JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 21, Number 2, April 2008, Pages 439–477
S 0894-0347(07)00581-4
Article electronically published on October 4, 2007

RANKIN-SELBERG WITHOUT UNFOLDING AND BOUNDS FOR SPHERICAL FOURIER COEFFICIENTS OF MAASS FORMS

ANDRE REZNIKOV

To Joseph Bernstein, as a small token of gratitude.

1. Introduction

In this paper we study periods of automorphic functions. We present a new method which allows one to obtain non-trivial spectral identities for weighted sums of certain periods of automorphic functions. These identities are modelled on the classical identity of R. Rankin [Ra] and A. Selberg [Se]. We recall that the Rankin-Selberg identity relates the weighted sum of Fourier coefficients of a cusp form ϕ to the weighted integral of the inner product of ϕ^2 with the Eisenstein series (e.g., formula (1.7) below).

We show how to deduce the classical Rankin-Selberg identity and similar new identities from the uniqueness principle in representation theory (also known under the following names: the multiplicity one property, Gelfand pair). The uniqueness principle is a powerful tool in representation theory; it plays an important role in the theory of automorphic functions.

We associate a non-trivial spectral identity to certain *pairs* of different *triples* of Gelfand subgroups. Namely, we associate a spectral identity (see formula (1.4) below) with two triples $\mathcal{F} \subset \mathcal{H}_1 \subset \mathcal{G}$ and $\mathcal{F} \subset \mathcal{H}_2 \subset \mathcal{G}$ of subgroups in a group $\mathcal{G}$ such that pairs $(\mathcal{G}, \mathcal{H}_i)$ and $(\mathcal{H}_i, \mathcal{F})$ for $i = 1,\ 2$ are strong Gelfand pairs having the same subgroup $\mathcal{F}$ in the intersection (for the notion of Gelfand pair that we use, see Section 1.1.3). We call such a collection $(\mathcal{G}, \mathcal{H}_1, \mathcal{H}_2, \mathcal{F})$ a *strong Gelfand formation.*

In the Introduction we explain our general idea and describe how to implement it in order to reprove the classical Rankin-Selberg formula. We also obtain a new anisotropic analog of the Rankin-Selberg formula. We present then an analytical application of these spectral identities towards non-trivial bounds for various Fourier coefficients of cusp forms. The novelty of our results lies mainly in the method, as we do not rely on the well-known technique of Rankin and Selberg

Received by the editors December 26, 2005.
2000 *Mathematics Subject Classification.* Primary 11F67, 22E45; Secondary 11F70, 11M26.
Key words and phrases. Representation theory, Gelfand pairs, periods, automorphic L-functions, subconvexity, Fourier coefficients of cusp forms.
The research for this paper was partially supported by a BSF grant, by the Minerva Foundation, by the Excellency Center "Group Theoretic Methods in the Study of Algebraic Varieties" of the Israel Science Foundation and the Emmy Noether Institute for Mathematics (the Center of Minerva Foundation of Germany). The paper was mostly written during one of the author's visits to MPIM at Bonn. It is a pleasure to thank MPIM for its excellent working atmosphere.

which is called "unfolding". Instead, we use the uniqueness of relevant invariant functionals which we explain below.

We would like to mention that the Rankin-Selberg method was revolutionized by H. Jacquet, J. Shalika and I. Piatetski-Shapiro who constructed integral representations for many automorphic L-functions (see the survey paper [Bu] and references therein). It is not yet clear what the relation is between these works and their points of view.

1.1. **Periods and Gelfand pairs.** We briefly review some well-known notions and constructions from representation theory and the theory of automorphic functions needed to formulate our identities.

Taking periods is the classical technique to studying automorphic functions. It goes back, at least, to Hecke and Maass. In the classical language it means the following. Let ϕ be an automorphic function on Y and let ψ be an automorphic function on a cycle $C \subset Y$, where the cycle C is equipped with a measure dc. Then one can consider the period defined via the integral

$$\int_C \phi(c)\psi(c)\ dc\ .$$

It is well known that in the modern language of automorphic representations this construction leads to the following setup which we are going to use throughout the paper.

1.1.1. *Automorphic representations.* Let $\mathcal{G}$ be a real reductive Lie group (e.g., $SL_2(\mathbb{R})$), let $\Gamma_{\mathcal{G}} \subset \mathcal{G}$ be a lattice and let $X_{\mathcal{G}} = \Gamma_{\mathcal{G}} \setminus \mathcal{G}$ be the corresponding automorphic space equipped with a $\mathcal{G}$-invariant measure $\mu_{\mathcal{G}}$ (which we will always normalize to have the total mass one). We denote by $L^2(X_{\mathcal{G}}) = L^2(X_{\mathcal{G}}, \mu_{\mathcal{G}})$ the corresponding unitary representation of $\mathcal{G}$ and by $C^\infty(X_{\mathcal{G}})$ its smooth part.

Let $(\pi, \mathcal{G}, V)$ be an abstract unitary irreducible representation of $\mathcal{G}$. We will work solely with the spaces of smooth vectors. Hence by (π, V) we mean the representation in the space of smooth vectors equipped with the invariant Hermitian form. (Usually, one denotes a unitary representation by $(\pi, \mathcal{G}, L)$ where L is the corresponding Hilbert space and the space of smooth vectors is denoted by $V = L^\infty$.)

A pair (π, ν), where $\nu : V \to L^2(X_{\mathcal{G}})$ is an isometric $\mathcal{G}$-morphism, is called an *automorphic* representation of $\mathcal{G}$ on $X_{\mathcal{G}}$. It is known that $\nu : V \to C^\infty(X_{\mathcal{G}})$ and we will denote by $V_\pi^{aut} = V_{(\pi,\nu)} \subset C^\infty(X_{\mathcal{G}})$ the image of V under ν. We denote by $\nu^* : C^\infty(X_{\mathcal{G}}) \to V$ the adjoint map.

1.1.2. *Periods and representation theory.* The notion of a period has the following counterpart in the language of automorphic representations.

Let $\mathcal{H} \subset \mathcal{G}$ be a subgroup (not necessarily reductive, e.g., a unipotent subgroup of $SL_2(\mathbb{R})$) and let $X_{\mathcal{H}} \subset X_{\mathcal{G}}$ be a closed orbit of $\mathcal{H}$ (e.g., $X_{\mathcal{H}} = \mathcal{H} \cap \Gamma_{\mathcal{G}} \setminus \mathcal{H}$). We fix an $\mathcal{H}$-invariant measure $\mu_{\mathcal{H}}$ on $X_{\mathcal{H}}$ and consider the unitary representation $L^2(X_{\mathcal{H}}, \mu_{\mathcal{H}})$ of $\mathcal{H}$. We denote by $r_{\mathcal{H}} = r_{X_{\mathcal{H}}} : C^\infty(X_{\mathcal{G}}) \to C^\infty(X_{\mathcal{H}})$ the corresponding restriction map.

Let (π, ν_π) be an automorphic representation of $\mathcal{G}$ on $X_{\mathcal{G}}$ and let (σ, ν_σ) be an automorphic representation of $\mathcal{H}$ on $X_{\mathcal{H}}$. This means that for the abstract irreducible unitary representation σ in the space of smooth vectors W equipped with the $\mathcal{H}$-invariant Hermitian form, we fix an isometry $\nu_\sigma : W \to L^2(X_{\mathcal{H}})$;

we allow ourselves to use the name automorphic even if $\mathcal{H}$ is not reductive. Let $\nu^* : C^\infty(X_\mathcal{H}) \to W$ be the adjoint map (induced by the $\mathcal{H}$-invariant Hermitian form on $L^2(X_\mathcal{H}, \mu_\mathcal{H})$).

We consider the $\mathcal{H}$-*equivariant* map $T^{aut}_{\pi,\sigma} = T^{aut}_{X_\mathcal{H},\pi,\sigma} = \nu^*_\sigma \circ r_\mathcal{H} \circ \nu_\pi : V \to W$ defined via the composition

$$T^{aut}_{\pi,\sigma} : V \xrightarrow{\nu_\pi} C^\infty(X_\mathcal{G}) \xrightarrow{r_\mathcal{H}} C^\infty(X_\mathcal{H}) \xrightarrow{\nu^*_\sigma} W \ .$$

We call this map the automorphic period map (or simply the period) associated to the collection (π, ν_π), (σ, ν_σ), $X_\mathcal{G}$, $X_\mathcal{H}$ and the choice of corresponding measures. This is the representation theoretic substitute for the classical period. It is well-defined if $X_\mathcal{H}$ is compact. Otherwise we have to assume that π or σ is cuspidal (i.e., rapidly decaying at infinity if $X_\mathcal{H}$ is not compact). For non-cuspidal data one has to introduce appropriate regularization in order to make sense out of the automorphic period.

Clearly $T^{aut}_{\pi,\sigma} \in \mathrm{Hom}_\mathcal{H}(V, W)$. We denote the vector space $\mathrm{Hom}_\mathcal{H}(V, W)$ by $\mathcal{P}(V, W)$ and call it the vector space of *periods* between π and σ. The space of periods is defined even if π or σ is not automorphic.

1.1.3. *Gelfand pairs and periods.* In many cases interesting periods are associated to some special subgroup (or representations). Many of these examples come from what is called the multiplicity one representations or Gelfand pairs (see [Gr] and references therein). In what follows, we will use the notion of Gelfand pairs for *real Lie groups*.

A pair $(\mathcal{G}, \mathcal{H})$ of a real Lie group $\mathcal{G}$ and a real Lie subgroup $\mathcal{H} \subset \mathcal{G}$ is called a *strong Gelfand pair* if for any pair of irreducible representations (π, V) of $\mathcal{G}$ and (σ, W) of $\mathcal{H}$ the dimension of the space $\mathrm{Hom}_\mathcal{H}(V, W)$ of $\mathcal{H}$-equivariant maps (i.e., the space of periods $\mathcal{P}(V, W)$) for the spaces of *smooth* vectors is at most *one*. (It is well known in the representation theory of reductive groups that the dimension of the space $\mathrm{Hom}_\mathcal{H}(V, W)$ for *infinite-dimensional* representations plays the role of the index for *finite-dimensional* representations.) The pair is called a Gelfand pair if the same holds for σ being the trivial representation.

One of the important observations in the theory of automorphic functions is that for Gelfand pairs, automorphic periods $T^{aut}_{\pi,\sigma}$ lead to certain interesting numbers or functions (e.g., Fourier coefficients, L-functions). This is based on the fact that in many cases one can construct another vector in the *one-dimensional* vector space $\mathcal{P}(V, W)$.

Namely, usually representations π and σ could be constructed explicitly in some *model* spaces of sections of various vector bundles over appropriate homogeneous manifolds (e.g., for $SL_2(\mathbb{R})$, in the spaces of homogeneous functions on the plane $\mathbb{R}^2 \setminus 0$; see Section 2). These models usually exist for all representations and not only for the automorphic ones. If $\dim \mathcal{P}(V, W) = 1$, using these models, one can construct an explicit non-zero map $T^{mod}_{\pi,\sigma} \in \mathcal{P}(V, W)$ by means of the corresponding kernel (i.e., construct this map as an integral operator with an explicit kernel). We call such a map a *model* period.

When $\dim \mathcal{P}(V, W) = 1$, such a choice of a non-zero model period $T^{mod}_{\pi,\sigma}$ gives rise to the automorphic coefficient of proportionality $a_{\pi,\sigma} = a_{\nu_\pi,\nu_\sigma} \in \mathbb{C}$ such that

$$T^{aut}_{\pi,\sigma} = a_{\pi,\sigma} \cdot T^{mod}_{\pi,\sigma} \ .$$

We would like to study these constants. In many cases these constants are related to interesting objects (e.g., Fourier coefficients of cusp forms, special values of L-functions, etc.). Of course, these constants depend on the choice of model periods $T^{mod}_{\pi,\sigma}$, normalization of measures, etc. In many cases we can find a way to canonically normalize norms of model maps in the adèlic setting (and hence define canonically if not the constants themselves then their absolute values). We hope to discuss this question elsewhere.

We explain now how one can sometimes obtain spectral identities involving certain coefficients $a_{\pi,\sigma}$.

1.1.4. *Triples of Gelfand subgroups and periods.* Let $\mathcal{G}$ be a real Lie group and let $\mathcal{F} \subset \mathcal{G}$ be a real Lie subgroup which is *not* a Gelfand subgroup. Suppose we still want to study a period of an automorphic representation π of $\mathcal{G}$ with respect to an $\mathcal{F}$-invariant closed cycle $X_{\mathcal{F}} \subset X_{\mathcal{G}}$ (endowed with an $\mathcal{F}$-invariant measure $\mu_{\mathcal{F}}$) and some automorphic representation χ of $\mathcal{F}$. For example, for the trivial representation of $\mathcal{F}$ we obtain the $\mathcal{F}$-invariant functional $I_{\mathcal{F}}$ on π given by the integral over the cycle $X_{\mathcal{F}}$:

$$I_{\mathcal{F}}(v) = \int_{X_{\mathcal{F}}} \nu_\pi(v) d\mu_{\mathcal{F}} \quad \text{for any } v \in V.$$

We cannot apply the idea described above directly. Instead in some cases we can obtain a spectral decomposition of this period.

Namely, suppose we can find an intermediate subgroup $\mathcal{H}$, $\mathcal{F} \subset \mathcal{H} \subset \mathcal{G}$, and an intermediate closed cycle $X_{\mathcal{F}} \subset X_{\mathcal{H}} \subset X_{\mathcal{G}}$ such that both pairs $(\mathcal{G}, \mathcal{H})$ and $(\mathcal{H}, \mathcal{F})$ are strong Gelfand pairs. We claim that this leads to the spectral decomposition of the functional $I_{\mathcal{F}}$. In what follows, we discuss for simplicity only the case of the trivial representation of $\mathcal{F}$. The case of a non-trivial representation χ is similar (and leads to other interesting identities).

Consider the space $L^2(X_{\mathcal{H}}) = L^2(X_{\mathcal{H}}, \mu_{\mathcal{H}})$ and assume that it has a decomposition into a direct sum (in general a direct integral)

$$L^2(X_{\mathcal{H}}) = \bigoplus_i \sigma_i \tag{1.1}$$

of irreducible automorphic representations $(\sigma_i, \nu_{\sigma_i} : W_i \to L^2(X_{\mathcal{H}}))$ of $\mathcal{H}$. This decomposition induces the spectral decomposition of the functional $I_{\mathcal{F}}$.

In fact, the inclusion $X_{\mathcal{F}} \subset X_{\mathcal{H}} \subset X_{\mathcal{G}}$ induces the period map $I^{aut}_{\pi,\sigma_i} \in \mathcal{P}(V, \mathbb{C})$ for every σ_i, via the following composition of maps:

$$I^{aut}_{\pi,\sigma_i} : V \xrightarrow{\nu_\pi} C^\infty(X_{\mathcal{G}}) \xrightarrow{r_{\mathcal{H}}} C^\infty(X_{\mathcal{H}}) \xrightarrow{\nu^*_{\sigma_i}} W_i \xrightarrow{\nu_{\sigma_i}} C^\infty(X_{\mathcal{H}}) \xrightarrow{r_{\mathcal{F}}} C^\infty(X_{\mathcal{F}}) \xrightarrow{I_{\mathcal{F}}} \mathbb{C}\ .$$

The spectral decomposition (1.1) gives rise to the decomposition $I_{\mathcal{F}} = \sum_{\sigma_i} I^{aut}_{\pi,\sigma_i}$. (Since in order to compute $I_{\mathcal{F}}(v)$ for a vector $v \in V$, we first can restrict $\nu_\pi(v)$ to the cycle $X_{\mathcal{H}}$, decompose it with respect to the action of $\mathcal{H}$ and then compute the integral over $X_{\mathcal{F}}$ for each component.) Note that the functional I^{aut}_{π,σ_i} is the composition of two automorphic periods: $I^{aut}_{\pi,\sigma_i} = T^{aut}_{\sigma_i,\mathbb{C}} \circ T^{aut}_{\pi,\sigma_i} : V \to W \to \mathbb{C}$.

We now use the strong Gelfand property for the triple $\mathcal{F} \subset \mathcal{H} \subset \mathcal{G}$, i.e., the fact that the product of dimensions $\dim \mathrm{Hom}_{\mathcal{H}}(V, W_i) \cdot \dim \mathrm{Hom}_{\mathcal{F}}(W_i, \mathbb{C}) \leq 1$ for all W_i. We choose model periods $T^{mod}_{\pi,\sigma_i} \in \mathrm{Hom}_{\mathcal{H}}(V, W_i)$, $T^{mod}_{\sigma_i,\mathbb{C}} \in \mathrm{Hom}_{\mathcal{F}}(W_i, \mathbb{C})$. As we explained above, this leads to the automorphic coefficients of proportionality: $T^{aut}_{\pi,\sigma_i} = a_{\pi,\sigma_i} \cdot T^{mod}_{\pi,\sigma_i}$ and $T^{aut}_{\sigma_i,\mathbb{C}} = b_{\sigma_i,\mathbb{C}} \cdot T^{mod}_{\sigma_i,\mathbb{C}}$. We denote them by $\gamma_{\pi,\sigma_i} = a_{\pi,\sigma_i} \cdot b_{\sigma_i,\mathbb{C}}$ and by $I^{mod}_{\pi,\sigma_i} = T^{mod}_{\sigma_i,\mathbb{C}} \circ T^{mod}_{\pi,\sigma_i} \in V^*$. With such notation we arrive at the spectral

decomposition of the functional $I_{\mathcal{F}} \in V^*$ which is associated to the triple of strong Gelfand subgroups $\mathcal{F} \subset \mathcal{H} \subset \mathcal{G}$

$$I_{\mathcal{F}} = \sum_{\sigma_i \text{ automorphic}} \gamma_{\pi,\sigma_i} \cdot I^{mod}_{\pi,\sigma_i} \ . \tag{1.2}$$

Remark 1.1. We note that for a non-compact cycle $X_{\mathcal{F}}$ there is no obvious way to write down the analog of the spectral decomposition (1.2) even if the initial representation π is cuspidal. This is because a priori the period $T^{aut}_{\sigma_i,\mathbb{C}}$ might not be defined for all σ_i (e.g., non-compact periods of Eisenstein series). Usually, one has to introduce an appropriate regularization procedure in order to define the corresponding periods. In this paper we only consider cycles $X_{\mathcal{F}}$ which are compact and hence will not face this problem.

1.2. **Rankin-Selberg type spectral identities.** Our main observation is that for a given pair of groups $\mathcal{F} \subset \mathcal{G}$ there might be different intermediate subgroups $\mathcal{H}$ as above leading to *different* spectral decompositions of the same functional $I_{\mathcal{F}}$ and hence to identities between the automorphic coefficients.

Let $\mathcal{G}$ be a real Lie group and let $\mathcal{F} \subset \mathcal{H}_i \subset \mathcal{G}$, $i = 1, 2$, be a collection of subgroups such that in the following commutative diagram each embedding is a strong Gelfand pair (i.e., pairs $(\mathcal{G}, \mathcal{H}_i)$ and $(\mathcal{H}_i, \mathcal{F})$ are strong Gelfand pairs)

(1.3)
$$\begin{tikzcd}
 & \mathcal{G} & \\
\mathcal{H}_1 \arrow[ur, hook] & & \mathcal{H}_2 \arrow[ul, hook'] \\
 & \mathcal{F} \arrow[ul, hook] \arrow[ur, hook'] &
\end{tikzcd}$$

We call such a collection of subgroups a *strong Gelfand formation*.

Let $\Gamma \subset \mathcal{G}$ be a lattice and let $X_{\mathcal{G}} = \Gamma \setminus \mathcal{G}$ be the corresponding automorphic space. Let $X_i = X_{\mathcal{H}_i} \subset X_{\mathcal{G}}$ and $X_{\mathcal{F}} \subset X_{\mathcal{G}}$ be closed orbits of $\mathcal{H}_i$ and $\mathcal{F}$, respectively, satisfying the commutative diagram of embeddings

$$\begin{tikzcd}
 & X_{\mathcal{G}} & \\
X_1 \arrow[ur, hook] & & X_2 \arrow[ul, hook'] \\
 & X_{\mathcal{F}} \arrow[ul, hook] \arrow[ur, hook'] &
\end{tikzcd}$$

assumed to be compatible with diagram (1.3). We endow each orbit (as well as $X_{\mathcal{G}}$) with a measure invariant under the corresponding subgroup (to explain the idea, we assume that all orbits are compact, and hence, these measures could be normalized to have mass one).

