CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

On the local L^2-Bound of the Eisenstein series

Subhajit Jana [ID] [1] and Amitay Kamber [ID] [2]

[1]School of Mathematical Sciences, Queen Mary University of London, Mile End Rd, London, E1 4NS, United Kingdom;
E-mail: s.jana@qmul.ac.uk (corresponding author).
[2]Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom;
E-mail: ak2356@dpmms.cam.ac.uk.

Abstract

We study the growth of the local L^2-norms of the unitary Eisenstein series for reductive groups over number fields, in terms of their parameters. We derive a *poly-logarithmic* bound on an average, for a large class of reductive groups. The method is based on Arthur's development of the spectral side of the trace formula, and ideas of Finis, Lapid and Müller.

As applications of our method, we prove the optimal lifting property for $\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ for square-free q, as well as the Sarnak–Xue [52] counting property for the principal congruence subgroup of $\mathrm{SL}_n(\mathbb{Z})$ of square-free level. This makes the recent results of Assing–Blomer [8] unconditional.

Contents

1. Introduction

Eisenstein series are ubiquitous in the theory of automorphic forms, in particular, they are crucially used by Langlands, and later by Arthur, in their seminal works on the development of the spectral decomposition of the space of automorphic forms of a reductive group and the trace formula (see [6, 36, 37, 38]. Many important questions in number theory and automorphic forms rely directly upon certain analytic properties of the Eisenstein series, for example, their complex analytic properties including the location and order of their poles and their various growth properties near the cusps.

The Eisenstein series, in contrast with *cusp forms*, do not decay near the cusps; in fact, they have polynomial growth. The rate of growth depends precisely on the constant terms of the Eisenstein series. In particular, the unitary Eisenstein series (barely!) fail to be square-integrable on the noncompact fundamental domain. However, it is quite natural to wonder how the L^2-norms of unitary Eisenstein series grow *locally* (i.e., on a fixed compact subset of the fundamental domain), in terms of their Archimedean and non-Archimedean parameters.

In this paper, we study the general local L^2-bound of a unitary Eisenstein series on a reductive group over a number field. We refer to [38, Section 5.3], where such questions are considered and the relations with other important automorphic and analytic questions, for example, location and order of the poles of the Eisenstein series and classification of the residual spectrum, are discussed. Informally, we prove *sharp* upper bounds for the averages of the squared local L^2-norms of the Eisenstein series over certain *short* families, for a large class of reductive algebraic groups; see Theorems 2 and 3. We refer to Section 1.1 for precise and detailed discussions of the problems and our results.

Our estimates of the local L^2-norms of the Eisenstein series are potentially useful to answer many interesting number-theoretic questions, in particular, the problems that are approached via the spectral theory of various *noncompact* arithmetic symmetric spaces. Such approaches often face serious analytic difficulties due to the existence of the continuous spectrum and require estimates for various local L^p-norms of the Eisenstein series. Our estimates and the methods to prove them seem to be utilizable to overcome the analytic hurdles that arise via the local L^2-norms of the Eisenstein series, at least in certain cases.

As a *proof-of-concept*, we prove the *optimal lifting property* for $SL_n(\mathbb{Z}/q\mathbb{Z})$ in Theorem 5 for square-free q, as well as the *Sarnak–Xue counting property* in Theorem 4, for the principal congruence subgroups of $SL_n(\mathbb{Z})$ of square-free level. Assing–Blomer recently proved those theorems *conditional* on the local L^2-bounds of the Eisenstein series (see [8]). Thus, we make their results unconditional. We refer to Section 1.3 for a detailed discussion. In a sequel [32], we prove *optimal Diophantine exponent* for $SL_n(\mathbb{Z}[1/p])$, as expected in a series of influential works by Ghosh–Gorodnik–Nevo (e.g. see [26]),

assuming *Sarnak's density hypothesis*. The proof of the main results in [32] crucially depends on (a slightly weaker version of) Theorem 1 below; see [32, Section 4.8]. We refer the reader to [32] for a detailed discussion of this problem and our proof.

Let us now discuss the motivating questions and our results.

1.1. Local L^2-bound

Let G be a reductive algebraic group over a number field F. Let $\mathbb{A}$ be the adele ring of F. Also let $\mathbb{A}_f$ be the finite adeles and $F_\infty := F \otimes \mathbb{R}$. Given a Levi subgroup M of G, let $\mathfrak{a}_M^*$ be the $\mathbb{R}$-vector space spanned by F-rational characters of M, and $(\mathfrak{a}_M^G)^*$ the $\mathbb{R}$-vector space spanned by characters that are trivial on G. Let $H_M : M(\mathbb{A}) \to \mathfrak{a}_M$ the natural map, and $M(\mathbb{A})^1$ the kernel of H_M. We denote by M_0 a minimal Levi subgroup.

Let P be a parabolic subgroup of G with Levi subgroup M. We denote $\mathfrak{a}_P = \mathfrak{a}_M$. Let φ be in $\mathcal{A}^2(P)$, which is the pre-Hilbert space of square-integrable automorphic forms on $N(\mathbb{A})M(F)\backslash G(\mathbb{A})$. Let λ be an element of $(\mathfrak{a}_{M,\mathbb{C}}^G)^* := (\mathfrak{a}_M^G)^* \otimes_\mathbb{R} \mathbb{C}$. Following Langlands in [37] (also see [6, Chapter 7]), we construct an Eisenstein series $\mathrm{Eis}(\varphi, \lambda)$ attached to φ and λ (see the details of the construction in Section 3.3).

Let φ be a *unit* vector (see Section 3.1 for the definition of the norm on $\mathcal{A}^2(P)$) with $\nu(\varphi)$ being a measurement of the complexity of φ. We will later use a more nuanced notion of complexity (both Archimedean and non-Archimedean) (see Section 4.1 for precise definitions). Finally, let $\lambda \in i(\mathfrak{a}_M^G)^*$ be purely imaginary, so that $\mathrm{Eis}(\varphi, \lambda)$ appears in the spectral decomposition of $L^2(G(F)\backslash G(\mathbb{A})^1)$. Here and elsewhere in the paper, $d\lambda$ will denote the real-valued Lebesgue measure on $i(\mathfrak{a}_M^G)^*$ as in, for example, [6].

