Improved sup-norm bounds for locally symmetric spaces

Christopher Lutsko

Abstract

Let $X = G/K$ be a symmetric space of noncompact type, of dimension n and rank r, and let $Y = \Gamma\backslash X$. Sarnak's local bound for an L^2-normalized spherical joint eigenfunction with regular tempered parameter of size T is $\|\phi\|_\infty \ll T^{(n-r)/2}$. We prove $o(T^{(n-r)/2})$ locally uniformly on every quotient. On finite-volume real hyperbolic manifolds this is uniform in the expanding cusp range $y \leq T^\beta$, $\beta < 1/2$. If the injectivity radius is bounded below, we prove the global estimate $T^{(n-r)/2}(\log T)^{-r/2}$.

1 Introduction

Let G be a connected, semisimple Lie group with finite center and maximal compact subgroup $K < G$. Let $X := G/K$ be a symmetric space of noncompact type and let $Y := \Gamma\backslash X$, where $\Gamma < G$ is a torsion-free, discrete subgroup. Let W denote the restricted Weyl group. Write

$$n := \dim X, \qquad r := \operatorname{rank} X, \qquad q := n - r.$$

Let $\mathbb{D}(X)$ denote the algebra of invariant differential operators on X. We say that $\phi \in L^2(Y)$ is a joint eigenfunction if it is a simultaneous eigenfunction of every operator in $\mathbb{D}(X)$. Via the Harish–Chandra isomorphism, the joint spectrum is parameterized by a W-invariant subset $\sigma \subset \mathfrak{a}^*_{\mathbb{C}} \cong \mathbb{C}^r$. To a joint eigenfunction we associate a spectral parameter $\lambda \in \sigma$, defined up to W, and put $T = 1 + \|\lambda\|$. A sequence of directions $\lambda/\|\lambda\|$ is called uniformly *regular* if it remains in a fixed compact subset of an open Weyl chamber.

In a letter to Caroline Morawetz [19], Peter Sarnak used a pretrace formula to show that

$$\|\phi\|_\infty \ll T^{q/2}, \tag{1.1}$$

on compact sets. We show that, along regular rays, this scale is never attained.

Theorem 1.1 (Strict Sarnak bound)**.** *Let $\phi_j \in L^2(Y)$ be an L^2-normalized spherical joint eigenfunction with tempered parameter $\lambda_j \in \mathfrak{a}^*$, and put $T_j = 1 + \|\lambda_j\|$. Assume that $T_j \to \infty$ and that the directions $\lambda_j/\|\lambda_j\|$ are uniformly regular. Then, on every compact subset $U \subset Y$,*

$$\|\phi_j\|_{L^\infty(U)} = o_U(T_j^{q/2}). \tag{1.2}$$

If, in addition, Y is a finite-volume real hyperbolic manifold, fix standard cusp coordinates and let Y_β be the compact core together with the parts of all cusps of height $y \leq T_j^\beta$. Then, for every $0 < \beta < 1/2$,

$$\|\phi_j\|_{L^\infty(Y_\beta)} = o_\beta(T_j^{q/2}). \tag{1.3}$$

In fact, we can improve the upper bound by a logarithmic factor as long as the injectivity radius is strictly positive.

Theorem 1.2 (Logarithmic improvement). *Under the spectral hypotheses of Theorem 1.1, suppose in addition that*

$$\inf_{x\in Y} \mathrm{inj}_Y(x) > 0. \tag{1.4}$$

Then

$$\|\phi_j\|_{L^\infty(Y)} \ll \frac{T_j^{q/2}}{(\log T_j)^{r/2}}. \tag{1.5}$$

The implied constant is uniform for spectral directions in a fixed compact regular set. In particular, (1.5) *holds on every compact quotient and on every convex cocompact real hyperbolic quotient.*

1.1 The sup-norm problem

For a general compact n-manifold and an L^2-normalized eigenfunction $-\Delta u = \lambda^2 u$, the local Weyl law gives

$$\|u\|_\infty \ll \lambda^{(n-1)/2}. \tag{1.6}$$

The round sphere shows that this can be sharp: zonal harmonics concentrate at a point. More generally, Sogge–Zelditch [20, 21] gave precise geometric criteria for maximal growth.

Negative curvature gives one model for improvement. Bérard propagated the wave kernel to logarithmic time on compact manifolds without conjugate points, which yields $\|u\|_\infty \ll \lambda^{(n-1)/2}(\log\lambda)^{-1/2}$ [1, 7]; Keeler proved a corresponding two-point Weyl law [11]. On a rank-r symmetric space, resolving the full joint spectrum first replaces the exponent $(n-1)/2$ by Sarnak's $(n-r)/2$. Theorem 1.2 recovers Bérard's bound in rank 1 and saves one factor of $(\log T)^{-1/2}$ for each spectral coordinate.

There is a different, arithmetic route to small sup norms. Amplification by Hecke operators gives power savings for special arithmetic eigenbases. On an arithmetic hyperbolic surface, Iwaniec–Sarnak proved

$$\|u\|_\infty \ll_\varepsilon T^{5/12+\varepsilon}, \tag{1.7}$$

against the local exponent $T^{1/2}$ [10]. Higher-rank amplification gives further power savings in particular groups and spectral or level regimes. Such estimates use arithmetic correspondences and a Hecke eigenfunction. Theorems 1.1 and 1.2 use neither: they apply to every spherical joint eigenfunction on every quotient in the stated geometric range. The two mechanisms are complementary.

The opposite problem of constructing large eigenfunctions is equally important. Arithmetic periods and theta lifts produce sequences with polynomially growing sup norms on compact locally symmetric spaces [2, 3].

Remark 1.3. The multiplier construction is useful beyond point evaluation. In particular, in work in progress by the author, the same approach is used to prove a fixed-window quantum ergodicity theorem in the eigenvalue aspect on a fixed compact quotient. This complements fixed-window results in the Benjamini–Schramm, or level, aspect due to Le Masson–Sahlsten [15] and, in higher rank, Brumley–Marshall–Matz–Peterson [4]. The microlocal framework for regular joint eigenfunctions on compact locally symmetric spaces was developed by Silberman–Venkatesh [18] and Hansen–Hilgert–Schröder [8].

1.2 Proof plan and the Green-kernel viewpoint

The starting point for this methodology is the observation that in rank 1, the Green function plays an important role in estimating spectral mass. This can be seen directly in [9, Chapter 5] and indirectly in the point-counting argument of Lax–Phillips [14], later developed by Kontorovich [12], by Kontorovich and the author [13], and in the author's spectral treatment of horospherical equidistribution [16]. However, a literal resolvent is poorly adapted to this problem: at the target eigenvalue it has a pole and its kernel has global support. In higher rank, there is also no canonical one-variable resolvent that resolves all r invariant operators simultaneously.

The central idea is to replace that resolvent by a positive, smoothed kernel. Write $\lambda = Tc$, where c is the regular spectral direction. We choose a W-invariant Paley–Wiener bump $\psi = \psi_{T,c,\epsilon}$ concentrated near the Weyl orbit of Tc and normalized by $\psi(Tc) = 1$. Taking its spherical inverse transform β, we set

$$\Phi = \beta^* * \beta, \qquad \widehat{\Phi} = |\psi|^2, \qquad \widehat{\Phi}(Tc) = 1. \tag{1.8}$$

Thus Φ has unit response on the eigenfunction under study, as a normalized Green kernel would, but it is positive and has finite propagation. Narrowing the joint window reduces its value at the identity by the volume of that window. The apparent cost is that the spatial support grows. Regular spherical oscillation is precisely what makes the nonidentity deck transforms small enough to pay for that growth. This trade—spectral narrowing against spatial propagation—is the proof.