We fix a decompositions $L^2(X_1) = \bigoplus_i \sigma_i$ into a direct sum (in general into a direct integral) of automorphic representations $(\sigma_i, \nu_{\sigma_i}, W_i)$ of $\mathcal{H}_1$ and similarly $L^2(X_2) = \bigoplus_j \tau_j$ for automorphic representations $(\tau_j, \nu_{\tau_j}, U_j)$ of $\mathcal{H}_2$.

Let (π, ν_π) be an automorphic representation of $\mathcal{G}$ and let $I_{\mathcal{F}} : V \to \mathbb{C}$ be the period defined by the integration over the cycle $X_{\mathcal{F}}$. As we explained in Section

1.1.4, two different triples $\mathcal{F} \subset \mathcal{H}_1 \subset \mathcal{G}$ and $\mathcal{F} \subset \mathcal{H}_2 \subset \mathcal{G}$ lead to two spectral decompositions of the period $I_{\mathcal{F}}$. Namely, let us choose the model periods $T^{mod}_{\pi,\sigma_i} \in \mathrm{Hom}_{\mathcal{H}_1}(V, W_i)$, $T^{mod}_{\sigma_i,\mathbb{C}} \in \mathrm{Hom}_{\mathcal{F}}(W_i, \mathbb{C})$ and similarly $T^{mod}_{\pi,\tau_j} \in \mathrm{Hom}_{\mathcal{H}_2}(V, U_j)$, $T^{mod}_{\tau_j,\mathbb{C}} \in \mathrm{Hom}_{\mathcal{F}}(U_j, \mathbb{C})$. This leads to the introduction of constants $T^{aut}_{\pi,\sigma_i} = a_{\pi,\sigma_i} T^{mod}_{\pi,\sigma_i}$, $T^{aut}_{\sigma_i,\mathbb{C}} = b_{\sigma_i,\mathbb{C}} T^{mod}_{\sigma_i,\mathbb{C}}$ and $T^{aut}_{\pi,\sigma_i} = c_{\pi,\tau_j} T^{mod}_{\pi,\tau_j}$, $T^{aut}_{\tau_j,\mathbb{C}} = d_{\tau_j,\mathbb{C}} T^{mod}_{\tau_j,\mathbb{C}}$ as in Section 1.1.3. We denote them by $\gamma_{\pi,\sigma_i} = a_{\pi,\sigma_i} \cdot b_{\sigma_i,\mathbb{C}}$, $I^{mod}_{\pi,\sigma_i} = T^{mod}_{\sigma_i,\mathbb{C}} \circ T^{mod}_{\pi,\sigma_i} \in V^*$ and similarly by $\delta_{\pi,\tau_j} = c_{\pi,\tau_j} \cdot d_{\tau_j,\mathbb{C}}$, $I^{mod}_{\pi,\tau_j} = T^{mod}_{\tau_j,\mathbb{C}} \circ T^{mod}_{\pi,\tau_j} \in V^*$. The spectral decomposition (1.2) for two triples of Gelfand pairs (1.3) and the corresponding orbits (1.2) implies the identity

$$\sum_{\sigma_i} \gamma_{\pi,\sigma_i} \cdot I^{mod}_{\pi,\sigma_i} = I_{\mathcal{F}} = \sum_{\tau_j} \delta_{\pi,\tau_j} \cdot I^{mod}_{\pi,\tau_j} \ . \tag{1.4}$$

We call such an identity the *Rankin-Selberg type spectral identity* or the *period identity* associated with the Gelfand formation $(\mathcal{G}, \mathcal{H}_1, \mathcal{H}_2, \mathcal{F})$, the corresponding orbits and the automorphic representation π. Note that the summation on the left in (1.4) is over the set of irreducible representations of $\mathcal{H}_1$ occurring in $L^2(X_1)$ and the summation on the right is over the set of irreducible representations of $\mathcal{H}_2$ occurring in $L^2(X_2)$. Since groups $\mathcal{H}_1$ and $\mathcal{H}_2$ might be quite different, the identity (1.4) is non-trivial in general. Surprisingly, even if $\mathcal{H}_1$ and $\mathcal{H}_2$ are conjugate in $\mathcal{G}$, the resulting identity is non-trivial in general (e.g., for $\mathcal{G} = PGL_2(\mathbb{R})$ and two unipotent subgroups intersecting over $\mathcal{F} = e$, this gives the Voronoĭ type summation formula for Fourier coefficients of cusp forms or of Eisenstein series).

The above identity is the identity between functionals on V. It is easy to translate it into the identities for weighted sums of coefficients γ's and δ's. Let $v \in V$ be a vector. It will play the role of a test function. As we explained in Section 1.1.3, in order to construct model periods, we have to consider model realizations of all corresponding representations. In particular, we can view $v \in V$ as a function on some manifold (or a section of a vector bundle). The resulting functionals I^{mod}_{π,σ_i} and I^{mod}_{π,τ_j} could be viewed as integral transforms on the spaces of such functions. Hence, we obtain for any $v \in V$ the identity

$$\sum_{\sigma_i} \gamma_{\pi,\sigma_i} \cdot I^{mod}_{\pi,\sigma_i}(v) = \sum_{\tau_j} \delta_{\pi,\tau_j} \cdot I^{mod}_{\pi,\tau_j}(v) \tag{1.5}$$

for the weighted sums of products of automorphic periods $\gamma_{\pi,\sigma_i} = a_{\pi,\sigma_i} \cdot b_{\sigma_i,\mathbb{C}}$ and $\delta_{\pi,\tau_j} = c_{\pi,\tau_j} \cdot d_{\tau_j,\mathbb{C}}$. The main point of (1.5) is that the weights $I^{mod}_{\pi,\sigma_i}(v)$ and $I^{mod}_{\pi,\tau_j}(v)$ could be computed in some explicit models without any reference to the automorphic picture. We will show below that as a special case, these identities include the classical Rankin-Selberg identity.

Remark 1.2. We note that one can associate a non-trivial spectral identity of a kind we described above to a pair of different *filtrations* of a group $\mathcal{G}$ by subgroups forming strong Gelfand pairs. Namely, we can associate a spectral identity to two filtrations $\{\mathcal{F} = G_0 \subset G_1 \subset \cdots \subset G_n = \mathcal{G}\}$ and $\{\mathcal{F} = H_0 \subset H_1 \subset \cdots \subset H_m = \mathcal{G}\}$ of subgroups in the same group $\mathcal{G}$ such that all pairs (G_{i+1}, G_i) and (H_{j+1}, H_j) are strong Gelfand pairs having the same intersection $\mathcal{F}$. One can also "twist" such an identity by a non-trivial character or an automorphic representation (χ, U_χ) of the group $\mathcal{F}$. In this case the resulting identity is not for an $\mathcal{F}$-invariant *functional* $I_{\mathcal{F}}$, but for an automorphic period *map* in the period space $\mathcal{P}(V_\pi, U_\chi)$.

1.2.1. *Bounds for coefficients.* The Rankin-Selberg type formulas (1.5) can be used in order to obtain bounds for the coefficients γ_{π,σ_i} and δ_{π,τ_j} (e.g., Theorems 1.3 and 1.5). To this end one has to study properties of the integral transforms defined by the functionals I^{mod}_{π,σ_i} and I^{mod}_{π,τ_j} on V. As mentioned in Section 1.1.3, the construction of model functionals involves explicit models of representations in some spaces of functions (or sections of vector bundles). The model periods T^{mod}_{π,σ_i} and T^{mod}_{π,τ_j} are then given as integral operators with explicit kernels and the same is true for the resulting model functionals I^{mod}_{π,σ_i} and I^{mod}_{π,τ_j}. These functionals could be defined for all unitary representations π of $\mathcal{G}$, σ of $\mathcal{H}_1$ and τ of $\mathcal{H}_2$. Hence we obtain a pair of integral transforms $h_\sigma = I^{mod}_{\pi,\sigma} : V^{mod} \to C(\hat{\mathcal{H}}_1)$, $v \mapsto h_\sigma(v) = I^{mod}_{\pi,\sigma}(v)$ (here $\hat{\mathcal{H}}_1$ is the unitary dual of $\mathcal{H}_1$ and V^{mod} is an explicit model of the representation V) and for the triple $(\mathcal{G}, \mathcal{H}_2, \mathcal{F})$ the transform $g_\tau(v) = I^{mod}_{\pi,\tau}(v)$. For the classical Rankin-Selberg identity this pair of transforms constitutes the pair of the Fourier and the Mellin transforms on the space of (smooth with certain decay at infinity) functions on the line $\mathbb{R}$. The latter is the model for the representation π of the principal series (see Section 3 for more details).

For applications, one needs to study analytical properties of these transforms. This is a problem in harmonic analysis which has nothing to do with the automorphic picture. We study the corresponding transforms, in the particular cases under consideration, in two technical lemmas, the lemmas in Sections 3.6 and 4.6, where some instance of what might be called an "uncertainty principle" for the pair of such transforms is established.

The idea behind the proofs of Theorems 1.3 and 1.5 is quite standard by now (and was learned by us from [Go]). It is based on the appropriate Rankin-Selberg type identity and the necessary analytic information for the corresponding integral transforms (e.g., the lemmas in Sections 3.6 and 4.6). Namely, we construct a family of test vectors $v_T \in V^{mod}$ parameterizes by the real parameter $T \geq 1$ such that when substituted into the Rankin-Selberg type identity (1.5), it will pick up the (weighted) sum of coefficients γ_{π,σ_i} for i in a certain "short" interval around T (i.e., the density $h_\sigma(v_T)$ is concentrated on $\hat{\mathcal{H}}_1$ around representations with the parameter of the representation σ close to T). We show then that the integral transform $g_\tau(v_T)$ of such a vector is a slowly changing function on $\hat{\mathcal{H}}_2$ and estimate its support and the size. This allows us to bound the right hand side in (1.5) using Cauchy-Schwartz inequality and the *mean value* (or convexity) bound for the coefficients δ_{π,τ_j} (e.g., bounds (4.9), (4.10)). A simple way to obtain these mean value bounds was explained by us in [BR3].

We note that in order to apply this idea to the identity (1.5) one needs to have some kind of a *positivity* which is not always easy to achieve. Namely, in order to bound a single coefficient γ_{π,σ_i}, we have to know that terms $\gamma_{\pi,\sigma_i} \cdot I^{mod}_{\pi,\sigma_i}(v)$ will not cancel each other in the sum (e.g., all terms are non-negative on one side of the identity). As a result of this constraint there are many identities from which it is not clear how to deduce bounds for the corresponding coefficients. In the examples that we consider in this paper, we choose representations $V = \mathcal{V} \otimes \bar{\mathcal{V}}$ of the group $\mathcal{G} = G \times G$ with $\mathcal{V}$ an irreducible unitary representation of some other group G. For such representations the necessary positivity is automatic.

In this paper we implement the above strategy for $G = PGL_2(\mathbb{R})$ and two cases: for the unipotent subgroup $N \subset G$ and for a compact subgroup $K \subset G$. The first case corresponds to the unipotent Fourier coefficients and the formula we obtain is

equivalent to the classical Rankin-Selberg formula. The second case corresponds to the spherical Fourier coefficients which were introduced by H. Peterson a long time ago, but the corresponding formula (see Theorem 1.4) has never appeared in print, to the best of our knowledge.

To relate these cases to the above discussion of Rankin-Selberg type spectral formulas, we set $\mathcal{G} = G \times G$, $\mathcal{H}_2 = \Delta G \hookrightarrow G \times G$ in both cases under consideration and $\mathcal{H}_1 = N \times N$, $\mathcal{F} = \Delta N \hookrightarrow N \times N \hookrightarrow G \times G$ for the unipotent Fourier coefficients and $\mathcal{H}_1 = K \times K$, $\mathcal{F} = \Delta K \hookrightarrow K \times K \hookrightarrow G \times G$ for the spherical Fourier coefficients. Strictly speaking, the uniqueness principle is only "almost" satisfied for the subgroup N, but the theory of the constant term of the Eisenstein series provides the necessary remedy in the automorphic setting (see Section 3.4).

We also illustrate *analytic applications* of these Rankin-Selberg type spectral identities. We prove non-trivial bounds for both types of these Fourier coefficients. While bounds for the unipotent coefficients (Theorem 1.3) are known (and even much better bounds are known for the Hecke-Maass forms), for the spherical case our bounds (Theorem 1.5) are new. As a corollary, we obtain a subconvexity bound for certain automorphic L-functions.

The method described above also lies behind the subconvexity for the triple L-function given in [BR4]. There the corresponding strong Gelfand formation consists of $\mathcal{G} = G \times G \times G \times G$ with $G = PGL_2(\mathbb{R})$, $\mathcal{F} = \Delta G$ and $H_i = G \times G$ with two different embeddings into $\mathcal{G}$. Recently it became clear that there are many strong Gelfand formations in higher rank groups. We hope to discuss the corresponding identities elsewhere.

1.3. **Unipotent Fourier coefficients of Maass forms.** Let $G = PGL_2(\mathbb{R})$ and denote by $K = PO(2)$ the standard maximal compact subgroup of G. Let $\mathbb{H} = G/K$ be the upper half plane endowed with a hyperbolic metric and the corresponding volume element $d\mu_{\mathbb{H}}$.

Let $\Gamma \subset G$ be a non-uniform lattice. We assume for simplicity that, up to equivalence, Γ has a unique cusp which is reduced at ∞. This means that the unique up to conjugation unipotent subgroup $\Gamma_\infty \subset \Gamma$ is generated by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (e.g., $\Gamma = PSL_2(\mathbb{Z})$). We denote by $X = \Gamma \setminus G$ the automorphic space and by $Y = X/K = \Gamma \setminus \mathbb{H}$ the corresponding Riemann surface (with possible conic singularities if Γ has elliptic elements). This induces the corresponding Riemannian metric on Y, the volume element $d\mu_Y$ and the Laplace-Beltrami operator Δ. We normalize $d\mu_Y$ to have the total *volume one*.

Let $\phi_\tau \in L^2(Y)$ be a Maass cusp form. In particular, ϕ_τ is an eigenfunction of Δ with the eigenvalue which we write in the form $\mu = \frac{1-\tau^2}{4}$ for some $\tau \in \mathbb{C}$. We will always assume that ϕ_τ is normalized to have L^2*-norm one*. We can view ϕ_τ as a Γ-invariant eigenfunction of the Laplace-Beltrami operator Δ on $\mathbb{H}$. Consider the classical Fourier expansion of ϕ_τ at ∞ given by (see [Iw])

$$\phi_\tau(x+iy) = \sum_{n \neq 0} a_n(\phi_\tau) \mathcal{W}_{\tau,n}(y) e^{2\pi i n x} \ . \tag{1.6}$$

Here $\mathcal{W}_{\tau,n}(y)e^{2\pi i n x}$ are properly normalized eigenfunctions of Δ on $\mathbb{H}$ with the same eigenvalue μ as that of the function ϕ_τ. The functions $\mathcal{W}_{\tau,n}$ are usually described in terms of the K-Bessel function. In Section 3.1 we recall the well-known description

of functions $\mathcal{W}_{\tau,n}$ in terms of certain matrix coefficients for unitary representations of G.

We note that from the group-theoretic point of view, the Fourier expansion (1.6) is a consequence of the decomposition of the function ϕ_τ under the natural action of the group N/Γ_∞ (commuting with Δ). Here N is the standard upper-triangular subgroup and the decomposition is with respect to the characters of the group N/Γ_∞ (see Section 3.1).

The vanishing of the zero Fourier coefficient $a_0(\phi_\tau)$ in (1.6) distinguishes cuspidal Maass forms (for Γ having several inequivalent cusps, the vanishing of the zero Fourier coefficient is required at each cusp).

The coefficients $a_n(\phi_\tau)$ are called the Fourier coefficients of the Maass form ϕ_τ and play a prominent role in analytic number theory.

One of the central problems in the analytic theory of automorphic functions is the following.

Problem. Find the best possible constants σ, ρ and C_Γ such that the following bound holds:

$$|a_n(\phi_\tau)| \leq C_\Gamma \cdot |n|^\sigma \cdot (1+|\tau|)^\rho \ .$$

In particular, one asks for constants σ and ρ which are *independent* of ϕ_τ (i.e., depend on Γ only; for a brief discussion of the question's history, see Section 1.5.4).

It is easy to obtain a polynomial bound for coefficients $a_n(\phi_\tau)$ using the boundedness of ϕ_τ on Y. Namely, G. Hardy and E. Hecke essentially proved that the bound

$$\sum_{|n|\leq T} |a_n(\phi_\tau)|^2 \leq C \cdot \max\{T, 1+|\tau|\}$$

holds for any $T \geq 1$, with the constant C depending on Γ only (see [Iw]). It would be very interesting to improve this bound for coefficients $a_n(\phi_\tau)$ in the range $|n| \ll |\tau|$. In the range $|n| \gg |\tau|$ the above bound is essentially sharp.

For a fixed τ, we have the bound $|a_n(\phi_\tau)| \leq C_\tau |n|^{\frac{1}{2}}$. This bound is usually called the standard bound or the Hardy/Hecke bound for the Fourier coefficients of cusp forms (in the n aspect).

The first improvements of the standard bound are due to H. Salié and A. Walfisz using exponential sums. Rankin [Ra] and Selberg [Se] independently discovered the so-called Rankin-Selberg unfolding method (i.e., formula (1.9) below) which allowed them to show that for any $\varepsilon > 0$, the bound $|a_n(\phi)| \ll |n|^{\frac{3}{10}+\varepsilon}$ holds. Their approach is based on the integral representation for the weighted sum of Fourier coefficients $a_n(\phi)$. To state it, we assume, for simplicity, that the so-called residual spectrum is trivial (i.e., the Eisenstein series $E(s,z)$ are holomorphic for $s \in (0,1)$; e.g, $\Gamma = PGL_2(\mathbb{Z})$). (The reader should also keep in mind that we use the normalization $\mathrm{vol}(Y) = 1$ and $\mathrm{vol}(\Gamma_\infty \setminus N) = 1$.) We then have

$$\sum_n |a_n(\phi)|^2 \hat{\alpha}(n) = \alpha(0) + \frac{1}{2\pi i} \int_{Re(s)=\frac{1}{2}} D(s,\phi,\bar{\phi}) M(\alpha)(s) ds \ , \tag{1.7}$$

where $\alpha \in C^\infty(\mathbb{R})$ is an appropriate test function with the Fourier transform $\hat{\alpha}$ and the Mellin transform $M(\alpha)(s)$,

$$D(s,\phi,\bar{\phi}) = \Gamma(s,\tau) \ \cdot \langle \phi\bar{\phi}, E(s)\rangle_{L^2(Y)} \ , \tag{1.8}$$

where $E(z,s)$ is an appropriate non-holomorphic Eisenstein series and $\Gamma(s,\tau)$ is given explicitly in terms of the Euler Γ-function (see Section 1.5.4).

The proof of (1.7), given by Rankin and Selberg, is based on the so-called unfolding trick, which amounts to the following. Let $E(s,z)$ be the Eisenstein series given by $E(s,z)=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} y^s(\gamma z)$ for $Re(s)>1$ (and analytically continued to a meromorphic function for all $s\in\mathbb{C}$). We have the following "unfolding" identity valid for $Re(s)>1$:

$$\begin{aligned}\langle\phi\bar{\phi},E(z,s)\rangle_{L^2(Y)} &= \int_{\Gamma\backslash\mathbb{H}}\phi(z)\bar{\phi}(z)\sum_{\gamma\in\Gamma_\infty\backslash\Gamma} y^s(\gamma z)d\mu_Y \\ &= \int_{\Gamma_\infty\backslash\mathbb{H}}\phi(z)\bar{\phi}(z)y^s(z)d\mu_{\mathbb{H}} = \int_0^\infty\left(\int_0^1\phi(x+iy)\bar{\phi}(x+iy)\ dx\right)y^{s-1}\ d^xy\ .\end{aligned} \tag{1.9}$$

The Mellin inversion formula, together with the Fourier expansion (1.6) for ϕ, leads to the Rankin-Selberg formula (1.7).

Using the strategy formulated in Section 1.2, in this paper we deduce the Rankin-Selberg formula (1.7) directly from the uniqueness principle in representation theory and hence avoid the use of the unfolding trick (1.9). One of the uniqueness results we are going to use is related to the unipotent subgroup $N\subset G$ such that $\Gamma_\infty\subset N$ (the so-called Γ-cuspidal unipotent subgroup). In fact, the definition of classical Fourier coefficients $a_n(\phi_\tau)$ is implicitly based on the uniqueness of N-equivariant functionals on an irreducible (admissible) representation of G (i.e., on the uniqueness of the so-called Whittaker functional). For this reason, we call the coefficients $a_n(\phi_\tau)$ the *unipotent* Fourier coefficients.

