Thin groups and superstrong approximation
MSRI Publications
Volume **61**, 2013

Harmonic analysis, ergodic theory and counting for thin groups

HEE OH

For a geometrically finite group Γ of $G = \mathrm{SO}(n, 1)$, we survey recent developments on counting and equidistribution problems for orbits of Γ in a homogeneous space $H\backslash G$ where H is trivial, symmetric or horospherical. Main applications are found in an affine sieve on orbits of thin groups as well as in sphere counting problems for sphere packings invariant under a geometrically finite group. In our sphere counting problems, spheres can be ordered with respect to a general conformal metric.

1. Introduction

In this article we discuss counting and equidistribution problems for orbits of thin groups in homogeneous spaces.

Let G be a connected semisimple Lie group and H a closed subgroup. We consider the homogeneous space $V = H\backslash G$ and fix the identity coset $x_0 = [e]$. Let Γ be a discrete subgroup of G such that the orbit $x_0\Gamma$ is discrete, and let $\{B_T : T > 1\}$ be a family of compact subsets of V whose volume tends to infinity as $T \to \infty$. Understanding the asymptotic of $\#(x_0\Gamma \cap B_T)$ is a fundamental problem which bears many applications in number theory and geometry. We refer to this type of counting problem as an archimedean counting as opposed to a combinatorial counting where the elements in $x_0\Gamma$ are ordered with respect to a word metric on Γ; both have been used in applications to sieves; see [Kowalski 2010; 2014].

When Γ is a lattice in G, i.e., when $\Gamma\backslash G$ admits a finite invariant measure, this problem is well understood for a large class of subgroups H, e.g., when H is a maximal subgroup: assuming that the boundaries of the B_T are sufficiently regular and that $H\cap\Gamma$ is a lattice in H, the value of $\#(x_0\Gamma\cap B_T)$ is asymptotically proportional to the volume of B_T, computed with respect to a suitably normalized G-invariant measure in V [Duke et al. 1993; Eskin and McMullen 1993; Eskin et al. 1996; Benoist and Oh 2012]. See also the survey paper [Oh 2010].

When Γ is a thin group, i.e., a Zariski dense subgroup which is not a lattice in G, it is far less understood in a general setting. In this article we focus on the case when G is the special orthogonal group $\mathrm{SO}(n, 1)$ and H is either a symmetric subgroup or a horospherical subgroup (the case of H being the trivial subgroup will be treated as well). In this case, we have a more or less satisfactory understanding for the counting problem for groups Γ equipped with a certain finiteness property, called the geometric finiteness (see Definition 2.1).

By the fundamental observation of Duke, Rudnick and Sarnak [Duke et al. 1993], the counting problem for $x_0\Gamma \cap B_T$ can be well approached via the following equidistribution problem:

Describe the asymptotic distribution of $\Gamma\backslash\Gamma Hg$ *in* $\Gamma\backslash G$ *as* $g \to \infty$.

The assumption that H is either symmetric or horospherical is made so that we can approximate the translate $\Gamma\backslash\Gamma Hg$ locally by the matrix coefficient function in the quasiregular representation space $L^2(\Gamma\backslash G)$. This idea goes back to Margulis's 1970 thesis (translated in [Margulis 2004]; also see [Kleinbock and Margulis 1996]), but a more systematic formulation in our setting is due to Eskin and McMullen [1993].

The fact that the trivial representation is contained in $L^2(\Gamma\backslash G)$ for Γ lattice is directly related to the phenomenon that, when $\Gamma\backslash\Gamma H$ is of finite volume, the translate $\Gamma\backslash\Gamma Hg$ becomes equidistributed in $\Gamma\backslash G$ with respect to a G-*invariant* measure as $g \to \infty$ in $H\backslash G$. When Γ is not a lattice, the minimal subrepresentation, that is, the subrepresentation with the slowest decay of matrix coefficients of $L^2(\Gamma\backslash G)$, is infinite-dimensional and this makes the distribution of $\Gamma\backslash\Gamma Hg$ much more intricate, and understanding it requires introducing several singular measures in $\Gamma\backslash G$. A key input is the work of Roblin [2003] on the asymptotic of matrix coefficients for $L^2(\Gamma\backslash G)$, which he proves using ergodic theoretic methods.

There is a finer distinction among geometrically finite groups depending on the size of their critical exponents. When the critical exponent δ of Γ exceeds $\frac{n-1}{2}$, work of Lax and Phillips [1982] implies a spectral gap for $L^2(\Gamma\backslash\mathbb{H}^n)$; which we call a *spherical* spectral gap for $L^2(\Gamma\backslash G)$ as it concerns only the spherical part of $L^2(\Gamma\backslash G)$. We formulate a notion of a spectral gap for $L^2(\Gamma\backslash G)$ which deals with both spherical and nonspherical parts, based on the knowledge of the unitary dual of G. Under the hypothesis that $L^2(\Gamma\backslash G)$ admits a spectral gap (this is known to be true if $\delta > n-2$), developing Harish-Chandra's work on harmonic analysis on G in combination with Roblin's work on ergodic theory, we obtain an effective version of the asymptotic of matrix coefficients for $L^2(\Gamma\backslash G)$ in [Mohammadi and Oh 2012b]. This enables us to state an effective equidistribution of $\Gamma\backslash\Gamma Hg$ in $\Gamma\backslash G$. Similar to the condition that $\Gamma\backslash\Gamma H$ is of finite volume in the case of Γ lattice, there is also a certain restriction on the size of the orbit $\Gamma\backslash\Gamma H$ in order to deduce such an equidistribution. A precise condition is that the skinning measure μ_H^{PS} of $\Gamma\backslash\Gamma H$, introduced in [Oh and Shah 2013], is finite; roughly speaking $|\mu_H^{\mathrm{PS}}|$ measures asymptotically the portion of $\Gamma\backslash\Gamma H$ which returns to a compact subset after flowed by the geodesic flow. When the skinning measure μ_H^{PS} is compactly supported, the passage from the asymptotic of the matrix coefficient to the equidistribution of $\Gamma\backslash\Gamma Hg$ can be done by the so-called usual thickening methods. However when μ_H^{PS} is not compactly supported, this step requires a genuinely different strategy from the lattice case via the study of the transversal intersections, carefully done in [Oh and Shah 2013].

The error term in our effective equidistribution result of $\Gamma\backslash\Gamma Hg$ depends only on the spectral gap data of Γ. This enables us to state the asymptotic of $\#(x_0\Gamma_d\gamma\cap B_T)$ effectively in a uniform manner for all $\gamma\in\Gamma$ and for any family $\{\Gamma_d<\Gamma\}$ of subgroups of finite index which has a uniform spectral gap, if they satisfy $\Gamma_d\cap H=\Gamma\cap H$. When Γ is a subgroup of an arithmetic subgroup of G with $\delta>n-2$, the work of Salehi Golsefidy and Varjú [2012], extending an earlier work of Bourgain, Gamburd and Sarnak [Bourgain et al. 2010a], provides a certain congruence family $\{\Gamma_d\}$ satisfying this condition.

A recent development on an affine sieve [Bourgain et al. 2011] then tells us that such a uniform effective counting statement can be used to describe the distribution of almost prime vectors, as well as to give a sharp upper bound for primes (see Theorem 6.6).

One of the most beautiful applications of the study of thin orbital counting problem can be found in Apollonian circle packings. We will describe this application as well as its higher-dimensional analogues. The ordering in counting circles can be done not only in the Euclidean metric, but also in general conformal metrics. It is due to this flexibility that we can also describe the asymptotic number of circles in the ideal triangle of the hyperbolic plane ordered by the hyperbolic area in the last section.

2. Γ-invariant conformal densities and measures on $\Gamma\backslash G$

In the whole article, let $(\mathbb{H}^n, d)$ be the n-dimensional real hyperbolic space with constant curvature -1 and $\partial(\mathbb{H}^n)$ its geometric boundary. Set $G := \mathrm{Isom}^+(\mathbb{H}^n) \simeq \mathrm{SO}(n,1)^\circ$. Let Γ be a discrete torsion-free subgroup of G. We assume that Γ is nonelementary, or equivalently, Γ has no abelian subgroup of finite index.

We review some basic geometric and measure theoretic concepts for Γ and define several locally finite Borel measures (Radon measures) on $\Gamma\backslash G$ associated to Γ-invariant conformal densities on $\partial(\mathbb{H}^n)$. When Γ is a lattice, these measures all coincide with each other, being simply a G-invariant measure. But for a thin subgroup Γ, they are all different and singular, and appear in our equidistribution and counting statements. General references for this section are [Ratcliffe 2006; Bowditch 1993; Patterson 1976; Sullivan 1979; 1984; Roblin 2003].

Limit set and geometric finiteness. We denote by $\delta_\Gamma = \delta$ the critical exponent of Γ, i.e., the abscissa of convergence of the Poincaré series $\sum_{\gamma\in\Gamma} e^{-sd(o,\gamma(o))}$ for $o \in \mathbb{H}^n$. We have $0 < \delta \le n-1$. The limit set $\Lambda(\Gamma)$ is defined to be the set of all accumulation points of $\Gamma(z)$ in the compactification $\overline{\mathbb{H}^n} = \mathbb{H}^n \cup \partial(\mathbb{H}^n)$, $z \in \mathbb{H}^n$. As Γ is discrete, $\Lambda(\Gamma)$ is contained in the boundary $\partial(\mathbb{H}^n)$.

Definition 2.1. (1) The convex core $C(\Gamma)$ of Γ is the quotient by Γ of the smallest convex subset of $\mathbb{H}^n$ containing all geodesics connecting points in $\Lambda(\Gamma)$.

(2) Γ is called geometrically finite (resp. convex cocompact) if the unit neighborhood of the convex core of Γ has finite volume (resp. compact).

A lattice is clearly a geometrically finite group and so is a discrete group admitting a finite sided convex fundamental domain in $\mathbb{H}^n$. An important characterization of a geometrically finite group is given in terms of its limit set. For this, we need to define: a point $\xi \in \Lambda(\Gamma)$ is called a parabolic limit point for Γ if ξ is a unique fixed point in $\partial\mathbb{H}^n$ for an element of Γ and a radial limit point if the projection of a geodesic ray ξ_t toward ξ in $\Gamma\backslash\mathbb{H}^n$ meets a compact subset for an unbounded sequence of time t.

Now Γ is geometrically finite if and only if $\Lambda(\Gamma)$ consists only of parabolic and radial limit points [Bowditch 1993]. For Γ geometrically finite, its critical exponent δ is equal to the Hausdorff dimension of $\Lambda(\Gamma)$, and is $n-1$ only when Γ is a lattice in G [Sullivan 1979].

Conformal densities. To define a conformal density, we first recall the Busemann function $\beta_\xi(x, y)$ for $x, y \in \mathbb{H}^n$ and $\xi \in \partial(\mathbb{H}^n)$:

$$\beta_\xi(x, y) = \lim_{t\to\infty} d(x, \xi_t) - d(y, \xi_t)$$

where ξ_t is a geodesic toward ξ. Hence $\beta_\xi(x, y)$ measures a signed distance between horospheres based at ξ passing through x and y (a horosphere based at ξ is a Euclidean sphere in $\mathbb{H}^n$ tangent at ξ).

Definition 2.2. A Γ-invariant conformal density of dimension $\delta_\mu > 0$ is a family $\{\mu_x : x \in \mathbb{H}^n\}$ of finite positive measures on $\partial(\mathbb{H}^n)$ satisfying

(1) $\gamma_* \mu_x = \mu_{\gamma(x)}$ for any $\gamma \in \Gamma$, and

(2) $\dfrac{d\mu_x}{d\mu_y}(\xi) = e^{\delta_\mu \beta_\xi(y,x)}$ for all $x, y \in \mathbb{H}^n$ and $\xi \in \partial(\mathbb{H}^n)$.

It is easy to construct such a density of dimension $n-1$, as we simply need to set m_x to be the $\mathrm{Stab}_G(x)$-invariant probability measure on $\mathbb{H}^n$. This is a unique up to scaling, and does not depend on Γ. We call it the Lebesgue density.

How about in other dimensions? A fundamental work of Patterson [1976], generalized by Sullivan [1979], shows the following by an explicit construction:

Theorem 2.3. *There exists a Γ-invariant conformal density of dimension $\delta_\Gamma = \delta$.*

Assuming that Γ is of divergence type, i.e., its Poincaré series diverges at $s = \delta$, Patterson's construction can be summarized as follows: Fixing $o \in \mathbb{H}^n$, for each $x \in \mathbb{H}^n$, consider the finite measure on $\overline{\mathbb{H}^n}$ given by

$$\nu_{x,s} = \frac{1}{\sum_{\gamma\in\Gamma} e^{-sd(o,\gamma(o))}} \sum_{\gamma\in\Gamma} e^{-sd(x,\gamma(o))} \delta_{\gamma(o)}.$$

Then ν_x is the (unique) weak-limit of $\nu_{x,s}$ as $s \to \delta^+$, and $\{\nu_x : x \in \mathbb{H}^n\}$ is the desired density of dimension δ.

In the following, we fix a Γ-invariant conformal density $\{\nu_x\}$ of dimension δ, and call it the Patterson–Sullivan density (or simply the PS density). It is known to be unique up to scaling, when Γ is of divergence type, e.g., geometrically finite groups.