The paper implements the idea in four steps. Section 2 fixes the notation and records the harmonic-analytic and geometric preliminaries, including the qualitative and quantitative spherical decay estimates and the cusp geometry. Section 3 then constructs the positive peak by spherical Paley–Wiener theory, derives the four kernel estimates, and proves the positive pretrace inequality. Section 4 first fixes the bandwidth and then lets it tend to zero, proving the compact-set little-o estimate. It next sums the rank-one parabolic translates to reach the expanding cusp range and finally chooses logarithmic propagation to prove Theorem 1.2. The last subsection isolates the Airy transition that prevents this thick-part argument from becoming an unrestricted global cusp estimate.

2 Notation and preliminaries

We collect the notation and background used in the construction before introducing the automorphic kernel. The identity element of G is denoted by e, and d_X is the invariant distance on $X = G/K$. For a function f on G, write

$$f^*(g) = \overline{f(g^{-1})}, \qquad (f_1 * f_2)(g) = \int_G f_1(gh^{-1}) f_2(h)\, \mathrm{d}h.$$

We write $E \Subset Z$ when E is a compact subset of Z, and $\operatorname{supp} f$ for the support of f. The notation $\widehat{f}$ always denotes the spherical transform of a K-bi-invariant function, while $\|\cdot\|$ denotes the metric norm on $\mathfrak{a}$ or its dual norm on $\mathfrak{a}^*$. Implied constants may depend on the symmetric space and on fixed compact regular sets; any additional dependence is displayed by a subscript.

2.1 Structure, measures, and spectral parameters

Fix a Cartan involution of G and write $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$. Choose a maximal abelian subspace $\mathfrak{a} \subset \mathfrak{p}$, put $A = \exp \mathfrak{a}$, and fix an Iwasawa decomposition $G = NAK$. The Iwasawa projection $\mathcal{A} : G \to \mathfrak{a}$ is determined by $g \in N \exp(\mathcal{A}(g))K$. The Cartan decomposition is $G = K \exp(\overline{\mathfrak{a}^+})K$. If $g = k_1 \exp H(g) k_2$ with $H(g) \in \overline{\mathfrak{a}^+}$, then

$$d_X(gK, K) = \|H(g)\|. \tag{2.1}$$

All norms on $\mathfrak{a}$ and $\mathfrak{a}^*$ below are the dual norms induced by the G-invariant metric on X.

Let $\Sigma^+ \subset \mathfrak{a}^*$ be the positive restricted roots, $m_\alpha = \dim \mathfrak{g}_\alpha$, and $W = N_K(A)/Z_K(A)$. Since X has no Euclidean factor, the restricted roots span $\mathfrak{a}^*$. Hence

$$\rho = \frac{1}{2} \sum_{\alpha \in \Sigma^+} m_\alpha \alpha, \qquad r = \dim \mathfrak{a}, \qquad q = \sum_{\alpha \in \Sigma^+} m_\alpha = \dim X - r. \tag{2.2}$$

We write $\mathfrak{a}^*_{\mathrm{reg}}$ for the complement of the root hyperplanes. A compact set $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ is always understood to lie in one open chamber. Thus there is a number $\eta_C > 0$ such that

$$|\alpha(c)| \geq \eta_C \quad (c \in C,\ \alpha \in \Sigma). \tag{2.3}$$

Haar measure on G is normalized compatibly with the Riemannian measure on $X = G/K$; the normalization affects only the fixed constants denoted by c_X. An L^2 spherical joint eigenfunction on Y is a smooth right K-invariant function $\phi \in L^2(\Gamma\backslash G)$ which is a joint eigenvector for $\mathcal{D}(X)$. By the Harish–Chandra isomorphism its eigencharacter is indexed by a W-orbit in $\mathfrak{a}^*_{\mathbb{C}}$. For a tempered eigenfunction we choose its representative $\lambda \in \overline{\mathfrak{a}^{*+}}$. Our hypotheses say that $\lambda = Tc$ with $c \in C$ after replacing T by $\|\lambda\|$; the use of $1 + \|\lambda\|$ in the statements only accommodates bounded parameters.

2.2 Spherical inversion, Paley–Wiener theory, and Plancherel growth

For $\lambda \in \mathfrak{a}^*$ the normalized elementary spherical function is

$$\varphi_\lambda(g) = \int_K \exp\bigl((\mathrm{i}\lambda - \rho)(\mathcal{A}(g^{-1}k))\bigr)\, \mathrm{d}k, \qquad \varphi_\lambda(e) = 1. \tag{2.4}$$

Changing g to g^{-1} or changing the sign convention for $\mathcal{A}$ gives the equivalent formulas found in the literature. What we use is that φ_λ is K-bi-invariant, $\varphi_{w\lambda} = \varphi_\lambda$, and, for real λ,

$$|\varphi_\lambda(g)| \leq \varphi_0(g) \leq 1. \tag{2.5}$$

The first inequality follows directly by taking absolute values in (2.4).

For $f \in C_c^\infty(K\backslash G/K)$ define

$$\widehat{f}(\lambda) = \int_G f(g)\varphi_{-\lambda}(g)\,\mathrm{d}g.$$

Spherical inversion and Plancherel's formula are

$$f(g) = c_X \int_{\mathfrak{a}^*} \widehat{f}(\lambda)\varphi_\lambda(g)|\mathbf{c}(\lambda)|^{-2}\,\mathrm{d}\lambda, \tag{2.6}$$

$$\|f\|_2^2 = c_X \int_{\mathfrak{a}^*} |\widehat{f}(\lambda)|^2|\mathbf{c}(\lambda)|^{-2}\,\mathrm{d}\lambda. \tag{2.7}$$

We integrate over all of $\mathfrak{a}^*$ and absorb $|W|^{-1}$ into c_X.

We use the following precise form of the spherical Paley–Wiener theorem. If F is a W-invariant entire function on $\mathfrak{a}^*_{\mathbb{C}}$ and, for every N, satisfies

$$|F(\lambda + \mathrm{i}\mu)| \leq C_N(1 + \|\lambda\| + \|\mu\|)^{-N}\mathrm{e}^{R\|\mu\|}, \tag{2.8}$$

then $F = \widehat{f}$ for a unique $f \in C_c^\infty(K\backslash G/K)$ supported where $d_X(gK, K) \leq R$. Conversely, the transform of every such f satisfies (2.8). This is Gangolli's Paley–Wiener theorem [6]; the equality between exponential type and metric support uses (2.1).

Two consequences of the Gindikin–Karpelevich product formula will be used repeatedly. First, globally on the real axis,

$$|\mathbf{c}(\lambda)|^{-2} \leq C_X(1 + \|\lambda\|)^q. \tag{2.9}$$

More explicitly, if Σ_0^+ denotes the indivisible positive roots and $\alpha_0 = \alpha/\langle\alpha,\alpha\rangle$, the product formula and Stirling's estimate give one-variable factors p_α satisfying

$$|\mathbf{c}(\lambda)|^{-2} = C_X \prod_{\alpha\in\Sigma_0^+} p_\alpha(\langle\lambda,\alpha_0\rangle), \qquad p_\alpha(s) \ll (1 + |s|)^{m_\alpha + m_{2\alpha}}, \tag{2.10}$$

with the matching asymptotic when $|s| \to \infty$; by convention $m_{2\alpha} = 0$ if 2α is not a root. Summing the exponents in (2.10) gives $\sum_{\alpha\in\Sigma_0^+}(m_\alpha + m_{2\alpha}) = q$. Second, for $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ and $\|u\| \leq T/2$,

$$|\mathbf{c}(Tc + u)|^{-2} \leq C_C T^q (1 + \|u\|)^q, \qquad c \in C. \tag{2.11}$$

For bounded u one has the sharper asymptotic

$$|\mathbf{c}(Tc + u)|^{-2} = T^q\big(\beta(c) + O_C(T^{-1}(1 + \|u\|)^{q+1})\big), \tag{2.12}$$

with β continuous and strictly positive on C. To see the exponent, apply Stirling's formula to each gamma quotient in the product formula. Every positive root contributes its multiplicity to the total degree, so the degree is the number q in (2.2). The same argument without asymptotic expansion gives (2.9) and (2.11).