We obtain a somewhat different (a slightly more "geometric") form of the Rankin-Selberg identity (1.7). In particular, we exhibit a connection between analytic properties of the function $D(s,\phi,\bar{\phi})$ and analytic properties of certain invariant functionals on irreducible unitary representations of G. This allows us to deduce subconvexity bounds for Fourier coefficients of Maass forms for a *general* Γ in a more transparent way (here we rely on ideas of A. Good [Go] and on our earlier results [BR1] and [BR3]). Namely, we prove the following bound for the Fourier coefficients $a_n(\phi_\tau)$.

Theorem 1.3. *Let ϕ_τ be a fixed Maass form of L^2-norm one. For any $\varepsilon>0$, there exists an explicit constant C_ε such that*

$$\sum_{|k-T|\le T^{\frac{2}{3}}}|a_k(\phi_\tau)|^2\le C_\varepsilon\cdot T^{\frac{2}{3}+\varepsilon}\ .$$

In particular, we have $|a_n(\phi_\tau)|\ll|n|^{\frac{1}{3}+\varepsilon}$. This is weaker than the Rankin-Selberg bound, but holds for *general lattices* Γ (i.e., not necessary a congruence subgroup). The bound in the theorem was first claimed in [BR1] and the analogous bound for holomorphic cusp forms was proved by Good [Go] by a different method. Here we give full details of the proof following a slightly different argument.

The main goal of this paper, however, is different. Our main new results deal with another type of Fourier coefficients associated with a Maass form. These Fourier coefficients, which we call spherical, are associated to a compact subgroup of G.

1.4. **Spherical Fourier coefficients.** When dealing with spherical Fourier coefficients, we assume, for simplicity, that $\Gamma\subset G$ is a co-compact subgroup and $Y=\Gamma\setminus\mathbb{H}$ is the corresponding compact Riemann surface. Let ϕ_τ be a norm one

eigenfunction of the Laplace-Beltrami operator on Y, i.e., a Maass form. We would like to consider a kind of Taylor series expansion for ϕ_τ at a point on Y. To define this expansion, we view ϕ_τ as a Γ-invariant eigenfunction on $\mathbb{H}$. We fix a point $z_0 \in \mathbb{H}$. Let $z = (r, \theta)$, $r \in \mathbb{R}^+$ and $\theta \in S^1$, be the geodesic polar coordinates centered at z_0 (see [He]). We have the following spherical Fourier expansion of ϕ_τ associated to the point z_0:

$$\phi_\tau(z) = \sum_{n\in\mathbb{Z}} b_{n,z_0}(\phi_\tau) P_{\tau,n}(r) e^{in\theta} \ . \tag{1.10}$$

Here functions $P_{\tau,n}(r)e^{in\theta}$ are properly normalized eigenfunctions of Δ on $\mathbb{H}$ with the same eigenvalue μ as that of the function ϕ_τ. The functions $P_{\tau,n}$ can be described in terms of the classical Gauss hypergeometric function or the Legendre function. In Section 4.2.1, we will describe special functions $P_{\tau,n}$ and their normalization in terms of certain matrix coefficients of irreducible unitary representations of G.

We call the coefficients $b_n(\phi_\tau) = b_{n,z_0}(\phi_\tau)$ the spherical (or anisotropic) Fourier coefficients of ϕ_τ (associated to a point z_0). These coefficients were introduced by H. Petersson and played a major role in recent works of Sarnak (e.g., [Sa]). It was discovered by J.-L. Waldspurger [Wa] that in certain cases these coefficients are related to special values of L-functions (see Section 1.5.1).

As in the case of the unipotent expansion (1.6), the spherical expansion (1.10) is the result of an expansion with respect to a group action. Namely, the expansion (1.10) is with respect to characters of the compact subgroup $K_{z_0} = \mathrm{Stab}_{z_0} G$ induced by the natural action of G on $\mathbb{H}$ (for more details, see Section 4).

The expansion (1.10) exists for *any* eigenfunction of Δ on $\mathbb{H}$. This follows from a simple separation of variables argument applied to the operator Δ on $\mathbb{H}$. For a proof and a discussion of the growth properties of coefficients $b_n(\phi)$ for a general eigenfunction ϕ on $\mathbb{H}$, see [He], [L]. For another approach which is applicable to Maass forms, see [BR2].

Under the normalization we choose, the coefficients $b_n(\phi_\tau)$ are bounded on the average. Namely, one can show that the bound

$$\sum_{|n|\le T} |b_n(\phi_\tau)|^2 \le C' \cdot \max\{T, 1 + |\tau|\}$$

holds for any $T \ge 1$, with the constant C' depending on Γ only (see [R]).

As our approach is based directly on the uniqueness principle, we are able to prove an analog of the Rankin-Selberg formula (1.7) with the group N replaced by a maximal compact subgroup of G. This is the main aim of the paper. We obtain an analog of the Rankin-Selberg formula (1.7) for the coefficients $b_n(\phi_\tau)$. Roughly speaking, our new formula amounts to the following (for the exact form, see formula (4.8)).

Theorem 1.4. *Let $\{\phi_{\lambda_i}\}$ be an orthonormal basis of $L^2(Y)$ consisting of Maass forms. Let ϕ_τ be a fixed Maass form.*

There exists an explicit integral transform $^\sharp : C^\infty(S^1) \to C^\infty(\mathbb{C})$, $u(\theta) \mapsto u^\sharp_\tau(\lambda)$, such that for all $u \in C^\infty(S^1)$, the following relation holds:

$$\sum_n |b_n(\phi_\tau)|^2 \hat{u}(n) = u(1) + \sum_{\lambda_i \ne 1} \mathcal{L}_{z_0}(\phi_{\lambda_i}) \cdot u^\sharp_\tau(\lambda_i) \ , \tag{1.11}$$

with some explicit coefficients $\mathcal{L}_{z_0}(\phi_{\lambda_i}) \in \mathbb{C}$ *which are independent of* u.

Here $\hat{u}(n) = \frac{1}{2\pi} \int\limits_{S^1} u(\theta) e^{-in\theta} d\theta$ *and* $u(1)$ *is the value at* $1 \in S^1$.

The definition of the integral transform $^\sharp$ is based on the uniqueness of certain invariant trilinear functionals on irreducible unitary representations of G and is described explicitly in the formula (4.7). The trilinear functional was studied by us in [BR3] and [BR4]. The main point of the relation (1.11) is that the transform $u_\tau^\sharp(\lambda_i)$ depends only on the *parameters* λ_i and τ, but *not* on the choice of Maass forms ϕ_{λ_i} and ϕ_τ. The coefficients $\mathcal{L}_{z_0}(\phi_{\lambda_i})$ are essentially given by the product of the triple product coefficients $\langle \phi_\tau^2, \phi_{\lambda_i} \rangle_{L^2(Y)}$ and the values of Maass forms ϕ_{λ_i} at the point z_0. In some special cases both types of these coefficients are related to L-functions (see [W], [JN], [Wa] and Section 1.5.1).

A formula similar to (1.11) holds for a non-uniform lattice Γ as well and includes the contribution from the Eisenstein series (see the formula in the remarks in Section 4.5). Also, a similar formula holds for holomorphic forms. We intend to discuss it elsewhere.

The new formula (1.11) allows us to deduce the following bound for the *spherical* Fourier coefficients of Maass forms.

Theorem 1.5. *Let* Γ *be as above and* ϕ_τ *is a fixed Maass form of* L^2*-norm one. For any* $\varepsilon > 0$*, there exists an explicit constant* D_ε *such that*

$$\sum_{|k-T| \le T^{\frac{2}{3}}} |b_k(\phi_\tau)|^2 \le D_\varepsilon \cdot T^{\frac{2}{3}+\varepsilon} \ .$$

In particular, we have $|b_n(\phi_\tau)| \ll |n|^{\frac{1}{3}+\varepsilon}$ for any $\varepsilon > 0$. An analogous bound should hold for spherical Fourier coefficients of holomorphic cusp forms. We hope to return to this subject elsewhere.

The proof of the bound in the theorem follows from essentially the same argument as in the case of the unipotent Fourier coefficients, once we have the Rankin-Selberg type identity (1.11). In the proof we use bounds for triple products of Maass forms obtained in [BR3] and a well-known bound for the averaged value of eigenfunctions of Δ.

In special cases, the bound in the theorem could be interpreted as a subconvexity bound for some automorphic L-function (see Section 1.5.1).

1.5. **Remarks.**

1.5.1. *Special values of L-functions.* One of the reasons one might be interested in bounds for coefficients $b_k(\phi_\tau)$ is their relation to certain automorphic L-functions. It was discovered by J.-L. Waldspurger [Wa] that, in certain cases, the coefficients $b_k(\phi_\tau)$ are related to special values of L-functions. H. Jacquet constructed the appropriate relative trace formula which covers these cases (see [JN]). The simplest case of the formula of Waldspurger is the following. Let $z_0 = i \in SL_2(\mathbb{Z}) \setminus \mathbb{H}$ and $E = \mathbb{Q}(i)$. Let π be the automorphic representation which corresponds to ϕ_τ, Π its base change over E and $\chi_n(z) = (z/\bar{z})^{4n}$ the n-th power of the basic Grössencharacter of E. One has then, under appropriate normalization (for details, see [Wa], [JN], [MW]), the beautiful formula

$$|b_n(\phi_\tau)|^2 = \frac{L(\frac{1}{2}, \Pi \otimes \chi_n)}{L(1, Ad\pi)} \ . \tag{1.12}$$

Using this formula, we can interpret the bound in Theorem 1.5 as a bound on the corresponding L-functions. In particular, we have $|L(\frac{1}{2}, \Pi \otimes \chi_n)| \ll |n|^{2/3+\varepsilon}$. This gives a subconvexity bound (with the convexity bound for this L-function being $|L(\frac{1}{2}, \Pi \otimes \chi_n)| \ll |n|^{1+\varepsilon}$). The exponent in the bound corresponds to what is known as an H. Weyl type subconvexity bound for an L-function.

The subconvexity problem is a classical question in analytic theory of L-functions which received much attention in recent years (we refer to the survey [IS] for the discussion of subconvexity for automorphic L-functions). In fact, Y. Petridis and P. Sarnak [PS] recently considered more general L-functions. Among other things, they have shown that $|L(\frac{1}{2} + it_0, \Pi \otimes \chi_n)| \ll |n|^{\frac{159}{166}+\varepsilon}$ for any fixed $t_0 \in \mathbb{R}$ and any automorphic cuspidal representation Π of $GL_2(E)$ (not necessarily a base change). Their method is also spectral in nature although it uses Poincaré series and treats L-functions through (unipotent) Fourier coefficients of cusp forms. We deal directly with periods and the special value of L-functions only appear through the Waldspurger formula. Of course, our interest in Theorem 1.5 lies not so much in the slight improvement of the Petridis-Sarnak bound for these L-functions, but in the fact that we can give a general bound valid for *any* point z_0. (It is clear for a generic point or a cusp form which is not a Hecke form that coefficients b_n are not related to special values of L-functions.)

Recently, A. Venkatesh [V] announced (among other remarkable results) a slightly weaker subconvexity bound for coefficients $b_n(\phi_\tau)$ for a fixed ϕ_τ. His method seems to be quite different and is based on ergodic theory. In particular, it is not clear how to deduce the identity (1.11) from his considerations. On the other hand, the ergodic method gives a bound for Fourier coefficients for higher rank groups (e.g., on $GL(n)$) while it is not yet clear in what higher-rank cases one can develop Rankin-Selberg type formulas similar to (1.11) which would lead to bounds for the corresponding coefficients.

1.5.2. *Fourier expansions along closed geodesics.* There is one more case where we can apply the uniqueness principle to a subgroup of $PGL_2(\mathbb{R})$. Namely, we can consider closed orbits of the diagonal subgroup $A \subset PGL_2(\mathbb{R})$ acting on X. It is well known that such an orbit corresponds to a closed geodesic on Y (or to a geodesic ray starting and ending at cusps of Y). Such closed geodesics give rise to Rankin-Selberg type formulas similar to ones we considered for closed orbits of subgroups N and K. In special cases the corresponding Fourier coefficients are related to special values of various L-functions (e.g., the standard Hecke L-function of a Hecke-Maass form which appears for a geodesic connecting cusps of a congruence subgroup of $PSL(2, \mathbb{Z})$). In fact, in the language of representations of adèle groups, which is appropriate for arithmetic Γ, the case of closed geodesics corresponds to real quadratic extensions of $\mathbb{Q}$ (e.g., twisted periods along Heegner cycles) while the anisotropic expansions (at Heegner points) which we considered in Section 1.4 correspond to imaginary quadratic extensions of $\mathbb{Q}$ (e.g., twisted "periods" at Heegner points).

In order to prove an analog of Theorems 1.3 and 1.5 for the Fourier coefficients associated to a closed geodesic, one has to face certain technical complications. Namely, for orbits of the diagonal subgroup A one has to consider contributions from representations of discrete series, while for subgroups N and K this contribution vanishes. It is more cumbersome to compute a contribution from discrete series as these representations do not have nice geometric models. Hence, while the proof

of an analog of Theorem 1.4 for closed geodesics is straightforward, one has to study invariant trilinear functionals on discrete series representations more closely in order to deduce bounds for the corresponding coefficients. We hope to return to this subject elsewhere.

1.5.3. *Dependence on the eigenvalue.* From the proof that we present, it follows that the constants C_ε and D_ε in Theorems 1.4 and 1.5 satisfy the bound

$$C_\varepsilon,\ D_\varepsilon \leq C(\Gamma)\cdot(1+|\tau|)\cdot|\ln\varepsilon|\ ,$$

for any $0<\varepsilon\leq 0.1$ and some explicit constant $C(\Gamma)$ depending on the lattice Γ only.

1.5.4. *Historical remarks.* The question of the size of Fourier coefficients of cusp forms was posed (in the n aspect) by S. Ramanujan for holomorphic forms (i.e., the celebrated Ramanujan conjecture established in full generality by P. Deligne for the holomorphic Hecke cusp form for congruence subgroups) and was extended by H. Petersson to include Maass forms (i.e., the Ramanujan-Petersson conjecture for Maass forms). In recent years the τ aspect of this problem also turned out to be important.

Under the normalization that we have chosen, it is expected that the coefficients $a_n(\phi_\tau)$ are at most slowly growing as $n\to\infty$ ([Sa]). Moreover, it is quite possible that the strong uniform bound of the form $|a_n(\phi_\tau)|\ll(|n|(1+|\tau|))^\varepsilon$ holds for any $\varepsilon>0$ (e.g., Ramanujan-Petersson conjecture for Hecke-Maass forms for congruence subgroups of $PSL_2(\mathbb{Z})$). We note, however, that the behavior of Maass forms and holomorphic forms in these questions might be quite different (e.g., high multiplicities of holomorphic forms).

Using the integral representation (1.7) and detailed information about Eisenstein series available only for *congruence subgroups*, Rankin and Selberg showed that for a cusp form ϕ for a congruence subgroup of $PGL_2(\mathbb{Z})$ one has $\sum_{|n|\leq T}|a_n(\phi)|^2 = CT+O(T^{3/5+\varepsilon})$ for any $\varepsilon>0$. In particular, this implies that for any $\varepsilon>0$, $|a_n(\phi)|\ll|n|^{\frac{3}{10}+\varepsilon}$. Since their groundbreaking papers, this bound was improved many times by various methods (with the current record for Hecke-Maass forms being $7/64\approx 0.109...$ due to H. Kim, F. Shahidi and P. Sarnak [KSa]).

The approach of Rankin and Selberg is based on the integral representation of the Dirichlet series given for $Re(s)>1$, by the series $D(s,\phi,\bar\phi)=\sum_{n>0}\frac{|a_n(\phi)|^2}{n^s}$.

The introduction of the so-called Ranking-Selberg L-function $L(s,\phi\otimes\bar\phi) = \zeta(2s)D(s,\phi,\bar\phi)$ played an even more important role in the further development of automorphic forms than the bound for Fourier coefficients which Rankin and Selberg obtained.

Using integral representation (1.8), Rankin and Selberg analytically continued the function $L(s,\phi\otimes\bar\phi)$ to the whole complex plane and obtained an effective bound for the function $L(s,\phi\otimes\bar\phi)$ on the critical line $s=\frac12+it$ for Γ being a congruence subgroup of $SL_2(\mathbb{Z})$. From this, using standard methods in the theory of Dirichlet series, they were able to deduce bounds for Fourier coefficients of cusp forms. In fact, Rankin and Selberg appealed to the classical Perron formula (in the form given by E. Landau) which relates analytic behavior of a Dirichlet series with non-negative coefficients to partial sums of its coefficients. The necessary analytic properties of $L(s,\phi\otimes\bar\phi)$ are inferred from properties of the Eisenstein series through formula (1.8).

A small drawback of the original Rankin-Selberg argument is that their method is applicable to Maass (or holomorphic) forms coming from *congruence subgroups* only. The reason for such a restriction is the absence of methods which would allow one to estimate unitary Eisenstein series for general lattices Γ. Namely, in order to effectively use the Rankin-Selberg formula, (1.7) one would have to obtain *polynomial* bounds for the normalized inner product $D(s,\phi,\bar{\phi}) = \Gamma(s,\tau)\cdot\langle\phi\bar{\phi}, E(s)\rangle_{L^2(Y)}$. This turns out to be notoriously difficult because of the *exponential* growth of the factor $\Gamma(s,\tau) = \frac{2\pi^s\Gamma(s)}{\Gamma^2(s/2)\Gamma(s/2+\tau/2)\Gamma(s/2-\tau/2)}$, for $|s|\to\infty$, $s\in i\mathbb{R}$. For a congruence subgroup, the question could be reduced to known bounds for the Riemann zeta function or for Dirichlet L-functions, as was shown by Rankin and Selberg. The problem of how to treat general Γ was posed by Selberg in his celebrated paper [Se].

The breakthrough in this direction was achieved in works of Good [Go] (for holomorphic forms) and Sarnak [Sa] (in general) who proved non-trivial bounds for Fourier coefficients of cusp forms for a general Γ using spectral methods. The method of Sarnak was finessed in [BR1] by introducing various ideas from representation theory and was further extended in [KS]. The method of this paper is different and avoids the use of analytic continuation which is central for [Sa], [BR1] and [KS].

Special cases of the Rankin-Selberg spectral identities described in the Introduction were obtained before by a different method. The first (vague) attempt to write the above-mentioned formula for four copies of $G = PGL_2(\mathbb{R})$ and all representations coming from Eisenstein series was made by N. Kuznetsov [Kz]. His aim was to obtain a formula representing the eighth moment of the Riemann zeta function. Later Y. Motohashi [Mo1] obtained the formula for the fourth moment of the Riemann zeta function. This corresponds to our identity with $\mathcal{G} = G\times G$ for $G = PGL_2(\mathbb{R})$, $H_1 = T\times T$ where $T\subset G$ is the diagonal subgroup, $\mathcal{F} = \Delta T$ and $H_2 = \Delta G$. To obtain the fourth moment of the Riemann zeta function, Motohashi considers representations coming from Eisenstein series. This leads to considerable technical difficulties which one should not underestimate. Both Kuznetsov and Motohashi based their approach on the celebrated Bruggeman-Kuznetsov trace formula (applying it twice!). The setup we present here, even if it does not simplify the arguments, at least gives a more conceptual explanation for the terms appearing in these identities.

Many other cases of these identities appeared more recently (mostly stated implicitly as a tool for estimation of L-functions or other quantities). Among these are works of R. Bruggeman, V. Bykovskiĭ, A. Ivić, M. Jutila, P. Michel, A. I. Vinogradov, to name a few.

Finally, we would like to mention that recently R. Bruggeman, M. Jutila and Y. Motohashi (see [BM], [Mo2] and references therein) developed what they call the inner product method. It is based on the unfolding of an appropriate Poincaré or Petersson type series. The standard unfolding leads to the spectral expansion for the series of the type $\sum_k A_k(\phi)A_{k+h}(\bar{\phi})W(k)$, where ϕ is a Maass form and A_k are appropriate Fourier coefficients (e.g., unipotent or spherical Fourier coefficients we discussed above). The formulas obtained in such a way are special cases of our Rankin-Selberg type formula (3.6) for a *special* test vectors v. These vectors are constructed from certain functions on the upper-half plane. As a result, the

corresponding weights in the Rankin-Selberg type formulas are reminiscent of *exponential* weights considered by Selberg and Rankin. It seems that by using our approach one can avoid the difficult task of removing these unwanted weights.

