The Harmonic Analysis of Lattice Counting on Real Spherical Spaces

Bernhard Krötz[1], Eitan Sayag[2], Henrik Schlichtkrull

Received: August 28, 2014
Revised: December 17, 2015

Communicated by Patrick Delorme

Abstract. By the collective name of *lattice counting* we refer to a setup introduced in [10] that aims to establish a relationship between arithmetic and randomness in the context of affine symmetric spaces. In this paper we extend the geometric setup from symmetric to real spherical spaces and continue to develop the approach with harmonic analysis which was initiated in [10].

2010 Mathematics Subject Classification: 22E40, 22E46, 43A85
Keywords and Phrases: Lattice counting, homogeneous spaces, real spherical spaces, spectral analysis

1 Introduction

1.1 Lattice counting

Let us recall from Duke, Rudnick and Sarnak [10] the setup of lattice counting on a homogeneous space $Z = G/H$. Here G is an algebraic real reductive group and $H < G$ an algebraic subgroup such that Z carries an invariant measure. Further we are given a lattice $\Gamma < G$ such that its trace $\Gamma_H := \Gamma \cap H$ in H is a lattice in H.
Attached to invariant measures dh and dg on H and G we obtain an invariant measure $d(gH)$ on Z via Weil-integration:

$$\int_Z \Big(\int_H f(gh) dh \Big) \, d(gH) = \int_G f(g) \, dg \qquad (f \in C_c(G)) \, .$$

[1]Supported by ERC Advanced Investigators Grant HARG 268105
[2]Partially supported by ISF 1138/10 and ERC 291612

Likewise the measures dg and dh give invariant measures $d(g\Gamma)$ and $d(h\Gamma_H)$ on $Y := G/\Gamma$ and $Y_H := H/\Gamma_H$. We pin down the measures dg and dh and hence $d(gH)$ by the request that Y and Y_H have volume one.
Further we are given a family $\mathcal{B}$ of "balls" $B_R \subset Z$ depending on a parameter $R \geq 0$. At this point we are rather imprecise about the structure of these balls and content us with the property that they constitute an exhausting family of compact sets as $R \to \infty$.
Let $z_0 = H \in Z$ be the standard base point. The *lattice counting problem* for $\mathcal{B}$ consists of the determination of the asymptotic behavior of the density of $\Gamma \cdot z_0$ in balls $B_R \subset Z$, as the radius $R \to \infty$. By *main term counting* for $\mathcal{B}$ we understand the statement that the asymptotic density is 1. More precisely, with

$$N_R(\Gamma, Z) := \#\{\gamma \in \Gamma/\Gamma_H \mid \gamma \cdot z_0 \in B_R\}$$

and $|B_R| := \mathrm{vol}_Z(B_R)$ we say that main term counting holds if

$$N_R(\Gamma, Z) \sim |B_R| \quad (R \to \infty). \tag{1.1}$$

1.2 Relevant previous works

The main term counting was established in [10] for symmetric spaces G/H and certain families of balls, for lattices with Y_H compact. Furthermore, the main term counting in the case where Y_H is non-compact was proven using a hypothesis on regularization of periods of Eisenstein series, whose proof remains unpublished. In subsequent work Eskin and McMullen [11] removed the obstruction that Y_H is compact and presented an ergodic approach. Later Eskin, Mozes and Shah [12] refined the ergodic methods and discovered that main term counting holds for a wider class of reductive spaces: For reductive algebraic groups G, H defined over $\mathbb{Q}$ and arithmetic lattices $\Gamma < G(\mathbb{Q})$ it is enough to request that the identity component of H is not contained in a proper parabolic subgroup of G which is defined over $\mathbb{Q}$ and that the balls B_R satisfy a certain condition of *non-focusing*.
In these works the balls B_R are constructed as follows. All spaces considered are affine in the sense that there exists a G-equivariant embedding of Z into the representation module V of a rational representation of G. For any such embedding and any norm on the vector space V, one then obtains a family of balls B_R on Z by intersection with the metric balls in V. For symmetric spaces all families of balls produced this way are suitable for the lattice counting, but in general one needs to assume non-focusing in addition. In particular all maximal reductive subgroups satisfy all the conditions and hence fulfill the main term counting.

1.3 Real spherical spaces

In this paper we investigate the lattice counting for a real spherical space Z, that is, it is requested that the action of a minimal parabolic subgroups $P < G$ on

Z admits an open orbit. In addition we assume that H is reductive and remark that with our standing assumption that Z is unimodular this is automatically satisfied for a spherical space when the Lie algebra $\mathfrak{h}$ of H is self-normalizing (see [17], Cor. 9.10).

Our approach is based on spectral theory and is a natural continuation to [10]. We consider a particular type of balls which are intrinsically defined by the geometry of Z (and thus not related to a particular representation V as before).

1.3.1 Factorization of spherical spaces

In the spectral approach it is of relevance to get a control over intermediate subgroups $H < H^\star < G$ which arise in the following way: Given a unitary representation $(\pi, \mathcal{H})$ one looks at the smooth vectors $\mathcal{H}^\infty$ and its continuous dual $\mathcal{H}^{-\infty}$, the distribution vectors. The space $(\mathcal{H}^{-\infty})^H$ of H-invariant distribution vectors is of fundamental importance. For all pairs $(v, \eta) \in \mathcal{H}^\infty \times (\mathcal{H}^{-\infty})^H$ one obtains a smooth function on Z, a *generalized matrix-coefficient*, via

$$m_{v,\eta}(z) = \eta(g^{-1} \cdot v) \qquad (z = gH \in Z)\,. \tag{1.2}$$

The functions (1.2) are the building blocks for the harmonic analysis on Z. The stabilizer H_η in G of $\eta \in (\mathcal{H}^{-\infty})^H$ is a closed subgroup which contains H, but in general it can be larger than H even if π is non-trivial.

Let us call $Z^\star = G/H^\star$ a *factorization* of Z if $H < H^\star$ and $Z^\star$ is unimodular. For a general real spherical space Z the homogeneous spaces $Z_\eta = G/H_\eta$ can happen to be non-unimodular (see [19] for H the Iwasawa N-subgroup). However there is a large subclass of real spherical spaces which behave well under factorization. Let us call a factorization co-compact if $H^\star/H$ is compact and *basic* if (up to connected components) $H^\star$ is of the form $H_I := HI$ for a normal subgroup $I \triangleleft G$. Finally we call a factorization *weakly basic* if it is obtained by a composition of a basic and a co-compact factorization.

1.3.2 Wavefront spherical spaces

A real spherical space is called *wavefront* if the attached compression cone is a quotient of a closed Weyl-chamber. The relevant definitions will be recalled in Section 3. Many real spherical spaces are wavefront: all symmetric spaces and all Gross-Prasad type spaces $G \times H/H$ (see (3.2) - (3.4)) are wavefront.[3] The terminology *wavefront* originates from [24] because wavefront real spherical spaces satisfy the "wavefront lemma" of Eskin-McMullen (see [11], [18]) which is fundamental in the approach of [11] to lattice counting.

On the geometric side wavefront real spherical spaces enjoy the following property from [19]: All Z_η are unimodular and the factorizations of the type Z_η are precisely the weakly basic factorizations of Z.

[3] Also, if Z is complex, then of the 78 cases in the list of [4], the non-wavefront cases are (11), (24), (25), (27), (39-50), (60), (61)

On the spectral level wavefront real spherical spaces are distinguished by the following integrability property, also from [19]: The generalized matrix coefficients $m_{v,\eta}$ of (1.2) belong to $L^p(Z_\eta)$ for some $1 \leq p < \infty$ only depending on π and η.

1.3.3 Main term counting

In the theorem below we assume that Z is a wavefront real spherical space of reductive type. For simplicity we also assume that all compact normal subgroups of G are finite.
Using soft techniques from harmonic analysis and a general property of decay from [21], our first result (see Section 5) is:

Theorem A. Let $Z = G/H$ be as above, and assume that $Y = G/\Gamma$ is compact. Then main term counting (1.1) holds.

Since wavefront real spherical spaces satisfy the wavefront lemma by [18], Section 6, this theorem could also be derived with the ergodic method of [11]. In the current context the main point is thus the proof by harmonic analysis.
To remove the assumption that Y is compact and to obtain error term bounds for the lattice counting problem we need to apply more sophisticated tools from harmonic analysis. This will be discussed in the next paragraph with some extra assumptions on G/H.

1.4 Error Terms

The problem of determining the error term in counting problems is notoriously difficult and in many cases relies on deep arithmetic information. Sometimes, like in the Gauss circle problem, some error term is easy to establish but getting an optimal error term is a very difficult problem.
We restrict ourselves to the cases where the cycle H/Γ_H is compact.[4] To simplify the exposition here we assume in addition that $\Gamma < G$ is irreducible, i.e. there do not exist non-trivial normal subgroups G_1, G_2 of G and lattices $\Gamma_i < G_i$ such that $\Gamma_1\Gamma_2$ has finite index in Γ.
The error we study is measure theoretic in nature, and will be denoted here as $\mathrm{err}(R, \Gamma)$. Thus, $\mathrm{err}(R, \Gamma)$ measures the deviation of two measures on $Y = \Gamma\backslash G$, the counting measure arising from lattice points in a ball of radius R, and the invariant measure $d\mu_Y$ on Y. More precisely, with $\mathbf{1}_R$ denoting the characteristic function of B_R we consider the densities

$$F_R^\Gamma(g\Gamma) := \frac{\sum_{\gamma\in\Gamma/\Gamma_H} \mathbf{1}_R(g\gamma H)}{|B_R|}.$$

Then,

$$\mathrm{err}(R, \Gamma) = ||F_R^\Gamma - d\mu_Y||_1,$$

[4] After a theory for regularization of H-periods of Eisenstein series is developed, one can drop this assumption.

where $||\cdot||_1$ denotes the total variation of the signed measure. Notice that $|F_R^\Gamma(e\Gamma)-1| = \frac{|N_R(\Gamma,Z)-|B_R||}{|B_R|}$ is essentially the error term for the pointwise count (1.1).
Our results on the error term $\mathrm{err}(R,\Gamma)$ allows us to deduce results toward the error term in the smooth counting problem, a classical problem that studies the quantity

$$\mathrm{err}_{pt,\alpha}(R,\Gamma) = |B_R||F_{\alpha,R}^\Gamma(e\Gamma)-1|$$

where $\alpha \in C_c^\infty(G)$ is a positive smooth function of compact support (with integral one) and $F_{\alpha,R}^\Gamma = \alpha * F_R^\Gamma$. See Remark 7.2 for the comparison of $\mathrm{err}(R,\Gamma)$ with $\mathrm{err}_{pt,\alpha}(R,\Gamma)$.
To formulate our result we introduce the exponent $p_H(\Gamma)$ (see (6.2)), which measures the worst L^p-behavior of any generalized matrix coefficient associated with a spherical unitary representation π, which is H-distinguished and occurs in the automorphic spectrum of $L^2(\Gamma\backslash G)$. We first state our result for the non-symmetric case of triple product spaces, which is Theorem 8.2 from the body of the paper.

Theorem B. Let $Z = G_0^3/\operatorname{diag}(G_0)$ for $G_0 = \mathrm{SO}_e(1,n)$ and assume that H/Γ_H is compact. For all $p > p_H(\Gamma)$ there exists a $C = C(p) > 0$ such that

$$\mathrm{err}(R,\Gamma) \le C|B_R|^{-\frac{1}{(6n+3)p}}$$

for all $R \ge 1$. (In particular, main term counting holds in this case). Furthermore, in regards to smooth counting, for any $\alpha \in C_c^\infty(G)$ and for all $p > p_H(\Gamma)$ there exists a $C = C(p,\alpha) > 0$ such that

$$\mathrm{err}_{pt,\alpha}(R,\Gamma) \le C|B_R|^{1-\frac{1}{(6n+3)p}}$$

for all $R \ge 1$.

To the best of our knowledge this is the first error term obtained for a non-symmetric space. The crux of the proof is locally uniform comparison between L^p and L^∞ norms of generalized matrix coefficients $m_{v,\eta}$ which is achieved by applying the model of [3] and [9] for the triple product functional η in spherical principal series.
It is possible to obtain error term bounds under a certain technical hypothesis introduced in Section 6 and refered to as Hypothesis A. This hypothesis in turn is implied by a conjecture on the analytic structure of families of Harish-Chandra modules which we explain in Section 9.1. The conjecture and hence the hypothesis appear to be true for symmetric spaces but requires quite a technical tour de force. In general, the techniques currently available do not allow for an elegant and efficient solution. Under this hypothesis we show that:

Theorem C. Let Z be wavefront real spherical space for which Hypothesis A is valid. Assume also

- G is semisimple with no compact factors

- Γ is arithmetic and irreducible
- $\Gamma_H = H \cap \Gamma$ is co-compact in H.
- $p > p_H(\Gamma)$
- $k > \frac{rank(G/K)+1}{2} \dim(G/K) + 1$

Then, there exists a constant $C = C(p, k) > 0$ such that

$$\operatorname{err}(R, \Gamma) \leq C|B_R|^{-\frac{1}{(2k+1)p}}$$

for all $R \geq 1$. Moreover, if $Y = \Gamma\backslash G$ is compact one can replace the third condition by $k > \dim(G/K) + 1$.