2.3 Uniform decay of regular spherical functions

The analytic input is Marshall's root-product estimate. We state exactly the two consequences needed below, including the dependence on a growing group-variable compactum.

Proposition 2.1 (Root-product estimate with tracked constants)**.** *Let $C_1 \Subset \mathfrak{a}^*_{\mathrm{reg}}$. There are constants A and D depending only on X and C_1 such that, for $R \geq 1$, $t \geq 1$, $c \in C_1$, and $H \in \mathfrak{a}$ with $\|H\| \leq R$,*

$$|\varphi_{tc}(\exp H)| \leq D\mathrm{e}^{AR} \prod_{\alpha \in \Sigma^+} (1 + t|\alpha(H)|)^{-m_\alpha/2}. \tag{2.13}$$

Proof. For fixed R, this is [17, Theorem 1.3]. We track the constant in that proof. Write

$$B_R = \{H \in \mathfrak{a} : \|H\| \leq R\}.$$

The Harish–Chandra formula is

$$\varphi_{tc}(\exp H) = \int_K b(k, H)\mathrm{e}^{\mathrm{i}t\Theta(k,H,c)}\,\mathrm{d}k, \qquad \Theta(k, H, c) = c\big(\mathcal{A}(k \exp H)\big),$$

where $b(k, H) = \exp(\rho(\mathcal{A}(k \exp H)))$.

Choose highest-weight representations π_j whose highest weights ω_j span $\mathfrak{a}^*$, with K-invariant norms and highest-weight vectors v_j. After passing to a finite cover of G, if necessary,

$$\mathrm{e}^{\omega_j(\mathcal{A}(g))} = \frac{\|v_j\|}{\|\pi_j(g^{-1})v_j\|}.$$

For $g = k \exp H$ and $H \in B_R$,

$$\|\pi_j(g)\| + \|\pi_j(g^{-1})\| \ll \mathrm{e}^{CR}.$$

Applying the chain rule to the preceding formula gives, for every fixed N,

$$\|\Theta\|_{C^N(K\times B_R\times C_1)} + \|b\|_{C^N(K\times B_R)} \ll_N \mathrm{e}^{C_N R}. \tag{2.14}$$

Marshall's normal forms are obtained by a fixed number of implicit-function and Morse changes of variables. At a Weyl critical point, the Hessian in an α-root direction is

$$\frac{1}{2}\langle c, \alpha\rangle\big(1 - \mathrm{e}^{2\alpha(wH)}\big) = \frac{1}{2}\langle c, \alpha\rangle\alpha(wH)u(\alpha(wH)), \qquad u(s) = \frac{1 - \mathrm{e}^{2s}}{s}.$$

This is [17, Proposition 4.5]. Regularity of C_1 bounds $|\langle c, \alpha\rangle|$ away from zero, and

$$|u(s)| + |u(s)|^{-1} \ll \mathrm{e}^{CR}, \qquad |s| \leq CR.$$

We retain $\alpha(H)$ in the phase and absorb $u(\alpha(H))$ into the constant.

The same factor occurs at the other root strata. In the induction proving [17, Theorem 4.7], the pivot after division by the root coordinate is

$$-\langle c, \alpha\rangle \mathrm{e}^s \frac{\sinh s}{s} = \frac{1}{2}\langle c, \alpha\rangle u(s);$$

see [17, Lemma 4.11]. The final linear coordinate has derivative

$$\alpha(H^L)\langle \mathrm{Ad}(l^{-1})H_c, V_{-\alpha} - V_\alpha\rangle.$$

The root factor $\alpha(H^L)$ is again retained. By [17, Lemma 4.19], at least one of the remaining coefficients is nonzero on each resolved stratum. There are finitely many strata, and the variables l, $H^L/\|H^L\|$, and c range over compact sets. Let $\mathcal{M}_R \geq 1$ be the maximum of the required phase seminorms and inverse pivot bounds. The preceding estimates give

$$\mathcal{M}_R \ll \mathrm{e}^{CR}. \tag{2.15}$$

The chart size follows from the quantitative inverse-function theorem. Suppose that $\|DF(x_0)^{-1}\| \leq \mathcal{M}$ and $\|D^2F\| \leq \mathcal{M}$. On the ball of radius $(4\mathcal{M}^2)^{-1}$,

$$\|DF(x) - DF(x_0)\| \leq (4\mathcal{M})^{-1}.$$

The inverse-function theorem applies there, with inverse derivative at most $2\mathcal{M}$. Repeated differentiation gives polynomial dependence on $\mathcal{M}$. Since Marshall's construction has a fixed number of steps, (2.14) and (2.15) give normal-form charts of radius

$$\delta_R \gg \mathrm{e}^{-CR}, \tag{2.16}$$

with coordinate maps and inverses of C^N-norm $O_N(\mathrm{e}^{C_N R})$.

The resolved parameter space for $K/M \times B_R \times C_1$ has fixed dimension. Its direction variables range over compact sets, while its H-variables have size $O(R)$. A maximal $\delta_R/2$-separated set therefore has cardinality

$$O((1+R)^r \delta_R^{-D}) = O(\mathrm{e}^{CR}), \tag{2.17}$$

where D is fixed. The corresponding cover has bounded overlap. Rescaling a fixed bump function gives a subordinate partition with

$$\|\chi_j\|_{C^N} \ll_N \delta_R^{-C_N} \ll_N \mathrm{e}^{C_N R}.$$

In normal coordinates, van der Corput gives $(1 + t|\alpha(H)|)^{-1/2}$ in each root direction, hence $(1 + t|\alpha(H)|)^{-m_\alpha/2}$ on the α-root space. Away from the critical strata, a fixed number of integrations by parts gives the same bound. By the preceding estimates, each chart contributes at most

$$C\mathrm{e}^{CR} \prod_{\alpha\in\Sigma^+} (1 + t|\alpha(H)|)^{-m_\alpha/2}.$$

Summing over (2.17) and enlarging A proves (2.13). $\square$

Lemma 2.2 (Decay away from the identity)**.** *Let $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$, let $E \Subset G \setminus K$, and let $U \Subset \mathfrak{a}^*$. Then*

$$\sup_{c\in C,\, u\in U,\, g\in E} |\varphi_{Tc+u}(g)| \longrightarrow 0 \qquad (T \to \infty). \tag{2.18}$$

Proof. By Cartan decomposition, E maps to a compact set $B_E \subset \mathfrak{a}$ which does not contain 0. Hence there is $\iota_E > 0$ with $\|H\| \geq \iota_E$ for $H \in B_E$. Since the restricted roots span $\mathfrak{a}^*$, norm equivalence gives a constant $c_X > 0$ such that

$$\max_{\alpha\in\Sigma^+} |\alpha(H)| \geq c_X \|H\|. \tag{2.19}$$

For $u \in U$ write $Tc + u = T(c + u/T)$. Once T is large, the directions $c + u/T$ lie in a fixed compact regular set C_1. Apply Proposition 2.1 with a fixed ball containing B_E. Choose a root α_H

satisfying (2.19) and discard every factor in the product except its factor. If $m_0 = \min_{\alpha\in\Sigma^+} m_\alpha$, then

$$|\varphi_{Tc+u}(\exp H)| \le C_E(1+Tc_X\iota_E/2)^{-m_0/2} \le C_E' T^{-m_0/2}.$$