The paper is organized as follows. In Section 2, we quickly recall the notion of automorphic representations of G and describe the standard models of representations we will use.

In Section 3 we reprove the classical Rankin-Selberg formula and deduce bounds for the unipotent Fourier coefficients of Maass forms. The proof is based on the uniqueness of trilinear invariant functionals on irreducible unitary representations of G. We use the description of these functionals obtained in [BR3].

In Section 4 we apply the same strategy to the spherical Fourier coefficients. In fact, in this case the proof is less involved since we do not need the theory of the Eisenstein series in order to remedy the non-uniqueness of N-invariant functionals on irreducible representations of G. Section 4 contains our main new results and the reader might read this section independently of Section 3.

In the appendix we prove an asymptotic expansion of the model trilinear functional. We use this analysis in the proof of Theorem 1.5.

2. Representations of $PGL_2(\mathbb{R})$

We start with a short reminder about the connection between Maass forms and the representation theory of $PGL_2(\mathbb{R})$ which is due to Gelfand and Fomin (see [G2]).

2.1. **Models of representations.** All irreducible unitary representations of the group $G = PGL_2(\mathbb{R})$ are classified. For simplicity we consider those with a nonzero K-fixed vector (so-called representations of class one) since only these representations arise from Maass forms. These are the representations of the principal and the complementary series and the trivial representation. We will use the following standard model (or realization) for these representations.

For every complex number τ, consider the space V_τ of smooth even homogeneous functions on $\mathbb{R}^2 \setminus 0$ of homogeneous degree $\tau - 1$ (which means that $f(ax, ay) = |a|^{\tau-1} f(x, y)$ for all $a \in \mathbb{R} \setminus 0$). The representation (π_τ, V_τ) is induced by the action of the group $GL_2(\mathbb{R})$ given by the formula $\pi_\tau(g) f(x, y) = f(g^{-1}(x, y))|\det g|^{\frac{\tau-1}{2}}$. This action is trivial on the center of $GL_2(\mathbb{R})$ and hence defines a representation of G. The representation (π_τ, V_τ) is called *the representation of the generalized principal series.*

For explicit computations it is often convenient to pass from the plane model to a line model. Namely, the restriction of functions in V_τ to the line $(x, 1) \subset \mathbb{R}^2$ defines an isomorphism of the space V_τ with the space $C_\tau^\infty(\mathbb{R})$ of restrictions of smooth homogeneous functions (e.g., decaying at infinity as $|x|^{\tau-1}$). Hence we can think about vectors in V_τ as functions on $\mathbb{R}$.

In the line model the action of an element $\tilde{a} = diag(a, a^{-1})$, $a \in \mathbb{R}^\times$, in the diagonal subgroup is given by $\pi_\tau(\tilde{a}) f(x, 1) = f(a^{-1}x, a) = |a|^{\tau-1} f(a^{-2}x, 1)$; and the action of an element $\tilde{n} = \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}$ in the unipotent group is given by $\pi_\tau(\tilde{n}) f(x, 1) = f(x - n, 1)$.

When $\tau = it$ is purely imaginary, the representation (π_τ, V_τ) is pre-unitary and irreducible; the G-invariant scalar product in V_τ is given by $\langle f, g\rangle_{V_\tau} = \frac{1}{\pi}\int_{\mathbb{R}} f\bar{g}dx$. These representations are called *the principal series* representations.

When $\tau \in (-1,1)$, the representation (π_τ, V_τ) is called a representation of *the complementary series*. These representations are also pre-unitary and irreducible, but the formula for the scalar product is more complicated (see [G1]).

All these representations have K-invariant vectors. We fix a K-invariant unit vector $e_\tau \in V_\tau$ to be a function which is constant on the unit circle S^1 in $\mathbb{R}^2$ in the plane realization. Note that in the line model a K-fixed unit vector is given by $e_\tau(x) = c(1+x^2)^{\frac{\tau-1}{2}}$ with $|c|^2 = \pi^{-1}$ for $\tau \in i\mathbb{R}$.

Another realization, which we call circle or spherical model, is obtained by restricting functions in V_τ to the unit circle $S^1 \subset \mathbb{R}^2 \setminus 0$. In the circle model we have the isomorphism $V_\tau \simeq C^\infty_{even}(S^1)$ and for $\tau \in i\mathbb{R}$, the scalar product is given by $\langle f, g\rangle = \frac{1}{2\pi}\int_{S^1} f\bar{g}d\theta$ while the action of K is induced by the rotation of S^1.

Representations of the principal and the complimentary series exhaust all non-trivial irreducible pre-unitary representations of G of class one.

2.2. **Automorphic representations.** Every automorphic form ϕ generates (under the right translations by elements in G) an automorphic representation of the group G (see [G2]); this means that, starting from ϕ, we produce a smooth irreducible unitarizable representation of the group G in a space V and its realization $\nu : V \to C^\infty(X)$ in the space of smooth functions on the automorphic space $X = \Gamma\backslash G$. We will denote by V_τ the isomorphism class of the representation arising in this way from a Maass form $\phi = \phi_\tau$ with the eigenvalue $\mu = \frac{1-\tau^2}{4}$.

Suppose we are given a class one unitary representation (τ, V_τ) and an automorphic realization of it, $\nu : V_\tau \to C^\infty(X)$; we assume ν to be an isometric embedding. Such ν gives rise to an eigenfunction of the Laplacian on the Riemann surface $Y = X/K$ as before. Namely, if $e_\tau \in V_\tau$ is a unit K-fixed vector, then the function $\phi = \nu(e_\tau)$ is an L^2-normalized eigenfunction of the Laplacian on the space $Y = X/K$ with the eigenvalue $\mu = \frac{1-\tau^2}{4}$. This explains why τ is a natural parameter for describing Maass forms.

This construction gives a one-to-one correspondence between Maass forms and class one automorphic representations (and more generally between automorphic forms and automorphic representations of G). We use this correspondence to translate problems in automorphic forms into problems in representation theory.

3. Unipotent Fourier coefficients

3.1. **Whittaker functionals.** We start with the classical interpretation of Fourier coefficients $a_k(\phi_\tau)$ in terms of representation theory. Namely, we consider Whittaker functionals on irreducible unitary representations of G.

Let ϕ be a Maass form and $\nu : V \to C^\infty(X)$ the corresponding automorphic realization of the space of smooth vectors of an irreducible unitary representation of G.

Let $N \subset G$ be the standard upper-triangular unipotent subgroup. We denote by $\mathcal{N}$ the N-invariant closed cycle (a horocycle) $\Gamma_\infty \setminus N \subset X$. The cycle $\mathcal{N}$ is an orbit $\mathcal{N} = \bar{e}\cdot N \subset X$ of N, where $\bar{e}$ is the image of the identity element $e \in G$ under the natural projection $G \to X$. In what follows, we can choose any closed orbit of any unipotent subgroup of G. We endow $\mathcal{N}$ with the N-invariant measure dn of the *total mass one* and fix an identification $\Gamma_\infty \setminus N \simeq \mathbb{Z}\setminus\mathbb{R}$.

For $k \in \mathbb{Z}$, let $\psi_k : N \to \mathbb{C}$ be the additive character $\psi_k(t) = e^{2\pi ikt}$ of $N \simeq \mathbb{R}$ trivial on $\Gamma_\infty \simeq \mathbb{Z} \subset \mathbb{R}$. We consider the functional $l^a_k = l^{aut}_{\psi_k} : V \to \mathbb{C}$ defined by

the automorphic period

$$l_k^a(v) = \int_{\mathcal{N}} \nu(v)(n)\bar{\psi}_k(n)dn \quad \text{for any } v \in V \ .$$

The functional $l_k^a \in V^*$ is (N, ψ_k)-equivariant, i.e., $l_k^a(\pi(n)v) = \psi_k(n)l_k^a(v)$ for any $n \in N$ and $v \in V$. It is well known that for a *non-trivial* character ψ_k the space of functionals in V^* satisfying this property is one-dimensional. The automorphic representation (V, ν) is called cuspidal if $l_{\psi_0}^a \equiv 0$ (for any cuspidal subgroup Γ_N). We also have the standard Fourier expansion of cuspidal automorphic functions along $\mathcal{N}$:

$$\nu(v)(x) = \sum_{k \neq 0} l_k^a(\pi(g)v),$$

where g corresponds to x under the projection $p : G \mapsto \Gamma \setminus G = X$ (i.e., $p(g) = x$).

We now consider *model* Whittaker functionals. In the line model of the representation $V = V_\tau \subset C^\infty(\mathbb{R})$, we can construct a model Whittaker functional $l_k^m = l_{\psi_k}^{mod} : V \to \mathbb{C}$ by using the Fourier transform. Namely, let $v \in V \subset C^\infty(\mathbb{R})$ be a vector (i.e., a smooth function) of a compact support and let $\xi \in \mathbb{R}$. We define the model Whittaker functional by the integral

$$l_\xi^m(v) = \hat{v}(\xi) = \int_{\mathbb{R}} v(x)e^{-i\xi x}dx \ .$$

The functional l_ξ^m clearly extends to the whole space V by continuity.

The uniqueness of Whittaker functionals implies that the model and the automorphic functionals are proportional. Namely, for any $k \in \mathbb{Z} \setminus 0$, there exists a constant $a_k(\nu) \in \mathbb{C}$ such that

$$l_k^a = a_k(\nu) \cdot l_k^m \ .$$

A simple computation shows that under our normalization $|a_k(\nu)| = |a_k(\phi_\tau)|$. Namely, we have $l_\xi^m(e_\tau) = \int (1+t^2)^{\frac{\tau-1}{2}} \exp(-i\xi t)dt = \frac{\pi^{\frac{1}{2}}|\xi/2|^{-\tau/2}}{\Gamma(\frac{1-\tau}{2})} K_{-\tau/2}(\xi)$, where K_t denotes the K-Bessel function. Based on this, we choose in (1.6) the following normalization for the Whittaker functions:

$$\mathcal{W}_{\tau,k}(y) = l_{\psi_k}^{mod}\left(\pi_\tau\begin{pmatrix} y^{\frac{1}{2}} & \\ & y^{-\frac{1}{2}}\end{pmatrix} e_\tau\right) = \frac{\pi^{\frac{1}{2}}}{\Gamma\left(\frac{1-\tau}{2}\right)} \cdot y^{-\frac{1}{2}} \ K_{-\tau/2}(2\pi|k|y) \ .$$

Under such normalization we have $a_k(\phi) = a_k(\nu)$, and this is consistent with one of the classical normalizations for Fourier coefficients of Maass forms (see [Iw]).

3.2. **Weighted sums of coefficients.** We are interested in bounds for Fourier coefficients $a_k(\nu)$. To this end we consider bounds for weighted sums of the type

$$\sum_k |a_k(\nu)|^2 \hat{\alpha}(k),$$

where $\hat{\alpha}$ is a non-negative weight function. There is a simple geometric way to construct these sums.

Let $\bar{V}$ be the representation which is complex conjugate to V; it is also an automorphic representation with the realization $\bar{\nu} : \bar{V} \to C^\infty(X)$. We only consider the case of representations of the principal series, i.e., we assume that $V = V_\tau$, $\bar{V} = V_{-\tau}$ for some $\tau \in i\mathbb{R}$; the case of representations of the complementary series is similar.

Consider the space $E = V \otimes \bar{V}$. We identify it with a subspace of $C^\infty(\mathbb{R}^2)$ using the line realization $V \subset C^\infty(\mathbb{R})$. We have the corresponding automorphic realization $\nu_E = \nu \otimes \bar{\nu} : E = V \otimes \bar{V} \to C^\infty(X \times X)$.

Let $\Delta\mathcal{N} \subset \Delta X \subset X \times X$ be the diagonal copy of the cycle $\mathcal{N}$. We define the following automorphic N-invariant functional $l_{\Delta\mathcal{N}} : E \to \mathbb{C}$ by

$$l_{\Delta\mathcal{N}}(w) = \int_{\Delta\mathcal{N}} \nu_E(w)(n,n)dn$$

for any $w \in E$.

We have the obvious Plancherel formula

$$\begin{aligned} l_{\Delta\mathcal{N}}(w) &= \sum_k l_k^a \otimes \bar{l}_{-k}^a(w) \\ &= \sum_k |a_k(\nu)|^2 l_k^m \otimes \bar{l}_{-k}^m(w) = \sum_k |a_k(\nu)|^2 \hat{w}(k,-k) \end{aligned} \tag{3.1}$$

for any $w \in E \subset C^\infty(\mathbb{R}^2)$ (here $\hat{w}$ denotes the standard Fourier transform on $\mathbb{R}^2$).

Varying the vector $w \in E$, we obtain different weighted sums $\sum_k |a_k(\nu)|^2 \hat{\alpha}(k)$ with a weight function $\hat{\alpha}(k) = \hat{w}(k,-k)$. The weight function might be easily arranged to be non-negative as we will see below.

We now obtain another expression for the functional $l_{\Delta\mathcal{N}}$ using spectral decomposition of $L^2(X)$ and trilinear invariant functionals on irreducible representations of G. This will give an instance of the Rankin-Selberg type formula discussed in the Introduction which in fact coincides with the classical formula of Rankin and Selberg. We first discuss the spectral decomposition of $L^2(X)$ into irreducible unitary representations of G.

3.3. **Spectral decomposition and the Eisenstein series.** It is well known that $L^2(X) = L^2_{cusp}(X) \oplus L^2_{res}(X) \oplus L^2_{Eis}(X)$ decomposes into the sum of three closed G-invariant subspaces of cuspidal representations, representations associated to residues of Eisenstein series and the space generated by the unitary Eisenstein series (see [B]). The spaces $L^2_{cusp}(X)$ and $L^2_{res}(X)$ decompose discretely into a direct sum of irreducible unitary representations of G and $L^2_{Eis}(X)$ is a direct integral of irreducible unitary representations of the principal series. We assume for simplicity that the residual spectrum is trivial, i.e., $L^2_{res}(X) = \mathbb{C}$ is the trivial representation of G (e.g., Γ is a congruence subgroup of $PSL_2(\mathbb{Z})$).

We are interested in the spectral decomposition of the functional $l_{\Delta\mathcal{N}}$ defined as a period along the diagonal copy of a horocycle inside of $X \times X$. Hence, the space $L^2_{cusp}(X)$ will not appear in the final formula as by definition it consists of functions satisfying $\int_{\mathcal{N}} f(nx)dn = 0$ for almost all $x \in X$.

We will need the following basic facts from the theory of the Eisenstein series (see [Be], [B], [Ku]).

Let $B = AN$ be the Borel subgroup of G (i.e., the subgroup of the upper triangular matrices). We denote $\Gamma_B = \Gamma \cap B$, $\Gamma_N = \Gamma_\infty = \Gamma \cap N$ and $\Gamma_L = \Gamma_B/\Gamma_N$ which we assume, for simplicity, is trivial. Let $Aff = N \setminus G \simeq \{\mathbb{R}^2 \setminus 0\}/\{\pm 1\}$ be the basic affine space. The group G acts from the right on the space Aff and preserves an invariant measure μ_{Aff}. The subgroup B/N acts on Aff on the left and acts on μ_{Aff} by a character.

We denote $X_B = \Gamma_B N \setminus G$ and endow it with the measure μ_{X_B} compatible with the measure μ_X. We identify X_B with Aff (in general, one considers the space $\Gamma_L \setminus Aff$).

Let $\mathcal{A}(X_B)$ be the space of smooth functions of moderate growth on X_B. For a complex number $s \in \mathbb{C}$, we denote by $\mathcal{A}^s(X_B) \simeq \mathcal{A}^s(Aff) \simeq \mathcal{A}^s_{even}(\mathbb{R}^2 \setminus 0)$ the subspace of even homogeneous functions of homogeneous degree $s-1$. The subspace $\mathcal{A}^s(X_B)$ is G-invariant and for s purely imaginary it is isomorphic to the space of smooth vectors of a unitary class one representation of G.

In this setting one have the Eisenstein series operator

$$\mathbb{E} : \mathcal{A}(X_B) \to C^\infty(X)$$

given by $\mathbb{E}(f) = \sum_{\gamma \in \Gamma/\Gamma_B} \gamma \circ f$ and the conjugate constant term operator $C : C^\infty(X) \to \mathcal{A}(X_B)$ given by $C(\phi) = \int_{n \in N/\Gamma_N} n \circ \phi \; dn$. The operator $\mathbb{E}$ is only partially defined as the Eisenstein series not always convergent.

The operators $\mathbb{E}$ and C commute with the action of G. Hence we also have the operator $\mathbb{E}(s) = \mathbb{E}|_{\mathcal{A}^s(X_B)} : \mathcal{A}^s(X_B) \to C^\infty(X)$ (defined via analytic continuation for all $s \in i\mathbb{R}$) and the fundamental relation $C(s) \circ \mathbb{E}(s) = Id + I(s)$ where $I(s) : \mathcal{A}^s(X_B) \to \mathcal{A}^{-s}(X_B)$ is an intertwining operator which is unitary for $s \in i\mathbb{R}$. It is customary to write it in the form $I(s) = c(s) I_s$ where I_s is a properly normalized unitary intertwining operator satisfying $I_s \circ I_{-s} = Id$ and $c(s)$ is a meromorphic function, satisfying the functional equation $c(s)c(-s) = 1$. The operator I_s is constructed explicitly in a model of the representation V_s. We also have the functional equation $\mathbb{E}(s) = \mathbb{E}(-s) \circ I(s)$ for the Eisenstein series.

The spectral decomposition of $L^2_{Eis}(X)$ then reads

$$L^2_{Eis}(X) = \int_{i\mathbb{R}^+} \mathbb{E}(s)(\mathcal{A}^s(X_B)) \; ds \; .$$

This means, in particular, that for any $f \in C^\infty(X) \cap L^2(X)$, the Eisenstein component $f_{Eis} = pr_{Eis}(f)$ of f in the space $L^2_{Eis}(X)$ has the representation $f_{Eis} = \int_{i\mathbb{R}^+} \mathbb{E}(s) f_s \; ds$ for an appropriate smooth family of functions $f_s \in \mathcal{A}^s(X_B)$. We choose an orthonormal basis $\{e_i(s)\} \subset \mathcal{A}^s(X_B)$ and set

$$f_s = \sum_i \langle f, \mathbb{E}(s) e_i(s) \rangle_{L^2(X)} e_i(s)$$

for all $s \in i\mathbb{R}$. We have then a more symmetrical spectral decomposition

$$f_{Eis} = \frac{1}{2} \int_{i\mathbb{R}} \mathbb{E}(s) f_s \; ds$$

and the corresponding Plancherel formula $||f_{Eis}||^2_{L^2(X)} = \frac{1}{2} \int_{i\mathbb{R}} ||f_s||^2_{\mathcal{A}^s(X_B)} \; ds$.

3.4. **Trilinear invariant functionals.** We construct the spectral decomposition of $l_{\Delta\mathcal{N}}$ with the help of trilinear invariant functionals on irreducible unitary representations of G. We review the construction below (for a more detailed discussion see [BR3]).

Let $\nu : V \to C^\infty(X)$ be a cuspidal automorphic representation. Let $E = V \otimes \bar{V}$ and let ν_E be as above. Consider the space $C^\infty(X \times X)$. The diagonal $\Delta X \to X \times X$ gives rise to the restriction morphism $r_\Delta : C^\infty(X \times X) \to C^\infty(X)$. Let $\nu_W : W \to C^\infty(X)$ be an irreducible automorphic subrepresentation. We assume

that for any $w \in W$ the function $\nu_W(w)$ is a function of moderate growth on X. We define the G-invariant trilinear functional $l^{aut}_{E\otimes W} = l^{aut}_{\nu_E\otimes\nu_W}$ on $E \otimes \bar{W}$ via

$$l^{aut}_{E\otimes W}(v \otimes v' \otimes u) = \langle r_\Delta(v \otimes v'), u\rangle_{L^2(X)}$$

for any $v \otimes v' \in E$ and $u \in \bar{W}$. The cuspidality of V and the moderate growth condition on W ensure that $l^{aut}_{E\otimes W}$ is well-defined (i.e., the integral over the non-compact space X is absolutely convergent).

Next we use a general result from representation theory, claiming that such a G-invariant trilinear functional is unique up to a scalar (see [O], [Pr] and the discussion in [BR3]). Namely, we have the following.