The existence of a non-quantitative error term for symmetric spaces was established in [1] and improved in [14].
We note that in case of the hyperbolic plane our error term is still far from the quality of the bound of A. Selberg. This is because we only use a weak version of the trace formula, namely Weyl's law, and use simple soft Sobolev bounds between eigenfunctions on Y.

2 Reductive homogeneous spaces

In this section we review a few facts on reductive homogeneous spaces: the Mostow decomposition, the associated geometric balls and their factorizations. We use the convention that real Lie groups are denoted by upper case Latin letters, e.g A, B, C, and their Lie algebras by the corresponding lower case German letter $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$.
Throughout this paper G will denote an algebraic real reductive group and $H < G$ is an algebraic subgroup. We form the homogeneous space $Z = G/H$ and write $z_0 = H$ for the standard base point.
Furthermore, unless otherwise mentioned we assume that H is reductive in G, that is, the adjoint representation of H on $\mathfrak{g}$ is completely reducible. In this case we say that G/H is *of reductive type*.
Let us fix a maximal compact subgroup $K < G$ for which we assume that the associated Cartan involution θ leaves H invariant (see the references to [21], Lemma 2.1). Attached to θ is the infinitesimal Cartan decomposition $\mathfrak{g} = \mathfrak{k} + \mathfrak{s}$ where $\mathfrak{s} = \mathfrak{k}^\perp$ is the orthogonal complement with respect to a non-degenerate invariant bilinear form κ on $\mathfrak{g}$ which is positive definite on $\mathfrak{s}$ (if $\mathfrak{g}$ is semi-simple, then we can take for κ the Cartan-Killing form). Further we set $\mathfrak{q} := \mathfrak{h}^\perp$.

2.1 Mostow decomposition

We recall Mostow's polar decomposition:

$$K \times_{H\cap K} \mathfrak{q} \cap \mathfrak{s} \to Z, \quad [k, X] \mapsto k \exp(X) \cdot z_0 \tag{2.1}$$

which is a homeomorphism. With that we define

$$\|k\exp(X)\cdot z_0\|_Z = \|X\| := \kappa(X,X)^{\frac{1}{2}}$$

for $k \in K$ and $X \in \mathfrak{q}\cap\mathfrak{s}$.

2.2 Geometric balls

The problem of lattice counting in Z leads to a question of exhibiting natural exhausting families of compact subsets. We use balls which are intrinsically defined by the geometry of Z.
We define the *intrinsic ball* of radius $R>0$ on Z by

$$B_R := \{z \in Z \mid \|z\|_Z < R\}\,.$$

Write B_R^G for the intrinsic ball of $Z=G$, that is, if $g = k\exp(X)$ with $k \in K$ and $X \in \mathfrak{s}$, then we put $\|g\|_G = \|X\|$ and define B_R^G accordingly.
Our first interest is the growth of the volume $|B_R|$ for $R\to\infty$. We have the following upper bound.

Lemma 2.1. *There exists a constant $c>0$ such that:*

$$|B_{R+r}| \leq e^{cr}|B_R|$$

for all $R \geq 1, r\geq 0$.

Proof. Recall the integral formula

$$\int_Z f(z)\,dz = \int_K \int_{\mathfrak{q}\cap\mathfrak{s}} f(k\exp(X).z_0)\delta(X)\,dX\,dk, \tag{2.2}$$

for $f \in C_c(Z)$, where $\delta(Y)$ is the Jacobian at (k,Y) of the map (2.1). It is independent of k because dz is invariant. Then

$$|B_R| = \int_{X\in\mathfrak{q}\cap\mathfrak{s}, \|X\|<R} \delta(X)\,dX\,.$$

Hence it suffices to prove that there exists $c>0$ such that

$$\int_0^{R+r} \delta(tX)t^{l-1}\,dt \leq e^{cr}\int_0^R \delta(tX)t^{l-1}\,dt$$

for all $X \in \mathfrak{q}\cap\mathfrak{s}$ with $\|X\|=1$. Here $l=\dim\mathfrak{q}\cap\mathfrak{s}$. Equivalently, the function

$$R\mapsto e^{-cR}\int_0^R \delta(tX)t^{l-1}\,dt$$

is decreasing, or by differentiation,

$$\delta(RX)R^{l-1} \leq c\int_0^R \delta(tX)t^{l-1}\,dt$$

for all R. The latter inequality is established in [12, Lemma A.3] with c independent of X. ∎

Further we are interested how the volume behaves under distortion by elements from G.

Lemma 2.2. *For all $r, R > 0$ one has $B_r^G B_R \subset B_{R+r}$.*

To prove the lemma we first record that:

Lemma 2.3. *Let $z = gH \in Z$. Then $\|z\|_Z = \inf_{h\in H} \|gh\|_G$.*

Proof. It suffices to prove that $\|\exp(X)h\|_G \geq \|X\|$ for $X \in \mathfrak{q} \cap \mathfrak{s}$, $h \in H$, and by Cartan decomposition of H, we may assume $h = \exp(T)$ with $T \in \mathfrak{h} \cap \mathfrak{s}$. Thus we have reduced to the statement that

$$\|\exp(X)\exp(T)\|_G \geq \|\exp(X)\|_G$$

for $X \perp T$ in $\mathfrak{s}$. In order to see this, we note that for each $g \in G$ the norm $\|g\|_G$ is the length of the geodesic in $K\backslash G$ which joins the origin x_0 to $x_0 g$. More generally the geodesic between $x_0 g_1$ and $x_0 g_2$ has length $\|g_2 g_1^{-1}\|_G$. Hence $c = \|\exp(X)\exp(T)\|_G$ is the distance from $A = x_0 \exp(-T)$ to $B = x_0 \exp(X)$. As $X \perp T$ the points A and B form a right triangle with $C = x_0$. The hypotenuse has length c and the leg CB has length $a = \|\exp(X)\|$. As the sectional curvatures are non-positive we have $a^2 + b^2 \leq c^2$. In particular $a \leq c$. □

In particular, it follows that

$$\|gz\|_Z \leq \|z\|_Z + \|g\|_G \quad (z \in Z, g \in G) \tag{2.3}$$

and Lemma 2.2 follows.

Remark 2.4. Observe that the norm $\|\cdot\|_G$ on G depends on the chosen Cartan decomposition θ. However, by applying (2.3) with $Z = G$ one sees that the norm obtained with a conjugate θ' of θ will satisfy

$$\|g\|'_G \leq \|g\|_G + c, \qquad \|g\|_G \leq \|g\|'_G + c' \tag{2.4}$$

for all $g \in G$ with some constants $c, c' \geq 0$.
For the definition of $\|\cdot\|_Z$ we assumed that θ leaves H invariant. If instead we use the identity in Lemma 2.3 as the definition of $\|\cdot\|_Z$ then this assumption can be avoided. In any case, it follows that the norms on Z obtained from two different Cartan involutions will satisfy similar inequalities as (2.4). The corresponding families of balls are then also compatible,

$$B_R \subset B'_{R+c}, \qquad B'_R \subset B_{R+c'},$$

for all $R > 0$.

2.3 Factorization

By a (reductive) factorization of $Z = G/H$ we understand a homogeneous space $Z^\star = G/H^\star$ with $H^\star$ an algebraic subgroup of G such that

- $H^\star$ is reductive.
- $H \subset H^\star$.

A factorization is called *compact* if $Z^\star$ is compact, and *co-compact* if the fiber space $\mathcal{F} := H^\star/H$ is compact. It is called *proper* if $\dim H < \dim H^\star < \dim G$.

Lemma 2.5. *Let $Z = G/H \to Z^\star = G/H^\star$ be a factorization. Then the following assertions are equivalent:*

1. *$Z \to Z^\star$ is co-compact.*
2. *There exist a compact subgroup $K^\star < H^\star$ such that $K^\star H = H^\star$.*
3. *There exists a compact subalgebra $\mathfrak{k}^\star < \mathfrak{h}^\star$ such that $\mathfrak{h}^\star = \mathfrak{k}^\star + \mathfrak{h}$ and $\exp(\mathfrak{k}^\star) < H^\star$ compact.*

Proof. First (1) implies (2) by the Mostow decomposition of the reductive homogeneous space $H^\star/H$. Clearly (2) implies (3) as the multiplication map $K^\star \times H \to H^\star$ needs to be submersive by Sard's theorem. Finally, for (3) implies (1) we observe that $H^\star/H$ has finitely many components and $\exp(\mathfrak{k}^\star)H$ is compact and open in there. $\square$

Let $\mathcal{F} \to Z \to Z^\star$ be a factorization of Z. We write $B_R^\star$ and $\mathcal{B}_R^{\mathcal{F}}$ for the intrinsic balls in $Z^\star$ and $\mathcal{F}$, respectively.

Lemma 2.6. *We have $B_R^\star = B_R H^\star/H^\star$ and $B_R^{\mathcal{F}} = B_R \cap \mathcal{F}$.*

Proof. Follows from Lemma 2.3. $\square$

For a compactly supported bounded measurable function ϕ on Z we define the fiberwise integral

$$\phi^{\mathcal{F}}(gH^\star) := \int_{H^\star/H} \phi(gh^\star)\ d(h^\star H)$$

and recall the integration formula

$$\int_Z \phi(gH)\ d(gH) = \int_{Z^\star} \phi^{\mathcal{F}}(gH^\star)\ d(gH^\star) \tag{2.5}$$

under appropriate normalization of measures. Consider the characteristic function $\mathbf{1}_R$ of B_R and note that its fiber average $\mathbf{1}_R^{\mathcal{F}}$ is supported in the compact ball $B_R^\star$. We say that the family of balls $(B_R)_{R>0}$ *factorizes well to* $Z^\star$ provided for all compact subsets $Q \subset G$

$$\lim_{R\to\infty} \frac{\sup_{g\in Q} \mathbf{1}_R^{\mathcal{F}}(gH^\star)}{|B_R|} = 0\,. \tag{2.6}$$

Observe that for all compact subsets Q there exists an $R_0 = R_0(Q) > 0$ such that

$$\sup_{g\in Q} \mathbf{1}_R^{\mathcal{F}}(gH^\star) \leq |B_{R+R_0}^{\mathcal{F}}|$$

by Lemma 2.2. Thus the balls B_R factorize well provided

$$\lim_{R\to\infty} \frac{|B_{R+R_0}^{\mathcal{F}}|}{|B_R|} = 0\,. \tag{2.7}$$

for all $R_0 > 0$.

Remark 2.7. The condition that the balls B_R factorize well is closely related to the non-focusing condition (Definition 1.14 in [12]). Thus, in the case of semi-simple connected H, the non-focusing condition of the intrinsic balls is implied by the condition that they factorize well to all factorizations.

2.4 Basic factorizations

There is a special class of factorizations with which we are dealing with in the sequel. ¿From now on we assume that $\mathfrak{g}$ is semi-simple and write

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \ldots \oplus \mathfrak{g}_m$$

for the decomposition into simple ideals. For a reductive subalgebra $\mathfrak{h} < \mathfrak{g}$ and a subset $I \subset \{1, \ldots, m\}$ we define the reductive subalgebra

$$\mathfrak{h}_I := \mathfrak{h} + \mathfrak{g}_I = \mathfrak{h} + \bigoplus_{i\in I} \mathfrak{g}_i\,. \tag{2.8}$$

We say that the factorization is *basic* provided that $\mathfrak{h}^* = \mathfrak{h}_I$ for some I. Finally we call a factorization *weakly basic* if it is built from consecutive basic and co-compact factorizations, that is, there exists a sequence

$$\mathfrak{h}^\star = \mathfrak{h}^k \supset \cdots \supset \mathfrak{h}^0 = \mathfrak{h} \tag{2.9}$$

of reductive subalgebras such that for each i we have $\mathfrak{h}^i = (\mathfrak{h}^{i-1})_I$ for some I or $\mathfrak{h}^i/\mathfrak{h}^{i-1}$ is compact. The following lemma shows that in fact it suffices with $k \leq 2$.