This is uniform in c, u, H and tends to zero, proving (2.18). $\square$

Lemma 2.3 (Quantitative regular decay)**.** *Fix $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ and $\iota > 0$. There are $\delta > 0$ and constants A, M such that, whenever*

$$\iota \le \|H\| \le A_0 \log T, \qquad c \in C, \qquad \|u\| \le T^{1/2},$$

one has

$$|\varphi_{Tc+u}(\exp H)| \le C_{C,\iota,A_0}(1+\|u\|)^M \mathrm{e}^{A\|H\|} T^{-\delta}. \tag{2.20}$$

Proof. Write $Tc+u = T(c+u/T)$. For $\|u\| \le T^{1/2}$, the vector $c+u/T$ belongs, for all sufficiently large T, to a fixed compact regular set C_1 containing C in its interior. Apply Proposition 2.1 with $R_H = \max(1, \|H\|)$; the hypothesis on H ensures $R_H \le 1 + A_0 \log T$. By (2.19), some positive root satisfies $|\alpha(H)| \ge c_X \iota$. Keeping only that factor gives

$$|\varphi_{Tc+u}(\exp H)| \le D' \mathrm{e}^{A\|H\|}(1+Tc_X\iota)^{-m_0/2}, \qquad m_0 = \min_{\alpha>0} m_\alpha.$$

Thus (2.20) holds with $\delta = m_0/2$ and $M = 0$. Enlarging the constant covers the bounded range of T for which $c + u/T$ has not yet entered C_1. $\square$

2.4 Orbit geometry

We shall use the following elementary geometric observation.

Lemma 2.4 (Uniform local orbit bounds)**.** *Let $K_0 \Subset Y$ and $R < \infty$. There are $\iota > 0$ and $N < \infty$ such that, for every $x \in K_0$ and every lift $g_x K$,*

$$d_X(g_xK, \gamma g_xK) \ge \iota \quad (\gamma \ne e), \tag{2.21}$$

$$\#\{\gamma : d_X(g_xK, \gamma g_xK) \le R\} \le N. \tag{2.22}$$

If $\inf_Y \mathrm{inj}_Y > 0$, the constants are uniform for $K_0 = Y$; moreover one may take

$$N \le C_\iota \mathrm{e}^{h_X R} \tag{2.23}$$

for a constant h_X depending only on X.

Proof. For a torsion-free quotient,

$$2\,\mathrm{inj}_Y(x) = \inf_{\gamma\ne e} d_X(g_xK, \gamma g_xK).$$

The injectivity-radius function is continuous. It therefore has a positive minimum on K_0, proving (2.21) with $\iota = 2\min_{K_0} \mathrm{inj}_Y$.

Put $p = g_xK$. Distinct orbit points γp and $\gamma' p$ are at distance

$$d_X(\gamma p, \gamma' p) = d_X(p, \gamma^{-1}\gamma' p) \ge \iota.$$

Thus the balls $B_X(\gamma p, \iota/3)$ are disjoint. If $d_X(p, \gamma p) \le R$, each such ball lies in $B_X(p, R + \iota/3)$. Homogeneity of X now gives the explicit packing bound

$$\#\{\gamma : d_X(p, \gamma p) \le R\} \le \frac{\operatorname{vol} B_X(p, R + \iota/3)}{\operatorname{vol} B_X(p, \iota/3)}. \tag{2.24}$$

This proves (2.22). The polar-coordinate formula on a noncompact symmetric space shows $\operatorname{vol} B_X(p, s) \le C_X \mathrm{e}^{h_X s}$ for $s \ge 1$ (one may take any $h_X > 2\|\rho\|$). Inserting this in (2.24) proves (2.23). If $\inf_Y \operatorname{inj}_Y > 0$, the same ι works for every $x \in Y$, so every step is global. □

2.5 Real-hyperbolic cusp coordinates

For the global part of Theorem 1.1, let $X = \mathbb{H}^{q+1}$ in the upper-half-space model. A finite-volume quotient has a compact core and finitely many cusps. After passing to a finite cover of a cusp cross-section (which changes every estimate only by the degree of that cover), one cusp is

$$\mathcal{C} = (\Lambda \backslash \mathbb{R}^q) \times [y_0, \infty), \qquad ds^2 = y^{-2}(\,\mathrm{d}y^2 + \,\mathrm{d}z^2), \qquad \mathrm{d}\operatorname{vol} = y^{-q-1}\,\mathrm{d}z\,\mathrm{d}y, \tag{2.25}$$

where Λ is a full Euclidean lattice. The parabolic displacement is given exactly by

$$\cosh d((z, y), (z + v, y)) = 1 + \frac{|v|^2}{2y^2}, \qquad d((z, y), (z + v, y)) = 2 \operatorname{arsinh} \frac{|v|}{2y}. \tag{2.26}$$

In particular, for each fixed R,

$$d((z, y), (z + v, y)) \asymp_R \frac{|v|}{y} \quad \text{when } d((z, y), (z + v, y)) \le R. \tag{2.27}$$

Choose the initial horoball precisely invariant under its parabolic stabilizer Γ_∞. The depth function $b(z, y) = \log(y/y_0)$ is a Busemann function and therefore is 1-Lipschitz. If $\gamma \notin \Gamma_\infty$, the interiors of the chosen horoball and its γ-translate are disjoint. A path from (z, y) to $\gamma(z, y)$ must descend through depth zero in the first horoball and enter the second horoball from depth zero. The Lipschitz property on the two segments gives

$$d((z, y), \gamma(z, y)) \ge 2 \log(y/y_0). \tag{2.28}$$

Replacing the chosen cusp neighborhood by a boundedly different one gives the equivalent form $2 \log y - O(1)$ used later.

Take the nonnegative Laplacian. A tempered eigenfunction has eigenvalue $q^2/4 + T^2$. Expanding in the characters of the torus and solving the resulting ordinary differential equation gives

$$\phi(z, y) = a_0^+ y^{q/2 + \mathrm{i}T} + a_0^- y^{q/2 - \mathrm{i}T} + y^{q/2} \sum_{0 \ne m \in \Lambda^*} a_m K_{\mathrm{i}T}(2\pi |m| y) \mathrm{e}^{2\pi \mathrm{i} \langle m, z \rangle}. \tag{2.29}$$

Indeed, substituting $y^{q/2} F(2\pi |m| y)$ into the separated radial equation gives

$$u^2 F''(u) + u F'(u) - (u^2 - T^2) F(u) = 0,$$

whose solution decaying as $u \to \infty$ is $K_{\mathrm{i}T}(u)$. The other solution $I_{\mathrm{i}T}(u)$ grows exponentially and is excluded by L^2. For the constant term, put $s = \log y$. If $T > 0$, then

$$\begin{aligned} &\int_{\log y_0}^{S} |a_0^+ \mathrm{e}^{\mathrm{i}Ts} + a_0^- \mathrm{e}^{-\mathrm{i}Ts}|^2 \,\mathrm{d}s \\ &\qquad = (|a_0^+|^2 + |a_0^-|^2)(S - \log y_0) + O_T(1). \end{aligned}$$

This is the radial L^2 integral after the factors $y^{q/2}$ and $y^{-q-1}\,\mathrm{d}y$ are combined. It stays bounded as $S \to \infty$ only when

$$a_0^+ = a_0^- = 0 \tag{2.30}$$

for every L^2 eigenfunction with real T.

