0$ such that for all $H \in \overline{\mathfrak{a}^+}$ with $\beta_{\Delta_0}(H) \geq c$, where*

$$\beta_{\Delta_0}(H) = \min_{\alpha \in \Delta \setminus \Delta_0} \alpha(H),$$

and for every regular $\lambda \in \Omega$, we have

$$|f_\lambda(H) - \Phi_\lambda(H)| \leq C(1 + \|H\|_2)^m e^{-2\beta_{\Delta_0}(H)}.$$

We now describe the key periodicity property of $\Phi_\lambda(X)$. For this we need the following definition:

Definition 4.3. We say that $\lambda \in i\mathfrak{a}^*$ is *rational* if $i\left\langle w\lambda, X^0 \right\rangle \in \mathbb{Q}$ for all $w \in W$.

If λ is rational, we can choose $\tau_1 \in \mathbb{R}_{>0}$ such that $\tau_1 \left\langle w\lambda, X^0 \right\rangle \in 2i\pi\mathbb{Z}$ for all $w \in W$. Recall the definition of $\mathcal{C}$ from Lemma A.2.

Lemma 4.4. *Let $\lambda \in i\mathfrak{a}^*$ be rational and let τ_1 be defined as before. Set $\tau_n = n\tau_1$ for $n \in \mathbb{N}$. Then for every $Y \in \mathcal{C}$, $t \geq 0$, and $n \in \mathbb{N}$, we have*

$$\Phi_\lambda\left(Y + (t + \tau_n)X^0\right) = \Phi_\lambda\left(Y + tX^0\right).$$

Proof. By the definition of τ_1, it follows that $e^{\left\langle \lambda, (t+\tau_n)X^0 \right\rangle} = e^{\left\langle \lambda, tX^0 \right\rangle}$ for every $n \in \mathbb{N}$. Further, note that since $\left\langle \rho^M, X^0 \right\rangle = 0$ and $H_0\left(ke^{Y+tX^0}\right) = tX^0 + H\left(ke^Y\right)$ for $k \in K \cap M$, we have

$$\varphi_\lambda^M\left(e^{Y+tX^0}\right) = e^{\left\langle \rho^M + \lambda, tX^0 \right\rangle} \varphi_\lambda^M\left(e^Y\right) = e^{\left\langle \lambda, tX^0 \right\rangle} \varphi_\lambda^M\left(e^Y\right).$$

Thus

$$f_\lambda^M\left(Y + tX^0\right) = e^{\left\langle \lambda, tX^0 \right\rangle} f_\lambda^M(Y). \tag{36}$$

Applying these two formulas for every $w\lambda$, $w \in W^M \backslash W$, yields the assertion. □

4.3. Convergence of an integral

To take advantage of Lemma 4.4, we shall need to replace the integration domain $\mathcal{P}_t^+$ in equation (34) by the translated cone $\mathcal{C}$. Note that the Jacobian factor (22) can be alternatively written as

$$J(X) = 2^{-\dim \mathfrak{n}} \sum_{w \in W} \sigma(w) e^{2\langle w\rho, X \rangle}, \quad X \in \overline{\mathfrak{a}^+}, \tag{37}$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

where $\sigma(w) \in \{\pm 1\}$ denotes the signature of w. With this in mind, for $t > 0$ and $\lambda \in \mathcal{T}_\eta$, we put

$$I(t,\lambda) = \int_{\mathcal{C}^0} \Phi_\lambda\left(Y + tX^0\right) e^{\langle \rho, Y\rangle} dY$$

whenever it converges.

We now address this question of convergence. From the defining expression for Φ_λ in equation (35), it will be enough to understand the convergence of the integral

$$J(t,\lambda) = \int_{\mathcal{C}^0} f_\lambda^M\left(Y + tX^0\right) e^{\langle \rho, Y\rangle} dY.$$

Lemma 4.5. *There is $0 < \eta < 1$ such that for every $\lambda \in \mathcal{T}_\eta$ and $t \geq 0$, the integral defining $J(t,\lambda)$ converges absolutely.*

Proof. We first observe that it suffices to consider $t = 0$. Indeed, from equation (36) it follows that

$$J(t,\lambda) = e^{t\left\langle \lambda, X^0\right\rangle} J(\lambda), \tag{38}$$

where we have put $J(\lambda) = J(0,\lambda)$. There should be no confusion between the integral $J(\lambda)$ and the Jacobian factor $J(X)$ from equation (22), due to the different nature of their arguments.

Set $Y \in \mathcal{C}^0$ and write $Y = Y_M + Y^M$, with $Y_M \in \mathfrak{a}_M$ and $Y^M \in \mathfrak{a}^M$. Then

$$f_\lambda^M(Y) = e^{\left\langle \rho^M, Y\right\rangle} \varphi_\lambda^M\left(e^Y\right) = e^{\langle \lambda, Y_M\rangle} e^{\left\langle \rho^M, Y^M\right\rangle} \varphi_\lambda^M\left(e^{Y^M}\right). \tag{39}$$

Proposition B.1 then immediately yields the bound $\left|e^{\langle \lambda, Y_M\rangle}\right| \leq e^{-\eta C\langle \rho, Y\rangle}$ on the first factor on the right-hand side.

To bound the remaining two factors, we first let Y_+^M be a point in the W^M-orbit of Y^M lying in $\overline{\mathfrak{a}^{M,+}}$. Then by [19, Ch. X, §2, Proposition 2.2, Theorem 2.8], we have

$$\begin{aligned}\left|\varphi_\lambda^M\left(e^{Y^M}\right)\right| \leq \varphi_{\operatorname{Re}\lambda}^M\left(e^{Y^M}\right) &= \varphi_{\operatorname{Re}\lambda}^M\left(e^{Y_+^M}\right) \\ &\leq \max_{w \in W^M} e^{\left\langle w\operatorname{Re}\lambda, Y_+^M\right\rangle} \varphi_0^M\left(e^{Y_+^M}\right) \\ &\ll \max_{w \in W^M} e^{\left\langle w\operatorname{Re}\lambda, Y_+^M\right\rangle} e^{-\left\langle \rho^M, Y_+^M\right\rangle} \left(1 + \left\|Y_+^M\right\|_2\right)^{\dim \mathfrak{n}}.\end{aligned}$$

Now $\left\|Y_+^M\right\|_2 = \left\|Y^M\right\|_2$ and $\left\langle \rho^M, Y^M\right\rangle \leq \left\langle \rho^M, Y_+^M\right\rangle$. Thus there is a $c > 0$ such that

$$\left|e^{\left\langle \rho^M, Y^M\right\rangle} \varphi_\lambda^M\left(e^{Y^M}\right)\right| \leq c \max_{w \in W^M} e^{\left\langle w\operatorname{Re}\lambda, Y_+^M\right\rangle} \left(1 + \left\|Y^M\right\|_2\right)^{\dim \mathfrak{n}}.$$

Using the W-invariance of $\langle \cdot, \cdot\rangle$ and $Y^M = Y - Y_M$, we obtain

$$\max_{w \in W^M} \left\langle w\operatorname{Re}\lambda, Y_+^M\right\rangle = \max_{w \in W^M} \left\langle w\operatorname{Re}\lambda, Y^M\right\rangle \leq \max_{w \in W} |\langle w\operatorname{Re}\lambda, Y\rangle| + \max_{w \in W} |\langle w\operatorname{Re}\lambda, Y_M\rangle|.$$

Note that $w\operatorname{Re}\lambda = \operatorname{Re} w\lambda$ and $w\lambda \in \mathcal{T}_\eta$. Using Remark A.1, we have $|\langle \operatorname{Re}\lambda, Y\rangle| \ll -\eta\langle \rho, Y\rangle$ for all $\lambda \in \mathcal{T}_\eta$ and $Y \in \mathcal{C}^0$. From this and Proposition B.1, we deduce that there is a

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

constant $C' > 0$, depending only on d, such that

$$\max_{w \in W^M} \langle w \operatorname{Re} \lambda, Y_+^M \rangle \leq -\eta C' \langle \rho, Y \rangle.$$

Similarly, we can find $c > 0$ such that $\left\| Y^M \right\|_2 \leq c \|Y\|_2$. Hence if $\eta > 0$ is sufficiently small, then it follows from equation (39) and our estimates that for some $c' > 0$,

$$\left| f_\lambda^M(Y) \right| \leq c' e^{-\frac{1}{2}\langle \rho, Y \rangle} (1 + \|Y\|_2)^{\dim \mathfrak{n}}.$$

We therefore obtain

$$\int_{\mathcal{C}^0} \left| f_\lambda^M(Y) e^{\langle \rho, Y \rangle} \right| dY \leq \int_{\mathcal{C}^0} e^{\left\langle \frac{1}{2}\rho, Y \right\rangle} (1 + \|Y\|_2)^{\dim \mathfrak{n}} dY.$$

Using Remark A.1 we see that this last integral is finite. Thus the integral defining $J(\lambda)$, and hence $J(t, \lambda)$, converges absolutely. □

It follows that for $\lambda \in \mathcal{T}_\eta$ we have

$$I(t, \lambda) = \sum_{w \in W^M \backslash W} \frac{\mathbf{c}(w\lambda)}{\mathbf{c}^M(w\lambda)} J(w\lambda) e^{t\left\langle w\lambda, X^0 \right\rangle}, \tag{40}$$

where, we recall, $J(\lambda) = J(0, \lambda)$.

4.4. Main-term replacement

Having established that $I(t, \lambda)$ is well defined, we now show, using Proposition 4.2, that $h_t(\lambda)$ can be approximated by $I(t, \lambda)$.

Proposition 4.6. *There are $c, C > 0$ such that for regular $\lambda \in i\mathfrak{a}^*$,*

$$h_t(\lambda) = C I(t, \lambda) + O_\lambda\left(e^{-ct}\right).$$

Before we prove this proposition, we establish the following lemma (see Lemma A.2 for the definition of the $\beta_i^\vee$):

Lemma 4.7. *Set $0 < \eta < 1$ and define $\mathcal{E} = X^0 - \sum_{i=1}^r [0, \eta] \beta_i^\vee = \sum_{i=1}^r [1 - \eta, 1] \beta_i^\vee$ and $\mathcal{E}^+ = \mathcal{E} \cap \mathcal{P}^+$. Then if η is sufficiently small, we have the following:*

(i) *For every V in the closure $\overline{\mathcal{P}^+ \setminus \mathcal{E}^+}$ of $\mathcal{P}^+ \setminus \mathcal{E}^+$, we have*

$$\langle \rho, V \rangle \leq \left\langle \rho, X^0 \right\rangle - \delta \tag{41}$$

for some $\delta = \delta(\eta) > 0$ uniform in all V.