Lemma 2.8. *Let $Z \to Z^\star$ be a weakly basic factorization. Then there exists an intermediate factorization $Z \to Z_b \to Z^\star$ such that $Z \to Z_b$ is basic and $Z_b \to Z^\star$ co-compact.*

Proof. Let a sequence (2.9) of factorizations which are consecutively basic or compact be given. We first observe that two consecutive basic factorizations make up for a single basic factorization, and likewise two consecutive compact factorizations yield a single compact factorization by Lemma 2.5. Hence it suffices to prove that we can modify a string

$$\mathfrak{h}^{i+2} \supset \mathfrak{h}^{i+1} \supset \mathfrak{h}^i$$

with $\mathfrak{h}^{i+2}/\mathfrak{h}^{i+1}$ basic and $\mathfrak{h}^{i+1}/\mathfrak{h}^{i}$ compact to

$$\mathfrak{h}^{i+2} \supset \mathfrak{h}_b^{i+1} \supset \mathfrak{h}^i$$

with $\mathfrak{h}^{i+2}/\mathfrak{h}_b^{i+1}$ compact and $\mathfrak{h}_b^{i+1}/\mathfrak{h}^i$ basic.
We have $\mathfrak{h}^{i+2} = \mathfrak{h}^{i+1} + \mathfrak{g}_I$ for some I, and by Lemma 2.5 that $\mathfrak{h}^{i+1} = \mathfrak{h}^i + \mathfrak{c}$ with $\mathfrak{c}$ compact. Then $\mathfrak{h}_b^{i+1} := \mathfrak{h}^i + \mathfrak{g}_I$ is a reductive subalgebra and a basic factorization of $\mathfrak{h}^i$. Furthermore $\mathfrak{h}^{i+2} = \mathfrak{h}_b^{i+1} + \mathfrak{c}$. This establishes the lemma □

3 Wavefront real spherical spaces

We assume that Z is real spherical, i.e. a minimal parabolic subgroup $P < G$ has an open orbit on Z. It is no loss of generality to assume that $PH \subset G$ is open, or equivalently that $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$.
If L is a real algebraic group, then we write $L_{\rm n}$ for the normal subgroup of L which is generated by all unipotent element. In case L is reductive we observe that $\mathfrak{l}_{\rm n}$ is the sum of all non-compact simple ideals of $\mathfrak{l}$.
According to [20] there is a unique parabolic subgroup $Q \supset P$ with the following two properties:

- $QH = PH$.
- There is a Levi decomposition $Q = LU$ with $L_{\rm n} \subset Q \cap H \subset L$.

Following [20] we call Q a Z-adapted parabolic subgroup.
Having fixed L we let $L = K_L A_L N_L$ be an Iwasawa decomposition of L. We choose an Iwasawa decomposition $G = KAN$ which inflates the one of L, i.e. $K_L < K, A_L = A$ and $N_L < N$. Further we may assume that N is the unipotent radical of the minimal parabolic P.

Remark 3.1. It should be noted that the assumption on the Cartan decomposition θ, which was demanded in Section 2.2, may be overruled by the above requirement to K. However, it follows from Remark 2.4 that the balls B_R can still be defined, and that the difference does not disturb the lattice counting on Z.

Set $A_H := A \cap H$ and put $A_Z = A/A_H$. We recall that $\dim A_Z$ is an invariant of the real spherical space, called the real rank (see [20]).
In [18], Section 6, we defined the notion of *wavefront* for a real spherical space, which we quickly recall. Attached to Z is a geometric invariant, the so-called compression cone which is a closed and convex subcone $\mathfrak{a}_Z^-$ of $\mathfrak{a}_Z$. It is defined as follows. Write $\Sigma_\mathfrak{u}$ for the space of $\mathfrak{a}$-weights of the $\mathfrak{a}$-module $\mathfrak{u}$ and let $\overline{\mathfrak{u}}$ denote the corresponding sum of root spaces for $-\Sigma_\mathfrak{u}$. According to [20] there exists a linear map

$$T : \oplus_{\alpha\in\Sigma_\mathfrak{u}} \mathfrak{g}^{-\alpha} = \overline{\mathfrak{u}} \to \mathfrak{l}_H^\perp \oplus \mathfrak{u} \subset \oplus_{\beta\in\{0\}\cup\Sigma_\mathfrak{u}} \mathfrak{g}^\beta \tag{3.1}$$

such that $\mathfrak{h} = \mathfrak{l}\cap\mathfrak{h} + \{\overline{X} + T(\overline{X}) \mid \overline{X} \in \overline{\mathfrak{u}}\}$. Here $\mathfrak{l}_H^\perp$ denotes the orthocomplement of $\mathfrak{l}\cap\mathfrak{h}$ in $\mathfrak{l}$. For each pair α, β we denote by

$$T_{\alpha,\beta} : \mathfrak{g}^{-\alpha} \to \mathfrak{g}^{\beta}$$

the map obtained from T by restriction to $\mathfrak{g}^{-\alpha}$ and projection to $\mathfrak{g}^{\beta}$. Then $T = \sum_{\alpha,\beta} T_{\alpha,\beta}$ and by definition

$$\mathfrak{a}_Z^- = \{X \in \mathfrak{a} \mid (\alpha+\beta)(X) \geq 0,\ \forall \alpha, \beta \text{ with } T_{\alpha,\beta} \neq 0\}.$$

It follows from (3.1) that $\alpha + \beta$ vanishes on $\mathfrak{a}_H$ if $T_{\alpha,\beta} \neq 0$. Hence $\mathfrak{a}_Z^- \subset \mathfrak{a}_Z$. If one denotes by $\mathfrak{a}^- \subset \mathfrak{a}$ the closure of the negative Weyl chamber, then $\mathfrak{a}^- + \mathfrak{a}_H \subset \mathfrak{a}_Z^-$ and by definition Z is wavefront if

$$\mathfrak{a}^- + \mathfrak{a}_H = \mathfrak{a}_Z^-.$$

Let us mention that many real spherical spaces are wavefront; for example all symmetric spaces and all Gross-Prasad type spaces $Z = G \times H/H$ with (G, H) one of the following

$$(\mathrm{GL}_{n+1}(\mathbb{C}), \mathrm{GL}_n(\mathbb{C})),\ (\mathrm{GL}_{n+1}(\mathbb{R}), \mathrm{GL}_n(\mathbb{R})), \tag{3.2}$$
$$(\mathrm{GL}_{n+1}(\mathbb{H}), \mathrm{GL}_n(\mathbb{H})),\ (\mathrm{U}(p+1, q), \mathrm{U}(p, q)), \tag{3.3}$$
$$(\mathrm{SO}(n+1, \mathbb{C}), \mathrm{SO}(n, \mathbb{C})),\ (\mathrm{SO}(p+1, q), \mathrm{SO}(p, q))\,. \tag{3.4}$$

We recall from [18] the polar decomposition for real spherical spaces

$$Z = \Omega A_Z^- F \cdot z_0 \tag{3.5}$$

where

- Ω is a compact set of the type $F'K$ with $F' \subset G$ a finite set.
- $F \subset G$ is a finite set with the property that $F \cdot z_0 = T \cdot z_0 \cap Z$ where $T = \exp(i\mathfrak{a})$ and the intersection is taken in $Z_\mathbb{C} = G_\mathbb{C}/H_\mathbb{C}$.

3.1 Volume growth

Define $\rho_Q \in \mathfrak{a}^*$ by $\rho_Q(X) = \frac{1}{2}\mathrm{tr}(\mathrm{ad}_\mathfrak{u} X)$, $X \in \mathfrak{a}$. It follows from the unimodularity of Z and the local structure theorem that $\rho_Q|_{\mathfrak{a}_H} = 0$, i.e. $\rho_Q \in \mathfrak{a}_Z^* = \mathfrak{a}_H^\perp$.

Lemma 3.2. *Let $Z = G/H$ be a wavefront real spherical space. Then*

$$|B_R| \asymp \sup_{\substack{X \in \mathfrak{a} \\ \|X\| \leq R}} e^{2\rho_Q(X)} = \sup_{\substack{X \in \mathfrak{a}_Z^- \\ \|X\| \leq R}} e^{-2\rho_Q(X)}\,. \tag{3.6}$$

Here the expression $f(R) \asymp g(R)$ signifies that the ratio $\frac{f(R)}{g(R)}$ remains bounded below and above as R tends to infinity.

Proof. First note that the equality in (3.6) is immediate from the wavefront assumption.
Let us first show the lower bound, i.e. there exists a $C > 0$ such that for all $R > 0$ one has

$$|B_R| \geq C \sup_{\substack{X \in \mathfrak{a} \\ \|X\| \leq R}} e^{2\rho_Q(X)}.$$

For that we recall the volume bound from [19], Prop. 4.2: for all compact subsets $B \subset G$ with non-empty interior there exists a constant $C > 0$ such that $\mathrm{vol}_Z(Ba \cdot z_0) \geq Ca^{2\rho_Q}$ for all $a \in A_Z^-$. Together with the polar decomposition (3.5) this gives us the lower bound.
As for the upper bound let

$$\mathfrak{a}_R^- := \{X \in \mathfrak{a}^- \mid \|X\| \leq R\}.$$

Observe that $B_R \subset B_R' := KA_R^-K \cdot z_0$. In the sequel it is convenient to realize A_Z as a subgroup of A (and not as quotient): we identify A_Z with $A_H^\perp \subset A$. The upper bound will follow if we can show that

$$|B_R'| \leq C \sup_{\substack{X \in \mathfrak{a} \\ \|X\| \leq R}} e^{2\rho_Q(X)} \qquad (R > 0).$$

for some constant $C > 0$. This in turn will follow from the argument for the upper bound in the proof of Prop. 4.2 in [19]: in this proof we considered for $a \in A_Z^-$ the map

$$\Phi_a : K \times \Omega_A \times \Xi \to G, \quad (k, b, X) \mapsto kb \exp(\mathrm{Ad}(a)X)$$

where $\Omega_A \subset A$ is a compact neighborhood of $\mathbf{1}$ and $\Xi \subset \mathfrak{h}$ is a compact neighborhood of 0. It was shown that the Jacobian of Φ_a, that is $\sqrt{\det(d\Phi_a d\Phi_a^t)}$, is bounded by $Ca^{-2\rho_Q}$. Now this bounds holds as well for the right K-distorted map

$$\Psi_a : K \times \Omega_A \times K \times \Xi \to G, \quad (k, b, k', X) \mapsto kb \exp(\mathrm{Ad}(ak')X).$$

The reason for that comes from an inspection of the proof; all what is needed is the following fact: let $d := \dim \mathfrak{h}$ and consider the action of $\mathrm{Ad}(a)$ on $V = \bigwedge^d \mathfrak{g}$. Then for $a \in A^-$ we have

$$a^{-2\rho} \geq \sup_{\substack{v \in V, \\ \|v\| = 1}} \langle \mathrm{Ad}(a)v, v \rangle.$$

We deduce an upper bound

$$\mathrm{vol}_Z(K\Omega_A aK \cdot z_0) \leq Ca^{-2\rho}. \tag{3.7}$$

We need to improve that bound from ρ to ρ_Q on the right hand side of (3.7). For that let W_L be the Weyl group of the reductive pair $(\mathfrak{l}, \mathfrak{a})$. Note that $\rho_Q =$

$\frac{1}{|W_L|}\sum_{w\in W_L} w\cdot\rho$. Further, the local structure theorem implies that $L_{\rm n}\subset H$ and hence W_L can be realized as a subgroup of $W_{H\cap K}:=N_{H\cap K}(\mathfrak{a})/Z_{H\cap K}(\mathfrak{a})$. We choose Ω_A to be invariant under $N_{H\cap K}(\mathfrak{a})$ and observe that $a\in A_Z$ is fixed under $W_{H\cap K}$. Thus using the $N_{H\cap K}(\mathfrak{a})$-symmetry in the a-variable we refine (3.7) to

$$\mathrm{vol}_Z(K\Omega_A aK\cdot z_0)\leq Ca^{-2\rho_Q}\,.$$

The desired bound then follows. $\square$

Corollary 3.3. *Let $Z=G/H$ be a wavefront real spherical space of reductive type. Let $Z\to Z^\star$ be a basic factorization such that $Z^\star$ is not compact. Then the geometric balls B_R factorize well to $Z^\star$.*

Proof. As $Z\to Z^\star$ is basic we may assume (ignoring connected components) that $H^\star=G_IH$ for some I. Note that $\mathcal{F}=H^\star/H\simeq G_I/G_I\cap H$ is real spherical.

Let Q be the Z-adapted parabolic subgroup attached to P. Let $P_I=P\cap G_I$ and $G_I\supset Q_I\supset P_I$ be the $\mathcal{F}$-adapted parabolic above P_I and note that $Q_I=Q\cap G_I$. With Lemma 3.2 we then get

$$|B_R^{\mathcal{F}}|\asymp\sup_{\substack{X\in\mathfrak{a}_I\\ \|X\|\leq R}} e^{2\rho_{Q_I}(X)}\,,$$

which we are going to compare with (3.6).

Let $\mathfrak{u}_I$ be the Lie algebra of the unipotent radical of Q_I. Note that $\mathfrak{u}_I\subset\mathfrak{u}$ and that this inclusion is strict since $G/H^\star$ is not compact. The corollary now follows from (2.7). $\square$

3.2 Property I

We briefly recall some results from [19].

Let $(\pi,\mathcal{H}_\pi)$ be a unitary irreducible representation of G. We denote by $\mathcal{H}_\pi^\infty$ the G-Fréchet module of smooth vectors and by $\mathcal{H}_\pi^{-\infty}$ its strong dual. One calls $\mathcal{H}_\pi^{-\infty}$ the G-module of distribution vectors; it is a DNF-space with continuous G-action.

Let $\eta\in(\mathcal{H}_\pi^{-\infty})^H$ be an H-fixed element and $H_\eta<G$ the stabilizer of η. Note that $H<H_\eta$ and set $Z_\eta:=G/H_\eta$. With regard to η and $v\in\mathcal{H}^\infty$ we form the generalized matrix-coefficient

$$m_{v,\eta}(gH):=\eta(\pi(g^{-1})v)\qquad(g\in G)$$

which is a smooth function on Z_η.

