Journal de Théorie des Nombres
de Bordeaux 00 (XXXX), 000–000

Bounding hyperbolic and spherical coefficients of Maass forms

par Valentin Blomer, Farrell Brumley, Alex Kontorovich
et Nicolas Templier

Résumé. On développe une nouvelle méthode pour majorer les coefficients de Fourier hyperboliques et sphériques des formes de Maass définies par rapport à des réseaux uniformes généraux.

Abstract. We develop a new method to bound the hyperbolic and spherical Fourier coefficients of Maass forms defined with respect to arbitrary uniform lattices.

Let Σ be a compact hyperbolic surface, endowed with a metric of constant negative curvature -1. Let f be a non-constant eigenfunction of the Laplacian $(\Delta + \lambda)f = 0$ on Σ, normalized so that $||f||_2 = 1$.

Let C be a closed geodesic or a geodesic circle in Σ. (By a geodesic circle we mean the set of points of fixed positive distance from a given point. We require that this distance be less than the injectivity radius of Σ.) Give C the uniform measure of total length 1.

One can expand f in its Fourier expansion along C against a fixed orthonormal basis of characters $\{\psi_n\}_{n\in\mathbb{Z}}$ of $L^2(C)$. Moreover, since f is an eigenfunction for the Laplacian, a separation of variables argument shows that

$$f(z) = \sum_{n\in\mathbb{Z}} c_n(f)\mathcal{W}_{n,\lambda}(z) \tag{0.1}$$

for some coefficients $c_n(f)$, where $\mathcal{W}_{n,\lambda}$ is a properly normalized eigenfunction of Δ on the universal cover $\mathbb{H}$ having eigenvalue λ, transforming along C by ψ_n, and satisfying an (exponential) decay condition. The complex numbers $c_n(f) = c_n(f, C)$ are called the Fourier coefficients of f along C.

2010 *Mathematics Subject Classification.* Primary: 11F70, 22E45.

Mots clefs. Maass forms, Fourier coefficients, geodesics, periods, equidistribution, Sobolev norms, wave front lemma.

V.B. was supported by the Volkswagen Foundation and a Starting Grant of the European Research Council. F.B. is partially supported by the ANR grant ArShiFo ANR-BLANC-114-2010. A.K. is partially supported by NSF grants DMS-1209373, DMS-1064214 and DMS-1001252. N.T. is supported by NSF grant DMS-1200684.

They are hyperbolic or spherical according to whether C is a closed geodesic or a geodesic circle. Under assumptions of arithmeticity of Σ and f, they are related to L-functions via the work of Waldspurger. We shall be interested in their size as $|n|$ gets large.

A local argument shows that $c_n(f) = O(|n|^{1/2})$. If we relax the condition that Σ be hyperbolic, this local bound is actually sharp, as it is realized by the Fourier coefficients of spherical harmonics on S^2 along a great circle. In the hyperbolic setting of this paper, one expects that $c_n(f) = O_\varepsilon(|n|^\varepsilon)$. By a non-trivial bound we mean one which establishes $c_n(f) = O(|n|^{1/2-\delta})$ for some $\delta > 0$. The implied constants here and throughout depend on λ and geometric invariants of Σ. Our goal in this note is to present a new method to bound the coefficients $c_n(f)$ non-trivially. Our main result is the following theorem (see Section 1.5 for a more precise formulation).

Theorem 1. *Let $0 \leq \tau_1 < 1$ be such that $\min(\frac{1}{4}, \lambda_1) = (1 - \tau_1^2)/4$, where λ_1 is the smallest non-zero Laplacian eigenvalue on Σ. Then*

$$c_n(f) = O(|n|^{\frac{1}{2}-\delta})$$

for any

$$0 < \delta < \frac{1 - \tau_1}{34 - 2\tau_1}.$$

This result is not new: in fact it is (numerically) weaker than what is already known from the work of Reznikov [10], where an exponent of 1/3 is achieved, independently of the spectral gap of the lattice Γ.[1]

The novelty of the present paper is therefore in the technique, which, in Reznikov's argument, replaces microlocal analysis of f (or, rather, its representation theoretic counterpart: invariant trilinear forms) with an ergodic argument. The dynamical approach is considerably less deep, but may be adaptable to more general contexts; as *any* nontrivial bound suffices for many applications, the numerical value being irrelevant, we believe that such a method may be useful.

Acknowledgments. This paper owes a great deal to the work of Andre Reznikov, whose sequence of beautiful papers on automorphic periods [3, 10, 11], some in collaboration with Joseph Bernstein, were a source of inspiration for us. We are equally indebted to the ideas of Sarnak [12] and Venkatesh [14] on using the group action to deform test vectors. One of our motivations was to understand the link between the methods of

[1]In [10], the case where C is a closed geodesic is not treated, for reasons which are explained in that paper. Since the appearance of [10], Reznikov has subsequently developed the techniques to deal with this remaining case. Our Theorem 1 is, however, the first time a non-trivial bound on Fourier coefficients about a closed geodesic for a general compact hyperbolic surface has appeared *in print*.

Bernstein and Reznikov, on one hand, and Sarnak and Venkatesh, on the other. We thank the referee for his careful reading and helpful comments.

1. Representation theoretic viewpoint

By the uniformization theorem, we have $\Sigma \simeq \pi_1(\Sigma)\backslash\mathbb{H}$ where $\pi_1(\Sigma)$ is viewed (up to conjugacy) as a torsion free uniform lattice in the group of orientation preserving isometries $\mathrm{Isom}_+(\mathbb{H}) \simeq \mathrm{PSL}_2(\mathbb{R})$. We will work in a slightly more general context than in the introduction, allowing Σ to be an orbifold. Thus from now on we put $G = \mathrm{PSL}_2(\mathbb{R})$, we fix a uniform lattice Γ in G (or any conjugacy class of such), and we let $\Sigma = \Gamma\backslash\mathbb{H}$.

Let $K = \mathrm{PSO}_2(\mathbb{R})$ and endow the quotient G/K with the metric coming from the Killing form on $\mathrm{Lie}(G)$. We isometrically identify the left G-spaces $\mathbb{H}$ and G/K. The unit tangent bundle $T^1(\Sigma)$ of Σ (properly interpreted when Σ is an orbifold) can be identified with $\Gamma\backslash G$. Unlike the base space Σ itself, the quotient $\Gamma\backslash G$ admits an action by G, given by right translation. We denote $X = \Gamma\backslash G$ and let dg be the unique G-invariant probability measure on X.

1.1. Maass forms. There is a unique up to scaling G-invariant second order differential operator on G, called the Casimir operator, denoted by Ω. When Ω is restricted to right K-invariant functions, an appropriate choice of scaling recovers the Laplacian on $\mathbb{H} = G/K$. The Maass form f of the introduction will then be viewed as an L^2-normalized Ω-eigenfunction on G, left-invariant under Γ and right invariant under K.