(ii) *Let $\Delta^M \subseteq \Delta$ denote the subset of simple roots in M. Then there exists $c = c(\eta) > 0$ such that for every $X \in \mathcal{E}$ and $\alpha \in \Delta \setminus \Delta^M$, we have $\alpha(X) \geq c$.*

Proof. Since X^0 is the unique point in $\mathcal{P}^+$ at which the linear map $X \mapsto \langle \rho, X \rangle$ attains its maximum, and $\overline{\mathcal{P}^+ \setminus \mathcal{E}^+}$ is closed, it will suffice to show that every point in $\overline{\mathcal{P}^+ \setminus \mathcal{E}^+}$ is bounded away from X^0. But this is clear, since $\mathcal{P}^+ = \left(X^0 - \sum_{i=1}^r [0, \infty) \beta_i^\vee\right) \cap \mathfrak{a}^+$ by Lemma A.2. Note that this argument holds for any $0 < \eta < 1$, with the resulting constant δ of course depending on η.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

For the second part, we note that for every $\alpha \in \Delta \setminus \Delta^M$, we have $\alpha\left(X^0\right) = 2$. Hence if η is chosen sufficiently small, each such α stays bounded away from 0 on $\mathcal{E}$ so that (ii) follows. □

Proof of Proposition 4.6. Let

$$G(t,\lambda) = e^{-t\langle \rho, X^0\rangle} \int_{\mathcal{P}_t^+} f_\lambda(X) e^{\langle \rho, X\rangle} dX.$$

From equations (34) and (37), as well as the volume computation in Corollary A.4, we have

$$h_t(\lambda) = CG(t,\lambda) + O\left(e^{-c_1 t}\right)$$

for some $c_1, C > 0$. Next, let $\mathcal{E}^+$ be defined as in Lemma 4.7 with η sufficiently small. Put

$$H(t,\lambda) = \int_{t\mathcal{E}^+} f_\lambda(X) e^{\langle \rho, X - tX^0\rangle} dX.$$

Then $G(t,\lambda) = H(t,\lambda) + O\left(e^{-c_2 t}\right)$ for some absolute constant $c_2 > 0$, by Lemma 4.7(i).

Note that by Lemma 4.7, for any $X \in t\mathcal{E}$ we have $\alpha(X) \geq ct$ for $\alpha \in \Delta \setminus \Delta^M$. Using Lemma 4.7(ii) and Proposition 4.2, there exists $c_3 > 0$ such that

$$f_\lambda(X) - \Phi_\lambda(X) \ll_\lambda e^{-c_3 t}$$

for every $X \in t\mathcal{E}^+$ and every regular $\lambda \in i\mathfrak{a}^*$. Hence, for regular $\lambda \in i\mathfrak{a}^*$, we have

$$\begin{aligned} H(t,\lambda) &= \int_{t\mathcal{E}^+} \Phi_\lambda(X) e^{\langle \rho, X - tX^0\rangle} dX + O_\lambda\left(e^{-c_3 t}\right) \\ &= \int_{t\mathcal{E}^+ - tX^0} \Phi_\lambda\left(Y + tX^0\right) e^{\langle \rho, Y\rangle} dY + O_\lambda\left(e^{-c_3 t}\right) \end{aligned}$$

after a change of variables. Noting that $t\mathcal{E}^+ - tX^0 \subseteq \mathcal{C}^0$, we can then extend the integral to all of $\mathcal{C}^0$, incurring an additional error of $O_\lambda\left(e^{-c_4 t}\right)$, for some $c_4 > 0$. □

4.5. A nonvanishing result

We now show that $I(t,\lambda)$ is generically nonvanishing on $i\mathfrak{a}^*$. This will follow from the absolute convergence of $I(t,\lambda)$, the periodic behavior of Φ_λ along X^0 for rational λ, and a linear-independence-of-characters argument.

Proposition 4.8. *There exists a finite number of hyperplanes $\mathcal{H}_1, \ldots, \mathcal{H}_N \subseteq i\mathfrak{a}^*$ and a locally finite collection $\mathcal{D}$ of $(d-2)$-dimensional manifolds in $i\mathfrak{a}^*$ such that for every*

$$\lambda \in i\mathfrak{a}^* \setminus (\mathcal{H}_1 \cup \cdots \cup \mathcal{H}_N \cup \mathcal{D}),$$

there exists $t = t_\lambda \geq 0$ with $I(t,\lambda) \neq 0$.

Proof. Let $\eta > 0$ be sufficiently small and at most as large as in Lemma 4.5. We first claim that there exists a locally finite collection of $(d-2)$-dimensional complex manifolds $\mathcal{Z} \subseteq \mathcal{T}_\eta$ such that $J(t, w\lambda) \neq 0$ for all $\lambda \in \mathcal{T}_\eta \setminus \mathcal{D}$, $w \in W$, and $t \geq 0$. Indeed, for fixed $m \in M$,

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

the map $\lambda \mapsto \varphi_\lambda^M(m)$ is holomorphic in $\lambda \in \mathfrak{a}_\mathbb{C}^*$. Since the integral defining $J(t,\lambda)$ converges absolutely, $J(t,\lambda)$ is in fact a holomorphic function for $\lambda \in \mathcal{T}_\eta$. From the nonvanishing at $\lambda = 0$,

$$J(t,0) = \int_{\mathcal{C}^0} e^{\langle \rho+\rho^M, Y\rangle} \varphi_0^M\left(e^Y\right) dY \neq 0,$$

it follows that $J(t,\lambda)$ does not vanish identically in λ, hence its zero set $\mathcal{Z} = \{\lambda \in \mathcal{T}_\eta \mid J(t,\lambda) = 0\}$ is a locally finite union of $(d-2)$-dimensional complex manifolds. Put $\mathcal{D} = \mathcal{Z} \cap i\mathfrak{a}^*$.

By our choice of η, $\mathbf{c}(w\lambda)/\mathbf{c}^M(w\lambda)$ has no pole for regular λ in $\mathcal{T}_\eta$. Let $\mathcal{T}_\eta'$ denote the set of all regular $\lambda \in \mathcal{T}_\eta \setminus \mathcal{D}$ such that $\langle \lambda, w^{-1}X^0\rangle \neq \langle \lambda, v^{-1}X^0\rangle$ for all $w,v \in W^M \backslash W$, $w \neq v$. Then $\mathcal{T}_\eta'$ is in fact dense in $\mathcal{T}_\eta$, since $w^{-1}X^0 \neq v^{-1}X^0$ for all $w \neq v$, $w,v \in W^M\backslash W$, so that we can obtain $\mathcal{T}_\eta'$ by removing the hyperplanes $\{\lambda \mid \langle \lambda, w^{-1}X^0\rangle = \langle \lambda, v^{-1}X^0\rangle\}$ for each $v \neq w$ in $W^M \backslash W$.

Recall the expansion (40). By the definition of $\mathcal{T}_\eta'$, the phases $\langle w\lambda, X^0\rangle$, $w \in W^M\backslash W$, are pairwise distinct for $\lambda \in \mathcal{T}_\eta'$. We can therefore apply [16, Lemma 56] (see also [19, Ch. VIII, Lemma 0.1]) to the sum over $w \in W^M \backslash W$ to obtain, for each $\lambda \in \mathcal{T}_\eta'$,

$$\limsup_{t\to\infty} |I(t,\lambda)|^2 \geq \sum_{w \in W^M\backslash W} \left| \frac{\mathbf{c}(w\lambda)}{\mathbf{c}^M(w\lambda)} J(w\lambda)\right|^2.$$

Now for each $w \in W$ and $\lambda \in \mathcal{T}_\eta'$, the terms $\frac{\mathbf{c}(w\lambda)}{\mathbf{c}^M(w\lambda)} J(w\lambda)$ are nonzero, and hence for each $\lambda \in \mathcal{T}_\eta'$ there exists $t = t_\lambda$ such that $I(t,\lambda) \neq 0$. □

4.6. Final arguments

We now finish the proof of Proposition 4.1.

By enlarging Ω if necessary, we can assume that $\overline{\Omega^\circ} = \Omega$. Let $\mathcal{H}_1, \ldots, \mathcal{H}_N$ and $\mathcal{D}$ be as in Proposition 4.8. If necessary, we add in more hyperplanes such that the complement of their union in $i\mathfrak{a}^*$ consists of regular points only. Then the set of rational λ in

$$\Omega \cap (i\mathfrak{a}^* \setminus (\mathcal{H}_1 \cup \cdots \cup \mathcal{H}_N \cup \mathcal{D}))$$

is dense in Ω. Since $\lambda \mapsto \frac{1}{T}\int_0^T |h_t(\lambda)|^2$ is a continuous function, and Ω is compact, Proposition 4.1 follows from the next result:

Lemma 4.9. *For every rational*

$$\lambda \in i\mathfrak{a}^* \setminus (\mathcal{H}_1 \cup \cdots \cup \mathcal{H}_N \cup \mathcal{D}),$$

there exists $T_0 > 0$ depending on λ such that for all $T > T_0$, we have

$$\frac{1}{T}\int_0^T |h_t(\lambda)|^2 dt > 0.$$

Proof. Recall the definition of 'rational' from Definition 4.3. Let $\tau_1 \in \mathbb{R}_{>0}$ (depending on λ) and $\tau_n = n\tau_1$ for $n \in \mathbb{N}$ be as in the statement of Lemma 4.4. By Proposition 4.8 we can choose $t_0 \geq 0$ such that $I(t_0,\lambda) =: a \neq 0$. Write $t_n = t_0 + \tau_n$ for $n \in \mathbb{N}$. Note that by

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Lemma 4.4, we have $I(t+\tau_n,\lambda) = I(t,\lambda)$ for all $t \geq 0$ and all $n \in \mathbb{N}$. Thus $I(t_n,\lambda) = a$, and we can find a small $\epsilon > 0$ such that for every $n \in \mathbb{N}$ and every $s \in (t_n - \epsilon, t_n + \epsilon)$, we have

$$|I(s,\lambda)| \geq |a|/2 > 0.$$

From Lemma 4.6 we may choose $T_0 > 0$ so large that $|h_t(\lambda) - I(t,\lambda)| < |a|/4$ for all $t \geq T_0$. Let $N \in \mathbb{N}$ be such that $t_n - \epsilon \geq T_0$ for all $n \geq N$. Then for every $n \geq N$, we have

$$|h_s(\lambda)| \geq |a|/4$$

for all $s \in (t_n - \epsilon, t_n + \epsilon)$. We therefore get, for $T > 2T_0 + \tau_1$, that

$$\frac{1}{T}\int_0^T |h_t(\lambda)|^2 dt \gg \frac{1}{T}\sum_{n=N}^{(T-1-t_0)/\tau_1} \int_{-\epsilon}^{\epsilon} |h_{t_n+s}(\lambda)|^2 ds \gg_\lambda \frac{T-T_0}{T}|a|^2 > 0,$$

as we wanted to show. □

5. Geometric side

The aim of this section is to prove Theorem 2.3. The essential feature of the geometric side is the ergodic properties of the sets $gE_t \cap E_t$ in the quotient $\Gamma\backslash G$, as $t > 0$ and $g \in G$ vary.

5.1. A general bound

In this section, we return to the general setting of §3, with $S = G/K$ and Γ a given torsion-free cocompact lattice in G.