Theorem. *Let (π_j, V_j), where $j = 1, 2, 3$, be three irreducible smooth admissible representations of G. Then* $\dim \mathrm{Hom}_G(V_1 \otimes V_2 \otimes V_3, \mathbb{C}) \leq 1$.

This implies that the automorphic functional $l^{aut}_{E\otimes W}$ is proportional to an explicit "model" functional $l^{mod}_{E\otimes W}$ which we describe using explicit realizations of representations V and W of the group G; it is important that this model functional carry no arithmetic information. The model functional is defined on any three irreducible admissible representations of $PGL_2(\mathbb{R})$ regardless of whether or not these are automorphic.

Thus we can write

$$l^{aut}_{E\otimes W} = a_{\nu_E\otimes\nu_W} \cdot l^{mod}_{E\otimes W} \tag{3.2}$$

for some constant $a_{E\otimes W} = a_{\nu_E\otimes\nu_W}$ (somewhat abusing notation as this coefficient depends on the realizations ν_E and ν_W and not only on the isomorphism classes of E and W).

It turns out that the proportionality coefficient $a_{E\otimes W}$ above carries important "automorphic" information (e.g., essentially is equal to the Rankin-Selberg L-function) while the second factor carries no arithmetic information and can be evaluated using explicit realizations of representations V and W (see the Appendix in [BR3] for an example of such a computation). In what follows, we only need the case of W being an irreducible unitary representation of the principal series V_s, $s \in i\mathbb{R}$ (or the trivial representation).

Denote by l^{mod}_s the *model* trilinear form $l^{mod}_s : V \otimes \bar{V} \otimes V_s \to \mathbb{C}$ which we describe explicitly in Section 3.5 below. Any G-invariant form $l : V \otimes \bar{V} \otimes V_s \to \mathbb{C}$ gives rise to a G-intertwining morphism $T^l : V \otimes \bar{V} \to V_s^*$ which extends to a G-morphism $T^l : E \to \bar{V}_s$, where we identify the complex conjugate space $\bar{V}_s$ with the smooth part of the space V_s^* ($\bar{V}_s \simeq V_{-s}$ for $s \in i\mathbb{R}$). We apply this construction in order to describe the projection of E to the space orthogonal to cusp forms, namely to $\mathbb{C} \oplus L^2_{Eis}(X)$ (in general to $L^2_{res}(X) \oplus L^2_{Eis}(X)$).

We realize the irreducible principal series representation V_s in the space of homogenous functions on the plane $\mathcal{A}^s(\{\mathbb{R}^2 \setminus 0\}/\{\pm 1\}) \simeq \mathcal{A}^s(Aff) \simeq \mathcal{A}^s(X_B)$. This is a model suitable for the theory of Eisenstein series. For a chosen family of G-invariant functionals $l^{mod}_s = l_{E\otimes V_s} : E \otimes V_{-s} \to \mathbb{C}$ and the corresponding family of morphisms $T_s = T^{l^{mod}_s} : E \to V_s \simeq A^s(X_B)$, we have the proportionality coefficient $a(s) = a_\nu(s) = a_{E\otimes V_{-s}}$ defined by $l^{aut}_s = a(s)\ l^{mod}_s$ as in (3.2) and the corresponding spectral decomposition for any $w \in E$,

$$pr_{res\oplus Eis}(\nu_E(w)) = \langle r_\Delta(\nu_E(w)), 1\rangle \cdot 1 + \frac{1}{2}\int_{i\mathbb{R}} a(s)\mathbb{E}(s)(T_s(w))\ ds\ . \tag{3.3}$$

Note that (3.3) is symmetrical under the change $s \to -s$. This is achieved by choosing the model trilinear functionals $l_s^{mod} : E \otimes V_s \to \mathbb{C}$ to satisfy the relation $l_s^{mod} = l_{-s}^{mod} \circ I_s$ and the coefficients $a(s)$ to satisfy $a(s) = c(s)a(-s)$ (this is equivalent to the functional equation for the Rankin-Selberg L-function). We note also that $\langle r_\Delta(\nu_E(w)), 1\rangle = Tr(w)$ for any $w \in E$, viewed as an element in $V \otimes V^*$.

We use spectral decomposition (3.3) to obtain the spectral decomposition of the functional $l_{\Delta\mathcal{N}}$.

Namely, consider the functional $l_{\mathcal{N}} : C^\infty(X) \to \mathbb{C}$ given by the constant term along $\mathcal{N} \subset X$ (i.e., $l_{\Delta\mathcal{N}}(f) = C(f)|_{x=\bar{e}}$ for any $f \in C^\infty(X)$). As $l_{\mathcal{N}}$ vanishes on $L^2_{cusp}(X)$, we only have to understand its form on the space of the Eisenstein series (and on the space of residues). The pair (G, N) is not a Gelfand pair (the space of N-*invariant* functionals is two dimensional) and we cannot use the argument we used for the Whittaker functionals. However, the theory of the Eisenstein series provides the necessary remedy. Namely, consider the representation of the (generalized) principal series $\mathcal{A}^s$ realized in the space of even homogenous functions on $X_B \simeq \mathbb{R}^2\backslash 0$. The space of N-invariant functionals on $\mathcal{A}^s$ is generated by the functionals δ_s and δ_{-s}, where $\delta_s(v) = v(0,1)$ and $\delta_{-s}(v) = I_s(v)(0,1)$ (in fact, the functional δ_{-s} is given (up to a normalization constant) by the integral over the line $\{(1,x)|\ x \in \mathbb{R}\} \subset \mathbb{R}^2$). The basic theory of the constant term of the Eisenstein series then implies that

$$C(\mathbb{E}(s)(v))|_{x=\bar{e}} = \delta_s(v) + c(s)\delta_{-s}(v)\ .$$

Applying this to (3.3), we obtain the spectral decomposition

$$l_{\Delta\mathcal{N}}(\nu_E(w)) = l_{\mathcal{N}}(pr_{res\oplus Eis}(\nu_E(w))) = \frac{vol(\mathcal{N})}{vol(X)^{-\frac{1}{2}}} \cdot Tr(w) + \int_{i\mathbb{R}} a(s)\delta_s(T_s(w))\ ds\ .$$

Here we have used the functional equation for the constant term of the Eisenstein series

$$a(s)c(s) \cdot \delta_{-s}(T_s(w)) = a(-s) \cdot \delta_{-s}(T_{-s}(w))$$

and the assumption that the residual spectrum is trivial. Taking into consideration the Plancherel formula (3.1) and the normalization of measures $vol(X) = vol(\mathcal{N}) = 1$, we arrive at the identity

$$\sum_k |a_k(\nu)|^2 \hat{w}(k,-k) = Tr(w) + \int_{i\mathbb{R}} a(s)\delta_s(T_s(w))\ ds\ . \tag{3.4}$$

This is our form of the Rankin-Selberg formula. To give it a more familiar form similar to (1.7), we will make (3.4) more explicit by describing T_s and δ_s in the line model of V_s. We do this by choosing an explicit kernel for the model invariant trilinear functional l_s^{mod}.

3.5. **Model trilinear functionals.** We recall the construction of model trilinear functionals presented in [BR3]. There it was shown that in the line model of representations $V \simeq V_\tau$, $\bar{V} \simeq V_{-\tau}$ and V_{-s} the kernel

$$\begin{aligned} &K_{\tau,-\tau,-s}(x,y,z) \\ &\quad = |x-y|^{(-s-1)/2}|xz-1|^{(-2\tau+s-1)/2}|yz-1|^{(2\tau+s-1)/2} \end{aligned} \tag{3.5}$$

defines a non-zero trilinear G-invariant functional l_s^{mod} on $V_\tau \otimes V_{-\tau} \otimes V_{-s}$. This gives rise to the map $T_s : E_\tau \simeq V_\tau \otimes V_{-\tau} \to V_s$ given by the same kernel. Here the variable x corresponds to the representation V_τ and $y,\ z$ correspond to $V_{-\tau}$ and

V_{-s}, respectively. Note a certain asymmetry between V_τ, $V_{-\tau}$ and V_{-s}. This is because we choose for the first two representations the line model associated with the upper triangular subgroup and for the last representation the model associated with the lower triangular subgroup.

In the line model $V_s \subset C^\infty(\mathbb{R})$ the N-invariant functional δ_s is given by the evaluation at the point $z = 0$: $\delta_s(f) = f(0)$. Hence from (3.5) it follows that the composition $\delta_s(T_s)$ is given by the Mellin transform for any $w \in E \subset C^\infty(\mathbb{R} \times \mathbb{R})$,

$$\delta_s(T_s(w)) = \int_{\mathbb{R}^2} w(x,y)|x-y|^{(-s-1)/2} dxdy \ .$$

Plugging this into (3.4), we arrive at the "classical" Rankin-Selberg formula (we assume as before that the residual spectrum is trivial)

$$\sum_k |a_k(\nu)|^2 \hat{w}(k,-k) = Tr(w) + \int_{i\mathbb{R}} a(s) w^\flat(s) \ ds \ , \tag{3.6}$$

where we denoted

$$w^\flat(s) = \int w(x,y)|x-y|^{(-s-1)/2} dxdy \ . \tag{3.7}$$

This coincides with the Mellin transform $M(\alpha)(s)$ for $\alpha(t) = \int_{x-y=t} w(x,y)dl$. The transform $^\flat$ is defined for any smooth rapidly decreasing function w, at least for all $\mathrm{Re}(s) < 1$. In fact, it could be defined for all $\lambda \in \mathbb{C}$, by means of analytic continuation, but we will not need this. We will consider only the case $s \in i\mathbb{R}$ as we assume that the residual spectrum is trivial. In general, the residual spectrum could be treated similarly. We note also that $Tr(w) = \int w(x,x)dx = \alpha(0)$.

We can now rewrite the Rankin-Selberg formula in a more familiar form if we substitute the vector $w(x,y)$ by $\alpha(t) = \int_{x-y=t} w(x,y)dl$. We have

$$\sum_k |a_k(\nu)|^2 \hat{\alpha}(k) = \alpha(0) + \int_{i\mathbb{R}} a(s) \cdot M(\alpha)(s) \ ds \ , \tag{3.8}$$

where $\hat{\alpha}(\xi) = \hat{w}(\xi, -\xi)$. This coincides with the classical Rankin-Selberg formula.

3.5.1. *Remarks.* (1) Taking into account that $M(\alpha)(s) = \gamma(s) M(\hat{\alpha})(1-s)$, where $\gamma(s) = \frac{\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})}{\pi^{-\frac{1-s}{2}}\Gamma(\frac{1-s}{2})}$ (note that $|\gamma(s)| = 1$ for $s \in i\mathbb{R}$), we see that

$$\sum_k |a_k(\nu)|^2 \hat{\alpha}(k) = \alpha(0) + \int_{i\mathbb{R}} a(s)\gamma(s) M(\hat{\alpha})(s) \ ds \ . \tag{3.9}$$

This seems to have the advantage of being an identity for one function $\hat{\alpha}$ and not for two functions $\hat{\alpha}$ and $M(\alpha)$. In practice we find it easier to work with one master function α and to use the identity (3.8).

(2) We would like to point out one essential difference between the classical Rankin-Selberg formula (1.7) obtained via unfolding and formula (3.8) that we prove. The unfolding method provides an explicit relation between a choice of a model Whittaker functional on a cuspidal representation and the coefficient of proportionality $D(s, \phi, \bar{\phi})$ (i.e., essentially the Rankin-Selberg L-function). In the argument we presented, the coefficient of proportionality $a(s)$ in addition depends on the choice of the auxiliary model trilinear functional. One can use the Whittaker functional in order to define the model trilinear functional and hence eliminate this extra indeterminacy. We hope to return to this subject elsewhere.

3.6. **Proof of Theorem 1.3.** We are interested in getting a bound for the coefficients $a_n(\phi)$. The idea is to find a test vector $w \in V \otimes \bar{V}$, i.e., a function $w \in C^\infty(\mathbb{R} \times \mathbb{R})$, such that when substituted in the Rankin-Selberg formula (3.6), it will produce a weight $\hat{w}$ which is not too small for a given n ($|n| \to \infty$). Given such a vector, we have to estimate its spectral density, i.e., the transform $w^\flat$. One might be tempted to take w such that $\hat{w}$ is essentially a delta function (i.e., the weight $\hat{w}$ picks up just a few coefficient $a_n(\phi)$ in (3.6)). However, for such a vector we have no means to estimate the right hand side of the Rankin-Selberg formula (3.6) because the weight function $w^\flat$ is spread over an interval of the spectrum which is too long to use the maximum modulus bound (still, conjecturally, even for such functions the contribution on the right hand side of the Rankin-Selberg formula is small thanks to cancellations). The solution to this problem is well known in harmonic analysis. One takes a function which produces a weighted sum of the coefficients $|a_k(\phi)|^2$ in a certain range depending on n and such that its transform $w^\flat$ spreads over a shorter interval. For a certain kind of such test vectors w (namely, those for which the support of $\hat{w}$ is not too small), we give essentially a *sharp* bound for the value of $l_{\Delta\mathcal{N}}(w)$.

We now explain how to choose the required test vectors. Let χ be a smooth function on $\mathbb{R}$ with the support $supp(\chi) \subset [-\frac{1}{2}, \frac{1}{2}]$ and such that the Fourier transform satisfies $|\hat{\chi}(\xi)| \geq 1$ for $|\xi| \leq 1$. We consider the convolution $\psi = \chi * \chi'$, where $\chi'(x) = \bar{\chi}(-x)$. We have $supp(\psi) \subset [-1, 1]$, $\hat{\psi}(\xi) \geq 0$ for all ξ and $\hat{\psi}(\xi) \geq 1$ for $|\xi| \leq 1$.

Let $N \geq T \geq 1$ be two real numbers. We consider the test vector

$$w_{N,T}(x, y) = T \cdot e^{-iN(x-y)} \cdot \psi(T(x - y)) \cdot \psi(x + y) \ .$$

We have the following technical lemma describing properties of $w^\flat_{N,T}$ (where the transform $^\flat$ was defined in formula (3.7) and is essentially the Mellin transform in $|x - y|$).

Lemma. *For $w_{N,T}$ as above, the following bounds hold:*

(1) $|\int w_{N,T}(t, t)dt| \leq cT$,
(2) $\hat{w}_{N,T}(\xi, -\xi) \geq 0$ *for all* ξ,
(3) $\hat{w}_{N,T}(\xi, -\xi) \geq 1$ *for all* ξ *such that* $|\xi - N| \leq T$,
(4) $|w^\flat_{N,T}(s)| \leq cT|N|^{-\frac{1}{2}}$ *for* $|s| \leq N/T$,
(5) $|w^\flat_{N,T}(s)| \leq cT(1 + |s|)^{-3}$ *for* $|s| \geq N/T$,

for some fixed constant $c > 0$ which is independent of N and T.

Bounds (1)–(3) are immediate. Bounds (4) and (5) are standard, once we apply the stationary phase method or the van der Corput lemma to the integral of the form $w^\flat_{N,T}(s) = \hat{\psi}(0) \cdot T^{\frac{1}{2}+s/2} \cdot \int \psi(t)\ e^{-i\frac{N}{T}t}|t|^{-\frac{1}{2}-s/2}dt$ (see Section 3.8).

We return to the proof of Theorem 1.4. We will use the mean value (or convexity) bound

$$\int_0^A |a(it)|^2 dt \leq C_\tau A^2 \ln A \ , \tag{3.10}$$

proved in [BR1] for any $A \geq 1$. The constant C_τ satisfies the bound $C_\tau \leq C_\Gamma(1+|\tau|)$ with a constant C_Γ depending on Γ only.

We substitute the vector $w_{N,T}$ into the Rankin-Selberg formula (3.6) (and use $Tr(w) = \int w(t, t)dt$). Taking into account the convexity bound (3.10), bounds in

the lemma and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
&\sum_{|k-N|\le T} |a_k(\nu)|^2 \le \sum_k |a_k(\nu)|^2 \hat{w}_{N,T}(k) \\
&= \int w_{N,T}(t,t)\ dt + \int_{i\mathbb{R}} a(s) w^\flat_{N,T}(s)\ d|s| \\
&\le cT + \int_{|s|\le N/T} cT|N|^{-\frac12}|a(s)|\ d|s| + \int_{|s|\ge N/T} cT(1+|s|)^{-3}|a(s)|\ d|s| \\
&\le cT + cT|N|^{-\frac12}\left(\int_{|s|\le N/T} |a(s)|^2 d|s| \cdot \int_{|s|\le N/T} 1\ d|s|\right)^{\frac12} \\
&\qquad\qquad + \ cT\int_{|s|\ge N/T} (1+|s|)^{-3}(1+|a(s)|^2)\ d|s| \\
&\le cT + CT|N|^{-\frac12}\left(\frac{N}{T}\right)^{3/2+\varepsilon} + DT = c'T + CT^{-\frac12-\varepsilon}|N|^{1+\varepsilon}\ ,
\end{aligned}$$

for any $\varepsilon > 0$ and some constants $c',\ C,\ D > 0$.

Setting $T = N^{2/3}$, we obtain $\sum_{|k-N|\le N^{2/3}} |a_k(\nu)|^2 \le A_\varepsilon N^{2/3+\varepsilon}$ for any $\varepsilon > 0$. □

3.7. **Remarks.** (1) It is more customary to use formula (3.9). We find the geometric formula (3.6) more transparent. Following the argument of Good [Go], one usually argues as follows. For $R \ge 1$ and $Z \ge 1$, choose a test function $\alpha_{Z,R}(t) = \alpha_Z(t/R)$, where α_Z is smooth, supported in $(1-2/Z, 1+2/Z)$ and $\alpha|_{(1-1/Z,1+1/Z)} \equiv 1$. This means that the sum $\sum_k |a_k(\nu)|^2\alpha_{Z,R}(k)$ is essentially over k in the interval of size R/Z centered at R. The Mellin transform $M(\alpha_Z)(s) = \int_{\mathbb{R}^+} \alpha_Z(t)|t|^s d^x t$ of α_Z satisfies the simple bound

$$|M(\alpha_Z)(s)| \le cZ^{-1}$$

for any $|s|$ and the bound

$$|M(\alpha_Z)(s)| \le c|s|^{-1}\left(\frac{Z}{|s|}\right)^m$$

for any $m > 0$ and $|s| \ge 1$. This follows easily from integration by parts (we are only interested in $s \in i\mathbb{R}$). In particular, we have $|M(\alpha_Z)(s)| \le cZ^{\frac12+\varepsilon}|s|^{-3/2-\varepsilon}$ for $|s| \ge Z$. Using the average bound $\int_0^A |a(it)|^2 dt \le C_\tau A^2 \ln A$ and the Cauchy-Schwartz inequality, one obtains

$$\left|\int_{i\mathbb{R}} a(s)\gamma(s)M(\alpha_{Z,R})(s)ds\right| \le C_\varepsilon R^{\frac12+\varepsilon}Z^{\frac12+\varepsilon}$$

for any $\varepsilon > 0$. We arrive at $\sum_k |a_k(\nu)|^2\alpha_{Z,R}(k) \le R/Z + C_\varepsilon R^{\frac12+\varepsilon}Z^{\frac12+\varepsilon}$ and setting $Z = R^{1/3}$, we obtain the bound claimed.

(2) One might conjecture that for any $A \ge 1$, the bound

$$\int_A^{2A} |a(it)|^2 dt \ll_{\nu,\varepsilon} A^{1+\varepsilon} \tag{3.11}$$

holds for any $\varepsilon > 0$ (e.g., the Lindelöff conjecture on the average for the Rankin-Selberg L-function). It is easy to see that such a bound would lead to the bound $|a_n(\nu)| \ll_{\nu,\varepsilon} |n|^{\frac14+\varepsilon}$. We note that this bound is a natural barrier which for the

Rankin-Selberg method would be hard to overcome. Nevertheless, it was suggested by Sarnak [Sa] that for a general lattice $\Gamma \subset PGL_2(\mathbb{R})$ the Ramanujan-Petersson conjecture $|a_n(\phi_\tau)| \ll |n|^\varepsilon$ might hold.

3.8. **Proof of the lemma in Section 3.6.** We prove the following statement from which Lemma 3.6 immediately follows.

Lemma. *Let ψ be a fixed smooth function with a compact support in $[-1,1]$. For $s \in i\mathbb{R}$ and $\xi \in \mathbb{R}$, denote $\psi^\flat(\xi,s) = \int_{\mathbb{R}} \psi(t)e^{-i\xi t}|t|^{-\frac{1}{2}-s}dt$. There exists a constant $c > 0$ such that*

(1) $|\psi^\flat(\xi,s)| \le c(1+|\xi|)^{-\frac{1}{2}}$ *for* $|s| \le 2|\xi|$,
(2) $|\psi^\flat(\xi,s)| \le c(1+|s|)^{-3}$ *for* $|s| \ge 2|\xi|$.