We may write $\lambda = \frac{1-\tau^2}{4}$, where $\tau \in (-1,1) \cup i\mathbb{R}$. As τ and $-\tau$ give rise to the same λ, we shall henceforth assume that $\tau \in (0,1) \cup i\mathbb{R}_{\geq 0}$. The representation theoretic object associated to the Maass form f is a triplet (π, ν, e_0) given by

(1) an infinite dimensional irreducible unitary spherical representation (π, E_τ) of G on which the Casimir operator acts by the scalar $(1 - \tau^2)/4$. We denote by V_τ (or simply V if no confusion can arise) the space of smooth vectors of E_τ;

(2) a unitary G-intertwining morphism $\nu : V_\tau \to C^\infty(X)$, where $C^\infty(X)$ is equipped with the standard inner product given by integration over X with respect to the measure dg. The map ν is called an *automorphic realization* of V_τ;

(3) an L^2-normalized K-invariant vector $e_0 \in V_\tau$ such that $f = \nu(e_0)$.

Passing from the Maass form f to its associated representation theoretic triplet (π, ν, e_0) offers many advantages. We can, for example, isolate the global (or automorphic) ingredient ν. Moreover, we can in a sense deform e_0 by working with all vectors in the space V.

1.2. Homogeneous cycles. Closed geodesics and geodesic circles C on Σ can be realized as projections of closed orbits of one-parameter subgroups of G acting on X. This allows us to use the tools of Lie theory, homogeneous dynamics, and representation theory to approach the problem of bounding the coefficients $c_n(f)$. More precisely, there exists $x \in X$ and a subgroup H of G such that xH projects to C. We call H spherical or hyperbolic, accordingly to whether C is a geodesic circle or closed geodesic.

When C is a geodesic circle, then $H = g^{-1}Kg$ for some $g \in G$ (see [10, §4.1]). We have identified the upper-half plane $\mathbb{H}$ with G/K, where $K = \mathrm{PSO}_2(\mathbb{R})$, but as there is no preferred choice of maximal compact subgroup of G, we can instead identify $\mathbb{H}$ with the quotient G/K', for some other maximal compact subgroup K'. Doing so with $K' = gKg^{-1}$, the acting group H giving rise to C simplifies to $H = g^{-1}K'g = K$. We will henceforth assume that $H = K$ in the spherical case. We use the usual parametrization of the circle group

$$H = \left\{ h(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} : \theta \in \mathbb{R} \right\}, \tag{1.1}$$

where the bracket matrix refers to the class in $\mathrm{PSO}_2(\mathbb{R})$.

Now let us consider the case of C a closed geodesic. Once again we let $\mathbb{H} = G/K$. Fix the standard identification of the unit tangent bundle $T^1(\mathbb{H})$ of $\mathbb{H}$ with $\mathrm{PSL}_2(\mathbb{R})$ by associating the point $(i, \uparrow) \in T^1(\mathbb{H})$ with the identity element in $\mathrm{PSL}_2(\mathbb{R})$. Then the geodesic flow on $T^1(\mathbb{H})$ is given by the action of A, the group consisting of (classes of) diagonal matrices (see [1, II.§3]). When one changes the base point $(i, \uparrow)$ in $T^1(\mathbb{H})$, the geodesic flow is given by a conjugate of A. This reflects the fact that, as in the maximal compact case, there is no preferred choice of maximal split torus. In the hyperbolic case, we may therefore freely suppose that $H = rAr^{-1}$, where $r = \left[\begin{smallmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{smallmatrix}\right]$. We parametrize our group as

$$H = \left\{ h(\theta) = \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}. \tag{1.2}$$

The above choices of H allow us to use similar notation for both the spherical and hyperbolic cases.

We now fix H to be either (1.1) or (1.2), according to whether C is a geodesic circle or a closed geodesic. Recall that there is an $x \in X$ such that xH projects to C. If $x = \Gamma g$, then the stabilizer of x in H is given by $g^{-1}\Gamma g \cap H$. Now changing Γ to a conjugate lattice $g^{-1}\Gamma g$ gives a surface isometric to Σ, so we are free to suppose that x is the identity coset.

We have reduced the set-up to H being of the form (1.1) or (1.2) and the uniform lattice Γ of G being such that $\Gamma \cap H \backslash H$ is compact. We denote by X_H the H-orbit on the identity coset in X. We give X_H the natural measure coming from the probability Haar measure on $\Gamma \cap H \backslash H$.

1.3. Model representations. If we drop the automorphic realization ν from our triplet (π, ν, e_0) then we are left with an abstract representation (π, V_τ). Abstract unitary representations of semisimple Lie groups such as $\mathrm{PSL}_2(\mathbb{R})$ can often be realized concretely as spaces of sections of vector bundles on flag varieties. This is the case for our representation (π, V_τ), as we explain below.

Let $\mathbb{R}^2 - 0$ be the plane punctured at the origin. We view the elements of $\mathbb{R}^2 - 0$ as column vectors. Let $V_\tau^{mod} = C^\infty_{ev,-\tau-1}(\mathbb{R}^2 - 0)$ consist of smooth complex valued functions Φ on $\mathbb{R}^2 - 0$ which are even and homogeneous of degree $-\tau - 1$. These conditions mean that

$$\Phi \begin{pmatrix} ax \\ ay \end{pmatrix} = |a|^{-\tau-1} \Phi \begin{pmatrix} x \\ y \end{pmatrix}$$

for all $a \in \mathbb{R}^\times$. The natural action of $g \in G$ on $\mathbb{R}^2 - 0$ by left matrix multiplication by g^{-1} induces a G-action on the space of functions V_τ^{mod}. The resulting representation, denoted π^{mod}, is irreducible and given explicitly by

$$\begin{aligned} \left(\pi^{mod}(g).\Phi\right) \begin{pmatrix} x \\ y \end{pmatrix} &= \Phi \left(g^{-1} \begin{pmatrix} x \\ y \end{pmatrix} \right) \\ &= \Phi \begin{pmatrix} \delta x - \beta y \\ -\gamma x + \alpha y \end{pmatrix} \quad \text{if } g = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}. \end{aligned} \tag{1.3}$$

The one-dimensional space of K-invariants consists of rotationally invariant functions. There exists an intertwining $V_\tau \to V_\tau^{mod}$, unique up to scalars, which we denote by $v \mapsto \Phi_v$.

We now endow V_τ^{mod} with a G-invariant inner product, in such a way that $v \mapsto \Phi_v$ becomes an isometry. A very general way of doing this is to take any simple closed curve Q going around the origin and endow it with the measure dq corresponding to "area swept out": an open set U of Q is given the Lebesgue measure of the cone emanating from the origin and abutting at U. One then defines

$$\langle \Phi, \Phi \rangle = \int_Q \Phi(q)(\mathcal{I}\Phi)(q) dq, \tag{1.4}$$

where $\mathcal{I} : V_\tau^{mod} \to V_{-\tau}^{mod}$ is an isometric intertwining. The curve Q can close "at infinity" and the condition that the functions in V_τ^{mod} be even allows us to integrate over a curve confined to some half-plane.