For $g,h \in G$, let $N_R(g,h)$ denote the number of $\gamma \in \Gamma$ for which $d(g,\gamma h) \leq R$. This generalizes the notation $N_R(g)$ from equation (26) when $g = h$. Let $N_\Gamma(R) = \sup_{(g,h)} N_R(g,h)$.

Lemma 5.1. *There exists $c > 0$, independent of R and Γ, such that*

$$N_\Gamma(R) \ll \mathrm{InjRad}(\Gamma\backslash S)^{-\dim S} e^{cR}.$$

Proof. Note that $N_R(g,h) = N_R(g,\gamma h)$ for all $\gamma \in \Gamma$, so that we can assume that h is such that $d(g,h) = \min_{\gamma\in\Gamma} d(g,\gamma h)$. If $d(g,h) > R$, then $N_R(g,h) = 0$, so that we can assume $d(g,h) \leq R$.

Now if $\gamma \in \Gamma$ is such that $d(g,\gamma h) \leq R$, then

$$d(h,\gamma h) \leq d(g,h) + d(g,\gamma h) \leq 2R,$$

so that $N_R(g,h) \leq N_{2R}(h,h)$. Applying [1, Lemma 6.18] to $N_{2R}(h,h)$ gives the assertion of the lemma. □

The following lemma is an adaptation of [24, Lemma 5.1] to our setting. Recall the geodesic balls B_t, first introduced in equation (13) in the setting of $G = \mathrm{SL}_d(\mathbb{R})$, but more generally defined for all G using the same notation.

Lemma 5.2. *There is $c > 0$, depending only on G, such that the following holds. Let $R > 0$, and let $K \in C(G \times G)$ be invariant under the diagonal action of Γ and satisfy*

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

$\{g^{-1}h : (g,h) \in \operatorname{supp}(K)\} \subset B_R$. *Let* $\mathbf{K}$ *be an integral operator on* $L^2(\Gamma\backslash G)$ *with kernel* K. *Then*

$$\|\boldsymbol{K}\|_{HS}^2 \leq \int_{\Gamma\backslash G}\int_G |K(g,h)|^2 dgdh + O\left(\frac{e^{cR}}{\operatorname{InjRad}(\Gamma\backslash S)^{\dim S}} \operatorname{vol}((\Gamma\backslash G)_{\leq 2R})\,\|K\|_\infty^2\right).$$

Proof. By definition, we have

$$\|\mathbf{K}\|_{\mathrm{HS}}^2 = \int_{\Gamma\backslash G}\int_{\Gamma\backslash G}\left|\sum_{\gamma\in\Gamma} K(g,\gamma h)\right|^2 dgdh.$$

Let $D \subset G$ be a fundamental domain for Γ. We decompose D into its $2R$-thin and -thick parts $D = D_{\leq 2R} \cup D_{>2R}$, where, similar to equation (3), we have put

$$D_{\leq 2R} = \{x \in D : \operatorname{InjRad}_\Gamma(x) \leq 2R\}.$$

From the support condition on K, it follows that if $K(g,\gamma h) \neq 0$ then $d(g,\gamma h) \leq R$.

Now suppose that $h \in D_{>2R}$ and $\gamma,\delta \in \Gamma$, $\gamma \neq \delta$, are such that $d(g,\gamma h) \leq R$ and $d(g,\delta h) \leq R$. Then $d(\gamma h,\delta h) \leq d(g,\gamma h) + d(g,\delta h) \leq 2R$, but $d(\gamma h,\delta h) > 2R$, since $h \in D_{>2R}$, leading to a contradiction. Hence there exists at most one $\gamma \in \Gamma$ with $K(g,\gamma h) \neq 0$. In this way,

$$\left|\sum_{\gamma\in\Gamma} K(g,\gamma h)\right|^2 = \sum_{\gamma\in\Gamma} |K(g,\gamma h)|^2,$$

so that

$$\begin{aligned}
\int_{D_{>2R}}\int_D \left|\sum_{\gamma\in\Gamma} K(g,\gamma h)\right|^2 dgdh &= \int_{D_{>2R}}\int_D \sum_{\gamma\in\Gamma} |K(g,\gamma h)|^2 dgdh \\
&\leq \int_D\int_D \sum_{\gamma\in\Gamma} |K(g,\gamma h)|^2 dgdh \\
&= \int_G\int_D |K(g,h)|^2 dgdh \\
&= \int_{\Gamma\backslash G}\int_G |K(g,h)|^2 dhdg.
\end{aligned}$$

To deal with the $2R$-thin part, we use Cauchy–Schwarz and the support condition on K to obtain

$$\left|\sum_{\gamma\in\Gamma} K(g,\gamma h)\right|^2 \leq N_\Gamma(R) \sum_{\gamma\in\Gamma} |K(g,\gamma h)|^2.$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

The factor $N_\Gamma(R)$ is bounded by Lemma 5.1. Furthermore,

$$\begin{aligned}&\int_{(\Gamma\backslash G)_{\le 2R}}\int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma}|K(g,\gamma h)|^2 dgdh\\ &\le \sup_{g\in\Gamma\backslash G, h\in G}|K(g,h)|^2\int_{(\Gamma\backslash G)_{\le 2R}}\int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma}\mathbf{1}_{B_R}\left(g^{-1}\gamma h\right)dgdh.\end{aligned}$$

By unfolding and changing variables, we have

$$\int_{\Gamma\backslash G}\sum_{\gamma\in\Gamma}\mathbf{1}_{B_R}\left(g^{-1}\gamma h\right)dg=\int_G\mathbf{1}_{B_R}\left(g^{-1}h\right)dg=m_G(B_R)=c_C\int_{\mathcal{B}_R^+}J(X)dX,$$

the last equality coming from the Cartan measure decomposition (21). The Jacobian factor satisfies $J(X)\ll e^{2\langle\rho,X\rangle}\ll e^{2R\|\rho\|_2}$ for $X\in\mathcal{B}^+$, using equation (37). From this one deduces[2] that $m_G(B_R)\ll R^r e^{2R\|\rho\|_2}$, which is enough to complete the proof. □

5.2. Description of the kernel of $\mathbf{A}(\tau)$

We return to the setting of $G=\mathrm{SL}_d(\mathbb{R})$. We would like to apply Lemma 5.2 to our operator $\mathbf{A}(\tau)$ on $L^2(\Gamma\backslash G)$ from equation (10). For this we shall need a description of its kernel, which we shall denote by $A(\tau)$; it is a continuous function on $\Gamma\backslash(G\times G)$.

Lemma 5.3. *The kernel of* $\mathbf{A}(\tau)$ *is*

$$A(\tau)(g,h)=\frac{1}{\tau}\int_0^\tau\frac{1}{m_G(E_t)}\int_{gE_t\cap hE_t}a(x)dxdt. \tag{42}$$

Proof. By definition, the action of U_taU_t on $f\in L^2(\Gamma\backslash G)$ is given by

$$(U_taU_t)f(g)=\frac{1}{m_G(E_t)}\int_{E_t}a(gh_1)\int_{E_t}f(gh_1h_2)dh_2dh_1.$$

We set $x=gh_1$ and $h=xh_2$, apply Fubini, and use $E_t^{-1}=E_t$ to obtain

$$\begin{aligned}(U_taU_t)f(g)&=\frac{1}{m_G(E_t)}\int_{gE_t}a(x)\int_{xE_t}f(h)dhdx\\ &=\int_G\left(\frac{1}{m_G(E_t)}\int_{gE_t\cap hE_t}a(x)dx\right)f(h)dh.\end{aligned}$$

Averaging over t yields the claim. □

5.3. Support and sup of $\mathbf{A}(\tau)$

In order to bound the second term in Lemma 5.2, we shall need to estimate the support and the supremum of the kernel function $A(\tau)$. This is accomplished in the next result, which is the higher-rank generalization of [2, Lemma 26]. We recall the norm $|\cdot|$ on $G=\mathrm{SL}_d(\mathbb{R})$ introduced in §2.2 and defining the sets E_t.

[2] One could also quote the more precise result of Knieper, recalled in formula (16), but an exponential bound of any quality is all that is required in our application.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Proposition 5.4. *There exists a constant $b>0$ depending only on d such that $gE_t \cap hE_t$ is empty unless $\left|h^{-1}g\right| \le 2t+b$. In particular, $A(\tau)(g,h)=0$ unless $\left|h^{-1}g\right| \le 2\tau+b$. Further,*

$$\sup_{g,h\in G} |A(\tau)(g,h)|^2 \le \|a\|_\infty^2.$$

Proof. We begin by establishing the following:

Claim: For $x \in \mathrm{SL}_d(\mathbb{R})$, we have $|x_{ij}| \le e^{|x|}$. Moreover, there exists a constant $c>0$ depending only on d, such that for every $x=(x_{ij}) \in \mathrm{SL}_d(\mathbb{R})$ and every $i \in \{1,\dots,d\}$ there exists $j \in \{1,\dots,d\}$ with $|x_{ij}| \ge ce^{-|x|}$.

Proof of claim: Write $x = ke^Z l$, with $k=(k_{ij})$, $l=(l_{ij}) \in K = \mathrm{SO}(d)$, and $Z \in \overline{\mathfrak{a}^+}$. Let $\kappa_i = (k_{i1},\dots,k_{id})$ and $\lambda_j = (l_{1j},\dots,l_{dj}) \in \mathbb{R}^d$, for $i,j=1,\dots,d$. Let $\langle\cdot,\cdot\rangle$ denote the standard inner product on $\mathbb{R}^d$. Then $\kappa_1,\dots,\kappa_d$ and $\lambda_1,\dots,\lambda_d$ are both orthonormal bases of $\mathbb{R}^d$, and $x_{ij} = \langle \kappa_i, \lambda_j e^Z\rangle$. In particular, by Cauchy and Schwarz we have

$$|x_{ij}| = \left|\left\langle \kappa_i, \lambda_j e^Z\right\rangle\right| \le \left\|\lambda_j e^Z\right\|^{1/2} \le e^{\|Z\|_\infty} = e^{|x|},$$

proving the first statement.

For the second statement, we begin by observing that for every $j=1,\dots,d$ we have

$$\left\langle \kappa_i, \kappa_i e^Z\right\rangle = \sum_{j=1}^d e^{Z_j} k_{ij}^2 \ge \left(\min_j e^{Z_j}\right) \|\kappa_i\|^2 = \min_j e^{Z_j} \ge e^{-\|Z\|_\infty} = e^{-|x|}.$$

Writing κ_i as a linear combination $\kappa_i = a_1\lambda_1 + \cdots + a_d\lambda_d$, we obtain

$$e^{-|x|} \le \left\langle \kappa_i, \kappa_i e^Z\right\rangle \le \sum_{j=1}^d |a_j| \left|\left\langle \kappa_i, \lambda_j e^Z\right\rangle\right|.$$

Note that the a_i are uniformly bounded, since k and l are contained in a compact set. Hence there exist $c>0$ and $j \in \{1,\dots,d\}$ such that

$$|x_{ij}| = \left|\left\langle \kappa_i, \lambda_j e^Z\right\rangle\right| \ge ce^{-|x|},$$

as claimed.