We ask the following two questions with (presumably, strictly) decreasing levels of difficulty. Let Ω be a fixed compact set in a fundamental domain of $G(F)\backslash G(\mathbb{A})^1$. How do

$$\int_\Omega |\mathrm{Eis}(\varphi, \lambda)(g)|^2 \, dg \tag{1.1}$$

and

$$\int_{\|\lambda' - \lambda\| \le 1} \int_\Omega |\mathrm{Eis}(\varphi, \lambda')(g)|^2 \, dg \, d\lambda' \tag{1.2}$$

grow as $\nu(\varphi) + \|\lambda\| \to \infty$?

Here is our general expectation regarding the estimates of the above integrals.

$$\text{The integrals (1.1) and (1.2)} \ll_\Omega (\log(1 + \nu(\varphi) + \|\lambda\|))^{\dim \mathfrak{a}_M^G}, \tag{1.3}$$

as $\nu(\varphi) + \|\lambda\| \to \infty$.

Let φ be spherical at all places, and $\nu(\varphi)$ denote the size of the Laplace eigenvalue of φ. If $G = \mathrm{GL}_2$ and $\mathrm{Eis}(\varphi, \lambda)$ is the nonholomorphic spherical Eisenstein series, then it is a classical result that (1.1) indeed satisfies (1.3) (see Section 1.4). For $G = \mathrm{GL}_3$, Miller in [43] proved that (1.2) satisfies (1.3) (see Section 1.4 for more discussions). For $G = \mathrm{GL}_n$, Zhang in [56] showed that if $\mathrm{Eis}(\varphi, \lambda)$ is the maximal parabolic *degenerate* Eisenstein series, that is induced from the trivial representation of the Levi subgroup $\mathrm{GL}_{n-1} \times \mathrm{GL}_1$, then (1.1) satisfies (1.3)[1]. As far as we know, these are the only examples where a poly-logarithmic strength bound is known for all the Eisenstein series for either (1.1) or (1.2), at least for higher rank groups.

---

[1]In fact, he proved an asymptotic formula in λ in this case.

Remark 1.1. In this remark, we comment on a few topics in the existing literature that have similar flavour as the local L^2-bound question in this paper. One may wonder about the size of the analogous local (and global, for cuspidal forms) L^p-bound of automorphic forms for any $1 \le p \le \infty$. For compact manifolds, general polynomial bounds in $\nu(\varphi)$ were obtained by Sogge [54], and there have been many improvements in various aspects. On arithmetic locally symmetric spaces local L^∞-norms for cuspidal automorphic forms have been studied quite a lot, in particular, quite extensively for small rank groups. We refer to [13] for the case of $\mathrm{GL}_n(\mathbb{R})$, to [42] for general semisimple groups and the references therein. The proofs there are via the pretrace formula, which can also be made to work to prove similar bounds for the local L^∞-bounds of the short averages of the square of the Eisenstein series in case of noncompact spaces. We also remark that local L^∞-bound of general Eisenstein series may not satisfy (1.3) (see, e.g. [11, Theorem 2]). On the other hand, the local L^∞ bounds of the cusp forms (similarly, of an average of Eisenstein series) on GL_n as in [13] often do not hold globally (i.e. uniformly near the cusp) (see [14]).

1.2. Main results

In this section, we state our main results. Informally, we prove that, for a large class of groups, (1.2) indeed can be estimated *poly-logarithmically* in $\nu(\varphi) + \|\lambda\|$; as expected in (1.3). Before stating the most general result in our paper, we first give the *state-of-the-art* result for GL_n that is possible to state in classical language and without detailed notations.

Theorem 1. *Let $G = \mathrm{GL}_n$ and P be a parabolic subgroup of G. Let $\varphi_0 \in \mathcal{A}^2(P)$ be a unit that is spherical at all places. Moreover, suppose that φ_0 is an eigenfunction of the full Hecke algebra at the non-Archimedean places and the ring of invariant differential operators in the centre of the universal enveloping algebra at the Archimedean places. Then, for any $\lambda_0 \in i(\mathfrak{a}_P^G)^*$, we have*

$$\int_{\substack{\lambda \in i(\mathfrak{a}_P^G)^* \\ \|\lambda - \lambda_0\| \le 1}} \int_\Omega |\mathrm{Eis}(\varphi_0, \lambda)(x)|^2 \, \mathrm{d}x \, \mathrm{d}\lambda \ll_\Omega (\log(1 + \nu(\varphi_0) + \|\lambda_0\|))^{n-1},$$

where $\nu(\varphi_0)$ is the size of the Laplace eigenvalue of φ_0.

Remark 1.2. Actually, Theorem 1 is true for any φ that lies in the π-isotypic subspace $\mathcal{A}_\pi^2(P)$ for some representation π of $M(\mathbb{A})$ and is only required to be spherical at the Archimedean places and invariant by any $K \subset \mathrm{GL}_n(\mathbb{A}_f)$, with $\ll$ replaced by $\ll_K$. The dependency of the constant on K could, a priori, be large, for example, polynomial in level(K) (see Section 4.1 for the definition of level). One may expect a polylogarithmic dependence on the level, but we can only prove it when φ is associated with *unramified cuspidal datum*, as in Theorem 2.

Now, we describe our full results. First, we recall certain terminologies (that are fully explained in the body of the paper).

○ Following Langlands, a *cuspidal datum* χ is a $G(F)$-conjugacy class of pairs (L, σ), where L is a Levi subgroup of G defined over F and σ is a cuspidal representation of $L(\mathbb{A})$ (see Section 3.1). Moreover, we recall the notion of *unramifiedness* of a cuspidal datum, due to Arthur, which ensures that the datum has a *trivial* stabilizer (see [6, Section 15][2]).

○ We discuss certain estimates of various intertwining operators. The following terminologies are due to Finis–Lapid–Müller.

    – We say G satisfies *property (tempered winding number, TWN+)* if the first derivatives of the global normalization factors of the intertwining operators satisfy certain estimates (see Section 5.1).