We recall the following facts from [19] Thm. 7.6 and Prop. 7.7:

Proposition 3.4. *Let Z be a wavefront real spherical space of reductive type. Then the following assertions hold:*

1. *Every generalized matrix coefficient $m_{v,\eta}$ as above is bounded.*

2. Let $H < H^\star < G$ be a closed subgroup such that $Z^\star$ is unimodular. Then $Z^\star$ is a weakly basic factorization.

3. Let $(\pi, \mathcal{H})$ be a unitary irreducible representation of G and let $\eta \in (\mathcal{H}_\pi^{-\infty})^H$. Then:

(a) $Z \to Z_\eta$ is a weakly basic factorization.

(b) Z_η is unimodular and there exists $1 \leq p < \infty$ such that $m_{v,\eta} \in L^p(Z_\eta)$ for all $v \in \mathcal{H}_\pi^\infty$.

The property of $Z = G/H$ that (3b) is valid for all π and η as above is denoted *Property (I)* in [19]. Note that (1) and (3b) together imply $m_{v,\eta} \in L^q(Z_\eta)$ for $q > p$. Assuming Property (I) we can then make the following notation.

Definition 3.5. *Given π as above, define $p_H(\pi)$ as the smallest index ≥ 1 such that all K-finite generalized matrix coefficients $m_{v,\eta}$ with $\eta \in (\mathcal{H}_\pi^{-\infty})^H$ belong to $L^p(Z_\eta)$ for any $p > p_H(\pi)$.*

Notice that $m_{v,\eta}$ belongs to $L^p(Z_\eta)$ for all K-finite vectors v once that this is the case for some non-trivial such vector v, see [19] Lemma 7.2. For example, this could be the trivial K-type, if it exists in π.
It follows from finite dimensionality of $(\mathcal{H}_\pi^{-\infty})^H$ (see [23]) that $p_H(\pi) < \infty$. We say that π is H-tempered if $p_H(\pi) = 2$.
The representation π is said to be H-distinguished if $(\mathcal{H}_\pi^{-\infty})^H \neq \{0\}$. Note that if π is not H-distinguished then $p_H(\pi) = 1$.

4 Lattice point counting: setup

Let G/H be a real algebraic homogeneous space. We further assume that we are given a lattice (a discrete subgroup with finite covolume) $\Gamma \subset G$, such that $\Gamma_H := \Gamma \cap H$ is a lattice in H. We normalize Haar measures on G and H such that:

- $\mathrm{vol}(G/\Gamma) = 1$.
- $\mathrm{vol}(H/\Gamma_H) = 1$.

Our concern is with the double fibration

$$\begin{tikzcd} & G/\Gamma_H \arrow[dl] \arrow[dr] & \\ Z := G/H & & Y := G/\Gamma \end{tikzcd}$$

Fibre-wise integration yields transfer maps from functions on Z to functions on Y and vice versa. In more precision,

$$L^\infty(Y) \to L^\infty(Z),\ \phi \mapsto \phi^H;\ \phi^H(gH) := \int_{H/\Gamma_H} \phi(gh\Gamma)\ d(h\Gamma_H) \tag{4.1}$$

and we record that this map is contractive, i.e

$$\|\phi^H\|_\infty \leq \|\phi\|_\infty \qquad (\phi \in L^\infty(Y)) . \tag{4.2}$$

Likewise we have

$$L^1(Z) \to L^1(Y),\ f \mapsto f^\Gamma;\ f^\Gamma(g\Gamma) := \sum_{\gamma \in \Gamma/\Gamma_H} f(g\gamma H) , \tag{4.3}$$

which is contractive, i.e

$$\|f^\Gamma\|_1 \leq \|f\|_1 \qquad (f \in L^1(Z)) . \tag{4.4}$$

Unfolding with respect to the double fibration yields, in view of our normalization of measures, the following adjointness relation:

$$\langle f^\Gamma, \phi\rangle_{L^2(Y)} = \langle f, \phi^H\rangle_{L^2(Z)} \tag{4.5}$$

for all $\phi \in L^\infty(Y)$ and $f \in L^1(Z)$. Let us note that (4.5) applied to $|f|$ and $\phi = \mathbf{1}_Y$ readily yields (4.4).

We write $\mathbf{1}_R \in L^1(Z)$ for the characteristic function of B_R and deduce from the definitions and (4.5):

- $\mathbf{1}_R^\Gamma(e\Gamma) = N_R(\Gamma, Z) := \#\{\gamma \in \Gamma/\Gamma_H \mid \gamma \cdot z_0 \in B_R\}$.
- $\|\mathbf{1}_R^\Gamma\|_{L^1(G/\Gamma)} = |B_R|$.

4.1 Weak asymptotics

In the above setup, G/H need not be of reductive type, but we shall assume this again from now on. For spaces with property (I) and Y compact we prove analytically in the following section that

$$N_R(\Gamma, Z) \sim |B_R| \quad (R \to \infty) . \tag{MT}$$

For that we will use the following result of [21]:

Theorem 4.1. *Let $Z = G/H$ be of reductive type. The smooth vectors for the regular representation of G on $L^p(Z)$ vanish at infinity, for all $1 \leq p < \infty$.*

With notation from (4.3) we set

$$F_R^\Gamma := \frac{1}{|B_R|}\mathbf{1}_R^\Gamma.$$

We shall concentrate on verifying the following limit of weak type:

$$\langle F_R^\Gamma, \phi\rangle_{L^2(Y)} \to \int_Y \bar{\phi}\, d\mu_Y \quad (R \to \infty), \qquad (\forall \phi \in C_0(Y)) . \tag{wMT}$$

Here C_0 indicates functions vanishing at infinity.

Lemma 4.2. (wMT) $\Rightarrow$ (MT).

Proof. As in [10] Lemma 2.3 this is deduced from Lemma 2.1 and Lemma 2.2. ∎

5 Main term counting

In this section we will establish main term counting under the mandate of property (I) and Y being compact. Let us call a family of balls $(B_R)_{R>0}$ *well factorizable* if it factorizes well to all proper factorizations of type $Z \to Z_\eta$.

5.1 Main theorem on counting

Theorem 5.1. *Let G be semi-simple and H a closed reductive subgroup. Suppose that Y is compact and Z admits* (I). *If $(B_R)_{R>0}$ is well factorizable, then* (wMT) *and* (MT) *hold.*

Remark 5.2. In case $Z = G/H$ is real spherical and wavefront, then Z has (I) by Proposition 3.4. If we assume in addition that G has no compact factors and that all proper factorizations are basic, then the family of geometric balls is well factorizable by Corollary 3.3. In particular, Theorem A of the introduction then follows from the above.

The proof is based on the following proposition. For a function space $\mathcal{F}(Y)$ consisting of integrable functions on Y we denote by $\mathcal{F}(Y)_{\rm van}$ the subspace of functions with vanishing integral over Y.

Proposition 5.3. *Let $Z = G/H$ be of reductive type. Assume that there exists a dense subspace $\mathcal{A}(Y) \subset C_b(Y)^K_{\rm van}$ such that*

$$\phi^H \in C_0(Z) \quad \textit{for all } \phi \in \mathcal{A}(Y)\,. \tag{5.1}$$

Then (wMT) *holds true.*

Proof. We will establish (wMT) for $\phi \in C_b(Y)$. As

$$C_b(Y) = C_b(Y)_{\rm van} \oplus \mathbb{C}\mathbf{1}_Y,$$

and (wMT) is trivial for ϕ a constant, it suffices to establish

$$\langle F_R^\Gamma, \phi\rangle_{L^2(Y)} \to 0 \qquad (\phi \in C_b(Y)_{\rm van})\,. \tag{5.2}$$

We will show (5.2) is valid for $\phi \in \mathcal{A}(Y)$. By density, as F_R^Γ is K-invariant and belongs to $L^1(Y)$, this will finish the proof.
Let $\phi \in \mathcal{A}(Y)$ and let $\epsilon > 0$. By the unfolding identity (4.5) we have

$$\langle F_R^\Gamma, \phi\rangle_{L^2(Y)} = \frac{1}{|B_R|}\langle \mathbf{1}_R, \phi^H\rangle_{L^2(Z)}. \tag{5.3}$$

Using (5.1) we choose $K_\epsilon \subset Z$ compact such that $|\phi^H(z)| < \epsilon$ outside of K_ϵ. Then

$$\frac{1}{|B_R|}\langle \mathbf{1}_R, \phi^H\rangle_{L^2(Z)} = \int_{K_\epsilon} + \int_{Z-K_\epsilon} \frac{\mathbf{1}_R(z)}{|B_R|}\phi^H(z)\, d\mu_Z(z)\,.$$

By (4.2), the first term is bounded by $\frac{|K_\epsilon|\|\phi\|_\infty}{|B_R|}$, which is $\leq \epsilon$ for R sufficiently large. As the second term is bounded by ϵ for all R, we obtain (5.2). Hence (wMT) holds. □

Remark 5.4. It is possible to replace (5.1) by a weaker requirement: Suppose that an algebraic sum

$$\mathcal{A}(Y) = \sum_{j\in J} \mathcal{A}(Y)_j \tag{5.4}$$

is given together with a factorization $Z_j^\star = G/H_j^\star$ for each $j \in J$. Suppose that the balls B_R all factorize well to $Z_j^\star$, $j \in J$. Suppose further that ϕ^H factorizes to a function

$$\phi^{H_j^\star} \in C_0(Z_j^\star) \tag{5.5}$$

for all $\phi \in \mathcal{A}(Y)_j$ and all $j \in J$. Then the conclusion in Proposition 5.3 is still valid. In fact, using (2.5) the last part of the proof modifies to:

$$\begin{aligned}\frac{1}{|B_R|}\langle \mathbf{1}_R, \phi^H\rangle_{L^2(Z)} = \frac{1}{|B_R|}\langle \mathbf{1}_R^{\mathcal{F}}, \phi^{H_j^\star}\rangle_{L^2(Z_j^\star)} = \\ = \int_{K_\epsilon^\star} + \int_{Z_j^\star - K_\epsilon^\star} \frac{\mathbf{1}_R^{\mathcal{F}}(z)}{|B_R|}\phi^{H_j^\star}(z)\, d\mu_{Z_j^\star}(z)\end{aligned}$$

for $\phi \in \mathcal{A}(Y)_j$. As $\|1_R^{\mathcal{F}}\|_{L^1(Z_j^\star)} = |B_R|$, the second term is bounded by ϵ for all R. As the balls factorize well to $Z_j^\star$ we get the first term as small as we wish with (2.6).

5.2 The space $\mathcal{A}(Y)$

We now construct a specific subspace $\mathcal{A}(Y) \subset C_b(Y)_{\text{van}}^K$ and verify condition (5.5).
Denote by $\widehat{G}_s \subset \widehat{G}$ the K-spherical unitary dual.
As Y is compact, the abstract Plancherel-theorem implies:

$$L^2(G/\Gamma)^K \simeq \bigoplus_{\pi\in\widehat{G}_s} (\mathcal{H}_\pi^{-\infty})^\Gamma.$$

If we denote the Fourier transform by $f \mapsto f^\wedge$ then the corresponding inversion formula is given by

$$f = \sum_\pi a_{v_\pi, f^\wedge(\pi)}. \tag{5.6}$$

Here $a_{v_\pi, f^\wedge(\pi)}$ denotes a matrix coefficient for Y with $v_\pi \in \mathcal{H}_\pi$ normalized K-fixed and $f^\wedge(\pi) \in (\mathcal{H}_\pi^{-\infty})^\Gamma$, and the sum in (5.6) is required to include multiplicities. The matrix coefficients for Y are defined as in (1.2), that is

$$a_{v,\nu}(y) = \nu(g^{-1}\cdot v) \qquad (y = gH \in Y)\,. \tag{5.7}$$

for $v \in \mathcal{H}_\pi$ and $\nu \in (\mathcal{H}_\pi^{-\infty})^\Gamma$.
Note that $L^2(Y) = L^2(Y)_{\text{van}} \oplus \mathbb{C}\cdot \mathbf{1}_Y$. We define $\mathcal{A}(Y) \subset L^2(Y)_{\text{van}}^K$ to be the dense subspace of functions with finite Fourier support, that is,

$$\mathcal{A}(Y) = \text{span}\{a_{v,\nu} \mid \pi \in \widehat{G}_s \text{ non-trivial}, v \in \mathcal{H}_\pi^K, \nu \in (\mathcal{H}_\pi^{-\infty})^\Gamma\}.$$

Then $\mathcal{A}(Y) \subset L^2(Y)^{K,\infty}_{\rm van}$ is dense and since $C^\infty(Y)$ and $L^2(Y)^\infty$ are topologically isomorphic, it follows that $\mathcal{A}(Y)$ is dense in $C(Y)^K_{\rm van}$ as required.
The following lemma together with Remark 5.4 immediately implies Theorem 5.1.