Continuing with the proof of the proposition, suppose that there exists $y \in h^{-1}gE_t \cap E_t$. We can assume without loss that $h^{-1}g = e^Y$ for some suitable $Y = (Y_1,\dots,Y_n) \in \overline{\mathfrak{a}^+}$.

Since $y \in E_t$, the first part of the claim shows that $|y_{ij}| \le e^t$ for all $i,j=1,\dots,d$. Moreover, since $y \in e^Y E_t$, we can write $y = e^Y x$ for some $x \in E_t$. Then $y_{ij} = e^{Y_i} x_{ij}$, and the second part of the claim shows that for every i there is a j_i such that $|x_{ij_i}| \ge ce^{-t}$. Putting this together, we obtain $e^t \ge ce^{Y_i}e^{-t}$, so that $e^{Y_i} \le c^{-1}e^{2t}$.

By similar arguments, we also know that $|x_{ij}| \le e^t$ for all i,j, and that for every i there exists j_i' with $e^{Y_i}\left|x_{ij_i'}\right| = \left|y_{ij_i'}\right| \ge ce^{-t}$ – that is, $e_i^Y \ge ce^{-2t}$.

These two inequalities together imply the first assertion of the corollary, and hence also the assertion of the support of $A(\tau)$. The last assertion, on the sup of $A(\tau)$, is a direct consequence of the definition (42). □

Recall from §2.5(iii) that $E_t \subset B_{t\|X^0\|_2}$. It then follows from Proposition 5.4 that we can take $R = \|X^0\|(2\tau + b)$ in Lemma 5.2. Inserting this and the sup norm bound of Proposition 5.4 into Lemma 5.2, we obtain the second term in Theorem 2.3.

5.4. First term

To establish Theorem 2.3, it remains to estimate the first term in Lemma 5.2, for the operator $\mathbf{A}(\tau)$ with kernel $A(\tau)$.

For a measurable set $E \subseteq G$, we write $\rho_{\Gamma\backslash G}(E)$ for the action by convolution of the normalized characteristic function of E on $L^2(\Gamma\backslash G)$. Thus

$$\rho_{\Gamma\backslash G}(E)f(x) = \frac{1}{m_G(E)} \int_E f(xg)dg.$$

We shall in particular be interested in this operator for sets of the form $gE_t \cap E_t$.

We begin with the following elementary upper bound on the quantity

$$\mathcal{A}(\tau) = \int_{\Gamma\backslash G} \int_G |A(\tau)(g,h)|^2 dg dh.$$

Lemma 5.5. *For all $\tau > 0$, we have*

$$\mathcal{A}(\tau) \leq \frac{1}{\tau^2} \int_{E_{2\tau+b}} \left(\int_{\max\{0,(|g|-b)/2\}}^{\tau} \frac{m_G(gE_t \cap E_t)}{m_G(E_t)} \left\| \rho_{\Gamma\backslash G}(gE_t \cap E_t)a \right\|_{L^2(\Gamma\backslash G)} dt \right)^2 dg,$$

where $b > 0$ is as in Proposition 5.4.

Proof. From the definition (42) of $A(\tau)$, we see that

$$\mathcal{A}(\tau) = \frac{1}{\tau^2} \int_{\Gamma\backslash G} \int_G \left| \int_0^\tau \frac{1}{m_G(E_t)} \int_{gE_t \cap hE_t} a(x) dx dt \right|^2 dg dh.$$

Changing variables $x \mapsto hx$ and $g \mapsto h^{-1}g$, this is

$$\frac{1}{\tau^2} \int_{\Gamma\backslash G} \int_G \left| \int_0^\tau \frac{1}{m_G(E_t)} \int_{gE_t \cap E_t} a(hx) dx dt \right|^2 dg dh.$$

In view of Proposition 5.4, the integral over $g \in G$ may be truncated at $|g| \leq 2\tau + b$, and the lower range of the t integral may be truncated at $\max\{0,|g| - b)/2\}$. Furthermore,

$$\begin{aligned} \frac{1}{m_G(E_t)} \int_{gE_t \cap E_t} a(hx) dx &= \frac{m_G(gE_t \cap E_t)}{m_G(E_t)} \frac{1}{m_G(gE_t \cap E_t)} \int_{gE_t \cap E_t} a(hx) dx \\ &= \frac{m_G(gE_t \cap E_t)}{m_G(E_t)} \left(\rho_{\Gamma\backslash G}(gE_t \cap E_t)a \right)(h). \end{aligned}$$

We conclude the proof by an application of the Minkowski integral inequality. □

We briefly return to the level of generality of §3, and let G denote a connected noncompact simple Lie group with finite center. Let (π, V_π) be a unitary representation of G. We recall that the *integrability exponent* $q(\pi)$ of π is defined as

$$q(\pi) = \inf\{q > 0 : \langle \pi(g)v_1, v_2 \rangle \in L^q(G) \text{ for } v_1, v_2 \text{ in a dense subspace of } V_\pi\}.$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

A classical result of Borel and Wallach [6], Cowling [11], and Howe and Moore [18] states that π has a spectral gap precisely when $q(\pi) < \infty$; we can effectively take this as our *definition* of the spectral gap in our setting.

Nevo [29] (see also [15, Theorem 4.1]) has proved a mean ergodic theorem for measure-preserving actions G on a probability space (X,μ) whose associated unitary representation $L^2_0(X,\mu)$ has a spectral gap. We shall be interested in the action of G on $X = \Gamma\backslash G$ by right translation, where $\Gamma < G$ is a lattice. The associated unitary representation is then $\rho^0_{\Gamma\backslash G}$, the restriction of the right-regular representation $\rho_{\Gamma\backslash G}$ to the subspace $L^2_0(\Gamma\backslash G)$ of $L^2(\Gamma\backslash G)$ consisting of functions f with $\int_{\Gamma\backslash G} f = 0$. In this setting, they show the following result:

Proposition 5.6 (Nevo). *Let G be a connected noncompact simple Lie group with finite center. Let $\Gamma < G$ be a lattice. Then there are constants $\theta, C > 0$, depending on the integrability exponents of $\rho^0_{\Gamma\backslash G}$, such that for any measurable $E \subseteq G$ of finite measure, we have*

$$\left\|\rho^0_{\Gamma\backslash G}(E)\right\|_2 \leq C m_G(E)^{-\theta}.$$

A famous result of Kazhdan [20] states that when G furthermore has rank at least 2,

$$\sup_{\pi:\pi^G=0} q(\pi) < \infty,$$

the supremum running over all unitary representations of G having no G-invariant vectors. In this case, the constants of Proposition 5.6 can be taken independently of Γ. In particular, this is true of $G = \mathrm{SL}_d(\mathbb{R})$ for $d \geq 3$.

We thus return to $G = \mathrm{SL}_d(\mathbb{R})$, for $d \geq 3$, and again require $\Gamma < \mathrm{SL}_d(\mathbb{R})$ to be a uniform lattice. Recall that $a \in L^2_0(\Gamma\backslash G)$. We can therefore apply Proposition 5.6 to estimate the quantity

$$\left\|\rho_{\Gamma\backslash G}(gE_t \cap E_t)a\right\|_{L^2(\Gamma\backslash G)} = \left\|\rho^0_{\Gamma\backslash G}(gE_t \cap E_t)a\right\|_{L^2_0(\Gamma\backslash G)}$$

appearing in Lemma 5.5. Hence there is $\theta > 0$, depending only on d, such that

$$\mathcal{A}(\tau) \ll \frac{\|a\|^2_{L^2(\Gamma\backslash G)}}{\tau^2}\int_{E_{2\tau+b}}\left(\int_{\max\{0,(|g|-b)/2\}}^{\tau}\frac{m_G(gE_t\cap E_t)^{1-\theta}}{m_G(E_t)}dt\right)^2 dg.$$

To conclude the proof of Theorem 2.3, it suffices to show the following estimate:

Proposition 5.7. *Let $g \in G$. Then for $\tau \gg 1$ we have*

$$\int_{E_{2\tau+b}}\left(\int_{\max\{0,(|g|-b)/2\}}^{\tau}\frac{m_G(gE_t\cap E_t)^{1-\theta}}{m_G(E_t)}dt\right)^2 dg \ll \tau,$$

the implied constant depending on d.

The main point in the proof of this result is an estimate on the intersection volumes $m_G(gE_t \cap E_t)$, proved in the next subsection.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

5.5. Intersection volumes

We seek to prove the following estimate on the intersection volumes $m_G\left(e^Y E_t \cap E_t\right)$. See §2.5 for a discussion of some geometric aspects.

Proposition 5.8. *Set $Y \in \mathfrak{a}^+$ and $t > 0$. Then*

$$m_G\left(e^Y E_t \cap E_t\right) \ll e^{-\langle\rho,Y\rangle} m_G(E_t). \tag{43}$$

Proof. Recall the definition of the Frobenius norm $\|\cdot\|$ from §2.5(iv) and the Frobenius ball $\boldsymbol{E}_t$ from equation (19). We claim that

$$E_t \subseteq \boldsymbol{E}_{t+\log\sqrt{d}} \subseteq E_{t+\log\sqrt{d}}. \tag{44}$$

Note that $\|\cdot\|$ is bi-K-invariant, so that if $g = k_1 e^X k_2$, where $X = (X_1,\dots,X_d) \in \mathfrak{a}$, then $\|g\|^2 = e^{2X_1} + \cdots + e^{2X_d}$. Thus we may write $\boldsymbol{E}_t = K\exp(\boldsymbol{P}_t)K$, where

$$\boldsymbol{P}_t = \left\{X = (X_1,\dots,X_d) \in \mathfrak{a} : \max\left\{e^{2X_1} + \cdots + e^{2X_d}, e^{-2X_1} + \cdots + e^{-2X_d}\right\} \le e^{2t}\right\}.$$

We therefore wish to establish formula (44) for the sets $\boldsymbol{P}_t$ and $\mathcal{P}_t = t\mathcal{P}$, where $\mathcal{P}$ is defined in equation (14):

- ⊳ If $|X_i| \le t$ for all i, we have $\max\left\{e^{2X_1} + \cdots + e^{2X_d}, e^{-2X_1} + \cdots + e^{-2X_d}\right\} \le de^{2t}$, so that $\mathcal{P}_t \subseteq \boldsymbol{P}_{t+\log\sqrt{d}}$.
- ⊲ In the other direction, if $\max\left\{e^{2X_1} + \cdots + e^{2X_d}, e^{-2X_1} + \cdots + e^{-2X_d}\right\} \le e^{2t}$, then $|X_i| \le t$ for all i, so that $\boldsymbol{P}_t \subseteq \mathcal{P}_t$.