Denoting by Δ the hyperbolic Laplacian on $\mathbb{H}^n$, the PS density is closely related to the bottom of the spectrum of Δ for its action on smooth functions on $\Gamma\backslash\mathbb{H}^n$. If we set $\phi_0(x) := |\nu_x|$ for each $x \in \mathbb{H}^n$, then

$$\phi_0(x) = \int_{\xi\in\Lambda(\Gamma)} d\nu_x(\xi) = \int_{\xi\in\Lambda(\Gamma)} \frac{d\nu_x}{d\nu_o}(\xi)\, d\nu_o(\xi) = \int_{\xi\in\Lambda(\Gamma)} e^{\delta\beta_\xi(o,x)} d\nu_o(\xi).$$

Since $\gamma_*\nu_x = \nu_{\gamma(x)}$, we note that $\phi_0(\gamma(x)) = \phi_0(x)$, i.e., ϕ_0 is a function on $\Gamma\backslash\mathbb{H}^n$, which is *positive* everywhere! Furthermore:

(1) $\Delta(\phi_0) = \delta(n-1-\delta)\phi_0$.

(2) If Γ is geometrically finite, then $\phi_0 \in L^2(\Gamma\backslash\mathbb{H}^n)$ if and only if $\delta > \frac{n-1}{2}$.

Measures on $\Gamma\backslash G$ associated to a pair of conformal densities. For $v \in \mathrm{T}^1(\mathbb{H}^n)$, we denote by v^+ and v^- the forward and backward endpoints of the geodesic determined by v. Fixing $o \in \mathbb{H}^n$, the map $v \mapsto (v^+, v^-, s = \beta_{v^-}(o, v))$ gives a homeomorphism between $\mathrm{T}^1(\mathbb{H}^n)$ and $(\partial(\mathbb{H}^n) \times \partial(\mathbb{H}^n) - \text{diagonal}) \times \mathbb{R}$. Therefore we may use the coordinates $(v^+, v^-, s = \beta_{v^-}(o, v))$ of v in order to define measures on $\mathrm{T}^1(\mathbb{H}^n)$.

Let $\{\mu_x\}$ and $\{\mu'_x\}$ be Γ-invariant conformal densities on $\partial(\mathbb{H}^n)$ of dimensions δ_μ and $\delta_{\mu'}$ respectively. After [Roblin 2003], we define a measure $\tilde{m}^{\mu,\mu'}$ on $\mathrm{T}^1(\mathbb{H}^n)$ associated to $\{\mu_x\}$ and $\{\mu'_x\}$ by

$$d\tilde{m}^{\mu,\mu'}(v) = e^{\delta_\mu \beta_{v^+}(o,v)}\, e^{\delta_{\mu'} \beta_{v^-}(o,v)}\, d\mu_o(v^+)\, d\mu'_o(v^-)\, ds. \tag{2.4}$$

It follows from the Γ-invariant conformal properties of $\{\mu_x\}$ and $\{\mu'_x\}$ that the definition of $\tilde{m}^{\mu,\mu'}$ is independent of the choice of $o \in \mathbb{H}^n$ and that $\tilde{m}^{\mu,\mu'}$ is left Γ-invariant. Hence it induces a Radon measure $m^{\mu,\mu'}$ on the quotient space $\mathrm{T}^1(\Gamma\backslash\mathbb{H}^n) = \Gamma\backslash \mathrm{T}^1(\mathbb{H}^n)$.

We will lift the measure $m^{\mu,\mu'}$ to $\Gamma\backslash G$. This lift depends on the choice of subgroups K, M and $A = \{a_t\}$ of G. Here K is a maximal compact subgroup of G and M is the stabilizer of a vector $X_0 \in \mathrm{T}^1(\mathbb{H}^n)$ based at $o \in \mathbb{H}^n$ with $K = G_o$. Via the isometric action of G, we may identify the quotient spaces G/K and G/M with $\mathbb{H}^n$ and $\mathrm{T}^1(\mathbb{H}^n)$ respectively. Let $A = \{a_t\}$ be the one parameter subgroup of diagonalizable elements of G such that the right multiplication by a_t on G/M corresponds to the geodesic flow on $\mathrm{T}^1(\mathbb{H}^n)$ for time t.

By abuse of notation, we use the same notation $m^{\mu,\mu'}$ for the M-invariant extension of $m^{\mu,\mu'}$ on $\Gamma\backslash G/M = \Gamma\backslash \mathrm{T}^1(\mathbb{H}^n)$ to $\Gamma\backslash G$, that is, for $\Psi \in C_c(\Gamma\backslash G)$,

$$m^{\mu,\mu'}(\Psi) = \int_{x \in \Gamma\backslash G/M} \Psi^M(x)\, dm^{\mu,\mu'}(x)$$

where $\Psi^M(x) = \int_M \Psi(xm)\, dm$ for the probability M-invariant measure dm on M.

The measures $m^{\mu,\mu'}$ on $\Gamma\backslash G$ where μ and μ' are the PS-density $\{\nu_x\}$ or the Lebesgue density $\{m_x\}$ are of special importance. We name them as follows:

- Bowen–Margulis–Sullivan measure: $m^{\mathrm{BMS}} := m^{\nu,\nu}$
- Burger–Roblin measure: $m^{\mathrm{BR}} := m^{m,\nu}$
- Burger–Roblin $*$-measure: $m^{\mathrm{BR}}_* := m^{\nu,m}$
- Haar measure: $m^{\mathrm{Haar}} := m^{m,m}$.

For brevity, we refer to these as BMS, BR, BR_*, and Haar measures, respectively. As the naming indicates, m^{Haar} turns out to be a G-invariant measure. For $g \in G$, we use the notation $g^\pm$ for $(gM)^\pm$ with gM considered as a vector in $\mathrm{T}^1(\mathbb{H}^n)$.

It is clear from the definition that the supports of BMS, BR, BR_* measures are respectively given by $\{g \in \Gamma\backslash G : g^{\pm} \in \Lambda(\Gamma)\}$, $\{g \in \Gamma\backslash G : g^{-} \in \Lambda(\Gamma)\}$, and $\{g \in \Gamma\backslash G : g^{+} \in \Lambda(\Gamma)\}$. In particular, the support of BMS measure is contained in the convex core of Γ. Sullivan showed that $|m^{\mathrm{BMS}}| < \infty$ if Γ is geometrically finite.

The BMS, BR and BR_* measures are respectively invariant under A, N^+ and N^- where N^+ and N^- denote the expanding and contracting horospherical subgroups of G for a_t:

$$N^{\pm} = \{g \in G : a_t g a_{-t} \to e \text{ as } t \to \pm\infty\}.$$

The finiteness of m^{BMS} turns out to be a critical condition for the ergodic theory on $\Gamma\backslash G$.

Theorem 2.5. *Suppose that $|m^{\mathrm{BMS}}| < \infty$ and that Γ is Zariski dense.*

(1) m^{BMS} *is A-mixing*: *for any $\Psi_1, \Psi_2 \in L^2(\Gamma\backslash G)$,*

$$\lim_{t\to\infty} \int_{\Gamma\backslash G} \Psi_1(ga_t)\Psi_2(g)\, dm^{\mathrm{BMS}}(g) = \frac{1}{|m^{\mathrm{BMS}}|} m^{\mathrm{BMS}}(\Psi_1) \cdot m^{\mathrm{BMS}}(\Psi_2).$$

(2) *Any locally finite N^+-ergodic invariant measure on $\Gamma\backslash G$ is either supported on a closed N^+M_0-orbit where M_0 is an abelian closed subgroup of M or m^{BR}.*

(3) m^{BR} *is a finite measure if and only if Γ is a lattice in G.*

Claim (1) was first made in [Flaminio and Spatzier 1990]. However there is a small gap in their proof which is now fixed in [Winter $\geq$ 2012]. Winter also obtained Claim (1) in a general rank one symmetric space. For M-invariant functions, this claim was earlier proved by Babillot [2002] and in this case the Zariski density assumption is not needed. Claim (2) was first proved by Burger [1990] for a convex cocompact surface with critical exponent bigger than $1/2$. In a general case, Winter obtained Claim (2) from Roblin's work [2003] and Claim (1).

Claim (3) is proved in [Oh and Shah 2013], using Ratner's measure classification [1991] of finite measures invariant under unipotent flows. Namely, we show that m^{BR} is not one of those homogeneous measures that her classification theorem lists for finite invariant measures.

In the spirit of Ratner's measure classification theorem, we pose the following question:

Problem 2.6. Under the assumption of Theorem 2.5, let U be a connected unipotent subgroup of G or more generally a connected subgroup generated by unipotent one-parameter subgroups.

(1) Classify all locally finite U-invariant ergodic measures in $\Gamma\backslash G$.

(2) Describe the closures of U-orbits in $\Gamma\backslash G$.

The emphasis here is that we want to understand not only finite measures but all Radon measures. In general, this seems to be a very challenging question. We mention a recent related result from [Mohammadi and Oh 2012a]: if Γ is a convex cocompact subgroup of G and U is a connected unipotent subgroup of N^+ of dimension k, m^{BR} is U-ergodic if $\delta > (n-1)-k$. Precisely speaking, this is proved only for $n=3$ in [Mohammadi and Oh 2012a], but the methods of proof works for a general $n \geq 3$ as well.

3. Matrix coefficients for $L^2(\Gamma\backslash G)$

Let Γ be a discrete, torsion-free, nonelementary subgroup of $G = \mathrm{SO}(n,1)^\circ$. The right translation action of G on $L^2(\Gamma\backslash G, m^{\mathrm{Haar}})$ gives rise to a unitary representation, as m^{Haar} is G-invariant. For $\Psi_1, \Psi_2 \in L^2(\Gamma\backslash G)$, the matrix coefficient function is a smooth function on G defined by

$$g \mapsto \langle g.\Psi_1, \Psi_2\rangle := \int_{\Gamma\backslash G} \Psi_1(xg)\overline{\Psi_2(x)}\,dm^{\mathrm{Haar}}(x).$$

Understanding the asymptotic expansion of $\langle a_t.\Psi_1, \Psi_2\rangle$ (as $t \to \infty$) is a basic problem in harmonic analysis as well as a main tool in our approach to the counting problem.

The quality of the error term in this type of the asymptotic expansion usually depends on Sobolev norms of Ψ_i's. For $\Psi \in C^\infty(\Gamma\backslash G)$ and $d \in \mathbb{N}$, the d-th Sobolev norm of Ψ is given by $\mathcal{S}_d(\Psi) = \sum \|X(\Psi)\|_2$ where the sum is taken over all monomials X in some fixed basis of the Lie algebra of G of order at most d and $\|X(\Psi)\|_2$ denotes the L^2-norm of Ψ.

In order to describe our results as well as a conjecture, we begin by describing the unitary dual of G, i.e., the set of equivalence classes of all irreducible unitary representations of G.

The unitary dual $\hat{G}$. Let K be a maximal compact subgroup of G and fix a Cartan decomposition $G = KA^+K$. We parametrize $A^+ = \{a_t : t \geq 0\}$ so that a_t corresponds to the geodesic flow on $\mathrm{T}^1(\mathbb{H}^n) = G/M$ where $M := C_K(A^+)$ is the centralizer of A^+ in K.

A representation $\pi \in \hat{G}$ is said to be *tempered* if for any K-finite vectors v_1, v_2 of π, the matrix coefficient function $g \mapsto \langle \pi(g)v_1, v_2\rangle$ belongs to $L^{2+\epsilon}(G)$ for any $\epsilon > 0$. We write $\hat{G} = \hat{G}_{\mathrm{temp}} \cup \hat{G}_{\mathrm{nontemp}}$ as the disjoint union of tempered representations and nontempered representations.

The work of Hirai [1962] on the classification of $\hat{G}$ implies that nontempered part of the unitary dual $\hat{G}$ consists of the trivial representation, and complementary series representations $\mathcal{U}(\upsilon, s-n+1)$ parameterized by $\upsilon \in \hat{M}$ and $s \in I_\upsilon$, where $\hat{M}$ is the unitary dual of M and I_υ is an interval contained in $(\frac{n-1}{2}, n-1)$, depending on υ (see also [Knapp and Stein 1971, Propositions 49, 50]). Moreover $\mathcal{U}(\upsilon, s-n+1)$ is spherical if and only if υ is the trivial representation 1 of M. By choosing a Casimir operator $\mathcal{C}$ of the Lie algebra of G normalized so that it acts on $C^\infty(G)^K = C^\infty(\mathbb{H}^n)$ by the negative Laplacian, the normalization is made so that $\mathcal{C}$ acts on $\mathcal{U}(1, s-n+1)^\infty$ by the scalar $s(s-n+1)$.

Lattice case: $|m^{\mathrm{Haar}}| < \infty$. Set

$$L^2_0(\Gamma\backslash G) := \{\Psi \in L^2(\Gamma\backslash G) : \int \Psi \, dm^{\mathrm{Haar}} = 0\}.$$

When Γ is a lattice, we have $L^2(\Gamma\backslash G) = \mathbb{C} \oplus L^2_0(\Gamma\backslash G)$. It is well-known that there exists $\frac{n-1}{2} < s_0 < (n-1)$ such that $L^2_0(\Gamma\backslash G)$ does not contain any complementary series representation of parameter $s \geq s_0$ [Borel and Garland 1983]. This implies:

Theorem 3.1 [Shalom 2000; Kontorovich and Oh 2011, Proposition 5.3]. *Suppose Γ is a lattice in G, that is, $|m^{\mathrm{Haar}}| < \infty$. There exists $\ell \in \mathbb{N}$ such that for any $\Psi_1, \Psi_2 \in L^2(\Gamma\backslash G) \cap C(\Gamma\backslash G)^\infty$, we have, as $t \to \infty$,*

$$\langle a_t.\Psi_1, \Psi_2\rangle = \frac{1}{|m^{\mathrm{Haar}}|} m^{\mathrm{Haar}}(\Psi_1) m^{\mathrm{Haar}}(\Psi_2) + O(\mathcal{S}_\ell(\Psi_1)\mathcal{S}_\ell(\Psi_2) e^{(s_0-n+1)t})$$

where $m^{\mathrm{Haar}}(\Psi_i) = \int_{\Gamma\backslash G} \Psi_i(x)\, dm^{\mathrm{Haar}}(x)$ for $i = 1, 2$.