Lemma 5.5. *Assume that Y is compact and Z has* (I)*, and define $\mathcal{A}(Y)$ as above. Then there exists a decomposition of $\mathcal{A}(Y)$ satisfying* (5.4)-(5.5)*.*

Proof. The map $\phi \mapsto \phi^H$ from (4.1) corresponds on the spectral side to a map $(\mathcal{H}_\pi^{-\infty})^\Gamma \to (\mathcal{H}_\pi^{-\infty})^H$, which can be constructed as follows.
As H/Γ_H is compact, we can define for each $\pi \in \widehat{G}_s$

$$\Lambda_\pi : (\mathcal{H}_\pi^{-\infty})^\Gamma \to (\mathcal{H}_\pi^{-\infty})^H, \;\; \Lambda_\pi(\nu) = \int_{H/\Gamma_H} \nu \circ \pi(h^{-1})\; d(h\Gamma_H) \tag{5.8}$$

by $\mathcal{H}_\pi^{-\infty}$-valued integration: the defining integral is understood as integration over a compact fundamental domain $F \subset H$ with respect to the Haar measure on H; as the integrand is continuous and $\mathcal{H}_\pi^{-\infty}$ is a complete locally convex space, the integral converges in $\mathcal{H}_\pi^{-\infty}$. It follows from (5.8) that $(a_{v,\nu})^H = m_{v,\Lambda_\pi(\nu)}$ for all $v \in \mathcal{H}_\pi^\infty$ and $\nu \in (\mathcal{H}_\pi^{-\infty})^\Gamma$.
Let J denote the set of all factorizations $Z^\star \to Z$, including also $Z^\star = Z$ which we give the index $j_0 \in J$. For $j \in J$ we define $\mathcal{A}(Y)_j \subset \mathcal{A}(Y)$ accordingly to be spanned by the matrix coefficients $a_{v,\nu}$ for which $H_{\Lambda_\pi(\nu)} = H_j^\star$. Then (5.4) holds.
Let $\phi \in \mathcal{A}(Y)_{j_0}$, then it follows from (5.6) that

$$\phi^H = \sum_{\pi \neq \mathbf{1}} m_{v_\pi, \Lambda_\pi(\phi^\wedge(\pi))}\,. \tag{5.9}$$

Note that $H_\eta = H$ for each distribution vector $\eta = \Lambda_\pi(\phi^\wedge(\pi))$ in this sum, by the definition of $\mathcal{A}(Y)_{j_0}$. As Z has property (I) the summand $m_{v_\pi,\Lambda_\pi(\phi^\wedge(\pi))}$ is contained in $L^p(G/H)$ for $p > p_H(\pi)$, and by [19], Lemma 7.2, this containment is then valid for all K-finite generalized matrix coefficients $m_{v,\Lambda_\pi(\phi^\wedge(\pi))}$ of π. Thus $m_{v_\pi,\Lambda_\pi(\phi^\wedge(\pi))}$ generates a Harish-Chandra module inside $L^p(G/H)$. As $m_{v_\pi,\Lambda_\pi(\phi^\wedge(\pi))}$ is K-finite, we conclude that it is a smooth vector. Hence $\phi^H \in L^p(G/H)^\infty$, and in view of Theorem 4.1 we obtain (5.1).
The proof of (5.5) for $\phi \in \mathcal{A}(Y)_j$ for general $j \in J$ is obtained by the same reasoning, where one replaces H by $H_j^\star$ in (5.8) and (5.9). $\square$

This concludes the proof of Theorem 5.1.

6 L^p-bounds for generalized matrix coefficients

¿From here on we assume that $Z = G/H$ is wavefront and real spherical. Recall that we assumed that G is semi-simple and that we wrote $\mathfrak{g} = \mathfrak{g}_1 \oplus \ldots \oplus \mathfrak{g}_m$ for the decomposition of $\mathfrak{g}$ into simple factors. It is no big loss of generality to

assume that $G = G_1 \times \ldots \times G_m$ splits accordingly. We will assume that from now on.
Further we request that the lattice $\Gamma < G$ is irreducible, that is, the projection of Γ to any normal subgroup $J \subsetneq G$ is dense in J.
Let π be an irreducible unitary representation of G. Then $\pi = \pi_1 \otimes \ldots \otimes \pi_m$ with π_j and irreducible unitary representation of G_j. We start with a simple observation.

Lemma 6.1. *Let $(\pi, \mathcal{H})$ be an irreducible unitary representation of G and $0 \neq \nu \in (\mathcal{H}^{-\infty})^\Gamma$. If one constituent π_j of π is trivial, then π is trivial.*

Proof. The element ν gives rise to a G-equivariant injection

$$\mathcal{H}^\infty \hookrightarrow C^\infty(Y), \quad v \mapsto (g\Gamma \mapsto \nu(\pi(g^{-1})v)) . \tag{6.1}$$

Say π_j is trivial and let $J := \prod_{\substack{i=1 \\ i \neq j}}^m G_i$. Let Γ_J be the projection of Γ to J. Then (6.1) gives rise to a J-equivariant injection $\mathcal{H}^\infty \hookrightarrow C^\infty(J/\Gamma_J)$. As Γ_J is dense in J, the assertion follows. □

We assume from now on that the cycle $H/\Gamma_H \subset Y$ is compact. This technical condition ensures that the vector valued average map (5.8) converges.

Lemma 6.2. *Let $(\pi, \mathcal{H})$ be a non-trivial irreducible unitary representation of G. Let $\nu \in (\mathcal{H}_\pi^{-\infty})^\Gamma$ such that $\eta := \Lambda_\pi(\nu) \in (\mathcal{H}_\pi^{-\infty})^H$ is non-zero. Then H_η/H is compact.*

Proof. Recall from Proposition 3.4 that $Z \to Z_\eta$ is weakly basic, and from Lemma 2.8 that then there exists $H \subset H_b \subset H_\eta$ such that H_η/H_b is compact and $Z \to Z_b$ is basic. Hence $\mathfrak{h}_b = \mathfrak{h}_I$ for some I. As π is irreducible it infinitesimally embeds into $C^\infty(Z_\eta)$ and hence also to $C^\infty(Z_b)$ on which G_i acts trivially for $i \in I$. It follows that π_i is trivial for $i \in I$. Hence Lemma 6.1 implies $I = \emptyset$ and thus $\mathfrak{h}_b = \mathfrak{h}$. □

In the sequel we use the Plancherel theorem (see [15])

$$L^2(G/\Gamma)^K \simeq \int_{\widehat{G}_s}^{\oplus} \mathcal{V}_{\pi,\Gamma} \, d\mu(\pi) ,$$

where $\mathcal{V}_{\pi,\Gamma} \subset (\mathcal{H}_\pi^{-\infty})^\Gamma$ is a finite dimensional subspace and of constant dimension on each connected component in the continuous spectrum (parametrization by Eisenstein series), and where the Plancherel measure μ has support

$$\widehat{G}_{\Gamma,s} := \operatorname{supp}(\mu) \subset \widehat{G}_s .$$

Given an irreducible lattice $\Gamma \subset G$ we define (cf. Definition 3.5)

$$p_H(\Gamma) := \sup\{p_H(\pi) : \pi \in \widehat{G}_{\Gamma,s}\} \tag{6.2}$$

and record the following.

Lemma 6.3. *Assume that $G = G_1 \times \ldots \times G_m$ with all $\mathfrak{g}_i$ simple and non-compact. Then $p_H(\Gamma) < \infty$.*

Proof. For a unitary representation $(\pi, \mathcal{H})$ and vectors $v, w \in \mathcal{H}$ we form the matrix coefficient $\pi_{v,w}(g) := \langle \pi(g)v, w\rangle$. We first claim that there exists a $p < \infty$ (in general depending on Γ) such that for all non-trivial $\pi \in \widehat{G}_{\Gamma,s}$ one has $\pi_{v,w} \in L^p(G)$ for all K-finite vectors v, w. In case G has property (T) this follows (independently of Γ) from [7]. The remaining cases contain at least one factor G_i of $\mathrm{SO}_e(n,1)$ or $\mathrm{SU}(n,1)$ (up to covering) and have no compact factors by assumption. They are treated in [6].
The claim can be interpreted geometrically via the leading exponent $\Lambda_V \in \mathfrak{a}^*$ which is attached to the Harish-Chandra module of $\mathcal{H}$ (see [19], Section 6). The lemma now follows from Prop. 4.2 and Thm. 6.3 in [19] (see the proof of Thm. 7.6 in [19] how these two facts combine to result in integrability). $\square$

Let $1 \leq p < \infty$. Let us say that a subset $\Lambda \subset \widehat{G}_s$ is *L^p-bounded* provided that $m_{v,\eta} \in L^p(Z_\eta)$ for all $\pi \in \Lambda$ and $v \in \mathcal{H}_\pi^\infty$, $\eta \in (\mathcal{H}_\pi^{-\infty})^H$. By definition we thus have that $\widehat{G}_{\Gamma,s}$ is L^p-bounded for $p > p_H(\Gamma)$.
In this section we work under the following:

Hypothesis A: *For every $1 \leq p < \infty$ and every L^p-bounded subset $\Lambda \subset \widehat{G}_s$ there exists a compact subset $\Omega \subset G$ and constants $c, C > 0$ such that the following assertions hold for all $\pi \in \Lambda$, $\eta \in (\mathcal{H}_\pi^{-\infty})^H$ and $v \in \mathcal{H}_\pi^K$:*

$$\|m_{v,\eta}\|_{L^p(Z_\eta)} \leq C\|m_{v,\eta}\|_\infty\,, \tag{A1}$$

$$\|m_{v,\eta}\|_\infty \leq c\|m_{v,\eta}\|_{\infty,\Omega_\eta} \tag{A2}$$

where $\Omega_\eta = \Omega H_\eta/H_\eta$. Here $\|\cdot\|_{\infty,\omega}$ denotes the supremum norm taken on the subset ω.

In the sequel we are only interested in the following choice of subset $\Lambda \subset \widehat{G}_s$, namely

$$\Lambda := \{\pi \in \widehat{G}_{\Gamma,s} \mid \Lambda_\pi(\nu) \neq 0 \text{ for some } \nu \in \mathcal{V}_{\pi,\Gamma}\}\,. \tag{6.3}$$

An immediate consequence of Hypothesis A is:

Lemma 6.4. *Assume that $p > p_H(\Gamma)$. Then there is a $C > 0$ such that for all $\pi \in \widehat{G}_{\Gamma,s}$, $v \in \mathcal{H}_\pi^K$, $\nu \in (\mathcal{H}_\pi^{-\infty})^\Gamma$ and $\eta := \Lambda_\pi(\nu) \in (\mathcal{H}_\pi^{-\infty})^H$ one has*

$$\|\phi_\pi^H\|_{L^p(Z_\eta)} \leq C\|\phi_\pi\|_\infty$$

where $\phi_\pi(g\Gamma) := \nu(\pi(g^{-1})v)$.

Proof. Recall from (4.2), that integration is a bounded operator from $L^\infty(Y) \to L^\infty(Z)$. Hence the assertion follows from (A1). $\square$

Recall the Cartan-Killing form κ on $\mathfrak{g} = \mathfrak{k} + \mathfrak{s}$ and choose a basis $X_1, \ldots, X_l$ of $\mathfrak{k}$ and $X_1', \ldots, X_s'$ of $\mathfrak{s}$ such that $\kappa(X_i, X_j) = -\delta_{ij}$ and $\kappa(X_i', X_j') = \delta_{ij}$. With that data we form the standard Casimir element

$$\mathcal{C} := -\sum_{j=1}^{l} X_j^2 + \sum_{j=1}^{s} (X_j')^2 \in \mathcal{U}(\mathfrak{g}).$$

Set $\Delta_K := \sum_{j=1}^{l} X_j^2 \in \mathcal{U}(\mathfrak{k})$ and obtain the commonly used Laplace element

$$\Delta = \mathcal{C} + 2\Delta_K \in \mathcal{U}(\mathfrak{g}) \tag{6.4}$$

which acts on $Y = G/\Gamma$ from the left.
Let $d \in \mathbb{N}$. For $1 \leq p \leq \infty$, it follows from [2], Section 3, that Sobolev norms on $L^p(Y)^\infty \subset C^\infty(Y)$ can be defined by

$$||f||_{p,2d}^2 = \sum_{j=0}^{d} ||\Delta^j f||_p^2 .$$

Basic spectral theory allows one to define $\| \cdot \|_{p,d}$ more generally for any $d \geq 0$.
Let us define

$$s := \dim \mathfrak{s} = \dim G/K = \dim \Gamma \backslash G/K$$

and

$$r := \dim \mathfrak{a} = \operatorname{rank}_{\mathbb{R}}(G/K),$$

where $\mathfrak{a} \subset \mathfrak{s}$ is maximal abelian.
We denote by $C_b(Y)$ the space of continuous bounded functions on Y and by $C_b(Y)_{\mathrm{van}}$ the subspace with vanishing integral.