    – We say G satisfies *property (bounded degree, BD)* if the first derivatives of the local normalized intertwining operators satisfy certain estimates (see Section 5.2).

---

[2]The notion should not be confused with 'unramifiedness' of the local components of an automorphic representation.

○ For any Levi subgroup M, let $\Pi_2(M)$ denote the isomorphism class of representations of $M(\mathbb{A})$ appearing discretely in $L^2(M(F)\backslash M(\mathbb{A}))$ (see Section 3.1).

○ For a parabolic subgroup P with Levi subgroup M, χ a cuspidal datum and $\pi \in \Pi_2(M)$, let $\mathcal{A}^2(P)_{\chi,\pi}$ denote the (χ, π)-isotypic part of $\mathcal{A}^2(P)$ (see Section 3.1).

○ We fix (good) maximal compact subgroups K_∞ of $G(F_\infty)$ and K_f of $G(\mathbb{A}_f)$. For $\tau \in \widehat{K_\infty}$, which is the unitary dual of K_∞, we write $\mathcal{A}^2(P)^{\tau,K}$ for the space of functions in $\mathcal{A}^2(P)$ which are right K-invariant and belong to τ-isotypic subspace. We define $\mathcal{A}^2_{\chi,\pi}(P)^{\tau,K}$ similarly (see Section 3.1).

○ We bound the integrals over Ω in (1.1) and (1.2) via the inner product of truncated Eisenstein series. We work with a *truncation operator* Λ^T, as defined by Arthur, where T is a sufficiently dominant element in $\mathfrak{a}_{M_0}$. We also define $d(T)$ to be a measurement of how *dominant T is* (see Section 3.4 for details).

○ Finally, we fix certain measurements of the Archimedean complexity of a representation π of a real group by the size of its *Casimir eigenvalue*, denoted by $\nu(\pi)$ and certain measurement of the non-Archimedean complexity of an open compact subgroup K of $G(\mathbb{A}_f)$ by its *depth and level*, denoted by $\text{level}(K)$ (see Section 4.1). To give a sense of their sizes for $G = \text{GL}_2$, if π is principal series representations induced from $|.|^{it} \otimes |.|^{-it}$, then $\nu(\pi) \asymp 1 + t^2$, and if K is the principal congruence subgroup of level q, then $\text{level}(K) = q$.

Here is our first main theorem.

Theorem 2. *Assume that G satisfies properties (TWN+) and (BD). Let P be a parabolic subgroup of G with Levi subgroup M and $\pi_0 \in \Pi_2(M)$. Moreover, assume that $\chi := (M, \pi_0)$ is an unramified cuspidal datum of G. Let $\tau \in \widehat{K_\infty}$ and $K \subset K_f$ be an open-compact. Then*

$$\int_{\substack{\lambda \in i(\mathfrak{a}_M^G)^* \\ \|\lambda - \lambda_0\| \leq 1}} \|\Lambda^T \text{Eis}(\varphi_0, \lambda)\|_2^2 \, d\lambda \ll \left((1 + \|T\|) \, \log\left(1 + \nu(\pi_{0,\infty}) + \|\lambda_0\| + \nu(\tau) + \text{level}(K)\right) \right)^{\dim \mathfrak{a}_M^G},$$

for any unit $\varphi_0 \in \mathcal{A}^2_{\chi,\pi_0}(P)^{\tau,K}$ and $\lambda_0 \in i(\mathfrak{a}_M^G)^*$.

Remark 1.3. The implicit constant in Theorem 2 depends only on G. For G and χ as in the statement, Theorem 2 proves that the integral (1.2) grows poly-logarithmically in all the parameters, verifying (1.3). For GL_n, when Theorem 2 applies, it is more general than Theorem 1 in the level and weight aspect, since it gives poly-logarithmic dependence on the level and is not restricted to K_∞-invariant vectors.

Of course, Theorem 2 excludes a 'narrow' class of cuspidal data that are *not* unramified. For example, for $G = \text{GL}_n$, a Levi subgroup is of the form $M \cong \text{GL}_{n_1} \times \cdots \times \text{GL}_{n_k}$, and a representation $\pi \in \Pi_2(M)$ satisfies $\pi \cong \pi_1 \otimes \cdots \otimes \pi_k$, where $\pi_i \in \Pi_2(\text{GL}_{n_i})$. Theorem 2 will apply to π if and only if each π_i is cuspidal, and $\pi_i \not\cong \pi_j$ for $i \neq j$.

In Theorem 3, we give a result towards the expectation in (1.3) for G with the similar level of generalities considered in Theorem 2 and τ being trivial but for *all* cuspidal data χ. Before stating the theorem precisely, we introduce a new quantity below, which roughly measures the number of discrete series representations with specific $K_\infty K_f$-type that arises from a given cuspidal datum.

Definition 1.4. Let $\tau \in \widehat{K_\infty}$ and $K \subset K_f$ be an open-compact. For any cuspidal datum χ of G, we define

$$F(\chi; \tau, K) := \sum_{P \supset P_0} \dim\left(\mathcal{A}^2_\chi(P)^{\tau,K}\right) = \sum_{P \supset P_0} \sum_{\pi \in \Pi_2(M_P)} \dim\left(\mathcal{A}^2_{\chi,\pi}(P)^{\tau,K}\right).$$

Here, P_0 is some fixed minimal parabolic subgroup containing M_0 and $P \supset P_0$ are called *standard parabolic* subgroups.

Theorem 3. *Assume that G satisfies properties (TWN+) and (BD). Let $\chi := (M, \pi_0)$ be any cuspidal datum of G. Then*

$$\sum_{P \supset P_0} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_{\chi,\pi}(P)^{K_\infty K}} \int_{\substack{\lambda \in i(\mathfrak{a}_P^G)^* \\ \|\lambda - \lambda_0\| \leq 1}} \|\Lambda^T \operatorname{Eis}(\varphi, \lambda)\|_2^2 \, d\lambda$$

$$\ll \left((1 + \|T\|) \log(1 + \nu(\pi_{0,\infty}) + \|\lambda_0\| + \operatorname{level}(K)) \right)^{\dim \mathfrak{a}_M^G} F(\chi; \operatorname{triv}, K),$$

where $\mathcal{B}_{\chi,\pi}(P)^{K_\infty K}$ denotes an orthonormal basis of $\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}$ and $\lambda_0 \in i(\mathfrak{a}_M^G)^*$.