It is now enough to prove formula (43) with E_t replaced by $\boldsymbol{E}_t$. Indeed, suppose that

$$m_G\left(e^Y \boldsymbol{E}_t \cap \boldsymbol{E}_t\right) \ll e^{-\langle\rho,Y\rangle} m_G(\boldsymbol{E}_t) \tag{45}$$

for any $Y \in \mathfrak{a}$. We use formula (44) to bound $m_G(e^Y E_t \cap E_t)$ by $m_G(e^Y \boldsymbol{E}_{t+\log\sqrt{d}} \cap \boldsymbol{E}_{t+\log\sqrt{d}})$. Inserting the estimate (45) and the inclusions (44) again, we find

$$m_G\left(e^Y \boldsymbol{E}_{t+\log\sqrt{d}} \cap \boldsymbol{E}_{t+\log\sqrt{d}}\right) \ll e^{-\langle\rho,Y\rangle} m_G\left(\boldsymbol{E}_{t+\log\sqrt{d}}\right) \le e^{-\langle\rho,Y\rangle} m_G\left(E_{t+\log\sqrt{d}}\right).$$

Since $m_G\left(E_{t+\log\sqrt{d}}\right) \asymp m_G(E_t)$, by Corollary A.4, we have reduced formula (43) to formula (45).

We prove formula (45) using the Iwasawa decomposition $G = ANK$ from §3.1. Let $g = auk$ so that $\|g\| = \|au\|$ and $\left\|g^{-1}\right\| = \left\|u^{-1}a^{-1}\right\|$. Now $\|au\|^2$ just equals the sum of all the squares of the matrix entries of au, so that $\|au\| = \|g\| \le e^t$ immediately implies that $\|a\| \le e^t$ and $\|au - a\| \le e^t$. Similarly, $\left\|g^{-1}\right\| \le e^t$ implies $\left\|a^{-1}\right\| \le e^t$ and $\left\|u^{-1}a^{-1} - a^{-1}\right\| \le e^t$. Thus we obtain the inclusion

$$\boldsymbol{E}_t \subseteq \left\{g = auk \in G : \|a\|, \left\|a^{-1}\right\|, \|au - a\|, \left\|u^{-1}a^{-1} - a^{-1}\right\| \le e^t\right\}.$$

Applying the same argument to $e^Y g$, for $Y \in \mathfrak{a}$, and dropping the last two conditions,

$$e^Y \boldsymbol{E}_t \subseteq \left\{g = auk \in G : \left\|e^{-Y}a\right\|, \left\|e^Y a^{-1}\right\| \le e^t\right\}.$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

We conclude that $m_G\left(e^Y\boldsymbol{E}_t\cap\boldsymbol{E}_t\right)\le I(Y,t)$, where

$$I(Y,t)=\int_{\substack{\left\|e^{\pm A}\right\|\le e^t\\ \left\|e^{\pm(A-Y)}\right\|\le e^t}}\int_{\substack{\left\|e^Au-e^A\right\|\le e^t\\ \left\|u^{-1}e^{-A}-e^{-A}\right\|\le e^t}} du\,dA.$$

Changing variables $u\mapsto e^Au-e^A+I$ in the U-integration, we obtain

$$I(Y,t)=I_U(t)\int_{\substack{\left\|e^{\pm A}\right\|\le e^t\\ \left\|e^{\pm(A-Y)}\right\|\le e^t}} e^{-\langle\rho,A\rangle}dA,$$

where $I_U(t)=m_U\left(u\in U:\|u-I\|,\left\|u^{-1}-I\right\|\le e^t\right)$. Changing variables $A\mapsto A-Y$ in the A-integration and dropping a condition, we obtain

$$\begin{aligned}\int_{\substack{\left\|e^{\pm A}\right\|\le e^t\\ \left\|e^{\pm(A-Y)}\right\|\le e^t}} e^{-\langle\rho,A\rangle}dA&=e^{-\langle\rho,Y\rangle}\int_{\substack{\left\|e^{\pm(A+Y)}\right\|\le e^t\\ \left\|e^{\pm A}\right\|\le e^t}} e^{-\langle\rho,A\rangle}dA\\ &\le e^{-\langle\rho,Y\rangle}\int_{\left\|e^{\pm A}\right\|\le e^t} e^{-\langle\rho,A\rangle}dA.\end{aligned}$$

We write this last integral as $I_A(t)$. Using equation (20), this establishes

$$m_G\left(e^Y\boldsymbol{E}_t\cap\boldsymbol{E}_t\right)\le c_Ie^{-\langle\rho,Y\rangle}I_A(t)I_U(t). \tag{46}$$

We now go in reverse. For $A\in\mathfrak{a}$, we change variables $u\mapsto e^{-A}u-e^{-A}+I$ to obtain

$$e^{-\langle\rho,A\rangle}I_U(t)=m_U\left(u\in U:\left\|e^Au-e^A\right\|,\left\|u^{-1}e^{-A}-e^{-A}\right\|\le e^t\right),$$

so that

$$I_A(t)I_U(t)=\int_{\left\|e^{\pm A}\right\|\le e^t}\int_{\left\|(e^Au)^{\pm1}-e^{\pm A}\right\|\le e^t} du\,dA.$$

Again by the triangle inequality and equation (20), this gives

$$I_A(t)I_U(t)\le\iint_{\|(au)^{\pm1}\|\le 2e^t} da\,du=c_I^{-1}m_G\left(\boldsymbol{E}_{t+\log 2}\right)\asymp m_G(\boldsymbol{E}_t).$$

Combining this last inequality with formula (46) yields formula (45). □

5.6. Proof of Proposition 5.7

From Proposition 5.8 and Lemma A.4 we have, for $Y\in\mathfrak{a}^+$,

$$\frac{m_G\left(e^YE_t\cap E_t\right)^{1-\theta}}{m_G(E_t)}\ll e^{-(1-\theta)\langle\rho,Y\rangle}m_G(E_t)^{-\theta}\ll e^{-(1-\theta)\langle\rho,Y\rangle}e^{-t\theta\left\langle 2\rho,X^0\right\rangle},$$

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

so that, using equation (15),

$$\int_{E_{2\tau+b}} \left(\int_{\max\{0,(|g|-b)/2\}}^{\tau} \frac{m_G(gE_t \cap E_t)^{1-\theta}}{m_G(E_t)} dt \right)^2 dg$$
$$\ll \int_{\mathcal{P}_{2\tau+b}^+} \left(\int_{\max\{0,(\|Y\|_\infty - b)/2\}}^{\tau} e^{-t\theta \langle 2\rho, X^0 \rangle} dt \right)^2 e^{-(1-\theta)\langle 2\rho, Y \rangle} J(Y) dY.$$

Inserting $J(X) \ll e^{2\langle \rho, X \rangle}$ for $X \in \mathfrak{a}^+$ from equation (37) and evaluating the t-integral, we majorize the foregoing expression by

$$\int_{\mathcal{P}_{2\tau+b}^+} e^{-\max\{0,\|Y\|_\infty - b\}\theta \langle 2\rho, X^0 \rangle} e^{\theta \langle 2\rho, Y \rangle} dY.$$

We break up the last integral as $I_1 + I_2$, according to $\mathfrak{a}_1^+ = \mathcal{P}_b^+$ and

$$\mathfrak{a}_2^+(\tau) = \left\{ Y \in \mathfrak{a}^+ : b \le \|Y\|_\infty \le 2\tau + b \right\}.$$

Since $\mathfrak{a}_1^+$ is independent of τ (as is the integrand), we have $I_1 \ll 1$. Next, we have

$$I_2 \ll \int_{\mathfrak{a}_2^+(\tau)} e^{2\theta\left(\langle \rho, Y \rangle - \|Y\|_\infty \langle \rho, X^0 \rangle \right)} dY.$$

We will drop several of the conditions on Y in the course of the proof to simplify the integral.

Write $\tau' = 2\tau + b$. We distinguish the cases of d being even and odd. Define

$$s = \begin{cases} d/2 & \text{if } d \text{ is even,} \\ (d+1)/2 & \text{if } d \text{ is odd.} \end{cases} \tag{47}$$

We first assume that $d \ge 3$ is odd. We can write Y in the integral as

$$Y = (a_1, \dots, a_s, -a_1 - \dots - a_s + b_1 + \dots + b_s, -b_s, \dots, -b_1),$$

with $\tau' \ge a_1 \ge \dots \ge a_s \ge 0$ and $\tau' \ge b_1 \ge \dots \ge b_s \ge 0$. Then $\langle \rho, Y \rangle = \sum_{i=1}^{s} (s-i)(a_i + b_i)$, and using Lemma A.1, $\langle \rho, X^0 \rangle = 2A$ with $A = \sum_{i=1}^{s} (s-i)$. Moreover, $\|Y\|_\infty = \max\{a_1, b_1\}$. Hence

$$I_2(\tau) \ll \int \cdots \int e^{2\theta\left(\sum_{i=1}^{s} (s-i)(a_i+b_i) - 2A \max\{a_1, b_1\} \right)},$$

where the integral runs over

$$\left\{ a_1, \dots, a_s, b_1, \dots, b_s \middle| \begin{matrix} \tau' \ge a_1 \ge 0,\ a_1 \ge a_2 \ge 0,\ \dots,\ a_{s-1} \ge a_s \ge 0, \\ \tau' \ge b_1 \ge 0,\ b_1 \ge b_2 \ge 0,\ \dots,\ b_{s-1} \ge b_s \ge 0. \end{matrix} \right\}$$

We now integrate first a_s and b_s, and continue successively with $a_{s-1},\dots,a_2$ and $b_{s-1},\dots,b_2$, until only a_1 and b_1 remain to be integrated. In this way we obtain

$$I_2(\tau) \ll \iint_0^{\tau'} e^{2\theta A(a_1+b_1-2\max\{a_1,b_1\})} da_1 db_1 = \iint_0^{\tau'} e^{-2\theta A|a_1-b_1|} da_1 db_1$$
$$= 2\tau' \int_0^{\tau'} e^{-2\theta A x}\, dx \ll \tau.$$

If d is even, then $d \geq 4$ and $s \geq 2$. We can write $Y = (a_1,\dots,a_s,-b_s,\dots,-b_1)$, with $\tau' \geq a_1 \geq \cdots \geq a_s \geq 0$, $\tau' \geq b_1 \geq \cdots \geq b_s \geq 0$, and

$$a_1 + \cdots + a_s = b_1 + \cdots + b_s. \tag{48}$$

Then $\|Y\|_\infty = \max\{a_1,b_1\}$ and

$$\langle \rho, Y\rangle = \sum_{i=1}^{s}(s-i+1/2)(a_i+b_i) = \sum_{i=1}^{s}(s-i+1)a_i + \sum_{i=1}^{s-1}(s-i)b_i,$$

where we have used equation (48) to replace b_s. Furthermore, using Lemma A.1, we have $\left\langle \rho, X^0\right\rangle = A+B$, where we have put

$$A = \sum_{i=1}^{s}(s-i+1) \quad \text{and} \quad B = \sum_{i=1}^{s}(s-i).$$

Extending the domain of integration if necessary, we obtain

$$I_2(\tau) \leq \int\cdots\int e^{2\theta\left(\sum_{i=1}^{s}(s-i+1)a_i + \sum_{i=1}^{s-1}(s-i)b_i - (A+B)\max\{a_1,b_1\}\right)},$$

where the integral runs over

$$\left\{ a_1,\dots,a_s,b_1,\dots,b_{s-1} \,\middle|\, \begin{array}{l} \tau' \geq a_1 \geq 0,\ a_1 \geq a_2 \geq 0,\ \dots,\ a_{s-1} \geq a_s \geq 0, \\ \tau' \geq b_1 \geq 0,\ b_1 \geq b_2 \geq 0,\ \dots,\ b_{s-2} \geq b_{s-1} \geq 0 \end{array} \right\}.$$

As before, we successively integrate over $a_s,a_{s-1},\dots,a_2$ and $b_{s-1},b_{s-2},\dots,b_2$, obtaining

$$I_2(\tau) \ll \iint_0^{\tau'} e^{2\theta(Aa_1+Bb_1-(A+B)\max\{a_1,b_1\})} da_1 db_1 \leq \iint_0^{\tau'} e^{-2\theta B|a_1-b_1|} da_1 db_1 \ll \tau,$$

as desired. □

Appendix A. Cones and volumes

Recall the definition of $\mathcal{P}^+$ and E_t in §2.2. Our goal is to write $\mathcal{P}^+$ as the intersection of $\overline{\mathfrak{a}^+}$ with a suitable cone in $\mathfrak{a}$ and to calculate the asymptotic volume of the set E_t.