We will choose the curve defining the inner product according to the relevant group H. Let $X_H^{mod} \subset \mathbb{R}^2 - 0$ be the H-orbit passing through the origin $x_0 = \binom{1}{0}$. In the spherical case, this *model orbit* is the circle

$$\left\{ x_\theta = (\cos\theta, \sin\theta)^t : \theta \in [-\pi, \pi] \right\},$$

while in the hyperbolic case, it is the hyperbola

$$\left\{x_\theta = (\cosh\theta, \sinh\theta)^t : \theta \in \mathbb{R}\right\}.$$

We give X_H^{mod} the measure $\frac{1}{2\pi}d\theta$ in the spherical case and $\frac{1}{2}d\theta$ in the hyperbolic case. The formula (1.4), where $Q = X_H^{mod}$ and dq is the above choice of measure, is our choice of G-invariant inner product. In the spherical case, this corresponds to passing to the circle model. Another common choice of Q is the vertical line through x_0 with measure $\frac{1}{2}dx$; this is referred to as the line model. The hyperbolic model does not seem to be present in the literature.

We write this out explicitly as follows. When $\tau \in i\mathbb{R}$ we have $V_{-\tau} = \overline{V_\tau}$, the complex conjugate representation. In this case the G-invariant scalar product $\langle \Phi, \Phi \rangle_\tau$ is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\Phi(x_\theta)|^2 d\theta \quad \text{ or } \quad \frac{1}{2}\int_{-\infty}^{\infty} |\Phi(x_\theta)|^2 d\theta,$$

for H spherical or hyperbolic, respectively. When $\tau \in (0,1)$, the G-invariant scalar product $\langle \Phi, \Phi \rangle_\tau$ is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(x_\theta)(\mathcal{I}\Phi)(x_\theta) d\theta \quad \text{ or } \quad \frac{1}{2}\int_{-\infty}^{\infty} \Phi(x_\theta)(\mathcal{I}\Phi)(x_\theta) d\theta \tag{1.5}$$

where the intertwining map $(\mathcal{I}\Phi)(x_\theta)$ is (see [7, §1.3.2], where their s is our $-\tau$)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{\Phi(x_{\theta'})}|\sin(\theta-\theta')|^{\tau-1}d\theta' \quad \text{ or } \quad \frac{1}{2}\int_{-\infty}^{\infty} \overline{\Phi(x_{\theta'})}|\sinh(\theta-\theta')|^{\tau-1}d\theta'.$$

We will sometimes drop the subscript τ and write $\langle\, , \rangle$ if no confusion can arise.

1.4. Periods and multiplicity one. For a character χ of H denote by $\mathbb{C}_\chi$ the complex numbers viewed as a representation space for χ. Given a χ such that $\dim \mathrm{Hom}_H(V_\tau, \chi) = 1$, we may define a non-zero element of $\mathrm{Hom}_H(V, \mathbb{C}_\chi)$ by the model inner product

$$\ell_\chi^{mod}(v) = \frac{\langle \Phi_v, \chi \rangle_\tau}{\sqrt{\langle \chi, \chi \rangle_\tau^{\mathrm{reg}}}}.$$

We make some remarks on this definition. In the spherical case, we may view χ as a vector in V_τ^{mod} by extending it uniquely, by homogeneity, from the circle to a function on the punctured plane. Moreover, the inner product in the denominator requires no regularization in this case. In the hyperbolic case, however, one cannot view χ as an element of V_τ^{mod}. The inner product $\langle \Phi_v, \chi \rangle_\tau$ in the numerator is therefore taken to be the integral defined in the previous section; it is the Fourier transform of Φ_v along the hyperbola. As for the regularized inner product, it is given in the hyperbolic case by

$$\langle \Phi, \Phi \rangle_\tau^{\mathrm{reg}} = \lim_{\delta \to 0} \langle \Phi K_\delta, \Phi \rangle_\tau,$$

where K_δ is any good kernel, such as $K_\delta(\theta) = 2\delta^{-1/2} e^{-\pi\theta^2/\delta}$. When $\tau \in i\mathbb{R}$ we have $\langle \chi, \chi \rangle_\tau^{\mathrm{reg}} = 1$.

If, furthermore, χ descends to $\Gamma \cap H \backslash H = X_H$, then we define an element $\ell_\chi^{aut} \in \mathrm{Hom}_H(V, \mathbb{C}_\chi)$ by the integral

$$\ell_\chi^{aut}(v) = \int_{X_H} \nu(v)\overline{\chi}.$$

As the space $\mathrm{Hom}_H(V, \mathbb{C}_\chi)$ is one dimensional, the (H, χ)-equivariant functionals ℓ_χ^{aut} and ℓ_χ^{mod} are proportional. Thus,

$$\ell_\chi^{aut} = c_\chi(\nu) \ell_\chi^{mod}$$

for some complex number $c_\chi(\nu) \in \mathbb{C}$. This constant of proportionality depends, up to a scalar factor of modulus 1, on the choice of model representation π^{mod} and the automorphic realization ν.

1.5. A more precise formulation of Theorem 1. We are now in a position to state Theorem 1 more precisely.

Assume X_H projects to the fixed curve $C \subset \Sigma$ from the Introduction. Fix a group isomorphism of $\mathbb{Z}$ with the character group of $\Gamma \cap H \backslash H = X_H$, denoted $n \mapsto \chi_n$. Each χ_n can be naturally viewed as a function on X_H, and (since X_H was given volume 1) the set $\{\chi_n\}_{n \in \mathbb{Z}}$ forms an orthonormal basis of $L^2(X_H)$. We write ℓ_n^{aut} for $\ell_{\chi_n}^{aut}$ and similarly for the model functionals. This defines coefficients of proportionality $c_n(\nu)$. Since we will only work with the modulus of the Fourier coefficients, we abuse notation and write simply c_n. Then the statement of Theorem 1 is that $c_n = O(|n|^{\frac{1}{2} - \delta})$, for the value of δ given there.

To see how to obtain the original statement, in the more concrete setting of the Fourier expansion (0.1), we let $\{\psi_n\}_{n \in \mathbb{Z}}$ be the orthonormal basis of $L^2(C)$ given by the image of $\{\chi_n\}_{n \in \mathbb{Z}}$ under the projection down to Σ, and put

$$\mathcal{W}_{n,\lambda}(z) = \ell_n^{mod}\left(\begin{bmatrix} r^{1/2} & \\ & r^{-1/2} \end{bmatrix} e_0 \right) \psi_n(\theta),$$

where $z = (r, \theta) = \left[\begin{smallmatrix} r^{1/2} & \\ & r^{-1/2} \end{smallmatrix}\right] h(\theta).z_0 \in \mathbb{H}$ and $z_0 \in C$ is fixed.

1.6. Other periods. One could relax the restriction that the hyperbolic surface (or orbifold) Σ be compact, and instead ask only that it be of finite volume, i.e., remove the condition that the lattice Γ be uniform. In this more general context, one can still study the Fourier coefficients of a Maass form along a closed geodesic or hyperbolic circle. However, some of the Sobolev norm estimates that we quote in Section 3 have not yet been extended to such non-compact Σ. Moreover, since the equidistribution result Lemma 1 would naturally be stated for functions of compact support, one would need to control the error incurred when applied to Maass forms, perhaps assuming cuspidality.