In fact, both claims in the lemma are simple consequences of the van der Corput lemma (see [St], p. 332). One can also use the following Fourier transform argument.

To prove (1), we use the fact that the Fourier transform of $|t|^{-\frac{1}{2}-s}$ is equal to $\gamma(-\frac{1}{2}-s)|\xi|^{-\frac{1}{2}+s}$, where $\gamma(s)$ is defined in Section 3.5.1 and $|\gamma(-\frac{1}{2}-it)| = 1$. The Fourier transform of ψ satisfies $|\hat{\psi}(\xi)| \ll (1+|\xi|)^{-M}$ for any $M > 0$. Hence, the Fourier transform of $\psi(t)|t|^{-\frac{1}{2}-s}$, i.e., the convolution $\hat{\psi}(\xi) * |\xi|^{-\frac{1}{2}+s}$, is bounded by $c(1+|\xi|)^{-\frac{1}{2}}$ for some c and all $s \in i\mathbb{R}$. This proves (1).

To prove (2), it is enough to notice that under the condition $|s| \ge |\xi|$ the phase in the oscillating integral defining $\psi^\flat(\xi,s)$ has no stationary points. The resulting bound easily follows from integration by parts (see Appendix A for similar computations). □

4. Spherical Fourier coefficients

When dealing with the spherical Fourier coefficients, we assume, for simplicity, that the lattice Γ is co-compact. A general finite co-volume lattice could be treated analogously without any significant changes (see the remark in Section 4.5).

4.1. **Geodesic circles.** We start with the geometric origin of the spherical Fourier coefficients. We fix a maximal compact subgroup $K = PO(2) \subset G$ and the identification $G/K \to \mathbb{H}$, $g \mapsto g \cdot i$. Let $y \in Y$ be a point and let $p : \mathbb{H} \to \Gamma \setminus \mathbb{H} \simeq Y$ be the corresponding projection compatible with the distance function $\mathrm{d}(\cdot,\cdot)$ on Y and on $\mathbb{H}$. Let $R_y > 0$ be the injectivity radius of Y at y. For any $r < R_y$ we define the geodesic circle of radius r centered at y to be the set $\sigma(r,y) = \{y' \in Y | \mathrm{d}(y',y) = r\}$. Since the map p is a local isometry, we have that $p(\sigma_{\mathbb{H}}(r,z)) = \sigma(r,y)$ for any $z \in \mathbb{H}$ such that $p(z) = y$, where $\sigma_{\mathbb{H}}(r,z)$ is the corresponding geodesic circle in $\mathbb{H}$ (all geodesic circles in $\mathbb{H}$ are the Euclidian circles, though with a center different from z). We associate to any such circle on Y an orbit of a compact subgroup on X. Namely, let $K_0 = PSO(2) \subset K$ be the connected component of K. Any geodesic circle on $\mathbb{H}$ is of the form $\sigma_{\mathbb{H}}(r,z) = hK_0g \cdot i$ with $h,\ g \in G$ such that $h \cdot i = z$ and $hg \cdot i \in \sigma_{\mathbb{H}}(r,z)$ (i.e., an h-translation of a standard geodesic circle centered at $i \in \mathbb{H}$ and passing through $g \cdot i \in \mathbb{H}$). Note that the radius of the circle is given by the distance $\mathrm{d}(i, g \cdot i)$ and hence $g \notin K_0$ for a nontrivial circle. Given the geodesic circle $\sigma(r,y) \subset Y$, we choose a circle $\sigma_{\mathbb{H}}(r,z) \subset \mathbb{H}$ projecting onto $\sigma(r,y)$ and the corresponding elements $g,\ h \in G$. We denote by $K_\sigma = g^{-1}K_0g$ the corresponding compact subgroup and consider its orbit $\mathcal{K}_\sigma = hg \cdot K_\sigma \subset X$. Clearly we have $p(\mathcal{K}_\sigma) = \sigma$. We endow the orbit $\mathcal{K}_\sigma$ with the unique K_σ-invariant measure $d\mu_{\mathcal{K}_\sigma}$

of the *total mass one* (from the geometric point of view a more natural measure would be the length of σ).

We note that, in what follows, the restriction $r < R_y$ is not essential. From now on we assume that $\mathcal{K} \subset X$ is a connected orbit of a *connected* compact subgroup $K' \subset G$ (i.e., K' is conjugated to $PSO(2)$). The restriction $r < R_y$ simply means that the projection $p(\mathcal{K}) \subset Y$ is a smooth non-self-intersecting curve on Y. We also remark that it is well known that polar geodesic coordinates (r, θ) centered at a point $z_0 \in \mathbb{H} = G/K$ could be obtained from the Cartan KAK-decomposition of G (see [He]). This allows one to give a purely geometric construction of the functions $P_{n,\tau}$ from the Introduction (see Section 4.2.1).

4.2. **K'-equivariant functionals.** We fix a point $\dot{o} \in \mathcal{K}$. Let $\chi : K' \to S^1$ be a character. To such a character we associate a function $\chi_{\cdot}(\dot{o}k') = \chi(k')$, $k' \in K'$, on the orbit $\mathcal{K}$ and the corresponding functional on $C^\infty(X)$ given by

$$d^{aut}_{\chi,\mathcal{K}}(f) = \int_{\mathcal{K}} f(k)\bar{\chi}_{\cdot}(k)d\mu_{\mathcal{K}}$$

for any $f \in C^\infty(X)$.

The functional $d^{aut}_{\chi,\mathcal{K}}$ is χ-equivariant: $d^{aut}_{\chi,\mathcal{K}}(R(k')f) = \chi(k') \cdot d^{aut}_{\chi,\mathcal{K}}(f)$ for any $k' \in K'$, where R is the right action of G on the space of functions on X. For a given orbit $\mathcal{K}$ and a choice of a generator χ_1 of the cyclic group $\hat{K'} \simeq \mathbb{Z}$ of characters of the compact group K', we will use the shorthand notation $d^{aut}_n = d^{aut}_{\chi_n,\mathcal{K}}$, where $\chi_n = \chi_1^n$. The functions $\chi_{n\cdot}$, $n \in \mathbb{Z}$, form an orthonormal basis for the space $L^2(\mathcal{K}, d\mu_{\mathcal{K}})$ (since we normalized the measure by $\mu_{\mathcal{K}}(\mathcal{K}) = 1$).

Let $\nu : V \to C^\infty(X)$ be an irreducible automorphic representation. When it does not lead to confusion, we denote by the same letter the functional $d^{aut}_{\chi,\mathcal{K}} = d^{aut}_{\chi,\mathcal{K},\nu}$ on the space V induced by the functional $d^{aut}_{\chi,\mathcal{K}}$ defined above on the space $C^\infty(X)$. Hence we obtain an element in the period space $\mathcal{P}_{K'}(V, \chi) = \mathrm{Hom}_{K'}(V, \chi)$. We next use the well-known fact that this space is at most one-dimensional.

Let $V \simeq V_\tau$ be a representation of the generalized principal series. We have then $\dim \mathrm{Hom}_{K'}(V_\tau, \chi) \leq 1$ for any character χ of K' (i.e., the space of K'-types is at most one-dimensional for a maximal connected compact subgroup of G). In fact, $\dim \mathrm{Hom}_{K'}(V_\tau, \chi_n) = 1$ if and only if n is even.

To construct a model χ-equivariant functional on V_τ, we consider the circle model $V_\tau \simeq C^\infty_{even}(S^1)$ in the space of even functions on S^1 and the standard vectors (exponents) $e_n = \exp(in\theta) \in C^\infty(S^1)$ which form a basis of K-types for the *standard* maximal compact subgroup $K = PO(2)$. For any n such that $\dim \mathrm{Hom}_{K_0}(V_\tau, \chi_n) = 1$, the vector $e'_n = \pi_\tau(g^{-1})e_n$ defines a non-zero (χ_n, K')-equivariant functional on V_τ by the formula

$$d^{mod}_n(v) = d^{mod}_{\chi_n,\tau}(v) = \langle v, e'_n \rangle \ .$$

We call such a functional the *model* χ_n-equivariant functional on $V \simeq V_\tau$.

The uniqueness principle implies that there exists a constant $b_n(\nu) = b_{\chi_n,\mathcal{K}}(\nu)$ such that for any $v \in V$

$$d^{aut}_n(v) = b_n(\nu) \cdot d^{mod}_n(v) \ .$$

4.2.1. *Functions $P_{n,\tau}$.* We want to compare the coefficients $b_n(\nu)$ to the coefficients $b_n(\phi_\tau)$ we introduced in (1.10). In particular we describe the functions $P_{n,\tau}$ and their normalization. Let $h,\ g \in G$ and let $\mathcal{K} = hgK' \subset \Gamma \setminus G = X$ be the orbit

of the connected compact group $K' = g^{-1}K_0 g$ as above. Let $\nu : V_\tau \to C^\infty(X)$ be an automorphic realization and let $\phi_\tau = \nu(e_0) \in C^\infty(X)$ be the K-invariant vector which corresponds to a K-invariant vector $e_0 \in V_\tau$ of norm one, i.e., ϕ_τ is a Maass form. We define the function $P_{n,\tau}$ through the following matrix coefficient: $P_{n,\tau}(r)e^{in\theta} = \langle e_0, \pi_\tau(g^{-1}k^{-1})e_n\rangle_{V_\tau}$, where $(r,\theta) = z = hkg \cdot i \in \mathbb{H}$ for $k \in K_0$. It is well known that the matrix coefficient is an eigenfunction of the Casimir operator and hence $P_{n,\tau}(r)e^{in\theta}$ is an eigenfunction of Δ on $\mathbb{H}$. In fact, the functions $P_{n,\tau}$ are equal to the Legendre functions for the special choice of parameters (compare to [Iw]). Under such normalization of functions $P_{n,\tau}$, we have

$$|b_n(\nu)| = |b_n(\phi_\tau)| \ .$$

Let $\bar{V}$ be the complex conjugate representation; it is also an automorphic representation with the realization $\bar{\nu} : \bar{V} \to C^\infty(X)$. We only consider the case of representations of the principal series, i.e., we assume that $V = V_\tau$, $\bar{V} = V_{-\tau}$ for some $\tau \in i\mathbb{R}$; the case of representations of the complementary series can be treated similarly. Let $\{e_n\}_{n\in 2\mathbb{Z}}$ be a K-type orthonormal basis in V. We denote by $\{\bar{e}_n\}$ the complex conjugate basis in $\bar{V}$ and we denote by $\bar{d}_n^{aut/mod}$ the corresponding automorphic/model functionals on the conjugate space $\bar{V} \simeq V_{-\tau}$.

We introduce another notation for a K'-*invariant* functional on an irreducible automorphic representation $\nu_i : V_{\lambda_i} \to C^\infty(X)$ of class one.

Let $\chi_0 : K' \to 1 \in S^1 \subset \mathbb{C}$ be the trivial character of K'. We have as above

$$d^{aut}_{\chi_0,\mathcal{K},\nu_i}(v) = \int_{\mathcal{K}} \nu_i(v)(k)\bar{\chi}_0(k)d\mu_{\mathcal{K}} = b_0(\nu_i) \cdot \langle v, e_0'\rangle_{V_{\lambda_i}} \ ,$$

for any $v \in V_{\lambda_i}$.

We denote by $d_\lambda(v) = \langle v, e_0'\rangle_{V_\lambda}$ the corresponding model functional and by

$$\beta(\lambda_i) = b_0(\nu_i)$$

the proportionality coefficient (somewhat abusing notation, since the coefficient depends on the automorphic realization ν_i and not only on the isomorphism class V_{λ_i}).

We want to compare coefficients $\beta(\lambda_i)$ with more familiar quantities. Let $\mathcal{K} = x_0K' \subset X$ be an orbit of the compact group K'. Let $\nu_i : V_{\lambda_i} \to C^\infty(X)$ be an automorphic realization and let $\phi'_{\lambda_i} = \nu_i(e_0')$ be the K'-invariant vector which corresponds to a K'-invariant vector $e_0' \in V_{\lambda_i}$ of norm one. From the definition of $b_0(\nu_i)$ it follows that

$$\beta(\lambda_i) = \phi'_{\lambda_i}(x_0) \ . \tag{4.1}$$

Hence, under the normalization we choose, the coefficients $\beta(\lambda_i)$ coincide with the value at a point x_0 for Maass form ϕ'_{λ_i} on the Riemann surface $Y' = \Gamma \setminus G/g^{-1}Kg$.

Finally, we note that on the discrete series representations any K'-invariant functional is identically zero. This greatly simplifies the technicalities in what follows.

4.3. **$\Delta\mathcal{K}$-restriction.** Let $\Delta\mathcal{K} \subset \Delta X \subset X \times X$ be the diagonal copy of the cycle $\mathcal{K}$. We define the $\Delta K'$-invariant automorphic functional $d_{\Delta\mathcal{K}} : E = V \otimes \bar{V} \to \mathbb{C}$ by

$$d_{\Delta\mathcal{K}}(w) = \int_{\Delta\mathcal{K}} \nu_E(w)(k,k)d\mu_{\mathcal{K}}$$

for any $w \in E$.

Arguing as in Section 3.1, we have the Plancherel formula on $\mathcal{K}$

$$\begin{aligned} d_{\Delta\mathcal{K}}(w) &= \sum_n d_n^{aut} \otimes \bar{d}_{-n}^{aut}(w) \\ &= \sum_n |b_n(\nu)|^2 d_n^{mod} \otimes \bar{d}_{-n}^{mod}(w) = \sum_n |b_n(\nu)|^2 \hat{w}(n,-n) \ , \end{aligned} \tag{4.2}$$

where $\hat{w}(n,-n) = \langle w, e_n \otimes \bar{e}_{-n} \rangle_E$. Hence, choosing various vectors w, we can obtain a variety of weighted sums $\sum_n |b_n(\nu)|^2 \hat{\alpha}(n)$.

We now obtain another expression for the functional $d_{\Delta\mathcal{K}}$ using the spectral decomposition of $L^2(X)$ and trilinear invariant functionals as in Section 3.4.

4.4. **Anisotropic Rankin-Selberg formula. Proof of Theorem 1.4.** We assume that the space X is compact. This implies the discrete sum decomposition $L^2(X) = (\bigoplus_i (L_i, \nu_i)) \oplus (\bigoplus_\kappa (L_\kappa, \nu_\kappa))$ into irreducible unitary representations of G, where $\nu_i : L_i \simeq L_{\lambda_i} \to L^2(X)$ are unitary representations of class one (i.e., those which correspond to Maass forms on Y) and L_κ are representations of discrete series (i.e., those which correspond to holomorphic forms on Y). We fix such a decomposition and denote by $V_i \subset L_i$ the corresponding spaces of smooth vectors and by $\nu_i^* : C^\infty(X) \to V_i$ the adjoint map.

We fix $\nu : V \to C^\infty(X)$ an irreducible automorphic representation and denote by $\nu_E = \nu \otimes \bar{\nu} : E = V \otimes \bar{V} \to C^\infty(X \times X)$ the corresponding realization of E.

We use the notation from Section 3.4. Let $r_\Delta : C^\infty(X \times X) \to C^\infty(X)$ be the map induced by the imbedding $\Delta : X \hookrightarrow X \times X$. Let $\nu_i : V_{\lambda_i} \to C^\infty(X)$ be an irreducible automorphic representation. Composing r_Δ with the adjoint map $\nu_i^* : C^\infty(X) \to V_{\lambda_i}$, we obtain the trilinear ΔG-invariant map $T_i^{aut} : E \to V_{\lambda_i}$ and the corresponding automorphic trilinear functional l_i^{aut} on $E \otimes V_{\lambda_i}^*$ defined by $l_i^{aut}(v \otimes u \otimes w) = \langle r_\Delta(\nu_E(u \otimes v)), \bar{\nu}_i(w) \rangle$ (we identified $\bar{V}_{\lambda_i}$ with the smooth part of $V_{\lambda_i}^*$). Such a functional is clearly G-invariant, and hence we can invoke the uniqueness principle for trilinear functionals from Section 3.4.

To this end, we fix a model trilinear functional $l_{\lambda_i}^{mod} = l_{E \otimes V_{\lambda_i}^*}^{mod}$ (see Section 3.5 and formula (4.5) below; for a detailed discussion, see [BR3]) and the corresponding intertwining model map $T_{\lambda_i} = T_{\lambda_i}^{mod} : E \to V_{\lambda_i}$. This gives rise to the coefficient of proportionality which we denote by $a(\lambda_i) = a_{\nu_E \otimes \nu_i}$ (somewhat abusing notation by suppressing the dependence on ν_E and ν_i) such that

$$T_i^{aut} = \nu_i^*(r_\Delta) = a(\lambda_i) \cdot T_{\lambda_i} \ .$$

The integral $p_\mathcal{K}(f) = \int_\mathcal{K} f(k) d\mu_\mathcal{K}$ defines the period map $p_\mathcal{K} : C^\infty(X) \to \mathbb{C}$. Note that $p_\mathcal{K}$ vanishes on representations with no non-zero K'-invariant vectors, e.g., on representations of discrete series.

We have the basic relation

$$d_{\Delta\mathcal{K}} = (r_\Delta)_*(p_\mathcal{K}) \ .$$

This means that for any $w \in E$, we have $d_{\Delta\mathcal{K}}(w) = \int_\mathcal{K} r_\Delta(\nu_E(w)) d\mu_\mathcal{K}$. We also have the following spectral decomposition for any $w \in E$:

$$r_\Delta(w) = \sum_{\nu_i} \nu_i^*(r_\Delta(w)) \ , \tag{4.3}$$

where ν_i runs through the fixed decomposition of $L^2(X)$ into irreducible components.

We apply the functional $p_{\mathcal{K}}$ to each term in (4.3) and invoke the uniqueness principle for K'-invariant functionals on irreducible representations V_{λ_i} (i.e., that $d^{aut}_{\lambda_i} = \beta(\lambda_i)\cdot d_{\lambda_i}$; see Section 4.2.1). This, together with the Fourier expansion (4.2), implies two different expansions for the functional $d_{\Delta\mathcal{K}}$: one which is *"geometric"* (i.e., the Fourier expansion (4.2) along the orbit $\mathcal{K}$) and another one which is *"spectral"* (i.e., induced by the trilinear invariant functionals and the expansion (4.3)).

Namely, we have

$$\sum_n |b_n(\nu)|^2 \hat{w}(n,-n) = d_{\Delta\mathcal{K}}(w) = \sum_{\lambda_i} a(\lambda_i)\beta(\lambda_i)\cdot d_{\lambda_i}(T_{\lambda_i}(w)) \ , \tag{4.4}$$

where $\hat{w}(n,-n) = \langle w, e'_n \otimes \bar{e}'_{-n}\rangle_E$ for any $w \in E$, with $\{e'_n\}$ a basis of K'-types in V and $\{\bar{e}'_n\}$ the conjugate basis in $\bar{V}$.

This is our substitute for the Rankin-Selberg formula in the anisotropic case.

To make this formula explicit, we describe the model trilinear functional in the circle model of representations $V = V_\tau$, $\bar{V} = V_{-\tau}$ and V_{λ_i}. We assume for simplicity that $\tau \in i\mathbb{R}$ (i.e., V is a representation of the principal series) and that there is no exceptional spectrum for the lattice Γ (i.e., that $\lambda_i \in i\mathbb{R}$ for all $i > 0$, and hence $V^*_\lambda \simeq V_{-\lambda}$). The general case could be treated analogously.

To simplify formulas, we make the following remark. Formula (4.4) appeals only to automorphic representations of G and a choice of a (non-trivial) connected compact subgroup $K' \subset G$ (i.e., the choice of another compact subgroup K that we made in Section 4.1 is irrelevant). Since there is no *preferred* compact subgroup in G, we may assume without loss of generality that $K' = PSO(2)$ is the standard connected compact subgroup of G.