A.1. Cones

We will identify $\mathfrak{a}$ as before with the subspace of $\mathbb{R}^d$ consisting of all vectors $X = (X_1,\dots,X_d)$ with $X_1 + \cdots + X_d = 0$. We also identify $\mathfrak{a}$ with the set of all trace 0 diagonal matrices whenever convenient.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

Lemma A.1. *Let $X^0 \in \mathcal{P}$ be any point in $\mathcal{P}$ satisfying*

$$\left\langle X^0, \rho \right\rangle = \max_{X \in \mathcal{P}^+} \langle X, \rho \rangle.$$

Let $\{e_1, \dots, e_d\}$ denote the usual standard basis of $\mathbb{R}^d$. Then

$$X^0 = \begin{cases} e_1 + \cdots + e_s - e_{s+1} - \cdots - e_d & \text{if } d \text{ is even,} \\ e_1 + \cdots + e_{s-1} - e_{s+1} - \cdots - e_d & \text{if } d \text{ is odd,} \end{cases} \tag{49}$$

where s is defined in equation (47). In particular, X^0 is unique.

Proof. We write the half-sum of positive roots $\rho \in \mathfrak{a}^*$ in coordinates as $\rho = (\rho_1, \dots, \rho_d) \in \mathbb{R}^d \simeq \left(\mathbb{R}^d\right)^*$, where $\rho_1 + \cdots + \rho_d = 0$. We have $\rho_1 \geq \cdots \geq \rho_d$ and $\rho_i = -\rho_{d+1-i}$ for every $i \leq d/2$, so that if d is odd, then $\rho_{(d+1)/2} = 0$. Let $X = (X_1, \dots, X_d) \in \mathcal{P}$. Then

$$\langle X, \rho \rangle = \sum_{i \leq d/2} (X_i - X_{d+1-i}) \rho_i.$$

Since $X \in \mathcal{P}$, we have $-2 \leq X_i - X_{d+1-i} \leq 2$. Furthermore, $\rho_i > 0$ for $i \leq d/2$, so this sum is maximized for $X_i = 1$ and $X_{d+1-i} = -1$. This proves our assertion for d even. For d odd, we finally note that $X_1 + \cdots + X_d = 0$ forces $X_{(d+1)/2} = 0$. □

Lemma A.2. *There is a basis $\mu_1, \dots, \mu_{d-1} \in \mathfrak{a}^*$ such that $\mu_i\left(X^0\right) = 1$, $i = 1, \dots, d-1$, and $\mathcal{P}^+ = \mathcal{C} \cap \overline{\mathfrak{a}^+}$, where*

$$\mathcal{C} = X^0 + \mathcal{C}^0 \quad \text{and} \quad \mathcal{C}^0 = \{X \in \mathfrak{a} : \mu_i(X) \leq 0\}.$$

In particular, if $\{\beta_i^\vee\}$ is the basis in $\mathfrak{a}$ which is dual to $\{\mu_i\}$, then

$$\mathcal{C}^0 = \left\{ \sum_{i=1}^{d-1} x_i \beta_i^\vee : \forall\, i\ x_i \leq 0 \right\}$$

and $X^0 = \beta_1^\vee + \cdots + \beta_{d-1}^\vee$.

Proof. Note that

$$\mathcal{P}^+ = \{X \in \mathfrak{a} : X_1 \leq 1, -X_d \leq 1\} \cap \overline{\mathfrak{a}^+}. \tag{50}$$

We rewrite this using the system of fundamental weights $\varpi_i \in \mathfrak{a}^*$:

$$\mathcal{P}^+ = \{X \in \mathfrak{a} : \varpi_1(X) \leq 1, \varpi_{d-1}(X) \leq 1\} \cap \overline{\mathfrak{a}^+}. \tag{51}$$

Since $\varpi_1\left(X^0\right) = \varpi_{d-1}\left(X^0\right) = 1$, we can put $\mu_1 = \varpi_1$ and $\mu_{d-1} = \varpi_{d-1}$. The strategy is then to complete $\{\mu_1, \mu_{d-1}\}$ to a basis of linear forms $\{\mu_i\}$ in such a way that

(1) $\mu_i\left(X^0\right) = 1$ for all $i = 1, \dots, d-1$ and

(2) the conditions $X \in \overline{\mathfrak{a}^+}$ and $\mu_1(X), \mu_{d-1}(X) \leq 1$, imply $\mu_i(X) \leq 1$ for $i = 2, \dots, d-2$.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

In this case, property (1) shows that

$$\mathcal{C} = X^0 + \mathcal{C}^0 = \left\{X \in \mathfrak{a} \mid \forall\ i : \mu_i\left(X - X^0\right) \leq 0\right\} = \{X \in \mathfrak{a} \mid \forall\ i : \mu_i(X) \leq 1\},$$

and then property (2) combines with equation (51) to establish $\mathcal{P}^+ = \mathcal{C} \cap \overline{\mathfrak{a}^+}$.

Note that relative to the standard basis $\{e_i\}$ of $\mathbb{R}^d$, μ_1 is the (restriction to $\mathfrak{a}$ of the) first coordinate functional and μ_{d-1} is minus the (restriction to $\mathfrak{a}$ of the) last coordinate functional. Thus if the $\pm\mu_i$, for $i = 2,\ldots,d-2$, are a linearly independent subset of the remaining $d-2$ coordinate functionals, then (2) follows from the fact that $1 \geq X_1 \geq \cdots \geq X_d \geq -1$ implies $\pm X_i \leq 1$. To ensure property (1), we simply need to choose the sign suitably and avoid the $(d+1)/2$ coordinate for odd d. Thus for d odd we put

$$\mu_i = \begin{cases} \varpi_i - \varpi_{i-1} & \text{if } 2 \leq i \leq s-1, \\ \varpi_i - \varpi_{i+1} & \text{if } s \leq i \leq d-2, \end{cases}$$

and for d even we put

$$\mu_i = \begin{cases} \varpi_i - \varpi_{i-1} & \text{if } 2 \leq i \leq s, \\ \varpi_i - \varpi_{i+1} & \text{if } s+1 \leq i \leq d-2, \end{cases}$$

where s is defined in equation (47). Using the definition (49) of X^0, we quickly verify property (1). □

Remark A.1. Using the explicit description of the basis $\{\mu_i\}_{1\leq i\leq d-1}$, one can easily see that writing ρ as a linear combination of the μ_i results in all the coefficients being positive integers. Hence if $Y \in \mathcal{C}^0$, then $\langle \rho, Y\rangle < 0$.

A.2. Volumes

Proposition A.3. *Suppose $P \subseteq \overline{\mathfrak{a}^+}$ is a convex polytope. Let V_P denote the vertices of P, and suppose that $V_P \ni X \mapsto \langle \rho, X\rangle$ takes its maximum at exactly one vertex $v_0 \in V_P$. Then there exist $c,\delta > 0$ such that*

$$\int_{P_t} \sum_{w\in W} \sigma(w) e^{\langle w\rho, X\rangle} dX = c e^{t\langle \rho, v_0\rangle} \left(1 + O\left(e^{-t\delta}\right)\right) \tag{52}$$

as $t \to \infty$. Here $P_t = \{tX \mid X \in P\}$ and $\sigma(w)$ denotes the signature of $w \in W$.

Before starting the proof of this proposition, we note that

$$\sum_{w\in W} \sigma(w) e^{\langle w\rho, X\rangle} \geq 0$$

for all $X \in \overline{\mathfrak{a}^+}$ because of the convexity of the exponential function.

Proof. First note that $\langle w\rho, X\rangle \leq \langle \rho, X\rangle$ for all $X \in \overline{\mathfrak{a}^+}$, and in fact $\langle w\rho, X\rangle < \langle \rho, X\rangle$ if $X \in \mathfrak{a}^+$. If P is not simple, we divide it into disjoint simple polytopes $P^1,\ldots,P^r$ whose respective sets of vertices we denote by $V^1,\ldots,V^r$. The vertices V_P are contained in the union $\bigcup_j V^j$. By Brion's formula [7], there exist nonzero coefficients $a_v^j \in \mathbb{R}$, $j = 1,\ldots,r$,

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

$v \in V^j$, such that

$$\int_{P_t^j} e^{\langle w\rho, X\rangle} dX = \sum_{v \in V^j} a_v^j e^{t\langle w\rho, v\rangle}.$$

We can find $\delta > 0$ such that for every j and $v \in V^j$, $\langle w\rho, v\rangle \leq \langle \rho, v_0\rangle - \delta$ unless $w^{-1}v = v_0$. Hence,

$$\int_{P_t} \sum_{w\in W} \sigma(w) e^{\langle w\rho, X\rangle} dX = \left(\sum_{\substack{j,v,w:\\ w^{-1}v=v_0}} a_v^j \sigma(w) \right) e^{t\langle \rho, v_0\rangle}. \tag{53}$$

It remains to argue that this last sum in brackets is a positive number. If v_0 is such that $\langle w\rho, v\rangle = \langle \rho, v_0\rangle$ if and only if $w = 1$ and $v = v_0$, then the left-hand side of equation (52) tends to $+\infty$ as $t \to \infty$. Moreover, the sum on the right-hand side of equation (53) has exactly one nonzero term in that situation, which consequently must be positive. If v_0 does not satisfy this condition, we choose another polytope $P' \subseteq P$ with regular vertices (i.e., $V_{P'} \subseteq \mathfrak{a}^+$) that are sufficiently close to the original vertices of P. Then

$$\int_{P_t} \sum_{w\in W} \sigma(w) e^{\langle w\rho, X\rangle} dX \leq \int_{P'_t} \sum_{w\in W} \sigma(w) e^{\langle w\rho, X\rangle} dX,$$

and the right-hand side grows like $ce^{tc'}$ for suitable $c > 0$ and $\langle \rho, v_0\rangle \geq c' > \langle \rho, v_0\rangle - \delta$, provided P' is chosen sufficiently close to P. This proves that the sum in brackets on the right-hand side of equation (53) is a positive number. □

We apply the foregoing result to estimate the volume of E_t.