When Γ is non-uniform, one can also consider Fourier coefficients along closed horocycles. If Γ is a congruence subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ and the Maass form is an eigenfunction of all Hecke operators, bounds on the unipotent Fourier coefficients are directly related to the Ramanujan conjecture. For an arbitrary lattice Γ, non-trivial bounds were first obtained by Good [8] and Sarnak [13]. Later, Venkatesh [14] gave a more dynamical treatment which used the mixing properties of the geodesic flow together with an amplification technique. Our method is quite close to that of Venkatesh, but rather than amplifying we "shorten the interval". In other words, we establish a sharp upper bound for the mean value $\sum_{|n-N|\leq T} |c_n|^2$ where $T = N^{1-\delta}$ for some $\delta > 0$, and the bound in Theorem 1 follows after dropping all but one term. A bound for dyadic interval would only recover the trivial bound.

Finally we mention that in the non-compact case, one can also consider the Fourier coefficients along a divergent geodesic, i.e., one which leaves every compact as $t \to \pm\infty$. Here, one more difficulty arises: the cycle itself is not of finite volume. See the work of Oh-Shah [9] for interesting results in this setting. In the case of congruence subgroups, the period integral along the divergent geodesic $i\mathbb{R}_+$ of a Hecke-Maass form against a character is related to special value of an L-function by means of the classical Hecke integral.

2. Summation formula

For any smooth function F on G and any $1 \leq p \leq \infty$ let

$$\mathcal{S}_{p,d}(F) = \sum_{\mathrm{ord}(\mathcal{D})\leq d} ||\mathcal{D}F||_{L^p(\Gamma\backslash G)}$$

be the Sobolev (p, d)-norm. Here $\mathcal{D}$ ranges over all monomials of a fixed basis of $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$. The action of an element $X \in \mathfrak{g}$ is by right differentiation $(X.F)(g) = \frac{d}{dt}F(ge^{tX})\big|_{t=0}$, and one extends this action to monomials by composition.

It is a well-known phenomenon that the smoothness and regularity properties of test functions influence the quality of corresponding analytic estimates. The above norms $S_{p,d}(F)$ measure the L^p-norms of F along with its first few derivatives. A standard Sobolev norm in analytic number theory is $\mathcal{S}_{\infty,1}$ which provides pointwise bounds for both F and its first order derivatives and comes in handy in first order Taylor approximations.

Proposition 1 below also features a slightly more complicated norm that we now define. For any $0 < \varepsilon < \frac{1}{2}$, put

$$\mathcal{N}_\varepsilon(F) = \mathcal{S}_{2,1}(F)^{1/2+\varepsilon} ||F||_2^{1/2-\varepsilon}.$$

Note that $\mathcal{N}_\varepsilon(\lambda F) = \lambda \mathcal{N}_\varepsilon(F)$ for scalars $\lambda > 0$. The norm $\mathcal{N}_\varepsilon$ is a simpler substitute for the fractional Sobolev $(2, 1/2+\varepsilon)$-norm. It comes from effective mixing, imported from [14, Equation (9.6)] that we use below in the proof of Lemma 1. At the cost of a slightly weaker exponent in Theorem 1, we could majorize $\mathcal{N}_\varepsilon$ by the straightforward norm $\mathcal{S}_{2,1}$.

For $T > 0$ let

$$a(T) = \begin{pmatrix} T^{1/2} & \\ & T^{-1/2} \end{pmatrix},$$

and if $v \in V$ put

$$v_T = \pi(a(T))v.$$

The aim of this section is to prove the following result.

Proposition 1. *For any smooth vector $v \in V$ and for all $T > 1$ we have*

$$\sum_{n\in\mathbb{Z}} |c_n|^2 |\ell_n^{mod}(v_T)|^2 - \int_X |\nu(v)|^2$$
$$= O_\varepsilon\left(\frac{\mathcal{S}_{\infty,1}(|\nu(v)|^2)^{\frac{1}{2}+\varepsilon} \mathcal{N}_\varepsilon(|\nu(v)|^2)^{\frac{1}{2}}}{T^{(1-\tau_1)/4-\varepsilon}} + \frac{\mathcal{S}_{\infty,1}(|\nu(v)|^2)}{T} \right).$$

Here τ_1 is related to λ_1 as in Theorem 1.

Let $\ell_{X_H}^{aut} \in \mathrm{Hom}_H(V \otimes \overline{V}, \mathbb{C})$ be the linear form on $V \otimes \overline{V}$ given by the L^2-norm of the restriction of $\nu(v)$ to X_H. In other words

$$\ell_{X_H}^{aut}(v) = \int_{X_H} |\nu(v)|^2.$$

The idea of the proof of Proposition 1, which goes back to Sarnak [12], is to evaluate $\ell_{X_H}^{aut}(v_T)$ in two different ways.

On one hand, we have the decomposition

$$\ell_{X_H}^{aut} = \sum_{n\in\mathbb{Z}} |\ell_n^{aut}|^2$$

given by the Parseval formula for $L^2(X_H)$. Inserting the proportionality relation we obtain

$$\ell_{X_H}^{aut} = \sum_{n\in\mathbb{Z}} |c_n|^2 |\ell_n^{mod}|^2,$$

which we can then apply to the vector v_T to obtain the left hand-side of Proposition 1.

On the other hand,

$$\ell^{aut}_{X_H}(v_T) = \int_{X_H} |\nu(v)(ha(T))|^2 dh.$$

Proposition 1 then follows by an application of the following lemma to the case where $F = |\nu(v)|^2$.

Lemma 1. *For any $F \in C^\infty(\Gamma\backslash G)$ and all $T > 1$ we have*

$$\int\limits_{\Gamma\cap H\backslash H} F(ha(T))dh - \int\limits_{\Gamma\backslash G} F(g)dg = O_\varepsilon\Big(\frac{\mathcal{S}_{\infty,1}(F)^{\frac{1}{2}+\varepsilon}\mathcal{N}_\varepsilon(F)^{1/2}}{T^{(1-\tau_1)/4-\varepsilon}} + \frac{\mathcal{S}_{\infty,1}(F)}{T}\Big).$$

In words, Lemma 1 states that the $a(T)$-translate of the probability measure integrating over the homogeneous cycle X_H weak-* converges, in an effective way, to the uniform probability measure on $\Gamma\backslash G$.

We prove Lemma 1 following a well-known argument of Eskin and McMullen [6] that goes back to Margulis's thesis. The idea is that the $a(T)$ translate of a fattening of X_H remains uniformly close to a fattening of the translate. This is the wavefront lemma, see [6]; it is a reflection of the negative curvature of Σ. Thus the left hand side can be viewed as a matrix coefficient, up to a small error. Then uniform and effective versions of the Howe-Moore theorem provide the desired decay of matrix coefficients.

Before proceeding, we introduce some explicit coordinates and group decompositions. Let

$$A = \left\{ a(t) = \begin{pmatrix} t^{1/2} & \\ & t^{-1/2} \end{pmatrix} : \quad t > 0 \right\},$$
$$N = \left\{ n(x) = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} : \quad x \in \mathbb{R} \right\}.$$

Using the parametrization of (1.1) or (1.2), we write a general element of H as $h(z)$ for $z \in \mathbb{R}$,

A generalized (local) form of the Iwasawa decomposition grants the existence of a neighborhood U of the identity element $e \in G$ such that any $g \in U$ can be written uniquely as $g = han$, where $h \in H$, $a \in A$, and $n \in N$. We can in fact take U to be all of G when H is spherical, for in this case one obtains the standard Iwasawa decomposition $G = KAN$.