Remark 1.5. Note that the (P, π)-summand on the left-hand side above vanishes identically unless $\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K} \neq \{0\}$. In fact, the total number of nonzero (P, π, φ)-summands on the left-hand side is finite and equals to $F(\chi; \operatorname{triv}, K)$ (the λ-integral, however, may be empty if $\lambda_0 \in i(\mathfrak{a}_M^G)^* \subset i(\mathfrak{a}_{M_0}^G)^*$ is far from $i(\mathfrak{a}_P^G)^* \subset i(\mathfrak{a}_{M_0}^G)^*$). Thus, Theorem 3 *on average* is of the same strength as Theorem 2, that is, the integral in (1.2) *on average* satisfies the estimate in (1.3).

In particular, by the positivity of the left-hand side in Theorem 3, we obtain the following corollary.

Corollary 1.6. *Assume that G and χ are as in Theorem 3. Let P be a parabolic subgroup of G and $\pi \in \Pi_2(M_P)$, such that $\mathcal{A}_{\chi,\pi}^2(P) \neq \{0\}$. Then*

$$\int_{\substack{\lambda \in i(\mathfrak{a}_P^G)^* \\ \|\lambda - \lambda_0\| \leq 1}} \|\Lambda^T \operatorname{Eis}(\varphi_0, \lambda)\|_2^2 \, d\lambda$$

$$\ll \left((1 + \|T\|) \log(1 + \nu(\pi_\infty) + \|\lambda_0\| + \operatorname{level}(K)) \right)^{\dim \mathfrak{a}_M^G} F(\chi; \operatorname{triv}, K),$$

for any unit $\varphi_0 \in \mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}$ and $\lambda_0 \in i(\mathfrak{a}_P^G)^*$.

Remark 1.7. Let us focus on GL_n. In this case, understanding $F(\chi; \operatorname{triv}, K)$ boils down to a local question, as we know for which π it holds that $\mathcal{A}_{\chi,\pi}^2 \neq \{0\}$ thanks to [44]. Then one can show that when π_0 is cuspidal it holds that

$$F(\chi; \operatorname{triv}, K) \ll \dim \mathcal{A}_{\chi,\pi_0}^2(P)^{K_\infty K}.$$

However, when π_0 is not cuspidal, $F(\chi; \operatorname{triv}, K)$ can be a lot larger than $\dim \mathcal{A}_{\chi,\pi_0}^2(P)^{K_\infty K}$, which eventually leads to a nonoptimal result, when averaging over a basis of $\mathcal{A}_{\chi,\pi_0}^2(P)^{K_\infty K}$.

For general groups, the size of $F(\chi; \tau, K)$ has two components: the local component, namely, the dimension of the $K_\infty K_f$-type subspace and the global component, namely, the number and multiplicity of square-integrable representations (with a given $K_\infty K_f \cap M(\mathbb{A})$-type) that are attached to χ.

The local component could, potentially, be quite larger than $\log(\|\nu(\tau)\| + \operatorname{level}(K))$ but by uniform admissibility (see [10]) is bounded depending only on τ and K. This nonoptimal dependence on $\operatorname{level}(K)$ even for GL_n is the main weakness of Theorem 3 relative to Theorem 2. Also, Theorem 3 is restricted to $\tau = \operatorname{triv}$, but this should not be a major issue.

The global component should, at least conjecturally, be small, and, in fact, bounded (perhaps depending on K). Its understanding depends eventually upon the classification of the residual spectrum. In other words, it depends on how the residual spectrum arises from the residues of the Eisenstein series induced from cuspidal representations, which in turn asks for highly nontrivial analytic properties of the Eisenstein series. All of the above problems are still wide open for general reductive groups, and understanding them for classical groups seems like a very interesting question that we plan to address in future work.

Remark 1.8. One may wonder whether one can obtain an estimate for the pointwise bound of the local L^2-norms in (1.1) from corresponding bounds for their λ-averages in Corollary 1.6. In principle, one may attempt this via a contour-shifting argument and employing the functional equation of the Eisenstein series. This kind of approach is prevalent in analytic number theory: for example, bounding central L-value $L(1/2 + iT)$ for some automorphic representation for large T, by sharp estimation of $\int_{T-H}^{T+H} |L(1/2 + it)|^2 \, dt$ for some $H \leq T$, shifting the it-contour and applying functional equation of the L-function (see, e.g. [28]). Clearly, to apply the above contour-shifting method in our case, we need to crucially understand meromorphic properties of the Eisenstein series *away from the unitary axis*, in particular, the location of its poles and residues. On GL_n, this is essentially equivalent to understanding *strong* zero-free regions of various Rankin–Selberg L-functions. Our current knowledge of such zero-free regions is rather limited and only yields a polynomial strength bound in the question (1.1).

1.3. Applications

While estimating the local L^2-bound of the Eisenstein series is interesting in its own right, it is also very useful in applications to number theory and Diophantine approximation. As a matter of fact, we came to study L^2-bounds of the Eisenstein series while trying to understand certain number theoretic problems via the spectral theory of automorphic forms.

We already mentioned the sequel [32], where we use the results of this work to prove *optimal Diophantine exponent* for SL_n assuming *Sarnak's density hypothesis*.

Here, we give two other applications of the L^2-bounds of the Eisenstein series, concerning the principal congruence subgroups of $\mathrm{SL}_n(\mathbb{Z})$. Let $\Gamma(1) := \mathrm{SL}_n(\mathbb{Z})$, and let $\Gamma(q) := \ker(\Gamma(1) \xrightarrow{\mathrm{mod}\ q} \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z}))$ be its principal congruence subgroup. By strong approximation, $\Gamma(1)/\Gamma(q) \cong \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$. Let $\|\cdot\|$ be the Frobenius norm on $\mathrm{SL}_n(\mathbb{R})$, that is $\|g\|^2 = \sum_{i,j} |g_{i,j}|^2$ (as a matter of fact, one can consider any other norm as well).