We note that the constant function $1/\sqrt{|m^{\mathrm{Haar}}|}$ is a unit vector in $L^2(\Gamma\backslash G)$ and the main term $\frac{1}{|m^{\mathrm{Haar}}|} m^{\mathrm{Haar}}(\Psi_1) m^{\mathrm{Haar}}(\Psi_2)$ is simply the product of the projections of Ψ_1 and Ψ_2 to the minimal subrepresentation space, which is $\mathbb{C}$, of $L^2(\Gamma\backslash G)$.

Discrete groups with $|m^{\mathrm{BMS}}| < \infty$. When Γ is not a lattice, we have $L^2(\Gamma\backslash G) = L^2_0(\Gamma\backslash G)$. By the well-known decay of the matrix coefficients of unitary representations with no G-invariant vectors due to Howe and Moore [Howe and Moore 1979], we have, for any $\Psi_1, \Psi_2 \in L^2(\Gamma\backslash G)$,

$$\lim_{t\to\infty} \langle a_t.\Psi_1, \Psi_2\rangle = 0.$$

However we have a much more precise description on the decay of $\langle a_t.\Psi_1, \Psi_2\rangle$ due to Roblin:

Theorem 3.2 [Roblin 2003]. *Let Γ be Zariski dense with $|m^{\mathrm{BMS}}| < \infty$. For any $\Psi_1, \Psi_2 \in C_c(\Gamma\backslash G)$,*

$$\lim_{t\to\infty} e^{(n-1-\delta)t}\langle a_t.\Psi_1, \Psi_2\rangle = \frac{m^{\mathrm{BR}}(\Psi_1)\cdot m_*^{\mathrm{BR}}(\Psi_2)}{|m^{\mathrm{BMS}}|}.$$

Roblin proved this theorem for M-invariant functions using the mixing of the geodesic flow due to Babillot [2002]. His proof extends without difficulty to general functions, based on the A-mixing in $\Gamma\backslash G$ stated as in Theorem 2.5.

Geometrically finite groups with $\delta > \frac{n-1}{2}$. In this subsection, we assume that Γ is a geometrically finite, Zariski dense, discrete subgroup of G with $\delta > \frac{n-1}{2}$. Under this assumption, the works of Lax and Phillips [Lax and Phillips 1982] and Sullivan [Sullivan 1984] together imply that there exist only finitely many $(n-1) \geq s_0 > s_1 \geq \cdots \geq s_\ell > \frac{n-1}{2}$ such that the spherical complementary series representation $\mathcal{U}(1, s-n+1)$ occurs as a subrepresentation of $L^2(\Gamma\backslash G)$ and $s_0 = \delta$. In particular, there is no spherical complementary representation $\mathcal{U}(1, s-n+1)$ contained in $L^2(\Gamma\backslash G)$ for $\delta < s < s_1$; hence we have a *spherical* spectral gap for $L^2(\Gamma\backslash G)$.

Using the classification of $\hat{G}_{\mathrm{nontemp}}$, we formulate the notion of a spectral gap. Recall that a unitary representation π is said to be weakly contained in a unitary representation π' if any diagonal matrix coefficients of π can be approximated, uniformly on compact subsets, by convex combinations of diagonal matrix coefficients of π'.

Definition 3.3. We say that $L^2(\Gamma\backslash G)$ has a *strong spectral gap* if

(1) $L^2(\Gamma\backslash G)$ does not contain any $\mathcal{U}(\upsilon, \delta-n+1)$ with $\upsilon \neq 1$;

(2) there exists $s_0(\Gamma)$ with $\frac{n-1}{2} < s_0(\Gamma) < \delta$ and such that $L^2(\Gamma\backslash G)$ does not weakly contain any $\mathcal{U}(\upsilon, s-n+1)$ with $s \in (s_0(\Gamma), \delta)$ and $\upsilon \in \hat{M}$.

For $\delta \leq \frac{n-1}{2}$, the Laplacian spectrum of $L^2(\Gamma\backslash\mathbb{H}^n)$ is continuous; this implies that there is no spectral gap for $L^2(\Gamma\backslash G)$.

Conjecture 3.4 (spectral gap conjecture, [Mohammadi and Oh 2012b]). *If Γ is a geometrically finite and Zariski dense subgroup of G with $\delta > \frac{n-1}{2}$, $L^2(\Gamma\backslash G)$ has a strong spectral gap.*

If $\delta > \frac{n-1}{2}$ for $n = 2, 3$, or if $\delta > (n-2)$ for $n \geq 4$, then $L^2(\Gamma\backslash G)$ has a strong spectral gap. This observation follows from the classification of $\hat{G}_{\mathrm{nontemp}}$ which says that there is no nonspherical complementary series representation $\mathcal{U}(\upsilon, s-n+1)$ of parameter $n-2 < s < n-1$ [Hirai 1962].

Our theorems are proved under the following slightly weaker spectral gap property assumption:

Definition 3.5. [Mohammadi and Oh 2012b] We say that $L^2(\Gamma\backslash G)$ has a *spectral gap* if there exist $\frac{n-1}{2} < s_0 = s_0(\Gamma) < \delta$ and $n_0 = n_0(\Gamma) \in \mathbb{N}$ such that

(1) the multiplicity of $\mathcal{U}(\upsilon, \delta-n+1)$ contained in $L^2(\Gamma\backslash G)$ is at most $\dim(\upsilon)^{n_0}$ for any $\upsilon \in \hat{M}$;

(2) $L^2(\Gamma\backslash G)$ does not weakly contain any $\mathcal{U}(\upsilon, s-n+1)$ with $s \in (s_0, \delta)$ and $\upsilon \in \hat{M}$.

The pair $(s_0(\Gamma), n_0(\Gamma))$ will be referred to as the spectral gap data for Γ.

The spectral gap hypothesis implies that for $\Psi_1, \Psi_2 \in L^2(\Gamma\backslash G)$, the leading term of the asymptotic expansion of the matrix coefficient $\langle a_t.\Psi_1, \Psi_2\rangle$ is determined by $\langle a_t.P_\delta(\Psi_1), P_\delta(\Psi_2)\rangle$ where P_δ is the projection operator from $L^2(\Gamma\backslash G)$ to $\mathcal{H}_\delta^\dagger$, which is the sum of all complementary series representations $\mathcal{U}(\upsilon, \delta-n+1)$, $\upsilon \in \hat{M}$ occurring in $L^2(\Gamma\backslash G)$ as subrepresentations.

Building up on the work of Harish-Chandra on the asymptotic behavior of the Eisenstein integrals (see [Warner 1972a; 1972b]), we obtain an asymptotic formula for $\langle a_t v, w\rangle$ for all K-isotypic vectors $v, w \in \mathcal{H}_\delta^\dagger$. This extension alone does not quite explain the leading term of $\langle a_t P_\delta(\Psi_1), P_\delta(\Psi_2)\rangle$ in terms of Ψ_1 and Ψ_2; however, with the help of Theorem 3.2, we are able to prove this:

Theorem 3.6 [Mohammadi and Oh 2012b]. *Suppose that $L^2(\Gamma\backslash G)$ possesses a spectral gap. There exist $\eta_0 > 0$ and $\ell \in \mathbb{N}$ such that for any $\Psi_1, \Psi_2 \in C_c^\infty(\Gamma\backslash G)$, as $t \to \infty$,*

$$e^{(n-1-\delta)t}\langle a_t\Psi_1, \Psi_2\rangle = \frac{m^{\mathrm{BR}}(\Psi_1)\cdot m_*^{\mathrm{BR}}(\Psi_2)}{|m^{\mathrm{BMS}}|} + O(\mathcal{S}_\ell(\Psi_1)\mathcal{S}_\ell(\Psi_2)e^{-\eta_0 t}).$$

We reiterate that the leading term in Theorem 3.6 is from the ergodic theory and the error term is from the harmonic analysis and the spectral gap.

Effective mixing for m^{BMS}. Theorem 3.6 can be used to obtain an effective mixing for the BMS measure (i.e., an effective version of Theorem 2.5(1)) for geometrically finite, Zariski dense subgroups with a spectral gap [Mohammadi and Oh 2012b].

For geometrically finite groups with $\delta \le \frac{n-1}{2}$, we cannot expect Theorem 3.6 to hold for such groups. However Guillarmou and Mazzeo [2012] have established meromorphic extensions of the resolvents of the Laplacian and of the Poincaré series.

For Γ convex cocompact, Stoyanov obtained, via the spectral properties of Ruelle transfer operators, an effective mixing of the geodesic flow for the BMS measure, regardless of the size of the critical exponent (see [Stoyanov 2011]). We remark that the idea of using spectral estimates of Ruelle transfer operators in obtaining an effective mixing originated in [Dolgopyat 1998].

It will be interesting to see if the results in [Guillarmou and Mazzeo 2012] can be used to answer the question:

Problem 3.7. Prove an effective mixing of the geodesic flow for the BMS measure for all geometrically finite groups.

We close this section with an open problem:

Problem 3.8. Find a suitable analogue of Theorem 3.6 for a higher rank simple Lie group G such as $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$ or $\mathrm{SL}_3(\mathbb{R})$.

4. Distribution of Γ in G

For a family $\{B_T : T > 1\}$ of compact subsets in G, the study of the asymptotic of $\#(\Gamma \cap B_T)$ can be approached directly by Theorems 3.2 and 3.6. The link is given by the following function on $\Gamma\backslash G \times \Gamma\backslash G$:

$$F_T(g, h) := \sum_{\gamma\in\Gamma} \chi_{B_T}(g^{-1}\gamma h)$$

where χ_{B_T} denotes the characteristic function of B_T. Observing that $F_T(e, e) = \#(\Gamma \cap B_T)$, we will explain how the asymptotic behavior of $F_T(e, e)$ is related to the matrix coefficient function for $L^2(\Gamma\backslash G)$. For $g = k_1 a_t k_2 \in K A^+ K$, we have $dm^{\mathrm{Haar}}(g) = \xi(t)\, dt\, dk_1\, dk_2$ where $\xi(t) = e^{(n-1)t}(1 + O(e^{-\beta t}))$ for some $\beta > 0$.

Now assume that Γ admits a spectral gap, so that Theorem 3.6 holds. For any real-valued $\Psi_1, \Psi_2 \in C_c^\infty(\Gamma\backslash G)$, if we set $\Psi_i^k(g) := \Psi_i(gk)$, then we can deduce from Theorem 3.6 that

$$\begin{aligned}
&\langle F_T, \Psi_1 \otimes \Psi_2\rangle_{\Gamma\backslash G\times\Gamma\backslash G}\\
&\quad= \int_{g\in B_T} \langle \Psi_1, g.\Psi_2\rangle_{L^2(\Gamma\backslash G)}\, dm^{\mathrm{Haar}}(g)\\
&\quad= \int_{k_1 a_t k_2\in B_T} \langle \Psi_1^{k_1^{-1}}, a_t.\Psi_2^{k_2}\rangle \cdot \xi(t)\, dt\, dk_1\, dk_2\\
&\quad= \tfrac{1}{|m^{\mathrm{BMS}}|} \int_{k_1 a_t k_2\in B_T} \big(m_*^{\mathrm{BR}}(\Psi_1^{k_1^{-1}}) m^{\mathrm{BR}}(\Psi_2^{k_2}) e^{\delta t} + O(e^{(\delta-\eta)t})\big)\, dt\, dk_1\, dk_2 \qquad (4.1)
\end{aligned}$$

for some $\eta > 0$, with the implied constant depending only on the Sobolev norms of Ψ_i's.

Therefore, if $F_T(e, e)$ can be *effectively approximated* by $\langle F_T, \Psi_\epsilon \otimes \Psi_\epsilon\rangle$ for an approximation of identity Ψ_ϵ in $\Gamma\backslash G$, a condition which depends on the regularity of the boundary of B_T relative to the Patterson–Sullivan density, then $F_T(e, e)$ can be computed by evaluating the integral (4.1) for $\Psi_1 = \Psi_2 = \Psi_\epsilon$ and by taking ϵ to be a suitable power of e^{-t}.

To state our counting theorem more precisely, we need:

Definition 4.2. Define a Borel measure $\mathcal{M}_G$ on G as follows: for $\psi \in C_c(G)$,

$$\mathcal{M}_G(\psi) = \frac{1}{|m^{\mathrm{BMS}}|} \int_{k_1 a_t k_2 \in K A^+ K} \psi(k_1 a_t k_2) e^{\delta t} d\nu_o(k_1)\, dt\, d\nu_o(k_2^{-1})$$

here ν_o is the M-invariant lift to K of the PS measure on $\partial(\mathbb{H}^n) = K/M$ viewed from $o \in \mathbb{H}^n$ with $K = \mathrm{Stab}_G(o)$.

Definition 4.3. For a family $\{B_T \subset G\}$ of compact subsets with $\mathcal{M}_G(B_T)$ tending to infinity as $T \to \infty$, we say that $\{B_T\}$ is *effectively well-rounded with respect to* Γ if there exists $p > 0$ such that for all small $\epsilon > 0$ and $T \gg 1$:

$$\mathcal{M}_G(B^+_{T,\epsilon} - B^-_{T,\epsilon}) = O(\epsilon^p \cdot \mathcal{M}_G(B_T))$$

where $B^+_{T,\epsilon} = G_\epsilon B_T G_\epsilon$ and $B^-_{T,\epsilon} = \cap_{g_1, g_2 \in G_\epsilon} g_1 B_T g_2$. Here G_ϵ denotes a symmetric ϵ-neighborhood of e in G.

For the next two theorems, we assume that $\Gamma' < \Gamma$ is a subgroup of finite index and both Γ and Γ' have spectral gaps.