We now prove Lemma 1. We borrow heavily from the presentation of a similar result in [14, Lemma 9.4].

Proof. We begin by defining a smoothing function on G, depending on a small parameter $0 < \delta < 1$. We make use of the HAN coordinates in the neighborhood U. For the H-coordinate, write $p > 0$ (for "period") for the smallest positive real number such that $h(p) \in \Gamma$, and let $\eta \in C_c^\infty(\mathbb{R})$ be such that

(1) (a) $\eta(z) = 1$ for all $\delta \leqslant z \leqslant p - \delta$,
(b) $\eta(z) = 0$ for all $z \leqslant \delta/2$ and all $p - \delta/2 \leqslant z$,
(c) $\frac{1}{p}\int_0^p \eta(z)dz = 1$.

For the A^0- and N-coordinates, let $\alpha, \nu \in C_c^\infty(\mathbb{R})$ be real-valued functions satisfying

(2) $\mathrm{supp}(\alpha) \subset [-\delta, \delta]$ and $\int_{\mathbb{R}} \alpha(y)dy = 1$;

(3) $\mathrm{supp}(\nu) \subset [-1, 1]$ and $\int_{\mathbb{R}} \nu(x)dx = 1$.

For δ small enough the image of $(z, y, x) \in [0, p] \times [-\delta, \delta] \times [-1, 1]$ under $h(z)a(e^y)n(x)$ is contained in U, and we may define a function $\xi \in C_c^\infty(G)$ by

$$\xi(h(z)a(e^y)n(x)) = \eta(z)\alpha(y)\nu(x).$$

We obtain a Γ-invariant non-negative function on G by averaging over the group,

$$\Xi(g) = \sum_{\gamma\in\Gamma} \xi(\gamma g).$$

From the HAN decomposition of the neighborhood U as well as the fact that Γ intersects A and N only at e, we compute the volume of Ξ by unfolding to get

$$\int_{\Gamma\backslash G} \Xi(g)dg = \int_{\Gamma\cap H\backslash G} \xi(g)dg = \frac{1}{p}\int_0^p \eta(z)dz \int_{\mathbb{R}} \alpha(y)dy \int_{\mathbb{R}} \nu(x)dx = 1.$$

With these preliminaries out of the way, we now consider the matrix coefficient

$$\langle a(T) \cdot F, \Xi\rangle = \int_{\Gamma\backslash G} F\big(ga(T)\big)\Xi(g)dg$$

in the right regular representation of G on $L^2(\Gamma\backslash G)$. This is the probability measure averaging the $a(T)$ translate of F over a small $O(1) \times O(\delta) \times O(1)$ ball (in the $H \times A \times N$-coordinates) around the origin (which is the same as averaging F over the $a(T)$ translate of an $O(\delta) \times O(1)$ fattening (in the $A \times N$ coordinates) of the H-orbit X_H). A uniform version of the Howe-Moore theorem states (see [14, Equation (9.6)]) that

$$\langle a(T) \cdot F, \Xi\rangle = \int_{\Gamma\backslash G} F(g)dg \int_{\Gamma\backslash G} \Xi(g)dg + O_\varepsilon\big(\mathcal{N}_\varepsilon(F)\mathcal{N}_\varepsilon(\Xi)T^{-\frac{1}{2}(1-\tau_1)+\varepsilon}\big).$$

Using the volume normalization of Ξ and the bounds

$$\mathcal{N}_\varepsilon(\Xi) \ll (\delta^{-3/2})^{1/2+\varepsilon}(\delta^{-1/2})^{1/2-\varepsilon} = \delta^{-1-\varepsilon},$$

we find

$$\langle a(T) \cdot F, \Xi\rangle = \int_{\Gamma\backslash G} F(g)dg + O_\varepsilon\big(\mathcal{N}_\varepsilon(F)T^{-\frac{1}{2}(1-\tau_1)+\varepsilon}\delta^{-1-\varepsilon}\big). \tag{2.1}$$

On the other hand, we may unfold the integral to obtain

$$\langle a(T)\cdot F, \Xi\rangle = \int_{\Gamma\cap H\backslash G} F\big(ga(T)\big)\xi(g)dg$$

which is

$$\frac{1}{p}\int_0^p \int_{y\in\mathbb{R}}\int_{x\in\mathbb{R}} F\big(h(z)a(e^y)n(x)a(T)\big)\eta(z)\alpha(y)\nu(x)dxdydz.$$

Using the contraction relation

$$a(T)^{-1}n(x)a(T) = n(xT^{-1}), \tag{2.2}$$

we find

$$\begin{aligned}&\langle a(T)\cdot F, \Xi\rangle\\ &\quad = T\frac{1}{p}\int_0^p \int_{y\in\mathbb{R}}\int_{x\in\mathbb{R}} F\big(h(z)a(T)a(e^y)n(x)\big)\eta(z)\alpha(y)\nu(Tx)dxdydz.\end{aligned}$$

The right hand side is the probability measure averaging F over a $O(\delta)\times O(1/T)$-fattening (in the $A\times N$ directions) of the $a(T)$-translate of X_H. The contraction relation (2.2), as elementary as it may seem, encodes the negative curvature of Σ. Indeed, it is simply a group theoretic reformulation of the geometric fact that the geodesic flow admits a hyperbolic structure on the unit tangent bundle of Σ.

Continuing, note that for $x = O(T^{-1})$ and $y = O(\delta)$ and any $g\in G$ one has

$$F\big(ga(x)n(y)\big) = F(g) + O\big(\mathcal{S}_{\infty,1}(F)\max(T^{-1},\delta)\big),$$

the implied constant being independent of g by the compactness of $\Gamma\backslash G$. Since the x, y integration variables are constrained to these ranges (and α and ν have volume 1), we deduce

$$\langle a(T)\cdot F, \Xi\rangle = \int_0^p F\big(h(z)a(T)\big)\eta(z)dz + O\big(\mathcal{S}_{\infty,1}(F)\max(T^{-1},\delta)\big).$$

From the assumed properties of η, we have

$$\langle a(T)\cdot F, \Xi\rangle = \int_{\Gamma\cap H\backslash H} F\big(ha(T)\big)dh + O\big(\mathcal{S}_{\infty,1}(F)\max(T^{-1},\delta)\big). \tag{2.3}$$

Combining (2.1) and (2.3) and choosing

$$\delta = \frac{\mathcal{N}_\varepsilon(F)^{1/2}}{\mathcal{S}_{\infty,1}(F)T^{(1-\tau_1)/2}}$$

completes the proof. $\square$

The ergodic result could have been proved spectrally (in the style of Duke-Rudnick-Sarnak [5]), instead of appealing to the wave-front lemma as we have. Proceeding by spectral methods to prove Lemma 1 is not, as one might suspect, a repackaging of Reznikov's argument. Indeed, the spectral argument reduces to bounding matrix coefficients and the resulting bounds depend again on the spectral gap of the lattice, which is not a feature of Reznikov's argument.