Parseval on $\Lambda\backslash\mathbb{R}^q$, followed by the substitution $u = 2\pi|m|y$, gives

$$\int_{\mathcal{C}} |\phi|^2 \,\mathrm{d}\,\mathrm{vol} = \mathrm{vol}(\Lambda\backslash\mathbb{R}^q) \sum_{m\neq 0} |a_m|^2 I_T(m), \quad I_T(m) = \int_{2\pi|m|y_0}^{\infty} |K_{\mathrm{i}T}(u)|^2 \frac{\mathrm{d}u}{u}. \tag{2.31}$$

This identity is the source of the simultaneous control of all Fourier modes in the far cusp.

3 The automorphic kernel

We now give the construction used throughout the paper. Fix a compact set $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ contained in one open Weyl chamber. We write the tempered parameter as Tc, with $c \in C$ and $T \to \infty$. When $T = \|\lambda\|$, the vector c is simply the unit direction of λ. The boldface symbol $\mathbf{c}(\lambda)$ below denotes the Harish–Chandra c-function and should not be confused with the direction c.

Proposition 3.1 (Positive spherical peak)**.** *For every fixed $0 < \epsilon < 1$, all $T \geq T_0(C, \epsilon)$, and every $c \in C$, there is a K-bi-invariant Hermitian function $\Phi_{T,c,\epsilon} \in C_c^\infty(G)$ such that convolution by $\Phi_{T,c,\epsilon}$ is positive on every unitary representation and*

$$\widehat{\Phi}_{T,c,\epsilon}(Tc) = 1, \tag{3.1}$$

$$\operatorname{supp} \Phi_{T,c,\epsilon} \subset \{g : d_X(gK, K) \leq R_\epsilon\}, \tag{3.2}$$

$$\Phi_{T,c,\epsilon}(e) \ll_C \epsilon^r T^q, \tag{3.3}$$

$$\sup_{c \in C} \sup_{g \in E} |\Phi_{T,c,\epsilon}(g)| = o_{E,\epsilon}(T^q) \qquad (E \Subset G \setminus K). \tag{3.4}$$

Here R_ϵ is independent of T and c.

Proof. We use the spherical transform and Paley–Wiener theorem in the normalization recorded in Section 2. Choose a real, even, W-invariant Paley–Wiener bump ψ_0 with $\psi_0(0) = 1$. Concretely, take a real, even, W-invariant $h \in C_c^\infty(\mathfrak{a})$ with $\int_{\mathfrak{a}} h = 1$ and let ψ_0 be its Euclidean Fourier transform. If $\operatorname{supp} h$ is contained in the radius-R_0 ball, then for every N

$$|\psi_0(\lambda + \mathrm{i}\mu)| \leq C_N (1 + \|\lambda\| + \|\mu\|)^{-N} \mathrm{e}^{R_0 \|\mu\|}. \tag{3.5}$$

The translate $\psi_0((\lambda - Tc)/\epsilon)$ has spectral width ϵ and exponential type R_0/ϵ. Since the spherical transform must be W-invariant, define

$$\psi_{T,c,\epsilon}(\lambda) = d_{T,c,\epsilon}^{-1} \sum_{w \in W} \psi_0\left(\frac{w\lambda - Tc}{\epsilon}\right), \qquad d_{T,c,\epsilon} = \sum_{w \in W} \psi_0\left(\frac{T(wc - c)}{\epsilon}\right). \tag{3.6}$$

Because c ranges over a compact set of regular vectors, there is $d_C > 0$ such that $\|wc - c\| \geq d_C$ for $w \neq e$. The Schwartz bound on the real axis therefore gives, for every N,

$$d_{T,c,\epsilon} = 1 + O_{N,C}((T/\epsilon)^{-N}). \tag{3.7}$$

In particular, $d_{T,c,\epsilon}$ is bounded away from zero for $T \geq T_0(C, \epsilon)$. Reindexing the finite sum in (3.6) proves W-invariance, and substituting $\lambda = Tc$ shows directly that

$$\psi_{T,c,\epsilon}(Tc) = 1. \tag{3.8}$$

By spherical Paley–Wiener theory there is a unique $\beta_{T,c,\epsilon} \in C_c^\infty(K \backslash G / K)$ with

$$\widehat{\beta}_{T,c,\epsilon} = \psi_{T,c,\epsilon}, \qquad \operatorname{supp} \beta_{T,c,\epsilon} \subset \{g : d_X(gK, K) \leq R_0/\epsilon\}. \tag{3.9}$$

Set

$$\Phi_{T,c,\epsilon} = \beta_{T,c,\epsilon}^* * \beta_{T,c,\epsilon}, \qquad \beta^*(g) = \overline{\beta(g^{-1})}. \tag{3.10}$$

If B is right convolution by $\beta_{T,c,\epsilon}$, then right convolution by $\Phi_{T,c,\epsilon}$ is B^*B and is therefore positive. Moreover,

$$\widehat{\Phi}_{T,c,\epsilon}(\lambda) = |\psi_{T,c,\epsilon}(\lambda)|^2 \geq 0 \qquad (\lambda \in \mathfrak{a}^*). \tag{3.11}$$

We derive the four displayed properties in order.

Derivation of (3.1)**.** Equations (3.8) and (3.11) give $\widehat{\Phi}_{T,c,\epsilon}(Tc) = |\psi_{T,c,\epsilon}(Tc)|^2 = 1$.

Derivation of (3.2)**.** The product support inclusion for convolution and the triangle inequality on X give

$$\mathrm{supp}(\beta^* * \beta) \subset \{g : d_X(gK, K) \leq 2R_0/\epsilon\}.$$

Thus (3.2) holds with $R_\epsilon = 2R_0/\epsilon$.

Derivation of (3.3)**.** From (3.6), Cauchy–Schwarz in the finite Weyl sum, and (3.7),

$$|\psi_{T,c,\epsilon}(\lambda)|^2 \leq C_W \sum_{w \in W} |\psi_0((w\lambda - Tc)/\epsilon)|^2. \tag{3.12}$$

Spherical inversion at the identity, followed by the change of variables $w\lambda = Tc + \epsilon u$, therefore gives

$$\begin{aligned} \Phi_{T,c,\epsilon}(e) &= c_X \int_{\mathfrak{a}^*} |\psi_{T,c,\epsilon}(\lambda)|^2 |\mathbf{c}(\lambda)|^{-2} \, \mathrm{d}\lambda \\ &\leq C\epsilon^r \sum_{w \in W} \int_{\mathfrak{a}^*} |\psi_0(u)|^2 |\mathbf{c}(w^{-1}(Tc + \epsilon u))|^{-2} \, \mathrm{d}u. \end{aligned} \tag{3.13}$$

The Plancherel density is W-invariant. On $\|u\| \leq T/(2\epsilon)$, the regular growth estimate (2.11) bounds it by $C_C T^q (1 + \|u\|)^q$. On the complementary region we use the global bound (2.9). Taking the decay exponent in (3.5) larger than $2r + 2q + 2$ yields

$$\begin{aligned} &\epsilon^r \int_{\|u\| \leq T/(2\epsilon)} |\psi_0(u)|^2 |\mathbf{c}(Tc + \epsilon u)|^{-2} \, \mathrm{d}u \leq C_C \epsilon^r T^q, \\ &\epsilon^r \int_{\|u\| > T/(2\epsilon)} |\psi_0(u)|^2 |\mathbf{c}(Tc + \epsilon u)|^{-2} \, \mathrm{d}u \leq C_{C,\epsilon,N} T^{-N}. \end{aligned}$$

This proves (3.3). The factor ϵ^r is exactly the Jacobian of the shrinking window in the r joint spectral coordinates, while T^q is the regular Plancherel density at Tc.