The first problem is about *counting* the number of elements

$$\#\{\gamma \in \Gamma(q) \mid \|\gamma\| \leq R\}.$$

There are two regimes in which this problem is well understood. In the regime $R < Cq$, where $C = C(n)$ is some small constant, it is simple to show that

$$\#\{\gamma \in \Gamma(q) \mid \|\gamma\| \leq R\} = 1.$$

In the regime where q is fixed and $R \to \infty$ (or at least a large power of q), it holds that

$$\#\{\gamma \in \Gamma(q) \mid \|\gamma\| \leq R\} = c_n \frac{R^{n(n-1)}}{\#\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})} + O(R^{n(n-1)-\delta_n}) \tag{1.4}$$

with c_n an explicit constant, and $\delta_n > 0$ a small constant. This follows from the work of Duke, Rudnick and Sarnak [16, Theorem 1.10] and is based on spectral methods[3].

Sarnak and Xue ([52, Conjecture 2], see also [53]) made the following conjecture, which addresses the intermediate range when R is a small power of q:

$$\#\{\gamma \in \Gamma(q) \mid \|\gamma\| \leq R\} \ll_\epsilon \frac{R^{n(n-1)+\epsilon}}{|\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})|} + R^{n(n-1)/2}.$$

In a recent work, Assing and Blomer [8, Theorem 1.4] essentially confirm the above Sarnak–Xue conjecture showing that

---

[3]The result is for $q = 1$, but the method works for every q. See also [30, Corollary 5.2] for a generalized statement with a simpler proof. Recently, in [12], the authors significantly improved the error term of Equation (1.4) over [16].

$$\#\{\gamma \in \Gamma(q) \mid \|\gamma\| \le R\} \ll_\epsilon (Rq)^\epsilon \left(\frac{R^{n(n-1)}}{q^{n^2-1}} + R^{n(n-1)/2} \right), \tag{1.5}$$

conditional on a difficult conjecture, namely [8, Hypothesis 1], on the local L^2-bound of the Eisenstein series. This hypothesis is slightly weaker than the expected bound in (1.3) for the integral (1.1). In Theorem 4, we make Assing–Blomer's result *unconditional*, by employing the method of proving Theorem 3, while also using their work as a crucial input.

Theorem 4. *Let q be a square-free integer. Then Equation (1.5) holds for every $\epsilon > 0$ and every $R \ge 1$.*

The second application we have is the problem of *optimal lifting*. Recall that by strong approximation for SL_n, the map $\mathrm{SL}_n(\mathbb{Z}) \to \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ is onto. We may therefore ask, given $\overline{\gamma} \in \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$, to find a lift of it to $\gamma \in \mathrm{SL}_n(\mathbb{Z})$, with $\|\gamma\|$ minimal. Gorodnik–Nevo studied this problem in a general context and proved that for every $\overline{\gamma}$, it is possible to find a lift γ with $\|\gamma\| \ll_\epsilon q^{C_n+\epsilon}$ where C_n is some explicit constant (see [29, Theorem 1.1]). One may then wonder what is the smallest C_n so that the statement holds for it. A simple lower bound is $C_n \ge 1 + 1/n$, which follows from Equation (1.4) (with $q = 1$) and the fact that $|\mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})| \gg_\epsilon q^{n^2-1-\epsilon}$, but it is not the correct answer. In [34], by the second author and Pétér Varjú (which was completed after this work), it is shown that the correct answer is $C_n = 2$ (following weaker results of [51]). More precisely, the authors showed that for every q and $\overline{\gamma} \in \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$, there is a lift $\gamma \in \mathrm{SL}_n(\mathbb{Z})$ with $\|\gamma\| \ll_n q^2 \log q$. In the other direction, they show that for every $n, \epsilon > 0$, and for every q, there is $\overline{\gamma} \in \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$, such that every lift $\gamma \in \mathrm{SL}_n(\mathbb{Z})$ of $\overline{\gamma}$ satisfies $\|\gamma\| \gg_{n,\epsilon} q^{2-\epsilon}$.

In contrast to the results of [34], Sarnak [51] observed that when one considers *almost all* the elements instead of *all* the elements, the simple lower bound $C_n = 1 + 1/n$ should be the correct answer to the problem, and proved it for $n = 2$. This property is expected to hold in a general context, and we call it *optimal lifting*. We again refer to [27] for a detailed discussion of the general problem of optimal lifting, and to [33] for this specific problem.

In the recent work of Assing and Blomer [8, Theorem 1.4], they essentially solved the optimal lifting problem (see the statement of Theorem 5) but, again, *conditional* on [8, Hypothesis 1]. Once again, in Theorem 5, employing the method of proving Theorem 3 and crucially inputting one of the main results of [8], we make the above result of Assing–Blomer's *unconditional*.

Theorem 5. *For every $\epsilon > 0$, there is $\delta > 0$, such that for every square-free integer q, the number of elements in $\overline{\gamma} \in \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ without a lift $\gamma \in \mathrm{SL}_n(\mathbb{Z})$ with $\|\gamma\| \le q^{1+1/n+\epsilon}$ is at most $O(q^{n^2-1-\delta})$.*

We now add a few remarks regarding Theorems 4 and 5.

Remark 1.9. While Theorems 4 and 5 are limited to q square-free (except for $n = 2$, where Sarnak's proof holds for all q), our method can prove them for general q assuming Sarnak's density conjecture, which is implied by the generalized Ramanujan conjecture (GRC) for cusp forms of GL_n over $\mathbb{Q}$. More generally, unconditionally and independently of (most of) the work of Assing–Blomer, one can get slightly weaker results, with the exponents multiplied by $1 + 2\delta/(n-1)$, where $0 \le \delta \le (n-1)/2$ is the best approximation for the (GRC) for GL_n over $\mathbb{Q}$. Specifically, one has $\delta \le \frac{1}{2} - \frac{1}{1+n^2}$ (see [50]).