Theorem 4.4 [Mohammadi and Oh 2012b]. *If* $\{B_T\}$ *is effectively well-rounded with respect to* Γ, *then there exists* $\eta_0 > 0$ *such that for any* $\gamma \in \Gamma$,

$$\#(\Gamma'\gamma \cap B_T) = \frac{1}{[\Gamma : \Gamma']} \mathcal{M}_G(B_T) + O(\mathcal{M}_G(B_T)^{1-\eta_0})$$

where $\eta_0 > 0$ *depends only on a uniform spectral gap for* Γ *and* Γ', *and the implied constant is independent of* Γ' *and* γ.

If $\|\cdot\|$ is a norm on the space $M_{n+1}(\mathbb{R})$ of $(n+1) \times (n+1)$ matrices, then the norm ball $B_T = \{g \in \mathrm{SO}(n,1) : \|g\| \le T\}$ is effectively well-rounded with respect to Γ, and hence Theorem 4.4 applies.

Consider the set $B_T = \Omega_1 A_T \Omega_2$ where $\Omega_i \subset K$ and $A_T = \{a_t : 0 \le t \le T\}$. Then $\{B_T\}$ is effectively well-rounded with respect to Γ if there exists $\beta' > 0$ such that the PS-measures of the ϵ-neighborhoods of the boundaries of $\Omega_1 M/M$ and $\Omega_2^{-1} M/M$ are at most of order $\epsilon^{\beta'}$ for all small $\epsilon > 0$. These conditions on Ω_i are satisfied for instance if their boundaries in K/M are disjoint from the limit set $\Lambda(\Gamma)$. But also many (but not all) compact subsets with piecewise smooth boundary also satisfy this condition (see Section 7 of [Mohammadi and Oh 2012b]).

Hence one can deduce the following:

Theorem 4.5 [Mohammadi and Oh 2012b]. *If* $\{B_T = \Omega_1 A_T \Omega_2\}$ *is effectively well-rounded with respect to* Γ, *then, as* $T \to \infty$,

$$\#(\Gamma'\gamma \cap \Omega_1 A_T \Omega_2) = \frac{\Xi(\Gamma, \Omega_1, \Omega_2)}{[\Gamma : \Gamma']} e^{\delta T} + O(e^{(\delta - \eta_0)T}),$$

where $\Xi(\Gamma, \Omega_1, \Omega_2) := \dfrac{\nu_o(\Omega_1)\nu_o(\Omega_2^{-1})}{\delta \cdot |m^{\mathrm{BMS}}|}$ *and the implied constant is independent of* Γ' *and* $\gamma \in \Gamma$.

The fact that the only dependence of Γ' on the right hand side of the above formula is on the index $[\Gamma : \Gamma']$ is of crucial importance for our intended applications to an affine sieve (page 197). Bourgain, Kontorovich and Sarnak [Bourgain et al. 2010b] showed Theorem 4.5 for the case $G = \mathrm{SO}(2, 1)$ via an explicit computation of matrix coefficients using the Gauss hypergeometric functions, and Vinogradov [2012] generalized their method to $G = \mathrm{SO}(3, 1)$. We note that in view of Theorem 3.2, the noneffective versions of Theorems 4.4 and 4.5 hold for any Zariski dense Γ with $|m^{\mathrm{BMS}}| < \infty$; this was obtained in [Roblin 2003] (see also [Oh and Shah 2013]) with a different proof. For Γ lattice, Theorem 4.5 is known in much greater generality [Gorodnik and Oh 2007].

5. Asymptotic distribution of $\Gamma\backslash\Gamma H a_t$

The counting problem in $H\backslash G$ for H a nontrivial subgroup can be approached via studying the asymptotic distribution of translates $\Gamma\backslash\Gamma H a_t$ as mentioned in the introduction. We will assume in this section that H is either a symmetric subgroup or a horospherical subgroup of G. Recall that H is symmetric means that H is the subgroup of fixed points under an involution of G. In this case, we can relate the distribution of $\Gamma\backslash\Gamma H a_t$ to the matrix coefficient function of $L^2(\Gamma\backslash G)$.

We fix a generalized Cartan decomposition $G = HAK$ (for H horospherical, it is just an Iwasawa decomposition) where K is a maximal compact subgroup and A is a one-dimensional subgroup of diagonalizable elements. As before, we parametrize $A = \{a_t : t \in \mathbb{R}\}$ so that the right multiplication by a_t on $G/M = \mathrm{T}^1(\mathbb{H}^n)$ corresponds to the geodesic flow for time t.

When H is horospherical, we will assume that $H = N^+$, i.e, the expanding horospherical subgroup with respect to a_t.

Any symmetric subgroup H of G is known to be locally isomorphic to $\mathrm{SO}(k, 1) \times \mathrm{SO}(n-k)$ for some $0 \le k \le n-1$, and the H-orbit of the identity coset in G/K (resp. G/M) is (resp. the unit normal bundle to) a complete totally geodesic subspace $\mathbb{H}^k$ of $\mathbb{H}^n$ of dimension k. The right multiplication by a_t on G/M corresponds to the geodesic flow and the image of Ha_t in G/M represents the expansion of the totally geodesic subspace of dimension k by distance t.

Measures on $\Gamma\backslash\Gamma H$ associated to a conformal density. The leading term in the description of the asymptotic distribution of $\Gamma\backslash\Gamma H a_t$ turns out to be a new measure on $\Gamma\backslash\Gamma H$ associated to the PS density (see [Oh and Shah 2013]). We assume that $\Gamma\backslash\Gamma H$ is closed in $\Gamma\backslash G$ in the rest of this section.

For a Γ-invariant conformal density $\{\mu_x\}$ on $\partial(\mathbb{H}^n)$ of dimension δ_μ, define a measure $\tilde{\mu}_H$ on $H/(H \cap M)$ by

$$d\tilde{\mu}_H([h]) = e^{\delta_\mu \beta_{[h]^+}(o,[h])}\, d\mu_o([h]^+)$$

where $o \in \mathbb{H}^n$ and $[h] \in H/(H \cap M)$ is considered as an element of $G/M = \mathrm{T}^1(\mathbb{H}^n)$ under the injective map $H/(H \cap M) \to G/M$.

By abuse of notation, we use the same notation $\tilde{\mu}_H$ for the $H \cap M$-invariant lift of $\tilde{\mu}_H$ to H: for $\psi \in C_c(\Gamma\backslash G)$,

$$\tilde{\mu}_H(\psi) = \int_{H/(H\cap M)} \psi^{H\cap M}(x)\, d\tilde{\mu}_H(x)$$

where $\psi^{H\cap M}(x) = \int_{H\cap M} \psi(xm)\, d_{H\cap M}(m)$ for the $H \cap M$-invariant probability measure $d_{H\cap M}$. This definition is independent of the choice of $o \in \mathbb{H}^n$ and the measure $\tilde{\mu}_H$ is $H \cap \Gamma$-invariant from the left, and hence induces a measure μ_H on $(H \cap \Gamma)\backslash H$ or equivalently on $\Gamma\backslash\Gamma H$.

For the PS density $\{\nu_x\}$ and the Lebesgue density $\{m_x\}$, the following two locally finite measures on $\Gamma\backslash\Gamma H$ are of special importance:

- Skinning measure: $\mu_H^{\mathrm{PS}} = \nu_H$.
- H-invariant measure: $\mu_H^{\mathrm{Leb}} = m_H$.

We remark that μ_H^{PS} is different from μ_H^{Leb} in general even when $H \cap \Gamma$ is a lattice in H.

Finiteness of the skinning measure μ_H^{PS}. The finiteness of the skinning measure μ_H^{PS} turns out to be the precise replacement for the finiteness of the volume measure μ_H^{Leb}, in extending the equidistribution statement from Γ lattices to thin subgroups.

When is the skinning measure μ_H^{PS} finite? This question is completely answered in [Oh and Shah 2013] for Γ geometrically finite. First, when H is a horospherical subgroup, the support of μ_H^{PS} is compact. When H is a symmetric subgroup, i.e., isomorphic to $\mathrm{SO}(k, 1) \times \mathrm{SO}(n-k)$ locally, the answer to this question depends on the notion of the parabolic corank of $\Gamma \cap H$: let $\Lambda_p(\Gamma)$ denote the set of all parabolic limit points of Γ. For $\xi \in \Lambda_p(\Gamma)$, the stabilizer Γ_ξ has a free abelian subgroup of finite index, whose rank is defined to be the rank of ξ (or the rank of Γ_ξ).

Definition 5.1. The parabolic corank of $\Gamma \cap H$ in Γ is defined to be the maximum of the difference $\mathrm{rank}(\Gamma_\xi) - \mathrm{rank}(\Gamma \cap H)_\xi$ over all $\xi \in \Lambda_p(\Gamma) \cap \partial(\mathbb{H}^k)$.

Theorem 5.2 [Oh and Shah 2013]. *Let Γ be geometrically finite.*

(1) μ_H^{PS} *is compactly supported if and only if the parabolic corank of* $\Gamma \cap H$ *is zero.*

(2) $|\mu_H^{\mathrm{PS}}| < \infty$ *if and only if* δ *is bigger than the parabolic corank of* $\Gamma \cap H$.

As we show $\operatorname{rank}(\Gamma_\xi) - \operatorname{rank}(\Gamma \cap H)_\xi \le n-k$ for all $\xi \in \Lambda_p(\Gamma) \cap \partial(\mathbb{H}^k)$, we have:

Corollary 5.3 [Oh and Shah 2013]. *If $\delta > (n-k)$, then $|\mu_H^{\mathrm{PS}}| < \infty$.*

For instance, if we assume $\delta > \frac{n-1}{2}$, then $|\mu_H^{\mathrm{PS}}| < \infty$ whenever $k \ge (n+1)/2$.

Distribution of $\Gamma\backslash\Gamma H a_t$. We first recall:

Theorem 5.4. *Suppose that Γ is a lattice in G and that $H \cap \Gamma$ is a lattice in H. In other words, $|m^{\mathrm{Haar}}| < \infty$ and $|m_H^{\mathrm{Leb}}| < \infty$. Then there exist $\eta_0 > 0$ (depending only on the spectral gap for Γ) and $\ell \in \mathbb{N}$ such that for any $\Psi \in C_c^\infty(\Gamma\backslash G)$, as $t \to \infty$,*

$$\int_{\Gamma\backslash\Gamma H} \Psi(ha_t)\, d\mu_H^{\mathrm{Leb}}(h) = \frac{|\mu_H^{\mathrm{Leb}}|}{|m^{\mathrm{Haar}}|} \cdot m^{\mathrm{Haar}}(\Psi) + O(\mathcal{S}_\ell(\Psi) \cdot e^{-\eta_0 t}).$$

In fact, this theorem holds in much greater generality of any connected semisimple Lie group: the noneffective statement is due to Duke et al. [1993] (also see [Eskin and McMullen 1993]). For the effective statement, see [Duke et al. 1993] for the case when $H \cap \Gamma\backslash H$ is compact and [Benoist and Oh 2012] in general.

An analogue of Theorem 5.4 for discrete groups that are not necessarily lattices is this:

Theorem 5.5 [Oh and Shah 2013]. *Let Γ be Zariski dense with $|m^{\mathrm{BMS}}| < \infty$. Suppose $|\mu_H^{\mathrm{PS}}| < \infty$. Then for any $\Psi \in C_c(\Gamma\backslash G)$,*

$$\lim_{t\to\infty} e^{(n-1-\delta)t} \int_{h\in\Gamma\backslash\Gamma H} \Psi(ha_t) d\mu_H^{\mathrm{Leb}}(h) = \frac{|\mu_H^{\mathrm{PS}}|}{|m^{\mathrm{BMS}}|} m^{\mathrm{BR}}(\Psi).$$

Theorem 5.6 [Mohammadi and Oh 2012b]. *Let Γ be a geometrically finite Zariski dense subgroup of G with a spectral gap (e.g., $\delta > n-2$). Suppose $|\mu_H^{\mathrm{PS}}| < \infty$. Then there exist $\eta_0 > 0$ and $\ell \in \mathbb{N}$ such that for any $\Psi \in C_c^\infty(\Gamma\backslash G)$, as $t \to \infty$,*

$$e^{(n-1-\delta)t} \int_{h\in\Gamma\backslash\Gamma H} \Psi(ha_t)\, d\mu_H^{\mathrm{Leb}}(h) = \frac{|\mu_H^{\mathrm{PS}}|}{|m^{\mathrm{BMS}}|} m^{\mathrm{BR}}(\Psi) + O(\mathcal{S}_\ell(\Psi) \cdot e^{-\eta_0 t}).$$

In the case when μ_H^{PS} is compactly supported, we can show that there exists a compact subset $\mathcal{O}_H$ (depending on Ψ) of $\Gamma\backslash\Gamma H$ such that for any $t \in \mathbb{R}$, $\Psi(ha_t) = 0$ for all $h \notin \mathcal{O}_H$. Therefore

$$\int \Psi(ha_t)\, d\mu_H^{\mathrm{Leb}} = \int_{\mathcal{O}_H} \Psi(ha_t)\, d\mu_H^{\mathrm{Leb}}.$$

Now the assumption on H being either symmetric or horospherical ensures the wave front property of [Eskin and McMullen 1993], which can be used to establish, as $t \to \infty$,

$$\int_{\mathcal{O}_H} \Psi(ha_t)\, d\mu_H^{\text{Leb}} \approx \langle a_t \Psi, \rho_{\mathcal{O}_H,\epsilon}\rangle_{L^2(\Gamma\backslash G)} \tag{5.7}$$

where $\rho_{\mathcal{O}_H,\epsilon} \in C_c^\infty(\Gamma\backslash G)$ is an ϵ-approximation of $\mathcal{O}_H$. Therefore the estimates on the matrix coefficients in Theorems 3.2 and 3.6 can be used to establish Theorems 5.5 and 5.6.