3. Results on Sobolev norms

Let $\{e_n\}_{n\in 2\mathbb{Z}}$ be an orthonormal basis of K-types in V_τ. We write $\phi_\tau^{(n)} = \nu(e_n)$ for the corresponding automorphic functions.

We list some useful bounds that follow from the work of Bernstein and Reznikov.

Lemma 2. *For any $n \in 2\mathbb{Z}$, and any integer $d \geq 0$, one has*

$$\mathcal{S}_{2,d}(\phi_\tau^{(n)}) \ll_d (|\tau| + |n| + 1)^d$$

and

$$\mathcal{S}_{\infty,d}(\phi_\tau^{(n)}) \ll_d (|\tau| + |n| + 1)^{d+1/2},$$

where the implied constants depend only on d.

Proof. This follows from [2, Theorem 2.1, Corollary 2.4] and [4, Proposition 2.5.4]. □

Lemma 3. *For any $n, m \in 2\mathbb{Z}$, $\tau' \in (0,1) \cup i\mathbb{R}_{\geq 0}$, and any $D \geq 0$, one has*

$$\langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \ll_{\tau,D} \min\left(1, \frac{(|m| + |n| + 1)^{D+1}}{(1 + |\tau'|)^D}\right),$$

where the implied constants depend only on D and τ.

Proof. We first observe that

$$\langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \leq \|\phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}\|_2 \leq \|\phi_\tau^{(m)}\|_4 \|\phi_\tau^{(n)}\|_4 \ll_\tau 1,$$

the last estimate on the L^4-norm being taken from [3, Theorem 2.6].

On the other hand, since the Casimir operator is self-adjoint (and τ' is bounded away from 1), for any fixed integer $d \geq 0$ we have

$$\langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \ll_d \frac{1}{(1 + |\tau'|)^{2d}} \|\Delta^d \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}\|_2.$$

It follows from [2, Theorem 2.1, Corollary 2.4] and [4, Proposition 2.5.4], that $\|\phi_\tau^{(n)}\|_\infty \ll \mathcal{S}_{2,1}(\phi_\tau^{(n)})$. From these estimates, and the Leibniz rule for derivations, one infers that

$$\langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \ll_d (1 + |\tau'|)^{-2d} \sum_{d_1 + d_2 = 2d+1} \mathcal{S}_{\infty,d_1}(\phi_\tau^{(m)}) \mathcal{S}_{2,d_2}(\phi_\tau^{(n)}).$$

Using the estimates in Lemma 2, we conclude the proof. □

Corollary 1. *Let $N \geqslant 1$ be an integer, and let a_n be a sequence of complex numbers supported on the even integers. Put*

$$g = \Big|\sum_{|n|\le N} a_n \phi_\tau^{(n)}\Big|^2 \qquad \textit{and} \qquad A = \sum_{|n|\le N} |a_n|^2.$$

Then for any integer $d \geq 0$ we have

$$\mathcal{S}_{\infty,d}(g) \ll_{d,\tau} AN^{7/2+d+\varepsilon} \qquad \textit{and} \qquad \mathcal{S}_{2,d}(g) \ll_{d,\tau} AN^{3+d+\varepsilon}.$$

Proof. By spectral expansion

$$g = \sum_{|m|,|n|\le N} a_m \bar{a}_n \sum_{\tau'} \langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \phi_{\tau'}^{(n-m)},$$

where τ' runs over all spectral parameters (with multiplicity). We apply Lemmas 2 and 3 to the τ'-sum. Using Weyl's law $\#\{\tau' \le R\} \ll R^2$ we deduce the stated inequalities. □

Note that the proof actually gave the stronger upper bounds

$$\mathcal{S}_{\infty,d}(g) \ll_{d,\tau} \Big(\sum_{|n|\le N} |a_n|\Big)^2 N^{5/2+d+\varepsilon},\ \mathcal{S}_{2,d}(g) \ll_{d,\tau} \Big(\sum_{|n|\le N} |a_n|\Big)^2 N^{2+d+\varepsilon},$$

to which we then applied Cauchy-Schwarz. We shall not need these last estimates.

We also remark that the proof of Lemma 3 and hence implicitly the proof of Corollary 1 uses deep and intricate methods of representation theory via the L^4-norm estimate imported from [3]. As mentioned in the introduction, the philosophy of this paper is to prove non-trivial bounds for Fourier coefficients mainly by ergodic methods. However, the use of [3, Theorem 3] can easily be avoided by the slightly weaker elementary estimate

$$\langle \phi_\tau^{(m)} \overline{\phi}_\tau^{(n)}, \phi_{\tau'}^{(n-m)} \rangle \le \|\phi_\tau^{(m)}\|_\infty \ll_\tau (|m|+1)^{1/2}$$

(by Lemma 2) which weakens the bounds in Corollary 1 by a factor $N^{1/2}$. This would replace the number 34 by 38 in Theorem 1.

4. Test functions and estimates

We now choose an explicit test function v to insert into Proposition 1.

We shall assume throughout this section that the Maass form f is tempered; thus $\tau \in i\mathbb{R}_{\geq 0}$. (Note that this is not the same thing as to assume that the *lattice* Γ is tempered: τ_1 could very well be close to 1.) The calculations when f is non-tempered are similar.

4.1. General formulae. We begin by putting in place some notation.

For a smooth vector $v \in V_\tau$ let $\Phi = \Phi_v$ be the associated vector in the model space $C^\infty_{ev,-\tau-1}(\mathbb{R}^2 - 0)$. Denote by φ the restriction of Φ to X_H^{mod} (see §1.4).

We recall the fixed isomorphism $n \mapsto \chi_n$ between $\mathbb{Z}$ and the character group of $\Gamma \cap H \backslash H$ of Section 1.5. As in the proof of Lemma 1, there is a $p > 0$ such that $\chi_n(\theta) = e^{inp\theta}$. For notational convenience, we shall assume henceforth that $p = 1$.

In the spherical case, we have

$$\ell_n^{mod}(v) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi(\theta) e^{-in\theta} d\theta.$$

For $\cos\theta \neq 0$ we have $\varphi(\theta) = |\cos\theta|^{-\tau-1}\Phi\left(\begin{smallmatrix}1\\ \tan\theta\end{smallmatrix}\right)$. Thus, if Φ is supported away from the line $x = 0$ we have

$$\ell_n^{mod}(v) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\cos\theta|^{-\tau-1}\Phi\begin{pmatrix}1\\ \tan\theta\end{pmatrix} e^{-in\theta} d\theta.$$

In the hyperbolic case, we have

$$\ell_n^{mod}(v) = \frac{1}{2}\int_{-\infty}^{\infty} \varphi(\theta) e^{-in\theta} d\theta.$$

For all θ we have $\varphi(\theta) = (\cosh\theta)^{-\tau-1}\,\Phi\left(\begin{smallmatrix}1\\ \tanh\theta\end{smallmatrix}\right)$, so that

$$\ell_n^{mod}(v) = \frac{1}{2}\int_{-\infty}^{\infty} (\cosh\theta)^{-\tau-1}\,\Phi\begin{pmatrix}1\\ \tanh\theta\end{pmatrix} e^{-in\theta} d\theta.$$

In either case, for Φ supported in a small neighborhood of $\pm x_0$, the expression for $\ell_n^{mod}(v)$ closely approximates the Fourier transform of $y \mapsto \Phi\left(\begin{smallmatrix}1\\ y\end{smallmatrix}\right)$.