Remark 1.10. Conjecturally, similar 'Sarnak–Xue Counting' and 'Optimal Lifting' theorems should hold more generally, in particular, for congruence subgroups of arithmetic lattices in semisimple Lie groups. We refer to [27], where those conjectures are discussed at length (for the specific problem of $\Gamma(q)$ see [27, Section 2.6]). In particular, for *uniform (cocompact) lattices* one can prove that Sarnak–Xue counting and optimal lifting follow from *Sarnak's density conjecture* in automorphic forms, which is, presumably, a true approximation of the (false) *naive Ramanujan conjecture* (see [27]). For nonuniform lattices, this implication is unknown, because of the analytic difficulties coming from the Eisenstein series. The techniques of this paper allow us to essentially overcome those analytic difficulties, for nonuniform lattices in groups different than GL_n, as long as properties (TWN+) and (BD) are known. For principal congruence subgroups of $\mathrm{SL}_n(\mathbb{Z})$ of square-free level, Assing–Blomer proved Sarnak's density conjecture, which allows in combination with this work to deduce the two theorems.

Remark 1.11. We note, that in this paper, we do not attempt to prove [8, Hypothesis 1], and it remains an interesting and hard open question. We refer to Section 1.6 for a discussion.

1.4. Difficulties in higher rank

In this section, we try to describe the difficulties one faces while trying to estimate (1.1) and (1.2).

We start with a toy-example – estimating (1.1) for a low-rank group. We take $G = \mathrm{GL}_2$ over $\mathbb{Q}$, $M = \mathrm{GL}_1 \times \mathrm{GL}_1$, P the Borel subgroup in G, $\pi = \mathrm{triv} \times \mathrm{triv}$ and $\lambda = (it, -it)$ with $t \in \mathbb{R}$. We take $\varphi \in \mathcal{A}^2(P)_\pi$ to be the $\mathrm{O}_2(\mathbb{R})\mathrm{GL}_2(\hat{\mathbb{Z}})$-invariant unit vector. In the classical language, $\mathrm{Eis}(\varphi, \lambda)$ is the same as $E(\cdot, 1/2 + it)$. The classical Maass–Selberg relations (see [43, pp.153]) yield that

$$
\|\Lambda^T E(\cdot, 1/2 + it)\|_2^2 = \lim_{t' \to t} \langle \Lambda^T E(\cdot, 1/2 + it'), \Lambda^T E(\cdot, 1/2 + it) \rangle
$$

$$
= \lim_{t' \to t} \left(\frac{e^{iT(t'-t)}}{i(t'-t)} + \frac{\overline{c(it')}e^{iT(t'+t)}}{i(t'+t)} - \frac{c(it)e^{-iT(t'+t)}}{i(t'+t)} - \frac{c(it')\overline{c(it)}e^{-iT(t'-t)}}{i(t'-t)} \right)
$$

$$
= 2T - \frac{c'(it)}{c(it)} + \Re \left(\frac{c(it)e^{-2itT}}{t} \right) \ll T \log(1 + |t|).
$$

Here, $c(s)$ is the *scattering matrix* appearing in the constant term of $E(\cdot, 1/2 + it)$, given by $\frac{\xi(s)}{\xi(1+s)}$, where $\xi(s)$ is the completed Riemann zeta function. The last estimate follows from the standard zero-free region of the Riemann zeta function. Now, we explain the difficulties we face if we try to prove the statement given in (1.3), by generalizing the method above.

1. The case when $\mathrm{Eis}(\varphi, \lambda)$ is the minimal parabolic Eisenstein series for GL_n is a natural higher-rank analogue of the above. Higher rank Maass–Selberg relations are in this case more complicated than the GL_2 version. The number of summands in the formula is $(n!)^2$. Moreover, each summand will likely have (a priori, higher-order) poles as $\lambda' \to \lambda$. The manoeuvre of 'grouping terms' which cancel each other's poles, analogous to the GL_2 case, is, in fact, combinatorially quite involved. Such complications arise already for GL_3, as can be seen in [43].

2. Apparently, the next difficult case is when $\mathrm{Eis}(\varphi, \lambda)$ is induced from a cuspidal representation of a nonminimal parabolic, that is φ is a cusp form. Apart from solving the combinatorial problem that arises in the limit procedure, one also needs to know certain strong estimates of the logarithmic (a priori, high) derivatives of the intertwining operators. When the Eisenstein series is spherical everywhere, such estimates can be described via the zero-free region of various Rankin–Selberg L-functions for $\mathrm{GL}_{m_1} \times \mathrm{GL}_{m_2}$ with $m_1 + m_2 \leq n$ at the edge of the critical strips. However, the quality and generality of the zero-free region that is required to prove (1.3) is still out of reach.

3. We remark that Langlands (see [6, Proposition 15.3]) proved the Maass–Selberg relations, but only when the underlying representation π of $M(\mathbb{A})$ is *cuspidal*. Arthur, in [5], proved an approximate Maass–Selberg relation for a general discrete series π. However, the precision and generality of the approximation that we need to prove (1.3) is not available. In fact, the required precision in Arthur's approximate Maass–Selberg formula can be obtained only via the classification of discrete series representations, in particular, how the residual spectrum arises from the Eisenstein series induced from cusp forms. This, in turn, requires understanding deep meromorphic properties of the Eisenstein series, in particular, the locations, orders and residues of their poles.

4. When we move from GL_n to general reductive groups, all of the above difficulties persist prominently. First of all, our ignorance of the classification of the discrete series, equivalently the required meromorphic properties of a general Eisenstein series, hinders the very first step of estimating the L^2-norm via the Maass–Selberg relations. Even in the case of the Eisenstein series induced from cuspidal representations, the limit procedure requires a deep understanding of the meromorphic properties of the intertwining operators and the size of the (a priori, high) derivatives of the same — none of which is yet in the grasp of the current literature.

Primarily, this is why, in this paper, we try to estimate (1.2) for general Eisenstein series on general reductive groups *not* via the Maass–Selberg formula, rather via the spectral side of the Arthur–Selberg trace formula. We explain this in the next subsection.