Proposition 6.5. *Assume that*

1. *Z is a wavefront real spherical space,*
2. *$G = G_1 \times \ldots \times G_m$ with all $\mathfrak{g}_i$ simple and non-compact.*
3. *$\Gamma < G$ is irreducible and Y_H is compact,*
4. *Hypothesis A is valid.*

Let $p > p_H(\Gamma)$. Then the map

$$\mathrm{Av_H} : C_b^\infty(Y)_{\mathrm{van}}^K \to L^p(Z)^K ; \mathrm{Av_H}(\phi) = \phi^H$$

is continuous. More precisely, for all

1. *$k > s + 1$ if Y is compact.*
2. *$k > \frac{r+1}{2} s + 1$ if Y is non-compact and Γ is arithmetic*

there exists a constant $C = C(p,k) > 0$ *such that*

$$\|\phi^H\|_{L^p(Z)} \le C\|\phi\|_{\infty,k} \qquad (\phi \in C_b^\infty(Y)_{\mathrm{van}}^K)$$

Proof. For all $\pi \in \widehat{G}$ the operator $d\pi(\mathcal{C})$ acts as a scalar λ_π and we set

$$|\pi| := |\lambda_\pi| \ge 0\,.$$

Let $\phi \in C_b^\infty(Y)_{\mathrm{van}}^K$ and write $\phi = \phi_d + \phi_c$ for its decomposition in discrete and continuous Plancherel parts. We assume first that $\phi = \phi_d$.
In case Y is compact we have Weyl's law: There is a constant $c_Y > 0$ such that

$$\sum_{|\pi|\le R} m(\pi) \sim c_Y R^{s/2} \qquad (R \to \infty)\,.$$

Here $m(\pi) = \dim \mathcal{V}_{\pi,\Gamma}$. We conclude that

$$\sum_\pi m(\pi)(1+|\pi|)^{-k} < \infty \tag{6.5}$$

for all $k > s/2+1$. In case Y is non-compact, we let $\widehat{G}_{\mu,d}$ be the the discrete support of the Plancherel measure. Then assuming Γ is arithmetic, the upper bound in [16] reads:

$$\sum_{\substack{\pi\in\widehat{G}_{\mu,d}\\ |\pi|\le R}} m(\pi) \le c_Y R^{rs/2} \qquad (R > 0)\,.$$

For $k > rs/2+1$ we obtain (6.5) as before.
Let $p > p_H(\Gamma)$. As ϕ is in the discrete spectrum we decompose it as $\phi = \sum_\pi \phi_\pi$ and obtain by Lemmas 6.2 and 6.4

$$\|\phi^H\|_p \le \sum_\pi \|\phi_\pi^H\|_p \le C\sum_\pi \|\phi_\pi\|_\infty\,.$$

The last sum we estimate as follows:

$$\begin{aligned}\sum_\pi \|\phi_\pi\|_\infty &= \sum_\pi (1+|\pi|)^{-k/2}(1+|\pi|)^{k/2}\|\phi_\pi\|_\infty\\ &\le C\sum_\pi (1+|\pi|)^{-k/2}\|\phi_\pi\|_{\infty,k}\end{aligned}$$

with $C > 0$ a constant depending only on k (we allow universal positive constants to change from line to line). Applying the Cauchy-Schwartz inequality combined with (6.5) we obtain

$$\|\phi^H\|_p \le C\Big(\sum_\pi \|\phi_\pi\|_{\infty,k}^2\Big)^{\frac{1}{2}}$$

with $C > 0$. With Hypothesis (A2) we get the further improvement:

$$\|\phi^H\|_p \leq C\Big(\sum_\pi \|\phi_\pi\|^2_{\Omega,\infty,k}\Big)^{\frac{1}{2}}$$

where the Sobolev norm is taken only over the compact set Ω.
To finish the proof we apply the Sobolev lemma on $K\backslash G$. Here Sobolev norms are defined by the central operator $\mathcal{C}$, whose action agrees with the left action of Δ. It follows that $\|f\|_{\infty,\Omega} \leq C\|f\|_{2,k_1,\Omega}$ with $k_1 > \frac{s}{2}$ for K-invariant functions f on G. This gives

$$\|\phi^H\|_p \leq C(\sum_\pi ||\phi_\pi||^2_{\Omega,2,k+k_1})^{\frac{1}{2}} = C||\phi||_{\Omega,2,k+k_1} \leq C||\phi||_{\infty,k+k_1}$$

which proves the proposition for the discrete spectrum.
If $\phi = \phi_c$ belongs to the continuous spectrum, where multiplicities are bounded (see [15]), the proof is simpler. Let μ_c be the restriction of the Plancherel measure to the continuous spectrum. As this is just Euclidean measure on r-dimensional space we have

$$\int_{\widehat{G}_s} (1+|\pi|)^{-k}\ d\mu_c(\pi) < \infty \tag{6.6}$$

if $k > r/2$. We assume for simplicity in what follows that $m(\pi) = 1$ for almost all $\pi \in \operatorname{supp}\mu_c$. As $\sup_{\pi\in\operatorname{supp}\mu_c} m(\pi) < \infty$ the proof is easily adapted to the general case.
Let

$$\phi = \int_{\widehat{G}_s} \phi_\pi\ d\mu_c(\pi).$$

As $\|\phi^H\|_\infty \leq \|\phi\|_\infty$ we conclude with Lemma 6.4, (6.6) and Fubini's theorem that

$$\phi^H = \int_{\widehat{G}_s} \phi^H_\pi\ d\mu_c(\pi)$$

and, by the similar chain of inequalities as in the discrete case

$$\|\phi^H\|_p \leq C\|\phi\|_{\infty,k+k_1}$$

with $k > \frac{r}{2}$ and $k_1 > \frac{s}{2}$. This concludes the proof. □

7 Error term estimates

Recall $\mathbf{1}_R$, the characteristic function of B_R. The first error term for the lattice counting problem can be expressed by

$$\operatorname{err}(R,\Gamma) := \sup_{\substack{\phi\in C_b(Y)\\ \|\phi\|_\infty\leq 1}} \left|\left\langle \frac{\mathbf{1}^\Gamma_R}{|B_R|} - \mathbf{1}_Y, \phi\right\rangle\right| \qquad (R>0),$$

and our goal is to give an upper bound for $\operatorname{err}(R,\Gamma)$ as a function of R. According to the decomposition $C_b(Y) = C_b(Y)_{\mathrm{van}} \oplus \mathbb{C}\mathbf{1}_Y$ we decompose functions as $\phi = \phi_o + \phi_1$ and obtain

$$\operatorname{err}(R,\Gamma) = \sup_{\substack{\phi \in C_b(Y) \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_R^\Gamma, \phi_o\rangle|}{|B_R|} = \sup_{\substack{\phi \in C_b(Y) \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_R, \phi_o^H\rangle|}{|B_R|}.$$

Further, from $\|\phi_o\|_\infty \leq 2\|\phi\|_\infty$ we obtain that $\operatorname{err}(R,\Gamma) \leq 2\operatorname{err}_1(R,\Gamma)$ with

$$\operatorname{err}_1(R,\Gamma) := \sup_{\substack{\phi \in C_b(Y)_{\mathrm{van}} \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_R^\Gamma, \phi\rangle|}{|B_R|} = \sup_{\substack{\phi \in C_b(Y)_{\mathrm{van}} \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_R, \phi^H\rangle|}{|B_R|}.$$

7.1 Smooth versus non-smooth counting

Like in the classical Gauss circle problem one obtains much better estimates for the remainder term if one uses a smooth cutoff. Let $\alpha \in C_c^\infty(G)$ be a non-negative test function with normalized integral. Set $\mathbf{1}_{R,\alpha} := \alpha * \mathbf{1}_R$ and define

$$\operatorname{err}_\alpha(R,\Gamma) := \sup_{\substack{\phi \in C_b(Y)_o^K \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_{R,\alpha}^\Gamma, \phi\rangle|}{|B_R|} = \sup_{\substack{\phi \in C_b(Y)_o^K \\ \|\phi\|_\infty \leq 1}} \frac{|\langle \mathbf{1}_{R,\alpha}, \phi^H\rangle|}{|B_R|}.$$

Lemma 7.1. *Let $k > s+1$ if Y is compact and $k > \frac{r+1}{2}s + 1$ otherwise. Let $p > p_H(\Gamma)$ and q be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then there exists $C > 0$ such that*

$$\operatorname{err}_\alpha(R,\Gamma) \leq C\|\alpha\|_{1,k}|B_R|^{-\frac{1}{p}} \tag{7.1}$$

for all $R \geq 1$ and all $\alpha \in C_c^\infty(G)$.

Proof. First note that

$$\langle \mathbf{1}_{R,\alpha}, \phi^H\rangle = \langle \mathbf{1}_{R,\alpha}, (-\mathbf{1}+\Delta)^{k/2}(-\mathbf{1}+\Delta)^{-k/2}\phi^H\rangle.$$

With $\psi = (-\mathbf{1}+\Delta)^{-k/2}\phi$ we have $\|\psi\|_{\infty,k} \leq C\|\phi\|_\infty$ for some $C > 0$. We thus obtain

$$\begin{aligned} \operatorname{err}_\alpha(R,\Gamma) \quad &\leq C \quad \sup_{\substack{\psi \in C_b(Y)_o^K \\ \|\psi\|_{\infty,k} \leq 1}} \frac{|\langle \mathbf{1}_{R,\alpha}, (-\mathbf{1}+\Delta)^{k/2}\psi^H\rangle|}{|B_R|} \\ &\leq \quad \frac{C}{|B_R|} \sup_{\substack{\psi \in C_b(Y)_o^K \\ \|\psi\|_{\infty,k} \leq 1}} |\langle \mathbf{1}_{R,\alpha}, (-\mathbf{1}+\Delta)^{k/2}\psi^H\rangle| \end{aligned}$$

Moving $(-\mathbf{1}+\Delta)^{k/2}$ to the other side we get with Hölder's inequality and Proposition 6.5 that

$$\operatorname{err}_\alpha(R,\Gamma) \leq \frac{C}{|B_R|}\|(-\mathbf{1}+\Delta)^{k/2}\alpha * \mathbf{1}_R\|_q.$$

Finally,

$$\|(-\mathbf{1}+\Delta)^{k/2}\alpha * \mathbf{1}_R\|_q \leq C\|\alpha\|_{1,k}\|\mathbf{1}_R\|_q$$

and with $\|\mathbf{1}_R\|_q = |B_R|^{\frac{1}{q}}$, the lemma follows. $\square$

Remark 7.2. In the literature results are sometimes stated not with respect to $\mathrm{err}(R,\Gamma)$ but the pointwise error term $\mathrm{err}_{pt}(R,\Gamma) = |\mathbf{1}_R^\Gamma(\mathbf{1}) - |B_R||$. Likewise we define $\mathrm{err}_{pt,\alpha}(R,\Gamma)$. Let B_Y be a compact neighborhood of $\mathbf{1}\Gamma \in Y$ and note that

$$\mathrm{err}_{pt,\alpha}(R,\Gamma) \leq |B_R| \sup_{\substack{\phi\in L^1(B_Y)\\ \|\phi\|_1\leq 1}} |\langle \frac{\mathbf{1}_{R,\alpha}^\Gamma}{|B_R|} - \mathbf{1}_Y, \phi\rangle| \qquad (R>0).$$

The Sobolev estimate $\|\phi\|_\infty \leq C\|\phi\|_{1,k}$, for K-invariant functions ϕ on B_Y and with $k=\dim Y/K$ the Sobolev shift, then relates these error terms:

$$\mathrm{err}_{pt,\alpha}(R,\Gamma) \leq |B_R| \sup_{\substack{\phi\in C_b^\infty(Y)\\ \|\phi\|_{\infty,-k}\leq 1}} |\langle \frac{\mathbf{1}_R^\Gamma}{|B_R|} - \mathbf{1}_Y, \phi\rangle|\,.$$

We then obtain

$$\mathrm{err}_{pt,\alpha}(R,\Gamma) \leq C|B_R|^{1-\frac{1}{p}} \qquad (R>0)$$

in view of (7.1).