The case when μ_H^{PS} is not compactly supported turns out to be much more intricate, the main reason being that we are taking the integral with respect to μ_H^{Leb} as well as multiplying the weight factor $e^{(n-1-\delta)t}$ in the left hand side of Theorem 5.5, whereas the finiteness assumption is made on the skinning measure μ_H^{PS}. In this case, we first develop a version of thick-thin decomposition of the nonwandering set, that is, $\mathcal{W} := \{h \in \Gamma\backslash\Gamma H : ha_t \in \operatorname{supp}(\Psi)\}$, which resembles that of the support of μ_H^{PS}. This together with (5.7) takes care of the integral $\int \Psi_{ha_t} d\mu_H^{\text{Leb}}(h)$ over a thick part as well as a very thin part of $\mathcal{W}$. What is left is the integration over an intermediate range, which is investigated by comparing the two measures $(a_t)_*\mu_H^{\text{PS}}$ and $(a_t)_*\mu_H^{\text{Leb}}$ via the transversal intersections of the orbits $\Gamma\backslash\Gamma Ha_t$ with the weak-stable horospherical foliations . A key reason that this approach works is that these transversal intersections are governed by the topological properties of the orbit $\Gamma\backslash\Gamma Ha_t$, independent of the measures put on $\Gamma\backslash\Gamma H$.

In the special case of $n = 2, 3$ and H horospherical, Theorem 5.6 was proved in [Lee and Oh 2013] by a different method.

6. Distribution of Γ orbits in $H\backslash G$ and affine sieve

***Distribution of* Γ *orbits in* $H\backslash G$.** Let H be as in the previous section (i.e., symmetric or horospherical subgroup) and let $\{B_T\}$ be a family of compact subsets in $H\backslash G$ which is getting larger and larger as $T \to \infty$. We assume that $[e]\Gamma$ is discrete in $H\backslash G$. The study of the asymptotic of $\#([e]\Gamma \cap B_T)$ can be now approached by Theorems 5.5 and 5.6 via the following counting function on $\Gamma\backslash G$:

$$F_T(g) := \sum_{\gamma \in H\cap\Gamma\backslash\Gamma} \chi_{B_T}(\gamma g)$$

as $F_T(e) = \#([e]\Gamma \cap B_T)$. Depending on the subgroup H, we have $G = HA^+K$ or $G = HA^+K \cup HA^-K$. For the sake of simplicity, we will consider the sets B_T contained in HA^+K. We have for $g = ha_tk$, $dm^{\text{Haar}}(g) = \rho(t)\, d\mu^{\text{Leb}}(h)\, dt\, dk$ where $\rho(t) = e^{(n-1)t}(1 + O(e^{-\beta t}))$ for some $\beta > 0$. Let μ be a G-invariant measure on $H\backslash G$ normalized so that $dm^{\text{Haar}} = d\mu_H^{\text{Leb}} \otimes d\mu$ locally. Then, under the assumption that Γ has a spectral gap and $|\mu_H^{\text{PS}}| < \infty$, we deduce from

Theorem 5.6 that

$$\begin{aligned}\langle F_T, \Psi\rangle &= \int_{g\in B_T}\int_{h\in\Gamma\backslash\Gamma H} \Psi(hg)\, d\mu_H^{\mathrm{Leb}}(h)\, d\mu(g)\\ &= \int_{[e]a_tk\in B_T}\int_{h\in\Gamma\backslash\Gamma H} \Psi^k(ha_t) d\mu_H^{\mathrm{Leb}}(h)\rho(t)\, dt\, dk\\ &= \int_{[e]a_tk\in B_T}\left(\frac{|\mu_H^{\mathrm{PS}}|}{|m^{\mathrm{BMS}}|} m^{\mathrm{BR}}(\Psi^k)e^{\delta t} + O(e^{(\delta-\eta_0)t})\right) dt\, dk \qquad (6.1)\end{aligned}$$

for some $\eta_0 > 0$.

Therefore, similarly to the discussion in Section 4, if $F_T(e)$ can be *effectively approximated* by $\langle F_T, \Psi_\epsilon\rangle$ for an approximation of identity Ψ_ϵ in $\Gamma\backslash G$, a condition which depends on the regularity of the boundary of B_T relative to the PS density, then $F_T(e)$ can be computed by evaluating the integral (6.1) for $\Psi = \Psi_\epsilon$ and by taking ϵ a suitable power of e^{-t}.

For H horospherical or symmetric, we have

$$G = HA^+K$$

or

$$G = HA^+K \cup HA^-K$$

(as a disjoint union except for the identity element), where $A^\pm = \{a_{\pm t} : t \geq 0\}$.

Definition 6.2. Define a Borel measure $\mathcal{M}_{H\backslash G}$ on $H\backslash G$ as follows: for $\psi \in C_c(H\backslash G)$,

$$\mathcal{M}_{H\backslash G}(\psi) = \begin{cases} \dfrac{|\mu_H^{\mathrm{PS}}|}{|m^{\mathrm{BMS}}|}\displaystyle\int_{a_tk\in A^+K} \psi([e]a_tk)e^{\delta t} dt\, d\nu_o(k^{-1}) & \text{if } G = HA^+K,\\ \displaystyle\sum \dfrac{|\mu_{H,\pm}^{\mathrm{PS}}|}{|m^{\mathrm{BMS}}|}\int_{a_{\pm t}k\in A^\pm K} \psi([e]a_{\pm t}k)e^{\delta t} dt\, d\nu_o(k^{-1}) & \text{otherwise,}\end{cases}$$

where $o \in \mathbb{H}^n$ is the point fixed by K, ν_o is the right M-invariant measure on K, which projects to the PS-measure ν_o on $K/M = \partial(\mathbb{H}^n)$ and $\mu_{H,-}^{\mathrm{PS}}$ is the skinning measure on $\Gamma \cap H\backslash H$ in the negative direction.

Definition 6.3. For a family $\{B_T \subset H\backslash G\}$ of compact subsets with $\mathcal{M}_{H\backslash G}(B_T)$ tending to infinity as $T \to \infty$, we say that $\{B_T\}$ is *effectively well-rounded with respect to* Γ if there exists $p > 0$ such that for all small $\epsilon > 0$ and $T \gg 1$:

$$\mathcal{M}_{H\backslash G}(B_{T,\epsilon}^+ - B_{T,\epsilon}^-) = O(\epsilon^p \cdot \mathcal{M}_{H\backslash G}(B_T))$$

where $B_{T,\epsilon}^+ = B_T G_\epsilon$ and $B_{T,\epsilon}^- = \cap_{g\in G_\epsilon} B_T g$.

In the next two theorems 6.4 and 6.5, we assume that $\Gamma' < \Gamma$ is a subgroup of finite index with $H \cap \Gamma = H \cap \Gamma'$ and that both Γ and Γ' have spectral gaps.

Theorem 6.4 [Mohammadi and Oh 2012b]. *When H is symmetric, we assume that $|\mu_H^{\mathrm{PS}}| < \infty$. Suppose that $\{B_T\}$ is effectively well-rounded with respect to Γ. Then for any $\gamma \in \Gamma$, there exists $\eta_0 > 0$ (depending only on the uniform spectral gaps of Γ and Γ') such that*

$$\#([e]\Gamma'\gamma \cap B_T) = \frac{1}{[\Gamma : \Gamma']}\mathcal{M}_{H\backslash G}(B_T) + O(\mathcal{M}_{H\backslash G}(B_T)^{1-\eta_0})$$

with the implied constant independent of Γ' and $\gamma \in \Gamma$.

For a given family $\{B_T\}$, understanding its effective well-roundedness can be the hardest part of the work in general. However verifying the family of norm balls is effectively well-rounded is manageable (see Section 7 of [Mohammadi and Oh 2012b]). Consider an example of the family $\{B_T = [e]A_T\Omega\}$ where $\Omega \subset K$ and $A_T = \{a_t : 0 \le t \le T\}$. In this case, it is rather simple to formulate the effective well-rounded condition: there exists $\beta' > 0$ such that the PS-measure of the ϵ-neighborhood of $\partial(\Omega^{-1}M/M)$ is at most of order $\epsilon^{\beta'}$ for all small $\epsilon > 0$. As mentioned before, this holds when the boundary of $\Omega^{-1}M/M$ does not intersect the limit set $\Lambda(\Gamma)$ (see [loc. cit.] for a more general condition).

Then, setting

$$\Xi(\Gamma, \Omega) := \frac{|\mu_H^{\mathrm{PS}}| \cdot \nu_o(\Omega^{-1})}{\delta \cdot |m^{\mathrm{BMS}}|},$$

one can deduce this from (6.1):

Theorem 6.5 [Mohammadi and Oh 2012b]. *Under the same assumption on μ_H^{PS} and Γ', there exists $\eta_0 > 0$ (depending only on the uniform spectral gaps of Γ and Γ') such that for any $\gamma \in \Gamma$*

$$\#([e]\Gamma'\gamma \cap [e]A_T\Omega) = \frac{\Xi(\Gamma, \Omega)}{[\Gamma : \Gamma']}e^{\delta T} + O(e^{(\delta-\eta)T}),$$

with implied constant independent of Γ' and γ.

We note that in view of (4.1), the noneffective version of this theorem holds for any Zariski dense Γ with $|m^{\mathrm{BMS}}| < \infty$ and $|\mu_H^{\mathrm{PS}}| < \infty$, as proved in [Oh and Shah 2013].

Affine sieve. For applications to an affine sieve, we consider the case when the homogeneous space $H\backslash G$ is defined over $\mathbb{Z}$. More precisely, we assume that G is defined over $\mathbb{Z}$ and acts linearly and irreducibly on a finite-dimensional vector space W defined over $\mathbb{Z}$ in such a way that $G(\mathbb{Z})$ preserves $W(\mathbb{Z})$. Let $w_0 \in W(\mathbb{Z})$ be a nonzero vector such the stabilizer of w_0 is a symmetric subgroup or the stabilizer of the line $\mathbb{R}w_0$ is a parabolic subgroup. We set $V := w_0 G$. Let Γ be a geometrically finite Zariski dense subgroup of G with a spectral gap, which is contained in $G(\mathbb{Z})$.

Let F be an integer-valued polynomial on the orbit $w_0\Gamma$. Salehi Golsefidy and Sarnak [2013], generalizing [Bourgain et al. 2010a], showed that for some $R > 1$, the set of $\boldsymbol{x} \in w_0\Gamma$ such that $F(\boldsymbol{x})$ has at most R prime factors is Zariski dense in V.

For a square-free integer d, let $\Gamma_d < \Gamma$ be a subgroup which contains $\{\gamma \in \Gamma : \gamma \equiv e \bmod d\}$ and satisfies $\mathrm{Stab}_{\Gamma_d}(w_0) = \mathrm{Stab}_\Gamma(w_0)$. For instance, we can set $\Gamma_d = \{\gamma \in \Gamma : w_0\gamma \equiv w_0 \bmod d\}$.

We say the family $\{\Gamma_d\}$ has a uniform spectral gap if $\sup_d s_0(\Gamma_d) < \delta$ and $\sup_d n_0(\Gamma_d) < \infty$ (see Definition 3.5 for notation).

Salehi Golsefidy and Varjú [2012], generalizing [Bourgain et al. 2010a], showed that the family of Cayley graphs of Γ/Γ_d's (with respect to the projections of a fixed symmetric generating set of Γ) forms expanders as d runs through square-free integers with large prime factors. If $\delta > \frac{n-1}{2}$, the transfer property from the combinatorial spectral gap to the archimedean one established in [Bourgain et al. 2011] (see also [Kim 2012]) implies that the family $\{\Gamma_d : d \text{ square-free}\}$ has a uniform *spherical* spectral gap. Together with the classification of $\hat{G}$, it follows that $\{\Gamma_d\}$ admits a uniform spectral gap if $\delta > \frac{n-1}{2}$ for $n = 2, 3$ or if $\delta > n - 2$ for $n \geq 4$.

For the following discussion, we assume that there is a finite set of primes S such that the family Γ_d with d square-free with no prime factors in S admits a uniform spectral gap. Then we can apply Theorem 6.5 to Γ_d's. By a recent development on the affine linear sieve on homogeneous spaces ([Bourgain et al. 2010a], [Nevo and Sarnak 2010]), we are able to deduce: let F be an integer-valued polynomial on $w_0\Gamma$. Letting $F = F_1F_2 \cdots F_r$ be a factorization into irreducible polynomials, assume that all F_j's are integral on $w_0\Gamma$ and distinct from each other. Let λ be the log of the largest eigenvalue of a_1 on the $\mathbb{R}$-span of w_0G.

Theorem 6.6 [Mohammadi and Oh 2012b]. *For any norm $\|\cdot\|$ on V,*

(1) $\#\{\boldsymbol{x} \in w_0\Gamma : \|\boldsymbol{x}\| < T,\ F_j(\boldsymbol{x}) \text{ is prime for } j = 1, \ldots, r\} \ll \dfrac{T^{\delta/\lambda}}{(\log T)^r}$;

(2) *there exists $R = R(F, w_0\Gamma) \geq 1$ such that*

$$\#\{\boldsymbol{x} \in w_0\Gamma : \|\boldsymbol{x}\| < T,\ F(\boldsymbol{x}) \text{ has at most } R \text{ prime factors}\} \gg \frac{T^{\delta/\lambda}}{(\log T)^r}.$$

The significance of $T^{\delta/\lambda}$ in the above theorem is that for $H = \mathrm{Stab}_G(w_0)$,

$$\mathcal{M}_{H\backslash G}\{w \in w_0G : \|w\| < T\} \asymp \#\{\boldsymbol{x} \in w_0\Gamma : \|\boldsymbol{x}\| < T\} \asymp T^{\delta/\lambda}.$$

In view of the above discussion, it is also possible to state Theorem 6.6 for more general sets B_T, instead of the norm balls, which then provides a certain uniform distribution of almost prime vectors.