It was shown in [BR3] that in the circle model of class one representations, the kernel of $l^{mod}_{E\otimes V_{-\lambda}}$ on the space $E \otimes V_{-\lambda} \simeq C^\infty_{even}(S^1 \times S^1 \times S^1)$ is given by the following function in three variables $\theta,\ \theta',\ \theta'' \in S^1$:

$$\begin{aligned} &K_{\tau,-\tau,\lambda}(\theta,\theta',\theta'') \\ &\qquad = |\sin(\theta-\theta')|^{\frac{-1-\lambda}{2}} |\sin(\theta-\theta'')|^{\frac{-1-2\tau+\lambda}{2}} |\sin(\theta'-\theta'')|^{\frac{-1+2\tau+\lambda}{2}} \ . \end{aligned} \tag{4.5}$$

This also defines the kernel of the map $T_\lambda : E \to V_\lambda$ via the relation

$$\langle T_\lambda(w), v\rangle_{V_\lambda} = \frac{1}{(2\pi)^3}\int_{(S^1)^3} w(\theta,\theta')v(\theta')K_{\tau,-\tau,\lambda}(\theta,\theta',\theta'')d\theta d\theta' d\theta'' \ .$$

We have $d_\lambda(T_\lambda(w)) = \langle T_\lambda(w), e_0\rangle_{V_\lambda} = \frac{1}{(2\pi)^3}\int w(\theta,\theta')K_{\tau,-\tau,\lambda}(\theta,\theta',\theta'')d\theta d\theta' d\theta''$ for any vector $w \in C^\infty(S^1 \times S^1)$. It is clear from formula (4.4) that we can assume without loss of generality that the vector $w \in E$ is ΔK-invariant. Such a vector w can be described by a function of one variable; namely, we set $w(\theta,\theta') = u(c)$ for $u \in C^\infty(S^1)$ and $c = (\theta-\theta')/2$. We have then $\hat{w}(n,-n) = \hat{u}(n) = \frac{1}{2\pi}\int_{S^1} u(c)e^{-inc}dc$, i.e., $\hat{u}$ is the Fourier transform of u.

We introduce a new kernel

$$k_\lambda(c) = k_{\tau,\lambda}\left(\tfrac{\theta-\theta'}{2}\right) = \frac{1}{2\pi}\int_{S^1} K_{\tau,-\tau,\lambda}(\theta,\theta',\theta'')d\theta'' \tag{4.6}$$

and the corresponding integral transform

$$u^\sharp(\lambda) = u^\sharp_\tau(\lambda) = \frac{1}{(2\pi)^2}\int_{S^1} u(c)k_\lambda(c)dc \ , \tag{4.7}$$

suppressing the dependence on τ as we have fixed the Maass form ϕ_τ. The transform is clearly defined for any smooth function $u \in C^\infty(S^1)$, at least for $\lambda \in i\mathbb{R}$. In fact, it could be defined for all $\lambda \in \mathbb{C}$, by means of analytic continuation, but we will not need this.

Note that k_λ is the average of the kernel $K_{\tau,-\tau,\lambda}$ with respect to the action of ΔK, or, in other terms, is the pullback of the K-invariant vector $e_0 \in V_\lambda$ under the map T^*_λ, i.e., $k_\lambda = T^*_\lambda(e_0) \in E^*$. We also note that the contribution in (4.4) coming from the trivial representation (i.e., $\lambda = 1$) is equal to $u(0) = \frac{vol(\mathcal{K})}{vol(X)^{\frac{1}{2}}} \cdot u(0)$ under our normalization of measures $vol(X) = vol(\mathcal{K}) = 1$.

Formula (4.4) then takes the form

$$\sum_n |b_n(\nu)|^2 \hat{u}(n) = u(0) + \sum_{\lambda_i \neq 1} a(\lambda_i)\beta(\lambda_i) \cdot u^\sharp(\lambda_i) \ . \tag{4.8}$$

This formula is an anisotropic analog of the Rankin-Selberg formula (3.8) for the unipotent Fourier coefficients of Maass forms. We finish the proof of Theorem 1.4. □

4.5. **Remarks.** (1) The kernel function k_λ is not an elementary function, unlike in the case of the unipotent Fourier coefficients where its analog is given by $|x-y|^{-\frac{1}{2}-s}$. This is related to the fact that the N-invariant distribution δ_s on V_s is also χ-equivariant under the action of the full Borel subgroup $B = AN$ for an appropriate character χ of B which is trivial on N. The space of (B, χ)-equivariant distributions on E is one-dimensional for a generic χ. This is due to the fact that B has one open orbit for the diagonal action on the space $\mathbb{R} \times \mathbb{R}$ and the vector space E is modelled in the space of smooth functions on this space. It is then easy to write a non-zero B-equivariant functional on E by an essentially algebraic formula. We do not have a similar phenomenon for a maximal compact subgroup of G. We will obtain, however, an elementary formula for leading terms in the asymptotic expansion of k_λ as $|\lambda| \to \infty$ (see Section A.1).

(2) For Hecke-Maass forms, the proportionality coefficient $a(s)$ in the Rankin-Selberg formula (3.4) for the unipotent Fourier coefficients gives the Rankin-Selberg L-function (after multiplication by $\zeta(2s)$). In the anisotropic case we do not know how to express the coefficient $a(\lambda_i)$ in terms of an appropriate L-function. It is known that the value of $|a(\lambda_i)|^2$ is related to the special value of the triple L-function (see [W]), but not the coefficient itself. The same is true for the coefficient $\beta(\lambda_i)$ where in special cases $|\beta(\lambda_i)|^2$ is related to a certain automorphic L-function (see [Wa], [JN]). There still might be a way to normalize the product $a(\lambda_i)\beta(\lambda_i)$ in a canonical way. We hope to return to this subject elsewhere.

(3) For a non-uniform lattice Γ, the proof we gave above leads to the following formula analogous to (4.8) which includes the contribution from the Eisenstein series. Namely, for the similarly defined coefficients $a_k(s)$ and $\beta_k(s)$ corresponding to the Eisenstein series contribution, we have

$$\sum_n |b_n(\nu)|^2 \hat{u}(n) = u(0) + \sum_{\lambda_i \neq 1} a(\lambda_i)\beta(\lambda_i) \cdot u^\sharp(\lambda_i) + \sum_{\text{cusps}} \int_{i\mathbb{R}^+} a_k(s)\beta_k(s) \cdot u^\sharp(s) ds \ .$$

4.6. **Bounds for spherical Fourier coefficients. Proof of Theorem 1.5.** We follow the strategy of Section 3.6. We construct a ΔK-invariant test vector $w \in V \otimes \bar{V}$, i.e., a function $u \in C^\infty(S^1)$, such that when substituted into the

Rankin-Selberg formula (4.8) will produce a weight $\hat{u}$ which is not too small for a given n, $|n| \to \infty$. We then have to estimate the spectral density of such a vector, i.e., the transform $u^\sharp$. In fact, as in Section 3.6 we take a function which produces a weighted sum of the coefficients $|b_k(\phi)|^2$ for k in a short range depending on n and such that its transform $u^\sharp$ is spread over a relatively short range of λ's. For such test vectors w we give an essentially sharp bound for the value of $d_{\Delta\mathcal{K}}(w)$.

We have the following technical lemma.

Lemma. *For any integer N and a real number T such that $N \geq T \geq 1$, there exists a smooth function $u_{N,T} \in C^\infty(S^1)$ satisfying the following bounds:*

(1) $|u_{N,T}(0)| \leq \alpha T$,
(2) $\hat{u}_{N,T}(k) \geq 0$ *for all* k,
(3) $\hat{u}_{N,T}(k) \geq 1$ *for all* k *satisfying* $|k - N| \leq T$,
(4) $|u^\sharp_{N,T}(\lambda)| \leq \alpha T |N|^{-\frac{1}{2}}(1+|\lambda|)^{-\frac{1}{2}} + \alpha T(1+|\lambda|^{-5/2})$ *for* $|\lambda| \leq N/T$, $\lambda \in i\mathbb{R}$,
(5) $|u^\sharp_{N,T}(\lambda)| \leq \alpha T(1+|\lambda|)^{-5/2}$ *for* $|\lambda| \geq N/T$,

for some fixed constant $\alpha > 0$ independent of N and T.

The proof of this lemma is given in Appendix A. We construct the corresponding function $u_{N,T}$ by considering a function $u_{N,T}(c) = T^2 \cdot e^{-iNc} \cdot (\psi_T * \psi'_T)(c)$, where $*$ denotes the convolution in $C^\infty(S^1)$ and $\psi'_T(c) = \bar{\psi}_T(-c)$. Here $\psi_T(c) = \psi(T \cdot c)$, where $\psi \in C^\infty(S^1)$ is a *fixed* smooth function supported in a small neighborhood of $1 \in S^1$ and $T \cdot c$ means the obvious scaling-up of the angle parameter in S^1. Functions $u_{N,T}$ obviously satisfy bounds (1)–(3) and the verification of (4)–(5) is reduced to a routine application of the stationary phase method (similar to our computations in [BR4]). These bounds are similar to bounds in the lemma in Section 3.6 for the unipotent Fourier coefficients. There are two differences though. First, in the corresponding bound in (4) there is a factor of $(1+|\lambda|)^{-\frac{1}{2}}$ in the first term. This constitutes an important difference between K-invariant and N-invariant functionals on the representation E. The additional term $T(1+|\lambda|^{-\frac{5}{2}})$ comes from the estimate of the remainder in the stationary phase method and could be improved further (although it would not make a difference in what follows). The second (minor) difference is that the integral transform $^\flat$ is elementary (i.e., the Mellin transform) while the integral transform $^\sharp$ has its kernel given by a non-elementary function (essentially by the hypergeometric function). This slightly complicates computations.

Remark. We would like to point out that it is absolutely essential that in the integral $u^\sharp_{N,T}(\lambda) = \int_{S^1} u_{N,T}(c) k_\lambda(c)\, dc$ the support of the function $u_{N,T}$ does not contain points $\pm\pi/4$, $\pm 3\pi/4$. Otherwise the phase in the above oscillating integral possess *degenerate* critical points at these values of c for $N/T \asymp |\lambda|$. The presence of degenerate critical points changes drastically the behavior of the corresponding transform $^\sharp$. In particular, for $n \asymp |\lambda|$, the $\sharp$-transform of a *pure tensor* $e_n \otimes \bar{e}_{-n}$ does not satisfy the bound (4) in the above lemma. Namely, $(e_n \otimes \bar{e}_{-n})^\sharp(\lambda)$ have a sharp peak for $n \asymp |\lambda|$ of the order of $|\lambda|^{-\frac{5}{6}}$ (as oppose to $|\lambda|^{-1}$). This phenomenon is the starting point for the proof of the subconvexity bound for the triple L-function given in [BR4]. In the present paper, we choose test vectors to vanish in a neighborhood of these degenerate points $\pm\pi/4$, $\pm 3\pi/4$ in the model realization $V \otimes \bar{V} \simeq C^\infty(S^1 \times S^1)$. This allows us to avoid the more delicate analysis of

degenerate critical points. Note that our test vectors are not given by a finite combination of *pure tensors* of K-types.

We return to the proof of the theorem. In the proof we will use two bounds for the coefficients $a(\lambda_i)$ and $\beta(\lambda_i)$. Namely, it was shown in [BR3] that

$$\sum_{A\le|\lambda_i|\le 2A} |a(\lambda_i)|^2 \le aA^2 \ , \tag{4.9}$$

for any $A \ge 1$ and some explicit $a > 0$. The second bound that we will need is the bound

$$\sum_{A\le|\lambda_i|\le 2A} |\beta(\lambda_i)|^2 \le bA^2 \ , \tag{4.10}$$

valid for any $A \ge 1$ and some b. In disguise, this is the classical bound of L. Hörmander [Ho] for the average value at a point for eigenfunctions of the Laplace-Beltrami operator on a compact Riemannian manifold (e.g., Δ on Y). This follows from the normalization $|\beta(\lambda_i)|^2 = |\phi'_{\lambda_i}(x_0)|^2$ we have chosen in (4.1) for K'-invariant eigenfunctions. In fact, the bound (4.10) is standard in the theory of the Selberg trace formula (see [Iw]) and can also be easily deduced from simple geometric considerations of [BR3]. We note that both bounds are consistent with the convexity bounds for relevant L-functions.

We also assume for simplicity that there is no exceptional spectrum (i.e., $\lambda_i \in i\mathbb{R}$ for $i > 0$). A general case could be treated analogously by extending the lemma in Section 4.6 to cover $\lambda \in (-1, 1)$ or as in [BR4] where we treated the exceptional spectrum using simple considerations based on the Sobolev restriction theorem.

We plug a test function satisfying bounds (1)–(5) of the lemma in Section 4.6 into the Rankin-Selberg formula (4.8). From the Cauchy-Schwartz inequality, bounds (4.9) and (4.10), and summation by parts, we obtain

$$\begin{aligned}
\sum_{|k-N|\le T} |b_k(\nu)|^2 &\le \sum_k |b_k(\nu)|^2\ \hat{u}_{N,T}(k) = u_{N,T}(0) + \sum_{\lambda_i\ne 1} a(\lambda_i)\beta(\lambda_i)\ u^\sharp_{N,T}(\lambda_i) \\
&\le \alpha T + \sum_{|\lambda_i|\le N/T} \alpha T|N|^{-\frac12}(1+|\lambda_i|)^{-\frac12}|a(\lambda_i)\beta(\lambda_i)| \\
&\qquad + \sum_{\lambda_i\ne 1} \alpha T(1+|\lambda_i|)^{-\frac52}|a(\lambda_i)\beta(\lambda_i)| \\
&\le \alpha T + \alpha T|N|^{-\frac12}\cdot \sum_{|\lambda_i|\le N/T} (1+|\lambda_i|)^{-\frac12}\left(|a(\lambda_i)|^2+|\beta(\lambda_i)|^2\right) \\
&\qquad + T\sum_{\lambda_i\ne 1} (1+|\lambda_i|)^{-\frac52}\left(|a(\lambda_i)|^2+|\beta(\lambda_i)|^2\right) \\
&\le \alpha T + CT|N|^{-\frac12}\left(\frac{N}{T}\right)^{\frac32+\varepsilon} + DT = c'T + CT^{-\frac12-\varepsilon}|N|^{1+\varepsilon} \ ,
\end{aligned}$$

for any $\varepsilon > 0$ and some constants $c',\ C,\ D > 0$.

Setting $T = N^{\frac23}$, we obtain $\sum_{|k-N|\le N^{2/3}} |b_k(\nu)|^2 \le A_\varepsilon N^{\frac23+\varepsilon}$ for any $\varepsilon > 0$. □

Remark 4.1. One expects that bounds $|a(\lambda_i)| \ll |\lambda_i|^\varepsilon$ and $|\beta(\lambda_i)| \ll |\lambda_i|^\varepsilon$ hold for any $\varepsilon > 0$. In special cases this would be consistent with the Lindelöff conjecture for the corresponding L-functions. This, however, will not have a similar effect on the

bound in Theorem 1.5 for spherical Fourier coefficients $b_n(\phi_\tau)$. The reason for such a discrepancy is that the spectral measure of the Eisenstein series is much "smaller" than that of the cuspidal spectrum. Nevertheless, it is natural to conjecture that for general $\Gamma \subset PGL_2(\mathbb{R})$ and a point $y_0 \in Y$ the spherical Fourier coefficients satisfy the bound $|b_n(\phi_\tau)| \ll |n|^\varepsilon$. For a CM-point y_0 and a Hecke-Maass form, this would correspond to the Lindelöff conjecture for the special value of the corresponding L-function via the Waldspurger formula (1.12).

Appendix A. Asymptotic expansions

A.1. **Asymptotic expansion for the kernel k_λ.** We set $c = \frac{\theta-\theta'}{2}$ and consider the integral (4.6), Section 4.6:

$$\begin{aligned} k_\lambda(c) &= k_{\tau,\lambda}\left(\tfrac{\theta-\theta'}{2}\right) = \frac{1}{2\pi}\int_{S^1} K_{\tau,-\tau,\lambda}(\theta,\theta',\theta'')\, d\theta'' \\ &= \frac{1}{2\pi}\cdot|\sin(2c)|^{-\frac12-\frac{\lambda}{2}}\cdot\int_{S^1}|\sin(\theta''-c)|^{-\frac12-\tau+\frac{\lambda}{2}}|\sin(\theta''+c)|^{-\frac12+\tau+\frac{\lambda}{2}}d\theta'' \\ &= |\sin(2c)|^{-\frac12-\frac{\lambda}{2}}K_{\lambda,\tau}(c)\ , \end{aligned}$$

where the kernel $K_{-\tau,\tau,\lambda}$ is as in (4.5) and we denoted by $K_{\lambda,\tau}$ the integral

$$K_{\lambda,\tau}(c) = \frac{1}{2\pi}\cdot\int_{S^1}|\sin(t-c)|^{-\frac12-\tau+\frac{\lambda}{2}}|\sin(t+c)|^{-\frac12+\tau+\frac{\lambda}{2}}dt\ . \tag{A.1}$$

The kernel $K_{\lambda,\tau}$ is not given by an elementary function. We obtain an asymptotic formula for $K_{\lambda,\tau}$ by applying the stationary phase method to the integral (A.1). The asymptotic formula we obtain is valid for a *fixed* τ and is uniform for $\lambda \in i\mathbb{R}$ and $c \neq 0,\ \pm\pi/2,\ \pi$. We denote the set of exceptional points by $S = \{0,\ \pm\pi/2,\ \pi\} \subset S^1$. These points are singular because the integrand in (A.1) degenerates for these values of c. We have the following claim.

Claim. There are constants A, B and C such that for all $\lambda \in i\mathbb{R}$ and $c \notin S$,

$$K_{\lambda,\tau}(c) = m_\lambda(c) + m_\lambda(c+\pi/2) + r_\tau(\lambda,c)\ , \tag{A.2}$$

where the main term $m_\lambda(c)$ is a smooth function of λ and c ($c \notin S$) and for $|\lambda| \geq 1$ it is given by

$$m_\lambda(c) = |\lambda|^{-\frac12}\left(A + B|\lambda|^{-1} + C|\lambda|^{-1}\cos^2(c)\right)\cdot|\sin(c)|^\lambda\ . \tag{A.3}$$

The remainder $r_\tau(\lambda,c)$ satisfies the estimate

$$|r_\tau(\lambda,c)| = O\left((1+|\lambda|)^{-5/2} + [1+|\ln(|\sin(c)\cos(c)|)|]\cdot(1+|\lambda|)^{-10}\right) \tag{A.4}$$

with the implied constant in the O-term depending on τ only.

Hence we have

$$\begin{aligned} k_\lambda(c) &= |\sin(2c)|^{-\frac12-\frac{\lambda}{2}}K_{\lambda,\tau}(c) \\ &= M_\lambda(c) + M_\lambda(c+\pi/2) + |\sin(2c)|^{-\frac12-\frac{\lambda}{2}}r_\tau(\lambda,c)\ , \end{aligned} \tag{A.5}$$

with $M_\lambda(c) = |\lambda|^{-\frac12}\left[A + B|\lambda|^{-1} + C|\lambda|^{-1}\cos^2(c)\right]\cdot|\sin(2c)|^{-\frac12}|\sin(c)|^{\frac{\lambda}{2}}|\cos(c)|^{-\frac{\lambda}{2}}$.

A.2. **Proof.** The asymptotic expansion in the claim follows from the stationary phase method applied to the integral (A.1). We consider the asymptotic expansion consisting of two terms and a remainder. Since all functions are π-periodic, we consider only the interval $c \in [0, \pi]$. For $c \notin S$, the phase of the oscillating kernel in the integral (A.1) has two non-degenerate critical points at $t = 0$ and $t = \pi/2$. Hence, the asymptotic expansion is given by the sum of two terms. It turns out that these terms have the form $m_\lambda(c)$ and $m_\lambda(c + \pi/2)$ for the same function m_λ. Singularities of the amplitude at $t = c,\ \pi - c$ are responsible for the logarithmic term in the remainder. For $|\lambda| \to \infty$, this contribution from the singularities of the amplitude is of order $O((1 + |\lambda|)^{-k})$ for any $k > 0$ due to the fast oscillation of the phase at the same points. In fact, in [BR4] we gave a self-contained treatment of such (and more complicated) integrals based on the reduction to standard integrals and the use of the van der Corput lemma. Here we show how one can deduce necessary bounds from the stationary phase method.

Our computations are based on the following well-known form of the two-term asymptotic in the stationary phase method (see [Bo], [F], [St]). We also use the estimation of the corresponding remainder.