Derivation of (3.4)**.** Insert (3.12) into spherical inversion and make the same change of variables. Fix $L > 0$. On $\|u\| \leq L$, the normalized density $T^{-q}|\mathbf{c}(Tc + \epsilon u)|^{-2}$ is uniformly bounded, and the regular spherical decay proved in Lemma 2.2 gives

$$\sup_{c \in C,\, g \in E,\, \|u\| \leq L} |\varphi_{Tc + \epsilon u}(g)| \longrightarrow 0.$$

The W-invariance of both the Plancherel density and the spherical functions handles every Weyl peak identically. Thus the portion $\|u\| \leq L$, after division by T^q, tends to zero uniformly in c and g. On $L < \|u\| \leq T/(2\epsilon)$, use $|\varphi_\lambda(g)| \leq 1$, (2.11), and Schwartz decay; the normalized contribution is at most

$$C_C \epsilon^r \int_{\|u\| > L} |\psi_0(u)|^2 (1 + \|u\|)^q \, \mathrm{d}u,$$

which tends to zero as $L \to \infty$, uniformly in T. The remaining region $\|u\| > T/(2\epsilon)$ is $O_{C,\epsilon,N}(T^{-N})$ by (2.9). Letting first $T \to \infty$ and then $L \to \infty$ proves (3.4). □

3.1 The logarithmic-scale peak

Lemma 3.2 (Logarithmic spherical peak)**.** *Fix $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ and $\iota > 0$. There are $\delta > 0$ and $A, B > 0$ such that, for $1 \le R \le A^{-1} \log T$, the function $\Phi_{T,c,R^{-1}}$ constructed in Proposition 3.1 satisfies*

$$\Phi_{T,c,R^{-1}}(e) \le BT^q R^{-r}, \tag{3.14}$$

$$\operatorname{supp} \Phi_{T,c,R^{-1}} \subset \{g : d(gK, K) \le BR\}, \tag{3.15}$$

$$|\Phi_{T,c,R^{-1}}(g)| \le BT^{q-\delta} R^{-r} \mathrm{e}^{Ad(gK,K)} \tag{3.16}$$

whenever $c \in C$ and $\iota \le d(gK, K) \le BR$.

Proof. Use (3.6) with $\epsilon = R^{-1}$. The nonidentity terms in its normalizing denominator are $\psi_0(TR(wc - c))$, so (3.5) gives

$$d_{T,c,R^{-1}} = 1 + O_{N,C}((TR)^{-N}) \tag{3.17}$$

uniformly for $R \ge 1$. The exponential type of $\psi_{T,c,R^{-1}}$ is $R_0 R$; after forming $\beta^* * \beta$, Paley–Wiener therefore gives (3.15) with $B = 2R_0$.

For the identity term use (3.12) and put $w\lambda = Tc + u/R$. The Jacobian is R^{-r}. Split at $\|u\| = TR/2$. On the inner region, (2.11) bounds the density by $C_C T^q (1 + \|u\|)^q$; on the outer region, (2.9) and Schwartz decay apply. Consequently

$$\begin{aligned} \Phi_{T,c,R^{-1}}(e) &\le CR^{-r} \int_{\mathfrak{a}^*} |\psi_0(u)|^2 |\mathbf{c}(Tc + u/R)|^{-2} \, \mathrm{d}u \\ &\le C_C T^q R^{-r} + O_N(T^{-N}), \end{aligned}$$

which is (3.14) after increasing B.

Let $g = k_1 \exp H k_2$ with $\iota \le \|H\| \le BR$. In spherical inversion make the same change of variables and split at $\|u\| = T^{1/4}$. On the inner region the spectral perturbation is u/R, whose norm is at most $T^{1/4}$, and Lemma 2.3 gives

$$|\varphi_{Tc+u/R}(g)| \le C\mathrm{e}^{A\|H\|} T^{-\delta}.$$

Together with (2.11) this part is at most

$$CT^{q-\delta} R^{-r} \mathrm{e}^{A\|H\|} \int_{\mathfrak{a}^*} |\psi_0(u)|^2 (1 + \|u\|)^q \, \mathrm{d}u. \tag{3.18}$$

For $\|u\| > T^{1/4}$ we use $|\varphi_\lambda(g)| \le 1$ and (2.9). Given N, the Schwartz bound yields

$$R^{-r} \int_{\|u\| > T^{1/4}} |\psi_0(u)|^2 (1 + T + \|u\|/R)^q \, \mathrm{d}u = O_N(T^{-N}), \tag{3.19}$$

uniformly for $1 \le R \le A^{-1} \log T$. The finitely many Weyl peaks obey identical estimates. Equations (3.18) and (3.19) imply (3.16), after reducing δ if necessary to absorb the negligible tail and increasing A, B. $\square$

3.2 The quotient kernel and positivity

Let $\Phi \in C_c^\infty(K\backslash G/K)$. Its convolution operator on $L^2(\Gamma\backslash G)^K = L^2(Y)$ has smooth kernel

$$K_\Phi(x,y) = \sum_{\gamma\in\Gamma} \Phi(g_x^{-1}\gamma g_y), \tag{3.20}$$

where $x = \Gamma g_x K$ and $y = \Gamma g_y K$. The sum is locally finite. If $\Phi = \beta^* * \beta$, the operator is positive. We now prove directly, without assuming that Y is compact, the inequality that isolates one eigenfunction. Let B be right convolution by β. Its quotient kernel is

$$K_\beta(x,y) = \sum_{\gamma\in\Gamma} \beta(g_x^{-1}\gamma g_y).$$

For fixed x, the function $K_\beta(cdot\,, x)$ is smooth and belongs to $L^2(Y)$: it is supported in the image of a fixed metric ball about x, and proper discontinuity makes the defining sum locally finite. Unfolding the composition B^*B gives

$$K_\Phi(x,x) = \int_Y |K_\beta(z,x)|^2 \,\mathrm{d}z = \|K_\beta(\,\cdot\,,x)\|_2^2. \tag{3.21}$$

There is no convergence issue in the unfolding: before passing to the quotient, both convolution factors have compact support, so for fixed x, z only finitely many deck transformations occur.

If ϕ is an L^2-normalized spherical joint eigenfunction with parameter λ, spherical convolution acts on its one-dimensional joint character by $B^*\phi = \overline{\widehat{\beta}(\lambda)}\phi$. Therefore

$$\overline{\widehat{\beta}(\lambda)}\phi(x) = \langle \phi, K_\beta(\,\cdot\,,x)\rangle_{L^2(Y)}.$$

Cauchy–Schwarz, (3.21), and $\widehat{\Phi} = |\widehat{\beta}|^2$ give

$$\widehat{\Phi}(\lambda)|\phi(x)|^2 \le K_\Phi(x,x). \tag{3.22}$$

This Hilbert-space proof is the positive pretrace inequality. It remains valid in the presence of continuous spectrum because it never expands the continuous part; it only uses the given L^2 eigenfunction.