When Γ is an arithmetic subgroup of a simply connected semi-simple algebraic $\mathbb{Q}$-group G, and H is a symmetric subgroup, the analogue of Theorem 6.6 was obtained in [Benoist and Oh 2012]. Strictly speaking, Theorem 1.3 of that reference is stated only for a fixed group; however it is clear from its proof that the statement also holds uniformly over its congruence subgroups with the correct main term. Based on this, one can use the combinatorial sieve to obtain an analogue of Theorem 6.6, as was done for a group variety in [Nevo and Sarnak 2010]. Theorem 6.6 on lower bound for arithmetic was obtained in [Gorodnik and Nevo 2012] further assuming that $H \cap \Gamma$ is cocompact in H.

7. Application to sphere packings

In this section we will discuss counting problems for sphere packings in $\mathbb{R}^n$ as an application of an orbital counting problem for thin subgroups of $\mathrm{SO}(n, 1)$. By a sphere packing in the Euclidean space $\mathbb{R}^n$ for $n \geq 1$, we simply mean a union of (possibly intersecting) $(n-1)$-dimensional spheres; here an $(n-1)$-dimensional plane is regarded as a sphere of infinite radius.

Fixing a sphere packing $\mathcal{P}$ in $\mathbb{R}^n$, a basic problem is to understand the asymptotic of the number $\#\{S \in \mathcal{P} : \mathrm{Radius}(S) > t\}$ or equivalently $\#\{S \in \mathcal{P} : \mathrm{Vol}(S) > t\}$ where $\mathrm{Vol}(S)$ means the volume of the ball enclosed by S. For the sake of brevity, we will simply refer $\mathrm{Vol}(S)$ to the volume of S. We will consider this problem in a more general setting, that is, allowing the volume of the sphere S to be computed in various conformal metrics.

Let U be an open subset of $\mathbb{R}^n$ and f a positive continuous function on U. A conformal metric associated to the pair (U, f) is a metric on U of the form $f(x)\,dx$, where dx is the Euclidean metric on $\mathbb{R}^n$. Given a conformal metric (U, f), we set $\mathrm{Vol}_f(S) := \int_B f(x)^n dx$ where B is the ball enclosed by S. When S is a plane, we put $\mathrm{Vol}_f(S) = \infty$.

For any compact subset E of U and $t > 0$, set

$$N_t(\mathcal{P}, f, E) := \#\{S \in \mathcal{P} : \mathrm{Vol}_f(S) > t,\ E \cap S \neq \varnothing\}.$$

For simplicity, we will omit $\mathcal{P}$ in the notation $N_t(\mathcal{P}, f, E)$. When $(U, f) = (\mathbb{R}^n, 1)$, i.e., the standard Euclidean metric, we simply write $N_t(E)$ instead of $N_t(1, E)$. We call $\mathcal{P}$ *locally finite* if $N_t(E) < \infty$ for all E bounded.

In order to approach the problem of the computation of the asymptotic of $N_t(f, E)$, a crucial condition is that $\mathcal{P}$ admits enough symmetries of Mobius transformations of $\mathbb{R}^n$. By the Poincaré extension, the group $\mathrm{MG}(\mathbb{R}^n)$ of Mobius transformations of $\mathbb{R}^n$ can be identified with the isometry group of the upper half space $\mathbb{H}^{n+1} = \{(z, r) : z \in \mathbb{R}^n, r > 0\}$. We assume that $\mathcal{P}$ is invariant under a nonelementary discrete subgroup Γ of $G := \mathrm{Isom}^+(\mathbb{H}^{n+1})$. As before, let δ

denote the critical exponent of Γ and $\{\nu_x : x \in \mathbb{H}^{n+1}\}$ a Patterson–Sullivan density for Γ.

We recall from [Oh and Shah 2012]:

Definition 7.1 (Γ-skinning size of $\mathcal{P}$). For a sphere packing $\mathcal{P}$ invariant under Γ, define $0 \le \mathrm{sk}_\Gamma(\mathcal{P}) \le \infty$ as follows:

$$\mathrm{sk}_\Gamma(\mathcal{P}) := \sum_{i \in I} \int_{s \in \mathrm{Stab}_\Gamma(S_i^\dagger)\backslash S_i^\dagger} e^{\delta\beta_{s^+}(o,s)} d\nu_o(s^+)$$

where $o \in \mathbb{H}^{n+1}$, $\{S_i : i \in I\}$ is a set of representatives of Γ-orbits in $\mathcal{P}$, and $S_i^\dagger \subset \mathrm{T}^1(\mathbb{H}^{n+1})$ is the set of unit normal vectors to the convex hull of S_i.

Definition 7.2. For the pair (U, f), we define a Borel measure $\omega_{\Gamma,f}$ on U: for $\psi \in C_c(U)$ and for $o \in \mathbb{H}^{n+1}$,

$$\omega_{\Gamma,f}(\psi) = \int_{z \in U} \psi(z) f(z)^\delta e^{\delta\beta_z(o,(z,1))}\, d\nu_o(z).$$

Alternatively, we have the simple formula

$$d\omega_{\Gamma,f} = f(z)^\delta (|z|^2 + 1)^\delta d\nu_{e_{n+1}},$$

where $e_{n+1} = (0, \dots, 0, 1) \in \mathbb{H}^{n+1}$.

Example 7.3. (1) For the spherical metric $(\mathbb{R}^n, 2/(1 + |z|^2))$ (also called the chordal metric) on $\mathbb{R}^n$, $\mathrm{Vol}_f(S)$ is the spherical volume of the ball enclosed by S and $d\omega_{\Gamma,f} = 2^\delta \cdot d\nu_{e_{n+1}}$.

(2) For the hyperbolic metric $(\mathbb{H}^n = \{z \in \mathbb{R}^n : z_n > 0\}, 1/z_n)$, $\mathrm{Vol}_f(S)$ is the hyperbolic volume of the ball enclosed by S and

$$d\omega_{\Gamma,f} = \frac{(1 + |z|^2)^\delta}{z_n^\delta} d\nu_{e_{n+1}}.$$

Definition 7.4. By an *infinite bouquet of tangent spheres glued at a point* $\xi \in \mathbb{R}^n \cup \{\infty\}$, we mean a union of two collections, each consisting of infinitely many pairwise internally tangent spheres with the common tangent point ξ and their radii tending to 0, such that the spheres in each collection are externally tangent to the spheres in the other at ξ.

We denote by $v_n := \mathrm{Vol}(B(0, 1))$ the Euclidean volume of the unit ball $B(0, 1) := \{x \in \mathbb{R}^n : \|x\| < 1\}$; v_n is equal to $(2\pi)^{n/2}/(2 \cdot 4 \cdots n)$ if n is even and $2(2\pi)^{(n-1)/2}/(1 \cdot 3 \cdots n)$ if n is odd.

Theorem 7.5. *Let $\mathcal{P}$ be a locally finite sphere packing invariant under a geometrically finite group Γ with finitely many Γ-orbits. In the case of* $\delta \le 1$, *we also* *assume that $\mathcal{P}$ has no infinite bouquet of spheres glued at a parabolic fixed point of* Γ.

Then for any conformal metric (U, f) *and for any compact subset* E *of* U *whose boundary has zero Patterson–Sullivan density, as* $t \to 0$,

$$\lim_{t\to 0} N_t(f, E)t^{\delta/n} = \frac{\mathrm{sk}_\Gamma(\mathcal{P}) \cdot v_n^{\delta/n} \cdot \omega_{\Gamma,f}(E)}{\delta \cdot |m^{\mathrm{BMS}}|}.$$

The assumption on the nonexistence of an infinite bouquet in $\mathcal{P}$ is to ensure that the Γ-skinning size for $\mathcal{P}$ is finite.

In the case when $(U, f) = (\mathbb{R}^n, 1)$, Theorem 7.5 was proved in [Oh and Shah 2012] for the case of $n = 2$ and and the proof given there extends easily for any $n \geq 2$ using the general equidistribution result in [Oh and Shah 2013]. See also [Oh and Shah 2010] for the case when f defines the spherical metric. To give an interpretation of Theorem 7.5 as a special case of the orbital counting problem discussed at the beginning of Section 6, fixing H to be the stabilizer of a sphere S_0 in $\mathcal{P}$, we may think of $H\backslash G$ as the space of all totally geodesic planes in $\mathbb{H}^{n+1}$. Then a key point is to describe a particular subset $B_t(E)$ in $H\backslash G$ such that $N_t(E)$ is same as $\#[e]\Gamma \cap B_t(E)$.

The extension to a general conformal metric (U, f) is possible basically due to the uniform continuity of f on a compact subset E and a covering argument. We give a brief sketch as follows (the argument below was established in a discussion with Shah): denote by $Q_z(\eta)$ the cube $\{z' \in \mathbb{R}^n : \max_{1\leq i\leq n} |z_i - z_i'| \leq \eta\}$ centered at $z \in \mathbb{R}^n$ with radius η.

First, Theorem 7.5 for $f = 1$, together with the uniform continuity of f on E, implies that for any $\epsilon > 0$, there exists $\eta = \eta(\epsilon) > 0$ (depending only on E and f) such that for any cube $Q_z(\eta)$ centered at $z \in E$

$$\frac{N_t(f, Q_z(\eta))t^{\delta/n}}{v_n^{\delta/n}} = (1 + O(\epsilon))\frac{\mathrm{sk}_\Gamma(\mathcal{P})}{\delta \cdot |m^{\mathrm{BMS}}|} f(z)^\delta \omega_\Gamma(Q_z(\eta)). \tag{7.6}$$

Let k be the minimal integer such that a k-dimensional sphere, say, P charges a positive PS density. As the PS density is atom-free, we have $k > 0$ and the limit set $\Lambda(\Gamma)$ is contained in P (see [Roblin 2003, Proposition 3.1]). We cover $E \cap P$ with cubes $\{Q_z(\eta) : z \in I_\eta\}$ with disjoint interiors for a finite subset I_η of $E \cap P$. As each cube is centered at a point of P which is a sphere, the intersection of its boundary with P is contained in a $(k-1)$-dimensional sphere for all small $\eta > 0$. It follows that the boundary of each cube has zero PS density by the minimality assumption on k. Let $\mathscr{C}(\eta) := \{Q_z(\eta) : z \in \tilde{I}_\eta\}$ be a covering of E with $\tilde{I}_\eta \supset I_\eta$. Note that the boundary of each cube in $\mathscr{C}(\eta)$ has zero PS density. We can find compact subsets $E_\epsilon^\pm$ of U with $E_\epsilon^- \subset E \subset E_\epsilon^+$ and a positive integer m_ϵ so that

$\omega_{\Gamma,f}(E_\epsilon^+ - E_\epsilon^-) < \epsilon$ and E_ϵ^+ (resp. E) contains all cubes centered at E (resp. E_ϵ^-) and of size less than η possibly except at most m_ϵ number of such cubes.

We may also assume that $f(z) = (1 + O(\epsilon))f(z')$ for all $z' \in Q_z(\eta) \in \mathscr{C}(\eta)$ where the implied constant is uniform for all $z \in E$. We may now apply Equation (7.6) for $f = 1$ to each cube in the covering of $\mathscr{C}(\eta)$ to obtain that

$$\begin{aligned}\frac{N_t(f,E)t^{\delta/n}}{v_n^{\delta/n}} &= (1+O(\epsilon))\frac{\mathrm{sk}_\Gamma(\mathscr{P})}{\delta\cdot|m^{\mathrm{BMS}}|}\sum_{z\in\tilde{I}_\eta} f(z)^\delta\omega_\Gamma(Q_z(\eta))\\ &= (1+O(\epsilon))\frac{\mathrm{sk}_\Gamma(\mathscr{P})}{\delta\cdot|m^{\mathrm{BMS}}|}\omega_{\Gamma,f}(E_\epsilon^+) + O(\epsilon)\\ &= (1+O(\epsilon))\frac{\mathrm{sk}_\Gamma(\mathscr{P})}{\delta\cdot|m^{\mathrm{BMS}}|}\omega_{\Gamma,f}(E) + O(\epsilon).\end{aligned}$$

As $\epsilon > 0$ is arbitrary, this proves Theorem 7.5.

8. On Apollonian circle packings

Construction. In the case of Apollonian circle packings in the plane $\mathbb{R}^2 = \mathbb{C}$, Theorem 7.5 can be made more explicit as the measure $\omega_{\Gamma,f}$ turns out to be the δ-dimensional Hausdorff measure with respect to (U, f), restricted to $\Lambda(\Gamma)$.

We begin by recalling Apollonian circle packings, whose construction is based on the following theorem of Apollonius of Perga:

Theorem 8.1 (Apollonius, 200 BC). *Given* 3 *mutually tangent circles in the plane* (*with distinct tangent points*), *there exist precisely two circles tangent to all three circles.*

Consider four mutually tangent circles in the plane with distinct points of tangency. By Apollonius' theorem, one can add four new circles each of which is tangent to three of the given ones. Continuing to repeatedly add new circles tangent to three of the previous circles, we obtain an infinite circle packing, called an *Apollonian circle packing*. Figure 1 shows the first three generations of this procedure where each circle is labeled with its curvature (the reciprocal of its radius).