Corollary A.4. *There exist $c, \delta > 0$ such that we have*

$$m_G(E_t) = ce^{2t\langle \rho, X^0\rangle} \left(1 + O\left(e^{-t\delta}\right)\right)$$

as $t \to \infty$.

Proof. By Cartan decomposition, we have

$$m_G(E_t) = \int_{P_t} J(X) dX = 2^{-\dim \mathfrak{n}} \int_{P_t} \sum_{w\in W} \sigma(w) e^{2\langle w\rho, X\rangle} dX$$

where $\sigma(w)$ denotes the signature of w. The assertion of the lemma then follows from Proposition A.3. □

Appendix B. Angles and inner products

The purpose of this appendix is to prove the following result, used in the course of the proof of Lemma 4.5.

Recall the notation of that lemma. Let X^0 be as in equation (17) and let M denote its centralizer in G. For $Y \in \mathfrak{a}$ we write $Y = Y_M + Y^M$ for unique $Y_M \in \mathfrak{a}_M$ and $Y^M \in \mathfrak{a}^M$.

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

As before, for $\eta > 0$ we write $\mathcal{T}_\eta$ for the set of all $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ such that $\operatorname{Re}\lambda$ lies in the convex hull of the Weyl group orbit of $\eta\rho$. The cone $\mathcal{C}^0$ is defined in Lemma A.2.

Proposition B.1. *There are constants $C_1, C_2 > 0$, depending only on d, such that for every $Y \in \mathcal{C}^0$, $\lambda \in \mathcal{T}_\eta$, and $w \in W$, we have*

(i) $|\langle w\rho, Y_M\rangle| \le -C_1\langle \rho, Y\rangle$ *and*

(ii) $|\langle \operatorname{Re}\lambda, Y_M\rangle| \le -C_2\eta\langle \rho, Y_M\rangle$.

In particular, there is a constant $C > 0$, depending only on d, such that for all $Y \in \mathcal{C}^0$ and $\lambda \in \mathcal{T}_\eta$, we have

$$|\langle \operatorname{Re}\lambda, Y_M\rangle| \le -\eta C\langle \rho, Y\rangle.$$

To prove the proposition we shall work abstractly with linear forms on the Euclidean space $\mathbb{R}^d$, equipped with its standard inner product. Namely, we make the usual identification of the standard Cartan subalgebra $\mathfrak{a}_0$ of $\mathfrak{gl}_d(\mathbb{R})$ with $\mathbb{R}^d$, so that $\mathfrak{a} = \mathfrak{a}_0^G$ identifies with the trace 0 hyperplane

$$\mathcal{H} = \left\{X = (X_1, \dots, X_d) \in \mathbb{R}^d : X_1 + \cdots + X_d = 0\right\}.$$

We are interested in the linear form $L(Y) = \langle \rho, Y\rangle$ on $\mathbb{R}^d$ or $\mathcal{H}$. It will be convenient to define a cone $\mathcal{C}'$ in $\mathbb{R}^d$ such that $\mathcal{C}' \cap \mathcal{H}$ coincides with the cone $-\mathcal{C}^0$ of Proposition A.2. Property (i) will then follow from a similar maximizing property of L on $\mathcal{C}'$. Property (ii) requires bounds on the angles that the vectors in $-\mathcal{C}^0$ can form with X^0, which we deduce from the relative position of X^0 with $\mathcal{H}$.

B.1. Positive linear forms

Let $\{e_1, \dots, e_d\}$ denote the standard basis of $\mathbb{R}^d$.

Definition B.2. If d is even, let $\mathcal{C}' \subset \mathbb{R}^d$ be the closed orthant

$$\mathcal{C}' = \operatorname{cone}_{\mathbb{R}_{\ge 0}}\left\{e_1, \dots, e_{d/2}, -e_{d/2+1}, \dots, -e_d\right\}.$$

If d is odd, let $\mathcal{C}' = \mathcal{O}^+ \cup \mathcal{O}^- \subset \mathbb{R}^d$ be the union of the closed orthants

$$\mathcal{O}^+ = \operatorname{cone}_{\mathbb{R}_{\ge 0}}\{e_1, \dots, e_{(d+1)/2-1}, e_{(d+1)/2}, -e_{(d+1)/2+1}, \dots, -e_d\},$$
$$\mathcal{O}^- = \operatorname{cone}_{\mathbb{R}_{\ge 0}}\{e_1, \dots, e_{(d+1)/2-1}, -e_{(d+1)/2}, -e_{(d+1)/2+1}, \dots, -e_d\}.$$

Remark B.1. Recall the explicit description of $X^0 \in \mathbb{R}^d$ as given in equation (49) (in Appendix A). If V_d denotes the set of vertices of the unit cube $[-1,1]^d$, then $\pm X^0 \in V_d$ for d even and $\pm X^0 \pm e_{(d+1)/2} \in V_d$ for d odd. In particular, when d is odd, $-X^0$ lies on the edge of $[-1,1]^d$ connecting the vertices $X^0 \pm e_{(d+1)/2}$. We deduce that the origin of $\mathbb{R}^d$ is a vertex of $-X^0 + [-1,1]^d$ if d is even, and it is the midpoint of an edge of $-X^0 + [-1,1]^d$ if d is odd.

It is easy to see that when d is even, $-\mathcal{C}'$ is the unique orthant in $\mathbb{R}^d$ containing $-X^0 + [-1,1]^d$, and when d is odd, $-X^0 + [-1,1]^d \subset -\mathcal{C}'$.

Remark B.2. If d is even, then $\mathcal{C}'$ does not contain any nontrivial linear subspace, and if d is odd, the only nontrivial linear subspace of $\mathcal{C}'$ is the 1-dimensional space $\mathbb{R}e_{(d+1)/2}$.

Definition B.3. We call a linear form $L:\mathbb{R}^d \longrightarrow \mathbb{R}$ *positive* if it is nonnegative on $\mathcal{C}'$ and either positive on $\mathcal{C}'\setminus\{0\}$ if d is even or positive on $\mathcal{C}'\setminus \mathbb{R}e_{(d+1)/2}$ if d is odd.

Let $\mathcal{L}\subseteq\mathbb{R}^d$ denote the ray $\mathbb{R}_{\geq 0}X^0$. Since $X^0\in\mathcal{C}'$, and $\mathcal{C}'$ is a cone, we have $\mathcal{L}\subset\mathcal{C}'$.

We identify the space of linear forms on $\mathbb{R}^d$ with $\left(\mathbb{R}^d\right)^*\simeq\mathbb{R}^d$ via the dual basis to $\{e_1,\dots,e_d\}$. If L is a linear form, written as $L=(L_1,\dots,L_d)$ with respect to this identification, and $\sigma\in S_d$ is a permutation, we define the linear form σL by the rule $\sigma L:=\left(L_{\sigma(1)},\dots,L_{\sigma(d)}\right)$. We make the following identification:

Lemma B.4. *Let $L\in\left(\mathbb{R}^d\right)^*$ be a positive linear form. Let L' be a form in the closed convex hull of $\{\sigma L:\sigma\in S_d\}\subseteq\left(\mathbb{R}^d\right)^*$. Then for all $X\in\mathcal{L}$, we have $L'(X)\leq L(X)$.*

Proof. A linear form $L=(L_1,\dots,L_d)$ is positive if and only if

$$\begin{cases} L_1,\dots,L_{d/2}>0,\ L_{d/2+1},\dots,L_d<0, & \text{if } d \text{ is even},\\ L_1,\dots,L_{(d+1)/2-1}>0,\ L_{(d+1)/2}=0,\ L_{(d+1)/2+1},\dots,L_d<0, & \text{if } d \text{ is odd}.\end{cases}$$

It follows that for such L we have $L\left(X^0\right)=|L_1|+\cdots+|L_d|$, and hence $\sigma L\left(X^0\right)\leq L\left(X^0\right)$ for every $\sigma\in S_d$. Hence if L' is in the closed convex hull of $\{\sigma L:\sigma\in S_d\}$ in $\left(\mathbb{R}^d\right)^*$, we also have $L'\left(X^0\right)\leq L\left(X^0\right)$. □

B.2. Hyperplane intersections

We now put $\mathcal{C}''=\mathcal{C}'\cap\mathcal{H}$.

Lemma B.5. *$\mathcal{C}''$ is a convex cone in $\mathbb{R}^d$ of dimension $d-1$ that satisfies $\mathcal{C}''\cap(-\mathcal{C}'')=\{0\}$. Moreover, there exists $\gamma>0$ such that the angle between any nonzero vector in $\mathcal{C}''$ and the vector X^0 is at most $\pi/2-\gamma$.*

Proof. The first part of the lemma follows directly from the description of $\mathcal{C}'$ in Definition B.2, coupled with Remark B.2.

For the assertion on the bound of the angles, we proceed as follows: If d is even, then $\mathcal{C}'$ is just a single orthant in $\mathbb{R}^d$, hence the angle between any two vectors is bounded by $\pi/2$. Since X^0 does not lie on the boundary of $\mathcal{C}'$, there exists $\gamma>0$ such that the angle between X^0 and any vector in $\mathcal{C}'$ is bounded from above by $\pi/2-\gamma$.

For d odd, recall that $\mathcal{C}'=\mathcal{O}^+\cup\mathcal{O}^-$ and $X^0\in\mathcal{O}^+\cap\mathcal{O}^-$, hence the angle between X^0 and any vector in $\mathcal{C}'$ is at most $\pi/2$. Moreover, it is readily seen from the explicit descriptions of X^0 and $\mathcal{O}^\pm$ that the only vectors in $\mathcal{C}'$ that are orthogonal to X^0 are the vectors on the line $\mathbb{R}e_{(d+1)/2}$. Since $e_{(d+1)/2}\notin\mathcal{H}$ and $\mathcal{C}''=\mathcal{C}'\cap\mathcal{H}$ is closed, the angle between X^0 and nonzero vectors in $\mathcal{C}''$ must be bounded from above by $\pi/2-\gamma$ for some $\gamma>0$. □

Note that $\mathcal{L}\subset\mathcal{C}''$, since $X^0\in\mathcal{H}$.

Proposition B.6. *Let $\mathcal{M}\subseteq\mathcal{H}$ be a vector subspace that is orthogonal to $\mathcal{L}$. Then for every $X\in\mathcal{L}$, the section*

$$(X+\mathcal{M})\cap\mathcal{C}''$$

is compact. Moreover, if $\mathcal{M}$ *is of dimension* $d-2$*, then*

$$\mathcal{C}'' = \bigcup_{X \in \mathcal{L}} ((X+\mathcal{M}) \cap \mathcal{C}''),$$

and writing $X = rX^0$ *for a suitable* $r \geq 0$*, we have* $(X+\mathcal{M}) \cap \mathcal{C}'' = r\left(\left(X^0+\mathcal{M}\right) \cap \mathcal{C}''\right)$.