We denote $v_T = \pi(a(T)).v$ as in Section 2 and put $\Phi_T = \Phi_{v_T}$. Similarly to before, let φ_T be the restriction of Φ_T to the model orbit X_H^{mod}. Observe that on vertical lines $x = \text{const}$ the A-action is given by contraction:

$$\Phi_T\begin{pmatrix}x\\ y\end{pmatrix} = \Phi\left(a(T^{-1})\begin{pmatrix}x\\ y\end{pmatrix}\right) = T^{(1+\tau)/2}\Phi\begin{pmatrix}x\\ Ty\end{pmatrix}.$$

Thus we have the formula

$$\varphi_T(\theta) = |\cos\theta|^{-\tau-1}T^{(1+\tau)/2}\Phi\begin{pmatrix}1\\ T\tan\theta\end{pmatrix},$$

valid for $\cos\theta \neq 0$. The formula in the hyperbolic case is the same, but with hyperbolic trigonometric functions (and valid for all θ). In practice, we will take Φ supported in a small neighborhood of $\pm x_0$, in which case this formula is valid for all θ in the support of φ_T. Assuming this, we have

$$\ell_n^{mod}(v_T) = T^{(1+\tau)/2}\int_{-\pi}^{\pi} (\cos\theta)^{-\tau-1}\Phi\begin{pmatrix}1\\ T\tan\theta\end{pmatrix} e^{-in\theta} d\theta,$$

and similarly in the hyperbolic case.

4.2. Mean value estimates and convexity breaking. If we choose v such that Φ_v is a *fixed* smooth L^2-normalized compactly supported bump function around $\pm x_0$, then one can show that

$$|\ell_n^{mod}(v_T)|^2 = T^{-1} f(n/T)$$

for a fixed positive smooth function f of rapid decay. In this case the summation formula in Proposition 1 reads

$$\sum_{n\in\mathbb{Z}} |c_n|^2 f(n/T) = T + O_\varepsilon(T^{1-\frac{1}{4}(1-\tau_1-\varepsilon)}).$$

This is a smooth mean-square asymptotic for the coeffcients c_n. Despite the power savings error term, the smoothness of the sum does not allow one to recover anything more than the trivial bound $c_n = O(|n|^{1/2})$ on any individual coefficient.

In the important paper [14], Venkatesh observed that one can set up a correspondence between certain dynamical techniques, such as equidistribution and mixing, and classical techniques in analytic number theory for establishing subconvexity of special values of L-functions, such as mean value estimates and the amplification method of Friedlander-Iwaniec. He then used the dynamical reformulation to greatly expand the number of examples for which one can prove a subconvex bound. We now discuss the relation between the method of Venkatesh and ours.

In the general outline of [14, §1.3], a subconvexity result is proven in two steps. First one establishes a mean value estimate on the coefficients c_n one is interested in bounding. This corresponds to the equidistribution of a sequence of measures supported on expanding H-orbits. The next step is to amplify: this means that one seeks to add weights to the mean value which are peaked at a given coefficient. The dynamical equivalent of this is to find a joint eigenmeasure σ for the orthonormal basis $\{\psi_n\}$ of $L^2(C)$ with respect to which f is uncorrelated. The eigenvalues of σ on ψ_n are the weights in the mean value, and the independence of f with σ is expressed quantitatively by the mixing of the H-action on X. The tension between f, ψ_n, and σ yields bounds on c_n.

In our setting, the equidistribution result Lemma 1 is not for a sequence of measures supported on H-orbits, but rather of measures supported on *translates* of a fixed H-orbit. There do exist situations where the two notions coincide: since A normalizes N, if one translates an closed N orbit by the A-action, one obtains a continuous deformation of closed N orbits. In the spherical and hyperbolic cases, however, the A-translates we consider are not themselves H-orbits. Besides that, it is not the quantitative mixing of the H-action which enters into our Lemma 1, but rather that of the

A-action, which is in a sense orthogonal to H. (Of course, it doesn't even make sense to speak of the mixing of the H-action when H is compact.)

In place of amplification, which puts a spike at a given coefficient, we shorten the interval in the mean value sum. The dynamical way to do this is to move the test vector v at the same time as we hit it with $a(T)$. This is explained in the next subsection, where we define v as a function of a second parameter M.

4.3. Precise test function. Let α be a fixed non-zero test function that is supported in a fixed, but sufficiently small neighborhood of 0. For a parameter $M > 0$ let $v \in V_\tau$ be such that $\Phi = \Phi_v$ be the unique function in $C^\infty_{ev,-\tau-1}(\mathbb{R}^2 - 0)$ such that

$$\Phi \begin{pmatrix} 1 \\ y \end{pmatrix} = e^{iMy}\alpha(y).$$

Note that we are suppressing the dependence of v on M in the notation, but we will be careful to observe this dependence when it comes time to estimate various norms of v. In the spherical case, this yields

$$\varphi_T(\theta) = (\cos\theta)^{-\tau-1}T^{(1+\tau)/2}e^{iMT\tan\theta}\alpha(T\tan\theta).$$

In the hyperbolic case, we have

$$\varphi_T(\theta) = (\cosh\theta)^{-\tau-1}T^{(1+\tau)/2}e^{iMT\tanh\theta}\alpha(T\tanh\theta).$$

The point of choosing v as we have is that it is *not* an A-translate of any fixed vector in V. Were this not the case, we would essentially be back in the situation of the smooth mean value estimate described in §4.2.

If we switch notational conventions to write $M = N/T$ for a parameter $N \geq T$ then we very nearly recover Reznikov's choice of test vector (cf. [10, §4.6]), which he takes (using our normalizations) as

$$T^{\frac{(1+\tau)}{2}}e^{iN\theta}\alpha(T\theta).$$

Our innovation here is simply to write Reznikov's test vector as the $a(T)$ translate of another vector v, which, despite depending on a large parameter M, has polynomially controlled Sobolev norm. The ergodic result is then sufficient for a subconvex result.

4.4. Model integrals. We begin by observing that

$$\ell_n^{mod}(v_T) \gg T^{-1/2} \quad \text{for } n = MT + O(T), \tag{4.1}$$

where the implied $O(T)$ constant is taken sufficiently small. Indeed, in the spherical case, we have

$$\ell_n^{mod}(v_T) = T^{(1+\tau)/2}\frac{1}{2\pi}\int_{-\pi}^{\pi}(\cos\theta)^{-\tau-1}\alpha(T\tan\theta)e^{i(MT\tan\theta-n\theta)}d\theta,$$

and similarly for the hyperbolic case. The integral is supported on an interval of length $\asymp 1/T$ and for $n = MT + O(T)$ there is no essential cancellation in the integral. (Recall also that we are assuming that τ is purely imaginary.)