Apollonian packing and Hausdorff measure. We start by recalling:

Definition 8.2. For $s > 0$, the s-dimensional Hausdorff (also known as covering) measure $\mathscr{H}^s$ of a closed subset E of $\mathbb{R}^2$ is defined as follows:

$$\mathscr{H}^s(E) := \lim_{\epsilon\to 0}\ \inf\left\{\sum_{i\in I}\mathrm{diam}(D_i)^s : E \subset \bigcup_{i\in I} D_i,\ \mathrm{diam}(D_i) \le \epsilon\right\}.$$

The Hausdorff dimension of E is then given as

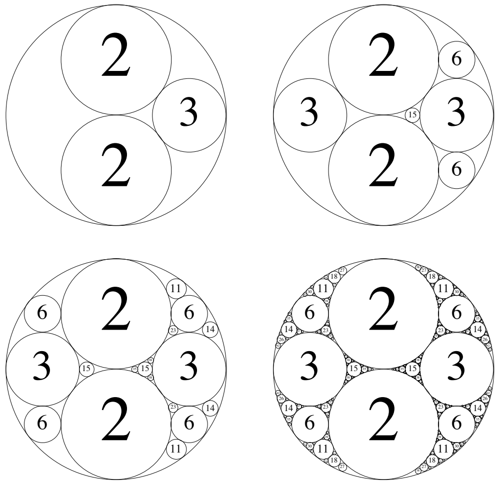

Figure 1. An Apollonian circle packing.

$$\dim_{\mathcal{H}}(E) := \sup\{s : \mathcal{H}^s(E) = \infty\} = \inf\{s \geq 0 : \mathcal{H}^s(E) = 0\}.$$

For s a positive integer, the s-dimensional Hausdorff measure is proportional to the usual Lebesgue measure on $\mathbb{R}^s$.

For an Apollonian circle packing $\mathcal{P}$, the residual set $\mathrm{Res}(\mathcal{P})$ is defined to be the closure of the union of all circles in $\mathcal{P}$. Its Hausdorff dimension, say α, is independent of $\mathcal{P}$ and known to be approximately 1.30568(8) [McMullen 1998].

Theorem 8.3. *Let $\mathcal{P}$ be any Apollonian circle packing. For any conformal metric (U, f) and for any compact subset $E \subset U$ with smooth boundary, we have*

$$\lim_{t\to 0} t^{\alpha/2} \cdot \#\{C \in \mathcal{P} : \mathrm{area}_f(C) > t,\, C \cap E \neq \varnothing\} = c_A \cdot \mathcal{H}^\alpha_f(\mathrm{Res}(\mathcal{P}) \cap E)$$

where $c_A > 0$ is independent of $\mathcal{P}$ and $d\mathcal{H}^\alpha_f(z) = f(z)^\alpha \cdot d\mathcal{H}^\alpha(z)$.

The symmetry group $\Gamma_{\mathcal{P}} := \{g \in \mathrm{PSL}_2(\mathbb{C}) : g(\mathcal{P}) = \mathcal{P}\}$ satisfies the following:

(1) $\Gamma_{\mathcal{P}}$ is geometrically finite.

(2) The limit set of $\Gamma_{\mathcal{P}}$ coincides with $\mathrm{Res}(\mathcal{P})$; in particular, its critical exponent is α.

(3) There are only finitely many $\Gamma_{\mathcal{P}}$-orbits of circles in $\mathcal{P}$.

Let $\nu_{\mathcal{P},j}$ denote the PS measure viewed from $j = (0, 0, 1) \in \mathbb{H}^3$ for the group $\Gamma_{\mathcal{P}}$. As $\Gamma_{\mathcal{P}}$ has no rank 2 parabolic limit points and $\alpha > 1$, Sullivan's work [Sullivan 1984] implies that the α-dimensional Hausdorff measure $\mathcal{H}^\alpha$ is a locally finite measure on $\mathrm{Res}(\mathcal{P})$ and that

$$\frac{1}{|\nu_{\mathcal{P},j}|}(|z|^2+1)^{\alpha}\, d\nu_{\mathcal{P},j} = d\mathcal{H}^{\alpha}.$$

Therefore Theorem 8.3 is a special case of Theorem 7.5. Moreover the constant c_A is given by

$$c_A = \frac{\pi^{\alpha/2}\cdot \mathrm{sk}_{\Gamma_{\mathcal{P}}}(\mathcal{P})\cdot |\nu_{\mathcal{P},j}|}{\alpha\cdot |m^{\mathrm{BMS}}|}$$

for any Apollonian circle packing $\mathcal{P}$. We propose to call c_A the Apollonian constant.

Problem 8.4. Compute (or estimate) c_A!

When $\mathcal{P}$ is a bounded Apollonian circle packing, the existence of the asymptotic formula $\#\{C\in\mathcal{P} : \mathrm{area}(C) > t, \} \sim c_{\mathcal{P}}\cdot t^{\alpha/2}$ was first shown in [Kontorovich and Oh 2011] without an error term, and later in [Lee and Oh 2013] with an error term (see also [Vinogradov 2012]). In view of [Oh and Shah 2012], Theorem 5.6 can be used to prove an effective circle count in a compact region E for general Apollonian packings, provided the boundary of E satisfies a regularity property.

There is also a beautiful arithmetic aspect of Apollonian circle packings which is entirely omitted in this article; see [Sarnak 2011; Oh 2011; Bourgain and Fuchs 2011; Bourgain and Kontorovich 2012].

Apollonian sphere packing for n = 3. Given $n+1$ mutually tangent spheres in $\mathbb{R}^n$ with disjoint interiors, it is known that there is a unique sphere, called a *dual sphere*, passing through their points of tangency and orthogonal to all $n+1$ spheres [Graham et al. 2006, Theorem 7.1]. Hence for $n+2$ mutually tangent spheres with disjoint interiors $S_1,\dots,S_{n+2}$ in $\mathbb{R}^n$, there are $n+2$ dual spheres, say, $\tilde{S}_1,\dots,\tilde{S}_{n+2}$. The Apollonian group $\mathcal{A}=\mathcal{A}(S_1,\dots,S_{n+2})$ is generated by the inversions with respect to $\tilde{S}_i$, $1\le i\le n+2$.

Only for $n=2$ or 3, the Apollonian group $\mathcal{A}$ is a discrete subgroup of $\mathrm{MG}(\mathbb{R}^n)$ [Graham et al. 2006, Theorem 4.1] and in this case its orbit $\mathcal{P} := \bigcup_{i=1}^{n+2}\mathcal{A}(S_i)$ consists of spheres with disjoint interiors. For $n=2$, $\mathcal{P}$ is an Apollonian circle packing. For $n=3$, $\mathcal{P}$ is called an Apollonian sphere packing. Note that $\mathcal{A}$ is geometrically finite and that $\mathcal{P}$ is locally finite, as the spheres in $\mathcal{P}$ have disjoint interiors, Hence Theorem 7.5 applies to $\mathcal{P}$. The critical exponent of $\mathcal{A}$ for $n=3$ has been estimated to be 2.473946(5) in [Borkovec et al. 1994].

Dual apollonian cluster ensemble for any n ≥ 2. Given $n+2$ mutually tangent spheres with disjoint interiors $S_1,\dots,S_{n+2}$ in $\mathbb{R}^n$, let $\mathcal{A}^*$ denote the group generated by the inversions with respect to S_i, $1\le i\le n+2$. The dual Apollonian group $\mathcal{A}^*$ is a discrete geometrically finite subgroup of $\mathrm{MG}(\mathbb{R}^n)$ for all $n\ge 2$

and the orbit $\mathcal{P} := \bigcup_{i=1}^{n+2} \mathcal{A}^*(S_i)$ is a sphere packing, in our sense, consisting of spheres nested in S_i's. We note that $\mathcal{P}$ is locally finite, as nested spheres are getting smaller and smaller and hence Theorem 7.5 applies to $\mathcal{P}$.

9. Packing circles of the ideal triangle in $\mathbb{H}^2$

Consider an ideal triangle $\mathcal{T}$ in $\mathbb{H}^2$ i.e., a triangle whose sides are hyperbolic lines connecting vertices on the boundary of $\mathbb{H}^2$. An ideal triangle exists uniquely up to hyperbolic congruences. Consider $\mathcal{P}(\mathcal{T})$ the circle packing of an ideal triangle by filling in largest inner circles. The notation $\overline{\mathcal{P}(\mathcal{T})}$ denotes the closure of $\mathcal{P}(\mathcal{T})$ and $\text{area}_{\text{Hyp}}(C)$ is the hyperbolic area of the disk enclosed by C.

Theorem 9.1 (packing circles of the ideal triangle). *Let $\mathcal{T}$ be the ideal triangle of $\mathbb{H}^2$. Then*

$$\lim_{t\to 0} t^{\alpha/2} \cdot \#\{C \in \mathcal{P}(\mathcal{T}) : \text{area}_{\text{Hyp}}(C) > t\} = c_A \cdot \int_{\overline{\mathcal{P}(\mathcal{T})}} y^{-\alpha}\, d\mathcal{H}^\alpha(z)$$

where c_A denotes the Apollonian constant.

Fix the Apollonian circle packing $\mathcal{P}_0$ generated by two vertical lines $x = \pm 1$ and the unit circle $\{|z| = 1\}$. The corresponding Apollonian group $\Gamma_0 = \Gamma(\mathcal{P}_0)$ is generated by the inversions with respect to horizontal lines $y = 0$ and $y = -2i$ and the circles $\{|z - (\pm 1 - i)| = 1\}$. We set $\mathbb{R}^2_+ := \{x + iy : y > 0\}$. Now for the conformal metric $(U, f) = (\mathbb{R}^2_+, 1/y) = \mathbb{H}^2$, we note that $\{C \in \mathcal{P}(\mathcal{T}) : \text{area}_{\text{Hyp}}(C) > t\} = \{C \in \mathcal{P}_0 : \text{area}_f(C) > t,\ C \cap \mathcal{T} \neq \varnothing\}$.

However Theorem 9.1 does not immediately follow from Theorem 8.3 since the ideal triangle $\mathcal{T}$ is not a compact subset of $\mathbb{H}^2$.

We need to understand the $\mathcal{H}^\alpha_f$-measure of neighborhoods of cusps in the triangle for $f = 1/y$. For the next two theorems, consider a conformal metric $(\mathbb{R}^2_+, f)$.

Theorem 9.2. *If $f(x + iy) \ll y^{-k}$ for some real number $k > \alpha^{-1}$ with implied constant independent of $|x| \le 1$, then for any $\eta > 0$,*

$$\mathcal{H}^\alpha_f\{|x| \le 1,\ y > \eta\} < \infty.$$

Moreover for any Borel subset $E \subset \{|x| \le 1,\ y > \eta\}$ (not necessarily compact) with smooth boundary,

$$\lim_{t\to\infty} t^{\alpha/2} \cdot \#\{C \in \mathcal{P}_0 : \text{area}_f(C) > t,\ C \cap E \neq \varnothing\} \sim c_A \cdot \mathcal{H}^\alpha_f(E).$$

Proof. It suffices to show the claim for $\eta = 1$, since $\{|x| \le 1, \eta \le y \le 1\}$ is a compact subset. So we put $\eta = 1$ and set $U_R := \{|x| \le 1,\ y > R\}$. Define $F_t(E) := \{C \in \mathcal{P}_0 : \text{area}_f(C) > t,\ C \cap E \neq \varnothing\}$ and $E_n := \{|x| \le 1,\ n \le y < n+1\}$. Then $F_t(U_1) = \bigcup_{n\ge 1} F_t(E_n)$.

For $C \in F_t(E_n)$, $C-(n-1)i \in F_t(E_1)$ and

$$\begin{aligned}\operatorname{area}(C-(n-1)i) &= \int_{C-(n-1)i} f(z)^2 dz \\ &= \int_C f(z+(n-1)i))^2 dz \ll n^{-2k}\operatorname{area}(C).\end{aligned}$$

Hence we get an injective map $F_t(E_n)$ to $F_{tn^{-2k}}(E_1)$ and hence for $R_0 \geq 1$,

$$\#F_t(U_{R_0}) = \sum_{n\geq R_0} \#F_{tn^{-2k}}(E_1).$$

By Theorem 8.3, for $\epsilon > 0$, there exists t_ϵ such that for all $t < t_\epsilon$,

$$\#F_t(E_1) \leq t^{-\alpha/2} c_A(\mathcal{H}^\alpha_f(E_1)+\epsilon)$$

and hence there exists $N_\epsilon > 1$ such that for all $t \leq 1$ and $n > N_\epsilon$,

$$\#F_{tn^{-2k}}(E_1) \leq t^{-\alpha/2} n^{-k\alpha} c_A(\mathcal{H}^\alpha_f(E_1)+\epsilon)$$

and hence

$$\#F_t(U_{N_\epsilon})t^{\alpha/2} \leq \left(\sum_{n\geq N_\epsilon} n^{-k\alpha}\right) c_A(\mathcal{H}^\alpha_f(E_1)+\epsilon).$$

Since $\left(\sum_{n\geq N_\epsilon} n^{-k\alpha}\right) < \infty$ as $k\alpha > 1$, if $N_\epsilon \gg 1$ is sufficiently large, we can make $\#F_t(U_{N_\epsilon})t^{\alpha/2} \leq \epsilon$ and $\mathcal{H}^\alpha_f(U_{N_\epsilon}) < \epsilon$. This implies $\mathcal{H}^\alpha_f(U_{N_\epsilon}) < \infty$. Moreover,

$$\limsup \#F_t(U_1)t^{\alpha/2} = \limsup \# \sum_{1\leq n<N_\epsilon} F_t(E_n)t^{\alpha/2} + O(\epsilon) = c_A \cdot \mathcal{H}^\alpha_f(U_1) + O(\epsilon).$$

As $\epsilon > 0$ is arbitrary, we have

$$\limsup \#F_t(U_1)t^{\alpha/2} = c_A \cdot \mathcal{H}^\alpha_f(U_1).$$

Similarly we have $\liminf \#F_t(U_1)t^{\alpha/2} = c_A \cdot \mathcal{H}^\alpha_f(U_1)$. Hence

$$\lim \#F_t(U_1)t^{\alpha/2} = c_A \cdot \mathcal{H}^\alpha_f(U_1).$$

In the same way, we can deduce the claim for any Borel set E using the fact that $\mathcal{H}^\alpha_f(E \cap U_{N_\epsilon}) = O(\epsilon)$. □

We note that the boundary of $\mathbb{R}^2_+$ meets with $\operatorname{Res}(\mathcal{P}_0)$ at three points $1, -1, \infty$ and these three points are in one Γ_0-orbit. Note that Γ_0 contains an element γ_0: $\gamma_0(x+yi) = x+(y+2)i$.