In a future work, in order to estimate (1.1) for GL_n using the Maass–Selberg relations, we plan to address some of the issues pointed to above and, in particular, to give a robust method of understanding the combinatorics and corresponding limit problem mentioned above.

Remark 1.12. There is one case in which the Maass–Selberg relations do yield a nice formula, and it is exactly the unramified case of Theorem 2. The Maass–Selberg relations, in this case, are given in Lemma 3.1 and essentially reduce to understanding a certain operator we denote by $\mathcal{M}_M^T(\lambda, P)$, defined by a certain limit. Arthur's method of (G, M)-*family* gives a way to calculate the limit in the formula. More precisely, Finis–Lapid–Müller [21], generalizing previous results of Arthur in the spherical case, essentially gave a formula for $\mathcal{M}_M^T(\lambda, P)$, given in Proposition 5.7. The formula is a polynomial in T with coefficients given by certain intertwining operators and their *first* derivatives, given along *linearly independent directions*. Then, when we take λ-average over a short interval around λ_0, we can use properties (TWN+) (and (BD) for general congruence subgroups) to deduce a bound on $\mathcal{M}_M^T(\lambda, P)$, given in Proposition 5.5, from which Theorem 2 follows. This technique can be traced back at least to [49] (for GL_n and a fixed level) and was formalized in [22].

1.5. High level sketch for the proof of Theorem 3

Our proof of Theorem 3 uses the Arthur–Selberg trace formula, or more precisely, Arthur's development of the spectral side of the trace formula and Finis–Lapid–Müller's refinement of it.

As we described in the previous subsection, it is quite difficult to estimate $\|\Lambda^T \mathrm{Eis}(\varphi_0, \lambda_0)\|_2^2$ pointwise in λ_0. Proving an estimate on a λ-average over a short interval around λ_0 seems to be more tractable. Moreover, in our proof, we also need to average over φ in an orthonormal basis, equivalence classes of associated parabolic subgroups and discrete series representations of the Levi of these parabolic subgroups.

For a sufficiently dominant truncation parameter T and a bi-$K_\infty K$-invariant test function f on $G(\mathbb{A})$, Arthur defined a distribution $J_\chi^T(f)$ on the spectral side, that (for $d(T)$ sufficiently large) essentially can be given by a weighted sum over $P \supset P_0$ of

$$\sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_{\chi,\pi}(P)} \int_{i(\mathfrak{a}_M^G)^*} \tilde{f}(\mu_\pi + \lambda) \|\Lambda^T \mathrm{Eis}(\varphi, \lambda)\|^2 \, d\lambda,$$

where μ_π is the Langlands parameter of π (which is necessarily spherical as an $M(\mathbb{A})$-a representation) and $\tilde{f}$ is the *spherical transform* of f.

We choose an f with *bounded support* via the Paley–Wiener theory and the theory of spherical inversion so that spherical transform $\tilde{f}$ localizes at $\mu_\pi + \lambda \approx \mu_{\pi_0} + \lambda_0$, and so that $J_\chi^T(f)$ majorizes the left-hand side of the estimate in Theorem 3. Our goal, now, is to estimate $J_\chi^T(f)$ satisfactorily.

Arthur, in his deep work to develop the spectral side of the trace formula, showed that $J_\chi^T(f)$ is a polynomial in T whose coefficients are given in terms of certain intertwining operators and their derivatives. More precisely, $J_\chi^T(f)$ essentially takes the form of a weighted sum over $P \supset P_0$ of

$$\sum_{\pi \in \Pi_2(M_P)} \sum_{L \supset M_P} \int_{i(\mathfrak{a}_L^G)^*} \tilde{f}(\lambda) \mathrm{tr}\left(\mathcal{M}_L^T(P, \lambda) \mid_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right) d\lambda.$$

The combinatorial formula of Finis–Lapid–Müller allows us to understand $\mathcal{M}_L^T(P, \lambda)$ in terms of a certain linear combination of various rank one intertwining operators and their first derivatives. We can then use the bound on $\mathcal{M}_L^T(P, \lambda)$ given in Proposition 5.5 that are yielded by properties (TWN+) and (BD), together with the Paley–Wiener theory to conclude the proof.

The fact that we only see the first derivative of the intertwiners in the above formula circumvents the difficulty that arises because of, a priori, high derivatives of the same in the Maass–Selberg formula, as described in Section 1.4. Moreover, the formula above, which is yielded by the theory of (G, L)-*families* due to Arthur and Finis–Lapid–Müller, bypasses the combinatorial complications that arise in the limit procedure in the Maass–Selberg formula, as described in Section 1.4.

It is worth noting that we do not use the geometric side of the trace formula, rather only use the two different expressions of the *spectral side* given in Proposition 6.1 (the spectral side of the pretrace formula) and Proposition 6.2 (the spectral side of the trace formula) (also see Proposition 6.3). We also remark that we crucially use the spectral decomposition for a specific cuspidal datum. This is presumably different than a usual proof of, for example, Weyl law (see, e.g. [20], [41]), where one does another sum over *all* cuspidal data in addition to the averages in Theorem 3. This extra average will not give us the result of the strength as in, for example, Theorem 1. The primary reason is that the contribution from the cusp forms will dominate. This way, we may only achieve a polynomial strength dependence on level(K) in Theorem 1 (and Remark 1.2) as opposed to a poly-logarithmic strength.

1.6. On the proof of the applications

The proofs of the applications do not use Theorems 3 or 2 directly, but instead use very similar methods. The reason that we cannot use the theorems directly is explained in Remark 1.7.

Instead, one can prove the theorems by proving a weighted version of Theorem 3, where representations are weighted by how far they are from being tempered. Our method allows such a modification, by using a function $h_{R,q}$ on $\mathrm{GL}_n(\mathbb{A})$ that is essentially the characteristic function of the R-radius ball at the Archimedean places and of $K(q)$ at the non-Archimedean places, where $K(q)$ is the principal congruence subgroup of level q of GL_n over the finite adeles (see Section 4.1 below for the definition). This requires increasing $d(T)$ sufficiently slowly so that Arthur's formulæ will still hold.