Proof. Follows from Lemma B.5. □

B.3. Application to linear forms

In the following lemma, we shall take L to be a positive linear form in the sense of Definition B.3, and $\mathcal{M} \subseteq \mathcal{H}$ a vector space of dimension $d-2$ which is orthogonal to $\mathcal{L}$:

Lemma B.7. *There exists a constant* $c > 0$ *such that* $L(X+Z) \leq cL(X)$ *for all* $X \in \mathcal{L}$ *and* $Z \in \mathcal{M}$ *with* $X+Z \in \mathcal{C}''$.

Proof. For fixed $X \in \mathcal{L}$, the set of all $Z \in \mathcal{M}$ such that $X+Z \in \mathcal{C}''$ is compact by Proposition B.6. In particular, there is $D \subseteq \mathcal{M}$ compact such that $\left(X^0+\mathcal{M}\right) \cap \mathcal{C}'' = X^0 + D$. Hence any $X+Z \in \mathcal{C}''$ with $X \in \mathcal{L}$ and $Z \in \mathcal{M}$ can be written as $X+Z = rX^0 + rZ'$ for some suitable $r \geq 0$ and $Z' \in D$. Putting

$$c = 1 + \frac{\max_{Z' \in D} L(Z')}{L\left(X^0\right)},$$

we then get $L(X+Z) = rL\left(X^0\right) + rL(Z') \leq cL(X)$. □

B.4. Proof of Proposition B.1

We now apply the previous results to our situation. Namely, we identify $\mathfrak{a}_0$ with $\mathbb{R}^d$ as usual so that $\mathfrak{a} = \mathfrak{a}_0^G$ is the hyperplane $\mathcal{H}$. The cone $\mathcal{C}'' = \mathcal{C}' \cap \mathcal{H}$ then coincides with the cone $-\mathcal{C}^0$ that we considered in Appendix A, and X^0 coincides with the vector of the same name from Appendix A.

We consider the linear form $L(Y) = \langle \rho, Y \rangle$, $Y \in \mathcal{C}$. Let $\mathfrak{a}_M \subseteq \mathfrak{a}$ be as before. Then $\mathbb{R}X^0 \subseteq \mathfrak{a}_M$. In fact, if d is even, then $\mathfrak{a}_M = \mathbb{R}X^0$ and $\mathfrak{a}^M$ is orthogonal to $\mathfrak{a}_M$ of dimension $d-2$. If d is odd, then $\mathfrak{a}_M$ has dimension 2. In that case, let $V \subseteq \mathfrak{a}_M$ be the orthogonal complement to the line $\mathbb{R}X^0$ so that $V \oplus \mathfrak{a}^M$ is orthogonal to $\mathbb{R}X^0$ of dimension $d-2$.

As before, if $Y \in \mathcal{C}''$, we write $Y = Y_M + Y^M$ for unique $Y_M \in \mathfrak{a}_M$ and $Y^M \in \mathfrak{a}^M$. If d is even, then $Y_M \in \mathbb{R}_{\geq 0}X^0$, and if d is odd, we can uniquely write $Y_M = Y_M' + Y_M''$ with unique $Y_M' \in \mathbb{R}_{\geq 0}X^0$ and $Y_M'' \in V$.

Having set up the notation, we now note that $\langle \rho, \cdot \rangle$ vanishes on V, and hence Proposition B.1(i) is an immediate consequence of Lemma B.4.

For part (ii), we first note that by the definition of $\mathcal{T}_\eta$, $\operatorname{Re}\lambda$ is contained in the convex hull of the Weyl group orbit of $\eta\rho$ so that $\left\langle \operatorname{Re}\lambda, wX^0 \right\rangle \leq \eta \left\langle \rho, X^0 \right\rangle$, by Lemma B.7, for every $w \in W$. This establishes part (ii) for d even. For d odd, we have $\eta\langle \rho, Y_M \rangle = \eta\langle \rho, Y_M' \rangle \geq \langle \operatorname{Re}\lambda, Y_M' \rangle$, since $\langle \rho, \cdot \rangle$ vanishes on V. Taking $Y_M' = X^0$, the possible Y_M'' lie in a compact set so that $\langle \operatorname{Re}\lambda, Y_M'' \rangle$ is bounded from above by ηc for c some absolute

https://doi.org/10.1017/S147474802100058X Published online by Cambridge University Press

constant. Scaling $Y'_M = X^0$ by a scalar, the compact set of Y''_M gets scaled by the same scalar. Hence there is $C > 0$ such that for any $Y \in \mathcal{C}''$, we have $\langle \mathrm{Re}\,\lambda, Y_M \rangle \leq \eta C \langle \rho, Y_M \rangle \square$

Acknowledgments The first author was supported by ANR grant 14-CE25.

Competing Interest. None.

References

[1] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault and I. Samet, On the growth of L^2-invariants for sequences of lattices in Lie groups, *Ann. of Math. (2)* **185**(3) (2017), 711–790.

[2] M. Abert, N. Bergeron and E. Le Masson, 'Eigenfunctions and random waves in the Benjamini–Schramm limit', Preprint, (2018), https://arxiv.org/abs/1810.05601.

[3] N. Arantharaman and E. Le Masson, Quantum ergodicity on large regular graphs, *Duke Math. J.* **164**(4) (2015), 723–765.

[4] M. B. Bekka and M. Mayer, *Ergodic theory and topological dynamics of group actions on homogeneous spaces. London Math. Soc. Lecture Note Series* **269**, pp. x–200 (Cambridge University Press, Cambridge, 2000).

[5] I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs, *Electron. J. Probab.* **6** (2001), 1–13.

[6] A. Borel and N. Wallach, *Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups*, *Annals of Mathematics Studies*, **94** (Princeton University Press, Princeton, NJ, 1980).

[7] M. Brion and M. Vergne, Residue formulae, vector partition functions and lattice points in rational polytopes, *J. Amer. Math. Soc.* **10** (1997), 797–833.

[8] S. Brooks, E. Le Masson and E. Lindenstrauss, Quantum ergodicity and averaging operators on the sphere, *Int. Math. Res. Not. IMRN* **19** (2016), 6034–6064.

[9] F. Brumley and S. Marshall, Lower bounds for Maass forms on semisimple groups, *Compos. Math.* **165**(5) (2020), 959–1003.

[10] Y. Colin de Verdière, Ergodicité et fonctions propres du laplacien, *Comm. Math. Phys.* **102**(3) (1985), 497–502.

[11] M. Cowling, Sur les coefficients des représentations unitaires des groupes de Lie simples, in *Analyse Harmonique sur les groupes de Lie II*, Lecture Notes in Mathematics, **739**, pp. 132–178 (Springer, Berlin, 1979).

[12] J. J. Duistermaat, J. A. C. Kolk and V. S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature, *Invent. Math.* **52**(1) (1979), 27–93.

[13] R. Gangolli, On the Plancherel formula and the Paley–Wiener theorem for spherical functions on semisimple Lie groups, *Ann. of Math (2)* **93** (1971), 150–165.

[14] R. Gangolli and V. S. Varadarajan, *Harmonic analysis of spherical functions on real reductive groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **101**, pp. xiv–365 (Springer, Verlag, Berlin, 1988).

[15] A. Gorodnik and A. Nevo, Quantitative ergodic theorems and their number theoretic applications, *Bull. Amer. Math. Soc. (N.S.)* **52**(1) (2015), 65–113.

[16] Harish-Chandra, Spherical functions on semisimple Lie groups I, *Amer. J. Math.* **79** (1958), 241–310.

[17] S. Helgason, *Groups and Geometric Analysis*, *Mathematical Surveys and Monographs*, **83** (American Mathematical Society, Providence, RI, 2000).

[18] R. E. Howe and C. C. Moore, Asymptotic properties of unitary representations, *J. Funct. Anal.* **32** (1979), 72–96.

[19] J. Jorgenson and S. Lang, *Spherical Inversion on $SL_n(\mathbb{R})$*, *Springer Monographs in Mathematics* (Springer-Verlag, New York, 2001).

[20] D. Kazhdan, On the connection of the dual space of a group with the structure of its closed subgroups, *Funct. Anal. Appl.* **1**(1) (1967), 63–65.

[21] G. Knieper, On the asymptotic geometry of nonpositively curved manifolds, *Geom. Funct. Anal.* **7**(4) (1997), 755–782.

[22] E. Lapid and W. Müller, Spectral asymptotics for arithmetic quotients of $SL_n(\mathbb{R})/\mathsf{SO}(n)$, *Duke. Math. J.* **149**(1) (2009), 117–155.

[23] E. Le Masson, Pseudo-differential calculus on homogeneous trees, *Ann. Henri Poincaré* **15**(9) (2014), 1697–1732.

[24] E. Le Masson and T. Sahlsten, Quantum ergodicity and Benjamini–Schramm convergence of hyperbolic surfaces, *Duke Math. J.* **166**(18) (2017), 3425–3460.

[25] E. Lindenstrauss, On quantum unique ergodicity for $\Gamma\backslash\mathcal{H}\times\mathcal{H}$, *Int. Math. Res. Not. IMRN* **17** (2001), 913–933.

[26] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity, *Ann. of Math. (2)* **163**(1) (2006), 165–219.

[27] G. A. Margulis, *Discrete Subgroups of Semisimple Lie Groups*, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, **17** (Springer-Verlag, Berlin, 1991).

[28] P. Nelson and A. Venkatesh, The orbit method and analysis of automorphic forms, *Acta Math.* **226**(1) (2021), 1–209.

[29] A. Nevo, Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups, *Math. Res. Lett.* **5**(3) (1998), 305–325.

[30] L. Silberman and A. Venkatesh, On quantum unique ergodicity for locally symmetric spaces, *Geom. Funct. Anal.* **17** (2007), 960–998.

[31] L. Silberman and A. Venkatesh, Entropy bounds for Hecke eigenfunctions on division algebras, in *Probabilistic Methods in Geometry, Topology and Spectral Theory*, Contemporary Mathematics, **739**, pp. 171–197 (Centre de Recherches Mathématique Proceedings, American Math Soc., Providence, RI, 2019).

[32] A. Šnirelman, Ergodic properties of eigenfunctions, *Uspekhi Mat. Nauk* **6**(180) (1974), 181–182.

[33] Z. Shem – Tov, Positive entropy using Hecke operators at a single place, *Int. Math. Res. Not. IMRN* (2020), https://doi.org/10.1093/imrn/rnaa235.

[34] P. C. Trombi and V. S. Varadarajan, Spherical transforms on semisimple Lie groups, *Ann. of Math.* **94** (1971), 246–303.

[35] S. Zelditch, Uniform distribution of eigenfunctions on compact hyperbolic surfaces, *Duke Math. J.* **55**(4) (1987), 919–941.

[36] S. Zelditch, Quantum ergodicity and mixing of eigenfunctions, in Encyclopedia of Mathematical Physics (Academic Press/Elsevier Science, *Oxford, UK*, 2006).