Theorem 9.3. *Let $(\mathbb{R}^2_+, f)$ be a conformal metric such that $f(x+iy) \asymp y^{-k}$ for $k \in \mathbb{R}$.*

(1) *If $\alpha^{-1} < k < 2-\alpha^{-1}$, we have $\mathcal{H}^\alpha_f(\mathbb{R}^2_+) < \infty$.*

(2) *If either $k < \alpha^{-1}$ or $k > 2-\alpha^{-1}$, we have $\mathcal{H}^\alpha_f(\mathbb{R}^2_+) = \infty$.*

Proof. Let $\gamma(z) = \frac{\bar z - i}{\bar z - 1 - i}$. Then $\gamma \in \Gamma_0$ and $\gamma(\infty) = 1$. Hence if we set $U_\eta := \{|x| \le 1,\, y > \eta\}$, $\gamma(U_\eta)$ is a neighborhood of 1 for all large $\eta > 1$. We will show that $\mathcal{H}^\alpha_f(\gamma(U_\eta)) < \infty$ if $k < 2 - \alpha^{-1}$.

We have

$$\mathcal{H}^\alpha_f(\gamma(U_\eta)) = \int_{z\in\gamma(U_\eta)} f(z)^\alpha \, d\mathcal{H}^\alpha(z) = \int_{w\in U_\eta} f(\gamma(w))^\alpha |\gamma'(w)|^\alpha \, d\mathcal{H}^\alpha(w).$$

Since $|\gamma'(w)| = |w - 1 + i|^{-2}$ and $\Im(\gamma(w)) \asymp y^{-1}$, we have

$$f(\gamma(w))|\gamma'(w)| \asymp y^{k-2}.$$

Hence by Theorem 9.2, if $2 - k > \alpha^{-1}$,

$$\mathcal{H}^\alpha_f(\gamma(U_\eta)) < \infty.$$

The remaining cases can be proved similarly and we leave them to the reader. □

Note that $\mathcal{T} := \mathcal{P}_0 \cap \mathbb{R}^2_+$ is a circle packing of the curvilinear triangle made by largest inner circles.

As $\alpha > 1$, Theorem 9.1 is a special case of the following:

Theorem 9.4. *Let* $(\mathbb{R}^2_+, f)$ *be a conformal metric such that* $f(x+iy) \asymp y^{-k}$ *where* $\alpha^{-1} < k < 2 - \alpha^{-1}$. *Then, for any* (*not necessarily compact*) *Borel subset* $E \subset \mathbb{R}^2_+$ *with smooth boundary, we have*

$$\lim_{t\to\infty} t^{\alpha/2} \cdot \#\{C \in \mathcal{T} : \text{area}_f(C) > t,\ C \cap E \ne \varnothing\} \sim c_A \cdot \mathcal{H}^\alpha_f(E).$$

Acknowledgments

This article are mostly based on the papers [Oh and Shah 2013] and [Mohammadi and Oh 2012b]. I am grateful to Nimish Shah and Amir Mohammadi for their collaborations. I would also like to thank Tim Austin for useful discussions concerning the last section.

References

[Babillot 2002] M. Babillot, "On the mixing property for hyperbolic systems", *Israel J. Math.* **129** (2002), 61–76.

[Benoist and Oh 2012] Y. Benoist and H. Oh, "Effective equidistribution of S–integral points on symmetric varieties", *Ann. Inst. Fourier* (*Grenoble*) **62**:5 (2012), 1889–1942.

[Borel and Garland 1983] A. Borel and H. Garland, "Laplacian and the discrete spectrum of an arithmetic group", *Amer. J. Math.* **105**:2 (1983), 309–335.

[Borkovec et al. 1994] M. Borkovec, W. de Paris, and R. Peikert, "The fractal dimension of the Apollonian sphere packing", *Fractals* **2**:4 (1994), 521–526.

[Bourgain and Fuchs 2011] J. Bourgain and E. Fuchs, "A proof of the positive density conjecture for integer Apollonian circle packings", *J. Amer. Math. Soc.* **24**:4 (2011), 945–967.

[Bourgain and Kontorovich 2012] J. Bourgain and A. Kontorovich, "On the strong density conjecture of Apollonian packings", preprint, 2012. arXiv 1205.4416

[Bourgain et al. 2010a] J. Bourgain, A. Gamburd, and P. Sarnak, "Affine linear sieve, expanders, and sum-product", *Invent. Math.* **179**:3 (2010), 559–644.

[Bourgain et al. 2010b] J. Bourgain, A. Kontorovich, and P. Sarnak, "Sector estimates for hyperbolic isometries", *Geom. Funct. Anal.* **20**:5 (2010), 1175–1200.

[Bourgain et al. 2011] J. Bourgain, A. Gamburd, and P. Sarnak, "Generalization of Selberg's $\frac{3}{16}$ theorem and affine sieve", *Acta Math.* **207**:2 (2011), 255–290.

[Bowditch 1993] B. H. Bowditch, "Geometrical finiteness for hyperbolic groups", *J. Funct. Anal.* **113**:2 (1993), 245–317.

[Burger 1990] M. Burger, "Horocycle flow on geometrically finite surfaces", *Duke Math. J.* **61**:3 (1990), 779–803.

[Dolgopyat 1998] D. Dolgopyat, "On decay of correlations in Anosov flows", *Ann. of Math.* (2) **147**:2 (1998), 357–390.

[Duke et al. 1993] W. Duke, Z. Rudnick, and P. Sarnak, "Density of integer points on affine homogeneous varieties", *Duke Math. J.* **71**:1 (1993), 143–179.

[Eskin and McMullen 1993] A. Eskin and C. McMullen, "Mixing, counting, and equidistribution in Lie groups", *Duke Math. J.* **71**:1 (1993), 181–209.

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

[Flaminio and Spatzier 1990] L. Flaminio and R. J. Spatzier, "Geometrically finite groups, Patterson–Sullivan measures and Ratner's rigidity theorem", *Invent. Math.* **99**:3 (1990), 601–626.

[Gorodnik and Nevo 2012] A. Gorodnik and A. Nevo, "Lifting, restricting and sifting integral points on affine homogeneous varieties", *Compos. Math.* **148**:6 (2012), 1695–1716.

[Gorodnik and Oh 2007] A. Gorodnik and H. Oh, "Orbits of discrete subgroups on a symmetric space and the Furstenberg boundary", *Duke Math. J.* **139**:3 (2007), 483–525.

[Graham et al. 2006] R. L. Graham, J. C. Lagarias, C. L. Mallows, A. R. Wilks, and C. H. Yan, "Apollonian circle packings: Geometry and group theory, III: Higher dimensions", *Discrete Comput. Geom.* **35**:1 (2006), 37–72.

[Guillarmou and Mazzeo 2012] C. Guillarmou and R. Mazzeo, "Resolvent of the Laplacian on geometrically finite hyperbolic manifolds", *Invent. Math.* **187**:1 (2012), 99–144.

[Hirai 1962] T. Hirai, "On irreducible representations of the Lorentz group of $n-$th order", *Proc. Japan Acad.* **38** (1962), 258–262.

[Howe and Moore 1979] R. E. Howe and C. C. Moore, "Asymptotic properties of unitary representations", *J. Funct. Anal.* **32**:1 (1979), 72–96.

[Kim 2012] I. Kim, "Counting, mixing and equidistribution of horospheres in geometrically finite rank one locally symmetric manifolds", (2012). arXiv 1103.5003

[Kleinbock and Margulis 1996] D. Y. Kleinbock and G. A. Margulis, "Bounded orbits of nonquasiunipotent flows on homogeneous spaces", pp. 141–172 in *Sinaĭ's Moscow Seminar on Dynamical Systems*, edited by L. A. Bunimovich et al., Amer. Math. Soc. Transl. Ser. 2 **171**, Amer. Math. Soc., Providence, RI, 1996.

[Knapp and Stein 1971] A. W. Knapp and E. M. Stein, "Intertwining operators for semisimple groups", *Ann. of Math.* (2) **93** (1971), 489–578.

[Kontorovich and Oh 2011] A. Kontorovich and H. Oh, "Apollonian circle packings and closed horospheres on hyperbolic 3–manifolds", *J. Amer. Math. Soc.* **24**:3 (2011), 603–648.

[Kowalski 2010] E. Kowalski, "Sieve in expansion", preprint, 2010. arXiv 1012.2793

[Kowalski 2014] E. Kowalski, "Sieves in discrete groups, especially sparse", pp. 107–140 in *Thin groups and superstrong approximation*, edited by H. Oh and E. Breuillard, Math. Sci. Res. Inst. Publ. **61**, Cambridge, New York, 2014.

[Lax and Phillips 1982] P. D. Lax and R. S. Phillips, "The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces", *J. Funct. Anal.* **46**:3 (1982), 280–350.

[Lee and Oh 2013] M. Lee and H. Oh, "Effective circle count for Apollonian packings and closed horospheres", *Geom. Funct. Anal.* **23**:2 (2013), 580–621.

[Margulis 2004] G. A. Margulis, *On some aspects of the theory of Anosov systems*, Springer, Berlin, 2004.

[McMullen 1998] C. T. McMullen, "Hausdorff dimension and conformal dynamics, III: Computation of dimension", *Amer. J. Math.* **120**:4 (1998), 691–721.

[Mohammadi and Oh 2012a] A. Mohammadi and H. Oh, "Ergodicity of unipotent flows and Kleinian groups", preprint, 2012. arXiv 1112.4024

[Mohammadi and Oh 2012b] A. Mohammadi and H. Oh, "Matrix coefficients, counting and primes for orbits of geometrically finite groups", preprint, 2012. To appear in *J. European Math. Soc.* arXiv 1208.4139

[Nevo and Sarnak 2010] A. Nevo and P. Sarnak, "Prime and almost prime integral points on principal homogeneous spaces", *Acta Math.* **205**:2 (2010), 361–402.

[Oh 2010] H. Oh, "Orbital counting via mixing and unipotent flows", pp. 339–375 in *Homogeneous flows, moduli spaces and arithmetic* (Pisa, 2007), edited by M. L. Einsiedler et al., Clay Math. Proc. **10**, Amer. Math. Soc., Providence, RI, 2010.

[Oh 2011] H. Oh, "Dynamics on geometrically finite hyperbolic manifolds with applications to Apollonian circle packings and beyond", pp. 1308–1331 in *Proceedings of the International Congress of Mathematicians, III* (Hyperabad, 2010), edited by R. Bhatia et al., Hindustan Book Agency, New Delhi, 2011.

[Oh and Shah 2010] H. Oh and N. Shah, "Counting visible circles on the sphere and Kleinian groups", preprint, 2010. arXiv 1004.2129

[Oh and Shah 2012] H. Oh and N. Shah, "The asymptotic distribution of circles in the orbits of Kleinian groups", *Invent. Math.* **187**:1 (2012), 1–35.

[Oh and Shah 2013] H. Oh and N. A. Shah, "Equidistribution and counting for orbits of geometrically finite hyperbolic groups", *J. Amer. Math. Soc.* **26**:2 (2013), 511–562.

[Patterson 1976] S. J. Patterson, "The limit set of a Fuchsian group", *Acta Math.* **136**:3-4 (1976), 241–273.

[Ratcliffe 2006] J. G. Ratcliffe, *Foundations of hyperbolic manifolds*, 2nd ed., Graduate Texts in Mathematics **149**, Springer, New York, 2006.

[Ratner 1991] M. Ratner, "On Raghunathan's measure conjecture", *Ann. of Math.* (2) **134**:3 (1991), 545–607.

[Roblin 2003] T. Roblin, *Ergodicité et équidistribution en courbure négative*, Mém. Soc. Math. Fr. (N.S.) **95**, Société Mathématique de France, Paris, 2003.

[Salehi Golsefidy and Sarnak 2013] A. Salehi Golsefidy and P. Sarnak, "The affine sieve", *J. Amer. Math. Soc.* **26**:4 (2013), 1085–1105.

[Salehi Golsefidy and Varjú 2012] A. Salehi Golsefidy and P. P. Varjú, "Expansion in perfect groups", *Geom. Funct. Anal.* **22**:6 (2012), 1832–1891.

[Sarnak 2011] P. Sarnak, "Integral Apollonian packings", *Amer. Math. Monthly* **118**:4 (2011), 291–306.

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

[Stoyanov 2011] L. Stoyanov, "Spectra of Ruelle transfer operators for axiom A flows", *Nonlinearity* **24**:4 (2011), 1089–1120.

[Sullivan 1979] D. Sullivan, "The density at infinity of a discrete group of hyperbolic motions", *Inst. Hautes Études Sci. Publ. Math.* 50 (1979), 171–202.

[Sullivan 1984] D. Sullivan, "Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups", *Acta Math.* **153**:3-4 (1984), 259–277.

[Vinogradov 2012] I. Vinogradov, "Effective bisector estimate with applications to Apollonian circle packings", preprint, 2012. arXiv 1204.5498

[Warner 1972a] G. Warner, *Harmonic analysis on semi-simple Lie groups, I*, Die Grundlehren der mathematischen Wissenschaften **188**, Springer, New York, 1972.

[Warner 1972b] G. Warner, *Harmonic analysis on semi-simple Lie groups, II*, Die Grundlehren der mathematischen Wissenschaften **189**, Springer, New York, 1972.

[Winter $\geq$ 2012] D. Winter, In preparation.

hee.oh@yale.edu

Mathematics Department, Yale University, New Haven, CT 06520, United States

Korea Institute for Advanced Study, Seoul, Korea