3.3 Automorphization and the two scales

For $x = \Gamma g_x K$, automorphizing the kernel gives the exact, locally finite diagonal sum

$$K_{T,c,\epsilon}(x,x) = \sum_{\gamma\in\Gamma} \Phi_{T,c,\epsilon}(g_x^{-1}\gamma g_x). \tag{3.23}$$

The Hilbert-space argument in Equation (3.22) shows that an L^2-normalized joint eigenfunction with parameter Tc satisfies

$$|\phi(x)|^2 \le K_{T,c,\epsilon}(x,x). \tag{3.24}$$

On a fixed compact subset of Y, only uniformly many terms of (3.23) occur. The identity term is $O(\epsilon^r T^q)$ by (3.3), and every nonidentity term is $o_\epsilon(T^q)$ by (3.4). Hence

$$|\phi(x)|^2 \le C\epsilon^r T^q + o_\epsilon(T^q).$$

Letting first $T \to \infty$ and then $\epsilon \downarrow 0$ is the qualitative mechanism behind Theorem 1.1.

For Theorem 1.2, put $R = \kappa \log T$ and $\epsilon = R^{-1}$. The same change of variables as in (3.13) gives the identity contribution $O(T^q R^{-r})$. The quantitative spherical estimate of Proposition 2.1, inserted into the preceding construction, gives Lemma 3.2 and hence

$$\sum_{\gamma \neq e} |\Phi_{T,c,R^{-1}}(g_x^{-1}\gamma g_x)| \ll T^{q-\delta} \mathrm{e}^{CR} R^{-r}.$$

When the injectivity radius is bounded below, orbit packing is uniform in x. Choosing κ sufficiently small makes the last display lower order than $T^q R^{-r}$. Thus the logarithmic saving comes from shrinking all r joint spectral coordinates, while regular spherical oscillation pays for the resulting logarithmic propagation radius.

4 Proof of the theorems

4.1 Compact sets and uniformly thick quotients

We first prove the compact-local assertion in Theorem 1.1.

Proposition 4.1 (Thick-part estimate)**.** *Under the spectral hypotheses of Theorem 1.1, for every $K_0 \Subset Y$,*

$$\|\phi_j\|_{L^\infty(K_0)} = o_{K_0}(T_j^{q/2}).$$

Proof. Let $C \Subset \mathfrak{a}^*_{\mathrm{reg}}$ contain all directions $c_j = \lambda_j/\|\lambda_j\|$. Replacing T_j by $\|\lambda_j\|$ changes nothing. Fix $\epsilon > 0$ and apply Proposition 3.1 with $T = T_j$, $c = c_j$. By (3.22),

$$|\phi_j(x)|^2 \le \Phi_{T_j,c_j,\epsilon}(e) + \sum_{\substack{\gamma \neq e \\ d_X(g_x K, \gamma g_x K) \le R_\epsilon}} |\Phi_{T_j,c_j,\epsilon}(g_x^{-1}\gamma g_x)|. \tag{4.1}$$

For $x \in K_0$, Lemma 2.4 puts every argument in the second line in a compact subset of $G \setminus K$ and bounds the number of terms uniformly. The identity term is at most $C\epsilon^r T_j^q$ by (3.3); the remaining sum is $o_{K_0,\epsilon}(T_j^q)$ by (3.4). Therefore

$$\limsup_{j \to \infty} T_j^{-q} \|\phi_j\|^2_{L^\infty(K_0)} \le C\epsilon^r.$$

Let $\epsilon \downarrow 0$. This proves the local assertion. □

Thus the local assertion of Theorem 1.1 holds without any volume hypothesis.

4.2 The finite-volume expanding-cusp range

We use the coordinates and the precisely invariant cusp neighborhood from the preceding subsection. We first make the rank-one kernel estimate explicit.

Lemma 4.2 (Rank-one peak kernel)**.** *Fix $0 < \epsilon < 1$. For the positive peak of Proposition 3.1 on $\mathbb{H}^{q+1}$,*

$$|\Phi_{T,c,\epsilon}(g)| \le C_\epsilon T^q (1 + T d_X(gK, K))^{-q/2} \tag{4.2}$$

whenever g belongs to its support.

Proof. In real rank one the sole indivisible positive root has total multiplicity q. Marshall's estimate (2.13), applied on the fixed ball of radius R_ϵ, becomes

$$|\varphi_s(\exp H)| \le C_\epsilon(1+s\|H\|)^{-q/2} \quad (s \asymp T,\ \|H\| \le R_\epsilon). \tag{4.3}$$

For the peak at T, put $s = T+\epsilon u$ in spherical inversion. On $|u| \le T/(2\epsilon)$ we have $s \asymp T$, so (4.3), (2.11), and Schwartz decay give

$$\begin{aligned}\epsilon \int_{|u|\le T/(2\epsilon)} &|\psi_0(u)|^2 |\varphi_{T+\epsilon u}(\exp H)| |\mathbf{c}(T+\epsilon u)|^{-2}\,\mathrm{d}u \\ &\le C_\epsilon T^q (1+T\|H\|)^{-q/2} \int_{\mathbb{R}} |\psi_0(u)|^2 (1+|u|)^q\,\mathrm{d}u.\end{aligned}$$

On $|u| > T/(2\epsilon)$ use $|\varphi_s| \le 1$, the polynomial density bound (2.9), and arbitrarily rapid decay of ψ_0. That tail is $O_{\epsilon,N}(T^{-N})$, which is absorbed by the right side of (4.2) because $\|H\| \le R_\epsilon$. The Weyl peak at $-T$ has the same estimate. □

We sum (4.2) over nonzero parabolic translations. Only vectors with $|v| \le C_\epsilon y$ occur, by (2.26) and the support condition. Write $\Phi_{T,c,\epsilon}(r)$ for the common value of this radial kernel on the Cartan double coset at distance r. The number of lattice vectors in a Euclidean ball of radius L is $O_\Lambda(1+L^q)$. Vectors with $0 < |v| \le y/T$ therefore contribute $O_\epsilon(y^q)$: when y/T is below the shortest lattice length the sum is empty, and otherwise their number is $O((y/T)^q)$ and each summand is $O(T^q)$. For $y/T < |v| \le C_\epsilon y$, use (2.27) and decompose into unit Euclidean shells:

$$\begin{aligned}\sum_{y/T<|v|\le C_\epsilon y} &T^q(1+T|v|/y)^{-q/2} \\ &\le C_\epsilon T^{q/2} y^{q/2} \sum_{0<|v|\le C_\epsilon y} |v|^{-q/2} \le C_{\epsilon,\Lambda} T^{q/2} y^q.\end{aligned} \tag{4.4}$$

The last inequality follows from $\sum_{k\le L} k^{q-1-q/2} = O(1+L^{q/2})$. We have proved

$$\sum_{0\ne v\in\Lambda} |\Phi_{T,c,\epsilon}(d((z,y),(z+v,y)))| \le C_\epsilon(y^q + T^{q/2}y^q) \le C_\epsilon T^{q/2} y^q. \tag{4.5}$$

By (2.28), no element outside the cusp stabilizer can occur in the kernel sum when $2\log(y/y_0) > R_\epsilon$. The complementary bounded-height portion of the cusp, together with the compact core, is a fixed compact subset of Y and is already covered by Proposition 4.1.

Fix $\beta < 1/2$. Combining (3.3), (4.5), and positivity shows that, uniformly for $y \le T^\beta$,

$$|\phi(z,y)|^2 \le C\epsilon T^q + C_\epsilon T^{q(1/2+\beta)} + o_\epsilon(T^q) = C\epsilon T^q + o_\epsilon(T^q). \tag{4.6}$$

Here $q(1/2+\beta) < q$, which is the strict power saving needed for the parabolic sum. Divide by T^q, first let $T\to\infty$ with ϵ fixed, and then let $\epsilon \downarrow 0$. Together with the compact-core estimate in Proposition 4.1, this proves (1.3) in every one of the finitely many cusps.