We return to the error bound in Lemma 7.1 and would like to compare $\mathrm{err}_1(R,\Gamma)$ with $\mathrm{err}_\alpha(R,\Gamma)$. For that we note (by the triangle inequality) that

$$|\,\mathrm{err}_1(R,\Gamma) - \mathrm{err}_\alpha(R,\Gamma)| \leq \sup_{\substack{\phi\in C_b(Y)_o^K\\ \|\phi\|_\infty\leq 1}} \frac{|\langle \mathbf{1}_{R,\alpha}^\Gamma - \mathbf{1}_R^\Gamma, \phi\rangle|}{|B_R|}\,.$$

Suppose that $\operatorname{supp}\alpha \subset B_\epsilon^G$ for some $\epsilon>0$. Then Lemma 2.2 implies that $\mathbf{1}_{R,\alpha}$ is supported in $B_{R+\epsilon}$, and hence

$$\begin{aligned} |\langle \mathbf{1}_{R,\alpha}^\Gamma - \mathbf{1}_R^\Gamma, \phi\rangle| &\leq \|\mathbf{1}_{R,\alpha}^\Gamma - \mathbf{1}_R^\Gamma\|_1 \\ &\leq \|\mathbf{1}_{R,\alpha} - \mathbf{1}_R\|_1 \\ &\leq |B_{R+\epsilon}|^{\frac{1}{2}}\|\mathbf{1}_{R,\alpha} - \mathbf{1}_R\|_2 \\ &\leq |B_{R+\epsilon}|^{\frac{1}{2}}|B_{R+\epsilon}\backslash B_R|^{\frac{1}{2}}\,. \end{aligned}$$

With Lemma 2.1 we get

$$|B_{R+\epsilon}\backslash B_R| \leq C\epsilon|B_R| \qquad (R\geq 1, \epsilon<1)\,.$$

Thus we obtain that

$$|\,\mathrm{err}_1(R,\Gamma) - \mathrm{err}_\alpha(R,\Gamma)| \leq C\epsilon^{\frac{1}{2}}\,.$$

Combining this with the estimate in Lemma 7.1 we arrive at the existence of $C > 0$ such that

$$\mathrm{err}_1(R, \Gamma) \leq C(\epsilon^{-k}|B_R|^{-\frac{1}{p}} + \epsilon^{\frac{1}{2}})$$

for all $R \geq 1$ and all $0 < \epsilon < 1$. The minimum of the function $\epsilon \mapsto \epsilon^{-k}c + \epsilon^{1/2}$ is attained at $\epsilon = (2kc)^{\frac{2}{2k+1}}$ and thus we get:

THEOREM 7.3. *Under the assumptions of Proposition 6.5 the first error term* $\mathrm{err}(R, \Gamma)$ *for the lattice counting problem on* $Z = G/H$ *can be estimated as follows: for all* $p > p_H(\Gamma)$ *and* $k > s + 1$ *for* Y *compact, resp.* $k > \frac{r+1}{2}s + 1$ *otherwise, there exists a constant* $C = C(p, k) > 0$ *such that*

$$\mathrm{err}(R, \Gamma) \leq C|B_R|^{-\frac{1}{(2k+1)p}}$$

for all $R \geq 1$.

REMARK 7.4. The point where we lose essential information is in the estimate (6.5) where we used Weyl's law. In the moment pointwise multiplicity bounds are available the estimate would improve. To compare the results with Selberg on the hyperbolic disc, let us assume that $p_H(\Gamma) = 2$. Then with $r = 1$ and $s = 2$ our bound is $\mathrm{err}(R, \Gamma) \leq C_\epsilon|B_R|^{-\frac{1}{14}+\epsilon}$ while Selberg showed $\mathrm{err}(R, \Gamma) \leq C_\epsilon|B_R|^{-\frac{1}{3}+\epsilon}$.

8 TRIPLE SPACES

In this section we verify our Hypothesis A for triple space $Z = G/H$ where $G = G' \times G' \times G'$, $H = \mathrm{diag}(G')$ and $G' = \mathrm{SO}_e(1, n)$ for some $n \geq 2$. Observe that $\mathrm{SO}_e(1, 2) \cong \mathrm{PSl}(2, \mathbb{R})$. We take $K' := \mathrm{SO}(n, \mathbb{R}) < G'$ as a maximal compact subgroup and set $K := K' \times K' \times K'$. Further we set $\mathfrak{s} := \mathfrak{s}' \times \mathfrak{s}' \times \mathfrak{s}'$. A maximal abelian subspace $\mathfrak{a} \subset \mathfrak{s}$ is then of the form

$$\mathfrak{a} = \mathfrak{a}_1' \times \mathfrak{a}_2' \times \mathfrak{a}_3'$$

with $\mathfrak{a}_i' \subset \mathfrak{s}'$ one dimensional subspaces. We recall the following result from [8].

PROPOSITION 8.1. *For the triple space the following assertion hold true:*

1. $G = KAH$ *if and only if* $\dim(\mathfrak{a}_1' + \mathfrak{a}_2' + \mathfrak{a}_3') = 2$.
2. *Suppose that all* $\mathfrak{a}_i'$ *are pairwise distinct. Then one has* PH *is open for all minimal parabolics* P *with Langlands-decomposition* $P = M_P A_P N_P$ *and* $A_P = A$.

We say that the choice of A is *generic* if all $\mathfrak{a}_i'$ are distinct and $\dim(\mathfrak{a}_1' + \mathfrak{a}_2' + \mathfrak{a}_3') = 2$.

The invariant measure dz on Z can then be estimated as

$$\int_Z f(z)\, dz \leq \int_K \int_A f(ka \cdot z_0) J(a) \ da\ dk \qquad (f \in C_c(Z), f \geq 0)$$

with

$$J(a) = \sup_{w \in W} a^{2w\rho} \tag{8.1}$$

by Lemma 3.2. Note that in this case the Weyl group W is just $\{\pm 1\}^3$.

8.1 Proof of the Hypothesis A

We first note that for all $\pi \in \widehat{G}_s$ the space of H-invariants

$$(\mathcal{H}_\pi^{-\infty})^H = \mathbb{C} I\,.$$

is one-dimensional, see [5], Thm. 3.1.
Write $\pi = \pi_1 \otimes \pi_2 \otimes \pi_3$ with each factor a K'-spherical unitary irreducible representation of G'. If we assume that $\pi \neq \mathbf{1}$ has non-trivial H-fixed distribution vectors, then at least two of the factors π_i are non-trivial.
Let v_i be normalized K'-fixed vectors of π_i and set $v = v_1 \otimes v_2 \otimes v_3$. Since Z is a multiplicity one space, the functional $I \in (\mathcal{H}_\pi^{-\infty})^H$ is unique up to scalars. Our concern is to obtain uniform L^p-bounds for the generalized matrix coefficients $f_\pi := m_{v,I}$:

$$f_\pi(g_1, g_2, g_3) := I(\pi_1(g_1)^{-1} v_1 \otimes \pi_2(g_2)^{-1} v_2 \otimes \pi_3(g_3)^{-1} v_3)\,,$$

when π belongs to the set Λ of (6.3).
We decompose $\Lambda = \Lambda_0 \cup \Lambda_1 \cup \{\mathbf{1}\}$ with $\Lambda_0 \subset \Lambda$ the set of $\pi \in \Lambda$ with all π_i non-trivial, and Λ_1 the set of π's with exactly one π_i to be trivial.
Consider first the case where $\pi \in \Lambda_1$, i.e. one π_i is trivial, say π_3. Then $\pi_2 = \pi_1^*$. We identify $Z \simeq G' \times G'$ via $(g, h) \mapsto (\mathbf{1}, g, h)H$ and obtain

$$f_\pi(g, h) = \langle \pi_1(g) v_1, v_1 \rangle\,,$$

a spherical function. Note that $Z_\eta \simeq G'$ and Hypothesis A follows from standard properties about K'-spherical functions on G'. To be more specific let $G' = N'A'K'$ be an Iwasawa-decomposition with middle-projection $\mathbf{a} : G' \to A'$, then

$$f_\pi(g, h) = \varphi_{\lambda_1}(g) := \int_{K'} \mathbf{a}(k'g)^{\lambda_1 - \rho'}\, dk'\,.$$

We use Harish-Chandra's estimates $|\varphi_\nu(a)| \leq a^\nu \varphi_0(a)$ and $\varphi_0(a) \leq C a^{-\rho}(1 + |\log a|)^d$ for $a \in A'$ in positive chamber. The condition of $\pi \in \Lambda_1$ implies that $\rho - \operatorname{Re}\lambda_1 > 0$ is bounded away from zero and Hypothesis A follows in this case.
Suppose now that $\pi \in \Lambda_0$, i.e. all π_i are non-trivial.
For a simplified exposition we assume that $n = 2$, i.e. $G' = \mathrm{PSl}(2, \mathbb{R})$, and comment at the end for the general case. Then $\pi_i = \pi_{\lambda_i}$ are principal series for some $\lambda_i \in i\mathbb{R}^+ \cup [0, 1)$ with $\mathcal{H}_{\pi_i}^\infty = C^\infty(\mathbb{S}^1)$ in the compact realization. Set $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ and set $\pi = \pi_\lambda$.

In order to analyze f_π we use $G = KAH$ and thus assume that $g = a = (a_1, a_2, a_3) \in A$. We work in the compact model of $\mathcal{H}_{\pi_i} = L^2(\mathbb{S}^1)$ and use the explicit model for I in [3]: for h_1, h_2, h_3 smooth functions on the circle one has

$$I(h_1 \otimes h_2 \otimes h_3) = \frac{1}{(2\pi)^3} \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} h_1(\theta_1) h_2(\theta_2) h_3(\theta_3) \cdot \\ \cdot \mathcal{K}(\theta_1, \theta_2, \theta_3) \, d\theta_1 d\theta_2 d\theta_3 \,,$$

where

$$\mathcal{K}(\theta_1, \theta_2, \theta_3) = |\sin(\theta_2 - \theta_3)|^{(\alpha-1)/2} |\sin(\theta_1 - \theta_3)|^{(\beta-1)/2} |\sin(\theta_1 - \theta_2)|^{(\gamma-1)/2} .$$

In this formula one has $\alpha = \lambda_1 - \lambda_2 - \lambda_3$, $\beta = -\lambda_1 + \lambda_2 - \lambda_3$ and $\gamma = -\lambda_1 - \lambda_2 + \lambda_3$ where $\lambda_i \in i\mathbb{R} \cup (-1, 1)$ are the standard representation parameters of π_i. According to to [5], Cor. 2.1, the kernel $\mathcal{K}$ is absolutely integrable.
Set

$$A' := \left\{ a_t := \begin{pmatrix} t & 0 \\ 0 & \frac{1}{t} \end{pmatrix} \mid t > 0 \right\} < G'$$

Then $A'_i = k_{\phi_i} A' k_{\phi_i}^{-1}$ with $\phi_i \in [0, 2\pi]$ and

$$k_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} .$$

Set $a_{t,i} = k_{\phi_i} a_t k_{\phi_i}^{-1}$.
Returning to our analysis of f_π we now take $h_i(t_i, \theta_i) = [\pi_1(a_{t_i,i}) v_i](\theta_i)$ and remark that

$$h_i(t_i, \theta_i) = \frac{1}{(t_i^2 + \sin^2(\theta_i - \phi_i)(\frac{1}{t_i^2} - t_i^2))^{\frac{1}{2}(1+\lambda_i)}} .$$

Let us set $|\pi| := \pi_{\operatorname{Re}\lambda_1} \otimes \pi_{\operatorname{Re}\lambda_2} \otimes \pi_{\operatorname{Re}\lambda_3}$. Our formulas then show

$$|f_\pi(a)| \leq f_{|\pi|}(a) \qquad (a \in A) . \tag{8.2}$$

Let $c_i := 1 - |\operatorname{Re}\lambda_i|$ for $i = 1, 2, 3$. The fundamental estimate in [22], Thm. 3.2, then yields a constant d, independent of π, and a constant $C = C(\pi) > 0$ such that for $a = (a_{t_1,1}, a_{t_2,2}, a_{t_3,3})$ one has

$$|f_\pi(a)| \leq C \frac{(1 + |\log t_1| + |\log t_2| + |\log t_3|)^d}{[\cosh \log t_1]^{c_1} \cdot [\cosh \log t_2]^{c_2} \cdot [\cosh \log t_3]^{c_3}} . \tag{8.3}$$

In view of (8.2) the constant $C(\pi)$ can be assumed to depend only on the distance of $\operatorname{Re}\lambda_i$ to the trivial representation. Looking at the integral representation of f_π with the kernel $\mathcal{K}$ we deduce a lower bound without the logarithmic factor, i.e. the bound is essentially sharp. Hence (8.1) together with the fact that all f_π for $\pi \in \Lambda_0$ are in $L^p(Z)$ for some $p < \infty$ implies that

$$\inf_{\pi \in \Lambda_0} c_i(\pi) > 0 \,. \tag{8.4}$$

We now claim

$$\sup_{\pi \in \Lambda_0} \|f_\pi\|_p < \infty \,, \tag{8.5}$$

and

$$\sup_{\pi \in \Lambda_0} \|f_\pi\|_\infty < \infty \,. \tag{8.6}$$

For $0 < \epsilon < 1$ set $\Lambda_{\epsilon,\mathbb{R}} = [0, 1-\epsilon] \times [0, 1-\epsilon] \times [0, 1-\epsilon]$ and $\Lambda_\epsilon := i\mathfrak{a}^* + \Lambda_\epsilon$. It follows from (8.4) that there exists an $\epsilon > 0$ such that $\Lambda_0 \subset \Lambda_\epsilon$. We prove the stronger inequalities with Λ_0 replaced by Λ_ϵ. In view of (8.2) and (8.3) we may replace by Λ_ϵ by $\Lambda_{\epsilon,\mathbb{R}}$. Let $\mathcal{E}_\epsilon$ be the eight element set of extreme points of $\Lambda_{\epsilon,\mathbb{R}}$. For fixed $a = a_t$ and $\theta = (\theta_1, \theta_2, \theta_3)$ we let $F_\lambda(a, \theta) = \mathcal{K}(\theta) h_1(t_1, \theta_1) h_2(t_2, \theta_2) h_3(t_3, \theta_3)$ and note that the assignment $\Lambda_{\epsilon,\mathbb{R}} \to \mathbb{R}_+$, $\lambda \mapsto F_\lambda(a, \theta)$ is convex. Therefore we get for all $\lambda \in \Lambda_\epsilon$ that

$$f_\lambda(a) \leq \sum_{\mu \in \mathcal{E}_\epsilon} f_\mu(a) \,.$$

In view of (8.3) the inequalities (8.5) and (8.6) then follow.
On the other hand for $g = \mathbf{1} = (\mathbf{1}, \mathbf{1}, \mathbf{1})$, the value $f_\pi(\mathbf{1})$ is obtained by applying I to the constant function $\mathbf{1} = \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1}$. This value has been computed explicitly by Bernstein and Reznikov in [3] as

$$\frac{\Gamma((\alpha+1)/4)\Gamma((\beta+1)/4)\Gamma((\gamma+1)/4)\Gamma((\delta+1)/4)}{\Gamma((1-\lambda_1)/2)\Gamma((1-\lambda_2)/2)\Gamma((1-\lambda_3)/2)}$$

where α, β, γ are as before and $\delta = -\lambda_1 - \lambda_2 - \lambda_3$. Stirling approximation,

$$|\Gamma(\sigma + it)| = \text{const.} e^{-\frac{\pi}{2}|t|} |t|^{\sigma - \frac{1}{2}} \left(1 + O(|t|^{-1})\right)$$

as $|t| \to \infty$ and σ is bounded, yields a lower bound for $f_\pi(\mathbf{1})$:

$$\inf_{\pi \in \Lambda_0} |f_\pi(\mathbf{1})| > 0 \,. \tag{8.7}$$

As $\|f_\pi\|_\infty \geq |f_\pi(\mathbf{1})|$ the assertion (A1) of Hypothesis A is readily obtained from (8.5) and (8.7). Likewise (A2) with $\Omega = \{\mathbf{1}\}$ follows from (8.6) and (8.7). In general for $G' = \mathrm{SO}_e(1, n)$ one needs to compute the Bernstein-Reznikov integral. This was accomplished in [9].