Let ϕ and f be smooth real valued functions on S^1. To state the stationary phase formula, we assume that ϕ has a unique non-degenerate critical point $t_0 \in S^1$. We consider the integral $I(\lambda) = \int_{S^1} f(t)e^{\lambda\phi(t)}dt$ for $\lambda \in i\mathbb{R}$. For $|\lambda| \geq 1$, we have the expansion

$$I(\lambda) = |\lambda|^{-\frac{1}{2}}(C_0 + C_1|\lambda|^{-1})e^{\lambda\phi(t_0)} + r(\lambda)\ , \tag{A.6}$$

where $C_0 = (2\pi)^{\frac{1}{2}} e^{i\cdot\mathrm{sign}(\phi''(t_0))\pi/4}|\phi''(t_0)|^{-\frac{1}{2}} f(t_0)$ and

$$C_1 = (\pi/2)^{\frac{1}{2}}\ e^{3i\cdot\mathrm{sign}(\phi''(t_0))\pi/4}\ |\phi''(t_0)|^{-\frac{3}{2}} \cdot[f'' - \phi^{(3)}f'/\phi'' - \phi^{(4)}f/4\phi'' + 5(\phi^{(3)})^2 f/12(\phi'')^2]_{t=t_0}\ ,$$

and the remainder satisfies $r(\lambda) = O((1 + |\lambda|)^{-5/2})$. The constant in the O-term is bounded for ϕ and f in a bounded, with respect to natural semi-norms, set in $C^\infty(S^1)$. If ϕ has a number of isolated non-degenerate critical points, then the asymptotic is given by the sum over these points of the corresponding contributions.

For $|\lambda| < 1$, we have the trivial bound: $|I(\lambda)| \leq \int |f|\ dt$.

A.2.1. *Leading terms.* We apply these formulas to compute leading terms in the asymptotic expansion of the integral (A.1). We set

$$\phi(t) = \ln|\sin(t - c)| + \ln|\sin(t + c)| \ \text{ and } \ f(t) = |\sin(t - c)|^{-\frac{1}{2}-\tau}|\sin(t + c)|^{-\frac{1}{2}+\tau}\ .$$

We have $\phi'(t) = \sin(2t)/\sin(t - c)\sin(t + c)$ and hence the phase ϕ has two critical points $t = 0$ and $t = \pi/2$, assuming that $c \neq 0,\ \pi/2,\ \pi$.

A straightforward computation gives for $t = 0$,

$$\phi''(0) = -2\sin^{-2}(c),\ \phi^{(3)}(0) = 0,\ \phi^{(4)}(0) = -4(1 + 2\cos^2(c))/\sin^4(c)$$

and $f(0) = |\sin(c)|^{-1},\ f''(0) = |\sin(c)|^{-3}(1+4\tau^2\cos^2(c))$, and similarly for $t = \pi/2$,

$$\phi''(\pi/2) = -2\cos^{-2}(c),\ \phi^{(3)}(\pi/2) = 0,\ \phi^{(4)}(\pi/2) = -4(1 + 2\sin^2(c))/\cos^4(c)$$

and $f(\pi/2) = |\cos(c)|^{-1},\ f''(\pi/2) = |\cos(c)|^{-3}(1 + 4\tau^2\sin^2(c))$.

Plugging this into (A.6), we see that for $c \neq 0,\ \pm\pi/2,\ \pi$,

$$K_{\lambda,\tau}(c) = m_\lambda(c) + m_\lambda(c + \pi/2) + r(\lambda, c)\ , \tag{A.7}$$

where $m_\lambda(c) = |\lambda|^{-\frac{1}{2}}\left(A + B|\lambda|^{-1} + C|\lambda|^{-1}\cos^2(c)\right)\cdot|\sin(c)|^\lambda$ with some explicit constants $A,\ B,\ C$.

A.2.2. *The remainder.* We need to estimate the remainder $r(\lambda, c) = r_\tau(\lambda, c)$ as c keeps away from the singular set $S \subset S^1$. We claim that

$$|r_\tau(\lambda, c)| = O\left((1+|\lambda|)^{-5/2} + [1 + |\ln(|\sin(c)\cos(c)|)|]\cdot(1+|\lambda|)^{-10}\right) .$$

To see this, we first note that the contribution to the remainder coming from integration over any *fixed* interval which does not include singularities of the phase and of the amplitude (i.e., points $t = \pm c,\ \pi \pm c$) is of order $O((1+|\lambda|)^{-5/2})$ (with the constant in the O-term depending on the proximity of the interval to these singular points and on τ).

The analysis of the integral is identical for all points in S; hence we treat only the case of $0 < c \le \frac{\pi}{4}$.

We use the appropriate partition of unity in order to separate different behavior of the kernel. Let $i_0, i_c, i_{-c}, i_{\pi/2}, i_{-\pi/2}, i_{\pi-c}, i_{\pi+c}, i_\pi \subset S^1 = \mathbb{R}/2\pi\mathbb{Z}$ be the following collection of closed *overlapping* intervals: $i_0 = [-3c/4, 3c/4],\ i_c = [c/4, \pi/3],\ i_{-c} = [-\pi/3, -c/4],\ i_\pi = \pi + i_0,\ i_{\pi-c} = \pi + i_{-c},\ i_{\pi+c} = \pi + i_c$ and similarly $i_{\pi/2} = [\pi/4, 3\pi/4],\ i_{-\pi/2} = [-\pi/4, -3\pi/4]$. Let $1 = \chi_0 + \chi_c + \chi_{-c} + \chi_{\pi/2} + \chi_{-\pi/2} + \chi_{\pi-c} + \chi_{\pi+c} + \chi_\pi$ be the corresponding partition of the unity on S^1 separating singular points $\pm c,\ \pi \pm c$, from the stationary points $0,\ \pi,\ \pm\pi/2$ (i.e., $supp(\chi_0) \subset i_0$, etc.). For χ_i as above, we denote by $I^i_{\lambda,\tau}$ the corresponding integral

$$I^i_{\lambda,\tau}(c) = \frac{1}{2\pi}\cdot\int_{S^1}|\sin(t-c)|^{-\frac{1}{2}-\tau+\frac{\lambda}{2}}|\sin(t+c)|^{-\frac{1}{2}+\tau+\frac{\lambda}{2}}\chi_i(t)dt .$$

We have $K_{\lambda,\tau}(c) = \sum I^i_{\lambda,\tau}(c)$. Due to symmetry, it is enough to deal with the integrals $I^0_{\lambda,\tau}, I^c_{\lambda,\tau}, I^{\pi/2}_{\lambda,\tau}$.

The integral $I^{\pi/2}_{\lambda,\tau}(c)$ falls under the standard stationary phase method and hence the remainder in the two-term asymptotic is of order $O([1+|\lambda|]^{-5/2})$ (with the constant in the O-term depending on τ). In fact the behavior of the integral $I^0_{\lambda,\tau}(c)$ is similar. This could be seen easily by scaling-up the variable t by c. In particular, integrals $I^{\pm\pi/2}_{\lambda,\tau}(c)$ and $I^{0,\pi}_{\lambda,\tau}(c)$ give rise to the leading terms in the asymptotic in the claim in Section A.1 and the remainder which is of the order $O([1+|\lambda|]^{-5/2})$.

The behavior of the remaining integral $I^c_{\lambda,\tau}$ is similar to the well-known Beta type integral of the form $B(\lambda,\chi) = \int|x+1|^{-\frac{1}{2}-\tau+\frac{\lambda}{2}}|x-1|^{-\frac{1}{2}+\tau+\frac{\lambda}{2}}\chi(x)\,dx$ for a compactly supported smooth function χ, vanishing in a *neighborhood* of the stationary point (i.e., near $x = 0$). (In fact, in [BR4] we showed how to reduce the integral $I^c_{\lambda,\tau}$ to the integral $B(\lambda,\chi)$ using an appropriate change of variable.) It follows from integration by parts that for such χ the integral $B(\lambda,\chi)$ is of the order $O([1+|\lambda|]^{-k})$ for any $k > 0$. A similar analysis is applicable to the integral $I^c_{\lambda,\tau}$.

For $|\lambda| \le 1$, the integral $I^c_{\lambda,\tau}$ is trivially of the order $O(|\ln(|\sin(c)\cos(c)|)|)$.

For $|\lambda| > 1$ and small c, we consider the two integrals

$$J^-_{\lambda,\tau}(c) = \frac{1}{2\pi}\cdot\int_{c/4}^{c}|\sin(t-c)|^{-\frac{1}{2}-\tau+\frac{\lambda}{2}}|\sin(t+c)|^{-\frac{1}{2}+\tau+\frac{\lambda}{2}}\chi_c(t)dt ,$$

$$J^+_{\lambda,\tau}(c) = \frac{1}{2\pi}\cdot\int_{c}^{\pi/3}|\sin(t-c)|^{-\frac{1}{2}-\tau+\frac{\lambda}{2}}|\sin(t+c)|^{-\frac{1}{2}+\tau+\frac{\lambda}{2}}\chi_c(t)dt .$$

Scaling-up the variable t by c, we see that the integral $J^-_{\lambda,\tau}(c)$ transforms into an integral of the form $\int_{\frac{1}{4}}^1 |f(x)|^{-\frac{1}{2}+\lambda/2}\psi(x)$, where $f(x) = f_c(x)$ is a monotone smooth function with derivatives $f^{(n)}$ bounded for $x \in [1/4, 1]$, uniformly in c, and satisfying $f(1) = 0$ and $f'(x) \geq 1$ for $x \in [1/4, 1]$. The function ψ is a smooth function also with, uniformly in c, bounded derivatives (depending on τ). Hence integration by parts implies that the integral $J^-_{\lambda,\tau}(c)$ is of the order $O([1 + |\lambda|]^{-k})$ for any $k > 0$ with the constant in the O-term independent of c.

Scaling-up the variable t by c, we see that the integral $J^+_{\lambda,\tau}(c)$ transforms into an integral of the form $\int_1^{\frac{\pi}{3c}} |(x-1)\cdot g(x)|^{-\frac{1}{2}+\lambda/2}\psi(x)$, where g is a monotone smooth function with, uniformly in c, bounded derivatives, satisfying $g(1) = 1$ and $g'(x) \geq 1$ for $x \in [1, \pi/3c]$, and ψ is a smooth function with, uniformly in c, bounded derivatives (depending on τ). To estimate such an integral, one breaks the interval $[1, \pi/3c]$ into $|\ln(c)|$ dyadic intervals. On each such interval we have the bound as above of the order $O([1 + |\lambda|]^{-k})$ for any $k > 0$ with the constant in the O-term independent of c. Hence the integral $J^+_{\lambda,\tau}(c)$ is of the order $O(|\ln(c)| \cdot [1 + |\lambda|]^{-k})$ for any $k > 0$. $\square$

A.3. **Proof of the lemma in Section 4.6.** We have to analyze the integral transform given by $u^\sharp_{N,T}(\lambda) = \int u_{N,T}(c)k_\lambda(c)dc$, where $u_{N,T}(c) = T^2 \cdot e^{-iNc} \cdot (\psi_T * \psi'_T)(c)$ with the parameters $N > T \geq 1$, $N \in \mathbb{Z}$, $T \in \mathbb{R}$. Here $\psi \in C^\infty(S^1)$ is a fixed smooth function of support in a small interval containing $1 \in S^1$, and $\psi_T(c) = \psi(T \cdot c)$, $\psi'_T(c) = \bar{\psi}_T(-c)$.

To analyze the asymptotic of $u^\sharp_{N,T}(\lambda)$, we consider the model integral

$$I(\lambda, N, T) = T \int e^{-iNc} |\sin(2c)|^{-\frac{1}{2}} |\sin(c)|^{\frac{\lambda}{2}} |\cos(c)|^{-\frac{\lambda}{2}} \chi(Tc)\ dc\ ,$$

where χ is a smooth function with $supp(\chi) \subset [-1, 1]$.

On the basis of the asymptotic expansion (A.5) for the kernel k_λ, we see that $u^\sharp_{N,T}(\lambda)$ is of the order $I(\lambda, N, T) \cdot (1 + |\lambda|)^{-\frac{1}{2}} + O(T(1 + |\lambda|)^{-5/2})$.

We claim that for $|\lambda| \leq N/T$, we have $|I(\lambda, N, T)| = O(TN^{-\frac{1}{2}})$ and for $|\lambda| > N/T$, we have $|I(\lambda, N, T)| = O(|\lambda|^{-k})$ for any $k > 0$. These bounds imply the claim in the lemma in Section 4.6.

To obtain desired bounds for $I(\lambda, N, T)$, we appeal again to the stationary phase method.

Namely, scaling-up by T the variable c in the integral $I(\lambda, N, T)$, we arrive at the integral

$$I_1(\lambda, N, T) = \int e^{-i\frac{N}{T}t} |\sin(\tfrac{2}{T}t)|^{-\frac{1}{2}} |\tan(\tfrac{t}{T})|^{\frac{\lambda}{2}} \chi(t)dt\ .$$

It is easy to see that for $|\lambda| \leq 1$, this integral is of the same order as the integral $T^{\frac{1}{2}} \int |t|^{-\frac{1}{2}+\frac{\lambda}{2}} e^{-i\frac{N}{T}t}\chi(t)dt$, which is of the order $O(TN^{-\frac{1}{2}})$. For $1 < |\lambda| \leq N/T$, the phase function in the integral I_1 has unique non-degenerate critical point and the contribution from the singularities of the amplitude is negligible. Hence, arguing as in Section A.2.2, we see that the integral I_1 is of the order $O(TN^{-\frac{1}{2}})$.

In fact, both cases follow immediately from the van der Corput lemma (see [BR4] for similar bounds).

For $|\lambda| > N/T$, the phase function has no critical points and hence we have $|I_1| = O([1 + |\lambda|]^{-k})$ for any $k > 0$. $\square$

Acknowledgments

This paper is a byproduct of a joint work with Joseph Bernstein. It was written under his insistence. It is a special pleasure to thank him for numerous discussions and for his constant encouragement and support over many years. I also would like to thank Peter Sarnak for stimulating discussions and support.

References

[Be] J. Bernstein, Eisenstein series, lecture notes. Park City, Utah (2004).

[BR1] J. Bernstein and A. Reznikov, Analytic continuation of representations and estimates of automorphic forms, Ann. of Math. (2) **150** (1999), no. 1, 329–352, arXiv: math.RT/9907202. MR1715328 (2001h:11053)

[BR2] ———, Sobolev norms of automorphic functionals, Int. Math. Res. Not. **2002**, no. 40, 2155–2174. MR1930758 (2003h:11058)

[BR3] ———, Estimates of automorphic functions, Mosc. Math. J. **4** (2004), no. 1, 19–37, 310. arXiv: math.RT/0305351. MR2074982 (2005f:11097)

[BR4] ———, Subconvexity of triple L-functions, preprint, arXiv: math.NT/0608555 (2006).

[B] A. Borel, *Automorphic forms on* $\mathrm{SL}_2(\mathbf{R})$, Cambridge Univ. Press, Cambridge, 1997. MR1482800 (98j:11028)

[Bo] V. A. Borovikov, *Uniform stationary phase method*, IEE, London, 1994. MR1325462 (96b:58111)

[Bu] D. Bump, The Rankin-Selberg method: an introduction and survey, in *Automorphic representations, L-functions and applications: progress and prospects*, 41–73, de Gruyter, Berlin. MR2192819 (2006k:11097)

[BM] R. Bruggeman, Y. Motohashi, A new approach to the spectral theory of the fourth moment of the Riemann zeta-function. J. Reine Angew. Math. 579 (2005), 75–114. MR2124019 (2005i:11108)

[F] M. Fedoruk, Asymptotic methods in analysis. *Analysis. I*, Encyclopaedia Math. Sci., 13, Springer, Berlin, 1989. MR1042759 (90j:00027)

[G1] I. M. Gelfand, M. I. Graev and N. Ya. Vilenkin, *Generalized functions. Vol. 5*, Academic Press, New York, 1966. MR0435835 (55:8786e)

[G2] I. M. Gelfand, M. I. Graev and I. I. Pyatetskii-Shapiro, *Representation theory and automorphic functions*, Saunders, Philadelphia, 1969. Reprint of the 1969 edition. Generalized Functions, 6. Academic Press, Inc., Boston, MA, 1990. MR1071179 (91g:11052)

[Go] A. Good, Cusp forms and eigenfunctions of the Laplacian, Math. Ann. **255** (1981), no. 4, 523–548. MR618183 (82i:10029)

[Gr] B. H. Gross, Some applications of Gelfand pairs to number theory, Bull. Amer. Math. Soc. (N.S.) **24** (1991), no. 2, 277–301. MR1074028 (91i:11055)

[He] S. Helgason, *Groups and geometric analysis*, Amer. Math. Soc., Providence, RI, 2000. MR1790156 (2001h:22001)

[Ho] L. Hörmander, *The analysis of linear partial differential operators. IV*, Springer, Berlin, 1985. MR781537 (87d:35002b)

[Iw] H. Iwaniec, *Spectral methods of automorphic forms*, Amer. Math. Soc., Providence, RI, 2002. MR1942691 (2003k:11085)

[IS] H. Iwaniec and P. Sarnak, Perspectives on the analytic theory of L-functions, Geom. Funct. Anal. **2000**, Special Volume, Part II, 705–741. MR1826269 (2002b:11117)

[JN] H. Jacquet and N. Chen, Positivity of quadratic base change L-functions, Bull. Soc. Math. France **129** (2001), no. 1, 33–90. MR1871978 (2003b:11048)

[KSa] H. Kim, P. Sarnak, Appendix to: H. H. Kim, Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2, J. Amer. Math. Soc. **16** (2003), no. 1, 139–183. MR1937203 (2003k:11083)

[KS] B. Krötz and R. J. Stanton, Holomorphic extensions of representations. I. Automorphic functions, Ann. of Math. (2) **159** (2004), no. 2, 641–724. MR2081437 (2005f:22018)

[Ku] T. Kubota, *Elementary theory of Eisenstein series*, Kodansha, Tokyo, 1973. MR0429749 (55:2759)

[Kz] N. V. Kuznetsov, Sums of Kloosterman sums and the eighth power moment of the Riemann zeta-function, in *Number theory and related topics (Bombay, 1988)*, 57–117, Tata Inst. Fund. Res., Bombay. MR1441327 (98c:11085)

[L] J. B. Lewis, Eigenfunctions on symmetric spaces with distribution-valued boundary forms, J. Funct. Anal. **29** (1978), no. 3, 287–307. MR512246 (80f:43020)

[M] H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, Math. Ann. **121** (1949), 141–183. MR0031519 (11:163c)

[MW] K. Martin and D. Whitehouse, Central L-values and toric periods for $GL(2)$, preprint (2007).

[Mo1] Y. Motohashi, *Spectral theory of the Riemann zeta-function*, Cambridge Univ. Press, Cambridge, 1997. MR1489236 (99f:11109)

[Mo2] ———, A note on the meanvalue of the zeta and L-functions, preprint, arXiv: math.NT/0401085.

[O] A. I. Oksak, Trilinear Lorentz invariant forms, Comm. Math. Phys. **29** (1973), 189–217. MR0340478 (49:5231)

[PS] Y. N. Petridis and P. Sarnak, Quantum unique ergodicity for $\mathrm{SL}_2(\mathcal{O})\backslash\mathbf{H}^3$ and estimates for L-functions, J. Evol. Equ. **1** (2001), no. 3, 277–290. MR1861223 (2003a:11060)

[Pr] D. Prasad, Trilinear forms for representations of GL(2) and local ϵ-factors, Compositio Math. **75** (1990), no. 1, 1–46. MR1059954 (91i:22023)

[Ra] R. A. Rankin, Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. II. The order of the Fourier coefficients of integral modular forms, Proc. Cambridge Philos. Soc. **35** (1939), 357–372. MR0000411 (1:69d)

[R] A. Reznikov, Norms of geodesic restrictions for eigenfunctions on hyperbolic surfaces and representation theory, preprint (2004). arXiv: math.AP/0403437.

[Sa] P. Sarnak, Integrals of products of eigenfunctions, Internat. Math. Res. Notices **1994**, no. 6, 251–261. MR1277052 (95i:11039)

[Se] A. Selberg, On the estimation of Fourier coefficients of modular forms, in *Proc. Sympos. Pure Math., Vol. VIII*, 1–15, Amer. Math. Soc., Providence, R.I. MR0182610 (32:93)

[St] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Univ. Press, Princeton, NJ, 1993. MR1232192 (95c:42002)

[V] A. Venkatesh, Sparse equidistribution problems, period bounds, and subconvexity, preprint (2005). arXiv: math.NT/0506224.

[Wa] J.-L. Waldspurger, Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie, Compositio Math. **54** (1985), no. 2, 173–242. MR783511 (87g:11061b)

[W] T. Watson, Thesis, Princeton, 2001.

Department of Mathematics, Bar-Ilan University, 52900 Ramat-Gan, Israel
E-mail address: reznikov@math.biu.ac.il