4.3 Proof of the logarithmic improvement

Assume (1.4), and let $\iota > 0$ be a uniform lower bound for the displacement of every nonidentity deck transformation. Let δ, A, B be given by Lemma 3.2, and let h_X be as in (2.23). Choose $\kappa > 0$ so small that $\kappa < A^{-1}$ and

$$\kappa B(A + h_X) < \frac{\delta}{2}, \tag{4.7}$$

and put $R = \kappa \log T$. Positivity, Lemma 3.2, and orbit packing give, uniformly for $x \in Y$,

$$\begin{aligned} |\phi(x)|^2 &\leq \Phi_{T,c,R^{-1}}(e) + \sum_{\substack{\gamma \neq e \\ d(x,\gamma x) \leq BR}} |\Phi_{T,c,R^{-1}}(g_x^{-1}\gamma g_x)| \\ &\ll \frac{T^q}{R^r} + \mathrm{e}^{h_X BR} \frac{T^{q-\delta}}{R^r} \mathrm{e}^{ABR} \ll \frac{T^q}{R^r}. \end{aligned} \tag{4.8}$$

Indeed, the second term divided by $T^q R^{-r}$ is $T^{-\delta+\kappa B(A+h_X)} \leq T^{-\delta/2}$ by (4.7). Taking square roots and using $R \asymp \log T$ proves (1.5).

Every compact quotient satisfies (1.4). A torsion-free convex cocompact real hyperbolic group is finitely generated and has compact convex core. Its translation-length spectrum is discrete away from zero, so $\ell_0 = \inf_{\gamma \neq e} \ell(\gamma) > 0$. Since $d_X(x, \gamma x) \geq \ell(\gamma)$ for every x, one has $\mathrm{inj}_Y(x) \geq \ell_0/2$ globally. This proves all assertions of Theorem 1.2.

4.4 Why the proof stops before the turning height

For completeness, we verify the Airy obstruction advertised in the introduction. It also explains why no hidden global finite-volume claim is contained in Theorem 1.1. Define

$$I_T(a) = \int_a^\infty |K_{\mathrm{i}T}(u)|^2 \frac{\mathrm{d}u}{u}.$$

The uniform Liouville–Green and Airy expansions for imaginary-order Bessel functions imply, for fixed $a_* > 0$ and $0 < c_0 < c_1 < 1$,

$$\mathrm{e}^{\pi T} I_T(a) \asymp \frac{1 + \log(T/a)}{T} \qquad (a_* \leq a \leq c_0 T), \tag{4.9}$$

$$\mathrm{e}^{\pi T} |K_{\mathrm{i}T}(T + sT^{1/3})|^2 \asymp T^{-2/3} |\mathrm{Ai}(2^{1/3} s)|^2 \qquad (|s| \leq s_0), \tag{4.10}$$

where $s_0 > 0$ is fixed and sufficiently small that the displayed Airy factor has no zero. These are direct consequences of Theorems 2.1 and 4.1 of [5]; (4.9) also follows by squaring the oscillatory expansion on $a \leq u \leq c_1 T$. In that range the phase derivative with respect to $\log u$ has magnitude $\sqrt{T^2 - u^2} \asymp T$. Nonstationary phase controls the cosine cross term, while the nonoscillatory term is $\asymp T^{-1}\, \mathrm{d}u/u$; integration gives (4.9). Formula (4.10) is the turning-point scaling; at $s = 0$ the Airy factor is nonzero.

Now take a lattice vector m and a height $y = T/(2\pi|m|)$, so that its Bessel argument is exactly T. If $2\pi|m|y_0 \leq c_0 T$, equations (4.9)–(4.10) give

$$\frac{|K_{\mathrm{i}T}(2\pi|m|y)|^2}{I_T(2\pi|m|y_0)} \asymp \frac{T^{1/3}}{1 + \log(T/(2\pi|m|y_0))}. \tag{4.11}$$

The right side is larger by $T^{1/3}$, up to a logarithm, than the $O(1)$ ratio predicted by replacing the transition value with its oscillatory average. If the shell $|2\pi|m|y - T| \leq T^{1/3}$ contains only finitely many lattice points, spatial lattice summation supplies no averaging that can cancel (4.11). Weighted Cauchy–Schwarz and the Parseval identity (2.31) therefore do not imply a global strict Sarnak bound. Additional arithmetic or geometric information on the individual Fourier coefficients would be required. This is the precise boundary of the present method.

Acknowledgements. The author used OpenAI's ChatGPT for research brainstorming, literature discovery, checking intermediate arguments and calculations, and editorial revision. ChatGPT was not treated as an author or as an independent source of mathematical authority. The author assumes full responsibility for all statements, proofs, citations, and errors.

We thank Simon Marshall for comments on an earlier draft.

References

[1] P. H. Bérard, *On the wave equation on a compact Riemannian manifold without conjugate points*, Math. Z. **155** (1977), 249–276.

[2] F. Brumley and S. Marshall, *Lower bounds for Maass forms on semisimple groups*, Compos. Math. **156** (2020), 959–1003.

[3] F. Brumley and S. Marshall, *Concentration properties of theta lifts on orthogonal groups*, arXiv:2309.06433, 2023.

[4] F. Brumley, S. Marshall, J. Matz, and C. Peterson, *Quantum ergodicity in the Benjamini–Schramm limit for locally symmetric spaces*, arXiv:2604.01075, 2026.

[5] T. M. Dunster, *Uniform asymptotic expansions for Bessel functions of imaginary order and their zeros*, arXiv:2412.12595, 2025.

[6] R. Gangolli, *On the Plancherel formula and the Paley–Wiener theorem for spherical functions on semisimple Lie groups*, Ann. of Math. (2) **93** (1971), 105–165.

[7] A. Hassell and M. Tacy, *Improvement of eigenfunction estimates on manifolds of nonpositive curvature*, Forum Math. **27** (2015), 1435–1451.

[8] S. Hansen, J. Hilgert, and M. Schröder, *Patterson–Sullivan distributions in higher rank*, Math. Z. **272** (2012), 607–643.

[9] H. Iwaniec, *Spectral methods of automorphic forms*, 2nd ed., Grad. Stud. Math. **53**, American Mathematical Society, Providence, RI, 2002.

[10] H. Iwaniec and P. Sarnak, *L^∞ norms of eigenfunctions of arithmetic surfaces*, Ann. of Math. (2) **141** (1995), 301–320.

[11] B. Keeler, *A logarithmic improvement in the two-point Weyl law for manifolds without conjugate points*, Ann. Inst. Fourier **74** (2024), 719–762.

[12] A. V. Kontorovich, *The hyperbolic lattice point count in infinite volume with applications to sieves*, Duke Math. J. **149** (2009), 1–36.

[13] A. Kontorovich and C. Lutsko, *Effective counting in sphere packings*, J. Assoc. Math. Res. **2** (2024), 15–52.

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

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

[16] C. Lutsko, *An abstract spectral approach to horospherical equidistribution*, Nonlinearity **38** (2025), Paper No. 105014.

[17] S. Marshall, *L^p norms of higher rank eigenfunctions and bounds for spherical functions*, J. Eur. Math. Soc. **18** (2016), 1437–1493.

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

[19] P. Sarnak, *Letter to C. Morawetz on bounds for eigenfunctions on symmetric spaces*, 2004, https://publications.ias.edu/sarnak/paper/480.

[20] C. D. Sogge and S. Zelditch, *Riemannian manifolds with maximal eigenfunction growth*, Duke Math. J. **114** (2002), 387–437.

[21] C. D. Sogge and S. Zelditch, *Focal points and sup-norms of eigenfunctions*, Rev. Mat. Iberoam. **32** (2016), 971–994.