Theorem 8.2. *Let $Z = G' \times G' \times G' / \operatorname{diag}(G')$ for $G' = \mathrm{SO}_e(1, n)$ and assume that H/Γ_H is compact. Then the first error term* $\operatorname{err}(R, \Gamma)$ *for the lattice counting problem on $Z = G/H$ can be estimated as follows: for all $p > p_H(\Gamma)$ there exists a $C = C(p) > 0$ such that*

$$\operatorname{err}(R, \Gamma) \leq C |B_R|^{-\frac{1}{(6n+3)p}}$$

for all $R \geq 1$.

8.2 Cubic lattices

Here we let $G_0 = \mathrm{SO}_e(1,2)$ with the quadratic Q form defining G_0 having integer coefficients and anisotropic over $\mathbb{Q}$, for example

$$Q(x_0, x_1, x_2) = 2x_0^2 - 3x_1^2 - x_2^2 \,.$$

Then, according to Borel, $\Gamma_0 = G_0(\mathbb{Z})$ is a uniform lattice in G_0.
Next let k be a cubic Galois extension of $\mathbb{Q}$. Note that k is totally real. An example of k is the splitting field of the polynomial $f(x) = x^3 + x^2 - 2x - 1$. Let σ be a generator of the Galois group of $k|\mathbb{Q}$. Let $\mathcal{O}_k$ be the ring of algebraic integers of k. We define $\Gamma < G = G_0^3$ to be the image of $G_0(\mathcal{O}_k)$ under the embedding

$$G_0(\mathcal{O}_k) \ni \gamma \mapsto (\gamma, \gamma^\sigma, \gamma^{\sigma^2}) \in G \,.$$

Then $\Gamma < G$ is a uniform irreducible lattice with trace $H \cap \Gamma \simeq \Gamma_0$ a uniform lattice in $H \simeq G_0$.

9 Outlook

We discuss some topics of harmonic analysis on reductive homogeneous spaces which are currently open and would have immediate applications to lattice counting.

9.1 A conjecture which implies Hypothesis A

Hypothesis A falls in the context of a more general conjecture about the growth behavior of families of Harish-Chandra modules.
We let $Z = G/H$ be a real spherical space. Denote by $A_Z^- \subset A_Z$ the compression cone of Z (see Section 3) and recall that wavefront means that $A^- A_H / A_H = A_Z^-$ which, however, we do not assume for the moment.
We use V to denote Harish-Chandra modules for the pair $(\mathfrak{g}, K)$ and V^∞ for their unique moderate growth smooth Fréchet globalizations. These V^∞ are global objects in the sense that they are G-modules whereas V is defined in algebraic terms. We write $V^{-\infty}$ for the strong dual of V^∞. We say that V is H-distinguished provided that the space of H-invariants $(V^{-\infty})^H$ is non-trivial.
It is no big loss of generality to assume that A_Z^- is a sharp cone, as the edge of this cone is in the normalizer of H and in particular acts on the finite dimensional space of H-invariants.
As A_Z^- is pointed it is a fundamental domain for the little Weyl group and as such a simplicial cone (see [17], Section 9). If $\mathfrak{a}_Z^- = \log A_Z^-$, then we write $\omega_1, \ldots, \omega_r$ for a set of generators (spherical co-roots) of $\mathfrak{a}_Z^-$.
Set $\overline{Q} := \theta(Q)$ where θ is the Cartan involution determined by the choice of K. Note that $V/\overline{\mathfrak{q}}V$ is a finite dimensional $\overline{Q}$ module, in particular a finite

dimensional A_Z-module. Let $\Lambda_1, \ldots, \Lambda_N \in \mathfrak{a}_Z^*$ be the $\mathfrak{a}_{Z,\mathbb{C}}$-weight spectrum. Then we define the H-spherical exponent $\Lambda_V \in \mathfrak{a}_Z^*$ of V by

$$\Lambda_V(\omega_i) := \max_{1\leq j\leq N} \operatorname{Re}\Lambda_j(\omega_i)\,.$$

Further attached to V is a "logarithmic" exponent $d \in \mathbb{N}$. Having this data we recall the main bound from [22]

$$|m_{v,\eta}(a\cdot z_0)| \lessapprox a^{\Lambda_V}(1+\|\log a\|)^{d_V} \qquad (a\in A_Z^-)\,.$$

Conjecture 9.1. *Fix a K-type τ, a constant $C>0$, and a compact subset $\Omega\subset G$. Then there exists a compact set $\Omega_A\subset A_Z^-$ such that for all Harish-Chandra modules V with $\|\Lambda_V\|\leq C$, all $v\in V[\tau]$ and all $\eta\in (V^{-\infty})^H$ one has*

$$\begin{aligned}&\max_{\substack{a\in A_Z^-\\ g\in\Omega}} |m_{v,\eta}(ga\cdot z_0)|a^{-\Lambda_V}(1+\|\log a\|)^{-d_V} = \\ &\max_{\substack{a\in\Omega_A\\ g\in\Omega}} |m_{v,\eta}(ga\cdot z_0)|a^{-\Lambda_V}(1+\|\log a\|)^{-d_V}\,.\end{aligned}$$

It is easily seen that this conjecture implies Hypothesis A if all the generalized matrix coefficients $m_{v,\eta}$ are bounded, as for example it is the case when Z is wavefront (see Proposition 3.4(1)).

Remark 9.2. It might well be that a slightly stronger conjecture is true. For that we recall that a Harish-Chandra module V has a unique minimal globalization, the analytic model V^ω. The space V^ω is an increasing union of subspaces V_ϵ for $\epsilon\to 0$. The parameter ϵ parametrizes left G-invariant neigborhoods $\Xi_\epsilon\subset G_\mathbb{C}$ of $\mathbf{1}$ which decrease with $\epsilon\to 0$. Further V_ϵ consists of those vectors $v\in V^\omega$ for which the orbit map $G\to V^\omega$, $g\mapsto g\cdot v$ extends to a holomorphic map on Ξ_ϵ. For fixed $\epsilon, C>0$ the strengthened conjecture would be that there exists a compact subset Ω_A such that for all Harish-Chandra modules V with $\|\Lambda_V\|\leq C$ and all $v\in V_\epsilon$ one has

$$\begin{aligned}&\max_{a\in A_Z^-} |m_{v,\eta}(a\cdot z_0)|a^{-\Lambda_V}(1+\|\log a\|)^{-d_V} = \\ &\max_{a\in\Omega_A} |m_{v,\eta}(a\cdot z_0)|a^{-\Lambda_V}(1+\|\log a\|)^{-d_V}\,.\end{aligned}$$

Note that the compact set Ω is no longer needed, as $\Omega\cdot V_\epsilon\subset V_{\epsilon'}$.

9.2 Spectral geometry of Z_η

In the general context of a reductive real spherical space it may be possible to establish both main term counting and the error term bound, with the arguments presented here for wavefront spaces, provided the following two key questions allow affirmative answers.

In what follows $Z = G/H$ is a real reductive spherical space and V denotes an irreducible Harish-Chandra module and $\eta \in (V^{-\infty})^H$.

Question A: *Is H_η reductive?*

Question B: *If for $v \in V$ the generalized matrix coefficient $m_{v,\eta}$ is bounded, then there exists a $1 \leq p < \infty$ such that $m_{v,\eta} \in L^p(Z_\eta)$.*

In this context we note that issues related to the well-factorization of the intrinsic balls in affine spherical spaces can possibly be resolved with similar methods to those applied here, using volume estimates as described in Theorem 7.17 of [13].

Acknowledgement. We are grateful to an anonymous referee for valuable comments that improved the presentation in the present version of the paper.

References

[1] Y. Benoist and H. Oh, *Effective equidistribution of S-integral points on symmetric varieties*, Ann. Inst. Fourier (Grenoble) 62 (2012), no. 5, 1889–1942.

[2] J. Bernstein and B. Krötz, *Smooth Fréchet globalizations of Harish-Chandra modules*, Israel J. Math. 199 (2014), 35–111.

[3] J. Bernstein and A. Reznikov, *Estimates of automorphic functions*, Moscow Math. J. 4 (1) (2004), 19–37

[4] P. Bravi and G. Pezzini, *The spherical systems of the wonderful reductive subgroups*, J. Lie Theory 25 (2015), 105–123.

[5] J.-L. Clerc and B. Ørsted, *Conformally invariant trilinear forms on the sphere.* Ann. Inst. Fourier 61 (2011), 1807–1838.

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

[7] M. Cowling, *Sur les coefficients des representations unitaires des groupes de Lie simples*, Lecture Notes in Mathematics 739 (1979), 132–178.

[8] T. Danielsen, B. Krötz and H. Schlichtkrull, *Decomposition theorems for triple spaces*, Geom. Dedicata 174 (2015), 145–154.

[9] A. Deitmar, *Invariant triple products*, Int. J. Math. Math. Sci. 2006, Art. ID 48274, 22 pp.

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

[11] A. Eskin and C. McMullen, *Mixing, counting, and equidistribution in Lie groups*, Duke Math. J. 71 (1993), no. 1, 181–209.

[12] A. Eskin, S. Mozes and N. Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. of Math. (2) 143 (1996), 253–299.

[13] A. Gorodnik and A. Nevo, The ergodic theory of lattice points, Princeton University press 2008.

[14] —, *Counting lattice points*, J. Reine Angew. Math. 663 (2012), 127–176.

[15] Harish-Chandra, Automorphic Forms on Semisimple Lie Groups, Springer LNM 62.

[16] L. Ji, *The Weyl upper bound on the discrete spectrum of locally symmetric spaces*, J. Diff. Geom. 51 (1) (1999), 97–147.

[17] F. Knop and B. Krötz, *Reductive group actions*, arXiv: 1604.01005

[18] F. Knop, B. Krötz, E. Sayag and H. Schlichtkrull, *Simple compactifications and polar decomposition for real spherical spaces*, Selecta Math. (N.S.) 21 (2015), 1071–1097.

[19] —, *Volume growth, temperedness and integrability of matrix coefficients on a real spherical space*, arXiv: 1407.8006. J. Funct. Anal. 271 (1) (2016), 12–36.

[20] F. Knop, B. Krötz and H. Schlichtkrull, *The local structure theorem for real spherical varieties*, Compositio Math. 151 (2015), 2145–2159.

[21] B. Krötz, E. Sayag and H. Schlichtkrull, *Vanishing at infinity on homogeneous spaces of reductive type*, arXiv: 1211.2781. Compositio Math., to appear.

[22] —, *Decay of matrix coefficients on reductive homogeneous space of spherical type*, Math. Z. 278 (2014), 229–249.

[23] B. Krötz and H. Schlichtkrull, *Multiplicity bounds and the subrepresentation theorem for real spherical spaces*, Trans. Amer. Math. Soc. 368 (2016), 2749-2762.

[24] Y. Sakellaridis and A. Venkatesh, *Periods and harmonic analysis on spherical varieties*, arXiv:1203.0039

Universität Paderborn
Institut für Mathematik
Warburger Straße 100
33098 Paderborn
Deutschland
bkroetz@gmx.de

Department of Mathematics
Ben Gurion University
of the Negev
P.O.B. 653
Be'er Sheva 84105, Israel
eitan.sayag@gmail.com

University of Copenhagen
Department of Mathematics
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
schlicht@math.ku.dk