4.5. Automorphic estimates. We apply the automorphic Sobolev estimates of Section 3.

Lemma 4. *Let $v \in V$ be such that $\Phi = \Phi_v$. We have*

(1) $\|\nu(v)\|_2^2 \ll 1$,

(2) $\mathcal{S}_{\infty,1}(|\nu(v)|^2) \ll M^{9/2+\varepsilon}$,

(3) $\mathcal{S}_{2,1}(|\nu(v)|^2) \ll M^{4+\varepsilon}$,

(4) $\mathcal{S}_{2,0}(|\nu(v)|^2) \ll M^{3+\varepsilon}$.

The implied constants depend only on τ and ε.

Proof. To prove part (1) we pass to the model side, via the isometry $\|\nu(v)\|_2^2 = \|\Phi\|_2^2$. The latter is $O(1)$, since the modulus $|\Phi|$ is independent of M.

For parts (2)-(4), we will appeal to Corollary 1 for an appropriate choice of parameter N. As an orthonormal basis of K-types $\{e_n\}_{n\in 2\mathbb{Z}}$, we take e_n such that the corresponding vector Ψ_n in the model space V_τ^{mod} restricts to the circle $S^1 = \{(\cos\theta, \sin\theta)^t : \theta \in \mathbb{R}\}$ as $e^{in\theta}$. By partial integration one has

$$a_n = \langle \Phi, \Psi_n \rangle_{L^2(S^1)} \ll_D (1 + |n|/M)^{-D}$$

for any $D \in \mathbb{N}_0$. This shows that the a_n are $O(1)$ for $n = O(M)$ and negligibly small thereafter. We may therefore insert $N = M^{1+\varepsilon}$ and $A = O(1)$ into Corollary 1 to produce the desired bounds. $\square$

Note that one has the average bound $\sum_n |\langle \Phi, \Psi_n \rangle_{L^2(S^1)}|^2 = \|\Phi\|^2 = O(1)$. This suggests that one can actually prove a sharper pointwise bound in the range $n = O(M)$ than that used in the above proof, but we have no need for such a strengthening.

4.6. Putting it all together. We now insert the above estimates into Proposition 1 to obtain

$$\frac{1}{T} \sum_{|n-TM| \ll T} |c_n|^2 \ll 1 + \left(\frac{M^4}{T^{(1-\tau_1)/4}} + \frac{M^{9/2}}{T} \right) (MT)^\varepsilon \tag{4.2}$$

Now, in order for the summation range on the left-hand side to be as short as possible, our interest is in taking M large relative to T. But for M too large, say $M \geq T^{\frac{3+\tau_1}{4}}$ – this is the range in which one has $M^{7/2}T^{-1} \geq M^3 T^{-\frac{1}{4}(1-\tau_1)}$ – the resulting bound on c_n is even worse than the trivial bound. So we may assume that $M < T^{\frac{3+\tau_1}{4}}$, in which case we have

$$\frac{1}{T} \sum_{|n-TM| \ll T} |c_n|^2 \ll 1 + M^{4+\varepsilon} T^{-\frac{1}{4}(1-\tau_1)+\varepsilon}.$$

Equalizing the terms on the right-hand side yields an optimal value of $M = T^{\frac{1}{16}(1-\tau_1)}$. Dropping all but one term on the left-hand side, we arrive at Theorem 1.

4.7. Non-tempered case. We add a few words about how the preceding calculations change when the Maass form f is non-tempered. In the following paragraphs we assume that $\tau \in (0,1)$.

If we rescale the test vector as

$$\tilde{v} = M^{\tau/2} v \quad \text{and} \quad \tilde{\Phi} = \Phi_{\tilde{v}} = M^{\tau/2} \Phi,$$

then all estimates given in the tempered setting remain valid when using the inner product (1.5). For example, in the situation of Lemma 4, part (1), we have

$$\|\nu(\tilde{v})\|_2^2 = \|\tilde{\Phi}\|_2^2 = M^\tau \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{e^{iM(\tan\theta - \tan\theta')}}{|\cos\theta\cos\theta'|^{1+\tau}} \frac{d\theta \, d\theta'}{|\sin(\theta-\theta')|^{1-\tau}}$$

in the spherical case (and similarly in the hyperbolic case). A straightforward computation shows that this is $O(1)$. Similarly, the upper bounds in the remaining parts of Lemma 4 are unchanged.

Regarding the model functional, since $\langle \chi_n, \chi_n \rangle_\tau^{\text{reg}} \sim_\tau |n|^{-\tau}$, one has

$$\ell_n^{mod}(\tilde{v}_T) \sim_\tau M^{\tau/2} |n|^{\tau/2} \langle \Phi_T, \chi_n \rangle_\tau.$$

The latter inner product is of size $T^{(-1+\tau)/2} n^{-\tau}$ for $|n - MT| \ll T$. One obtains (4.1) as before.

References

[1] M. B. Bekka and M. Mayer, *Ergodic theory and topological dynamics of group action on homogeneous spaces.* London Mathematical Society Lecture Note Series, 269. Cambridge University Press, Cambridge, 2000. x+200 pp.

[2] J. Bernstein and A. Reznikov, *Sobolev norms of automorphic functionals*, IMRN (2002), 2155–2174

[3] J. Bernstein and A. Reznikov, *Subconvexity bounds for triple L-functions and representation theory.* Ann. of Math. (2) **172** (2010), no. 3, 1679–1718.

[4] D. Bump, *Automorphic forms and representations.* Cambridge Studies in Advanced Mathematics **55**, Cambridge University Press 1996

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

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

[7] I. Gelfand, M. Graev, and I. Piatetski-Shapiro, *Representation Theory and Automorphic Functions.* W.B. Saunders Co., Philadelphia, 1969.

[8] A. Good, *Cusp forms and eigenfunctions of the Laplacian.* Math. Ann. **255** (1981), 523–548.

[9] H. Oh and N. Shah, *Limits of translates of divergent geodesics and integral points on one-sheeted hyperboloids*, Israel J. Math, to appear.

[10] A. Reznikov, *Rankin-Selberg without unfolding and bounds for spherical Fourier coefficients of Maass forms.* J. Amer. Math. Soc. **21** (2008), no. 2, 439–477.

[11] A. Reznikov, *Geodesic restrictions for the Casimir operator.* J. Funct. Anal. **261** (2011), no. 9, 2437–2460.

[12] P. Sarnak, *Fourth moments of Grössencharakteren zeta functions.* Comm. Pure Appl. Math. **38** (1985), no. 2, 167–178.

[13] P. Sarnak, *Integrals of products of eigenfunctions.* Internat. Math. Res. Notices 1994, no. 6, 251–260.

[14] A. Venkatesh, *Sparse equidistribution problems, period bounds and subconvexity.* Ann. of Math. (2) **172** (2010), no. 2, 989–1094.

Mathematisches Institut, Bunsenstr. 3-5, 37073 Göttingen, Germany
E-mail : `blomer@uni-math.gwdg.de`

Institut Galilée, Université Paris 13, 99 avenue J.-B. Clément, 93430 Villetaneuse, France
E-mail : `brumley@math.univ-paris13.fr`

Department of Mathematics, Yale University, New Haven, CT 06511 USA
E-mail : `alex.kontorovich@yale.edu`

Department of Mathematics, Fine Hall, Washington Road, Princeton, NJ 08544 USA
E-mail : `templier@math.princeton.edu`