In practice, we use the full distribution J^T of Arthur, over the entire spectrum. The idea is to compare two different formulæ of Arthur for $J^T(h_{R,q})$. On the one hand, the spectral side of the trace formula allows us to give a bound on $J^T(h_{R,q})$, given in Proposition 8.1. This is done by using the bound on $\mathcal{M}_L^T(\lambda, P)$ given in Proposition 5.5 to reduce the problem to a sum, which is bounded in the work of Assing–Blomer [8, Theorem 7.1].

The other formula for $J^T(h_{R,q})$, which requires that $d(T)$ will be sufficiently large (actually depending logarithmically on R), can be seen as a pretrace formula and leads to a bound on the usual kernel of $h_{R,q}$, given in Proposition 8.3. The bound on the kernel of $h_{R,q}$ by standard methods (e.g, [51, 27]) leads to the proofs of both Theorems 4 and 5.

We finally remark that similar ideas, of using two formulæ for $J^T(h)$, were recently used by Finis–Lapid to provide a remainder term for Weyl law for a large class of arithmetic noncompact locally symmetric spaces [20]. However, [20] compares the pretrace formula with the *geometric side* of the trace formula, not the spectral side.

1.7. Added in proof

After the completion of this article, Assing, Blomer and Nelson [9] extended the results of [8]. In particular, they show that in Theorems 4 and 5, the assumption that q is square-free can be removed.

1.8. Conventions

As usual in analytic number theory, we use ϵ to denote a small positive constant, whose actual value may change from line to line. We use Vinogradov's notations $\gg$ and $\ll$, whose implied constants are understood to depend only on the group G and the number field F.

2. General preliminaries

Let F denote a number field with integer ring $\mathfrak{o}$. The letter v will usually (unless mentioned otherwise) denote an arbitrary place of F, and F_v will denote the v-adic completion of F. Let G denote a reductive group defined over F. Let $\mathbb{A}$ denote the adele ring of F. Also let $\mathbb{A}_f$ denote the finite adeles and $F_\infty := F \otimes \mathbb{R}$.

Below, we briefly discuss the main ingredients and quantities in our proof. We primarily follow the notations of [21, Section 2.1] and [22, Section 4].

2.1. Brief structure theory of reductive groups

Let T_0 denote a fixed maximal F-split torus of G. Let M_0 denote the centralizer of T_0 in G, which is a minimal Levi subgroup. Let $\mathcal{L}$ denote the finite set of Levi subgroups defined over F containing M_0. Given $M \in \mathcal{L}$, let $\mathcal{L}(M) \subset \mathcal{L}$ denote the (finite) set of Levi subgroups containing M and $\mathcal{P}(M)$ denote the (finite) set of parabolic subgroups defined over F with Levi subgroup M. We also fix a minimal parabolic subgroup $P_0 \supset M_0$.

Let T_M denote the split part of the identity component of the centre of M. We denote $W(M) := N_G(M)/M$, where $N_G(M)$ denotes the normalizer of M in G. We denote the Weyl group of (G, T_0) by $W_0 := W(M_0)$. We can identify $W(M)$ as a subgroup of W_0.

Let $\mathfrak{a}_M^*$ denote the $\mathbb{R}$-vector space spanned by the lattice $X^*(M)$ of F-rational characters of M, and let $\mathfrak{a}_{M,\mathbb{C}}^* := \mathfrak{a}_M^* \otimes_{\mathbb{R}} \mathbb{C}$. For $M_1 \subset M_2$, it holds that $\mathfrak{a}_{M_2}^* \subset \mathfrak{a}_{M_1}^*$ with a well-defined complement $\mathfrak{a}_{M_1}^* = \mathfrak{a}_{M_2}^* \oplus (\mathfrak{a}_{M_1}^{M_2})^*$. Let $\mathfrak{a}_M$ denote the dual space of $\mathfrak{a}_M^*$, and we fix a pairing $\langle, \rangle$ between $\mathfrak{a}_M$ and $\mathfrak{a}_M^*$. For $P \in \mathcal{P}(M)$, we write $\mathfrak{a}_P = \mathfrak{a}_M$.

We choose a W_0-invariant inner product on $\mathfrak{a}_0 := \mathfrak{a}_{M_0}$, which also defines measures on $\mathfrak{a}_M$ and $\mathfrak{a}_M^*$ for all $M \in \mathcal{L}$. We normalize all measures as in [6, Section 7], similar to [21].

For $P \in \mathcal{P}(M)$, we write $M_P = M$. Let N_P denote the unipotent radical of P. Let $\Sigma_P \subset (\mathfrak{a}_P^G)^*$ denote the set of roots of T_M on the Lie algebra of N_P. Let Δ_P denote the subset of simple roots of Σ_P. Let $\Sigma_M := \cup_{P \in \mathcal{P}(M)} \Sigma_P$. Given two parabolics $P, Q \in \mathcal{P}(M)$, we say that P and Q are *adjacent along* $\alpha \in \Sigma_M$ and write $P|^\alpha Q$ if $\Sigma_P = \Sigma_Q \setminus \{-\alpha\} \cup \{\alpha\}$.

Let A_0 denote the identity component of $T_0(\mathbb{R})$. We let $A_M := A_0 \cap T_M(\mathbb{R})$. Let $H_M : M(\mathbb{A}) \to \mathfrak{a}_M$ denote the natural homomorphism defined by

$$e^{\langle \chi, H_M(m) \rangle} := |\chi(m)|_{\mathbb{A}} = \prod_v |\chi(m_v)|_v$$

for any $\chi \in X^*(M)$. Here, $|\cdot|_{\mathbb{A}}$ denotes the homomorphism from $\mathbb{A}^\times \to \mathbb{R}^+$ given by $|\cdot|_{\mathbb{A}} := \prod_v |\cdot|_v$, where $|\cdot|_v$ denotes the usual absolute value on F_v. Let $M(\mathbb{A})^1 \subset M(\mathbb{A})$ denote the kernel of H_M. It holds that $M(\mathbb{A}) = A_M \times M(\mathbb{A})^1$.

We fix a maximal compact subgroup K_∞ of $G(F_\infty)$ and a maximal open-compact subgroup K_f of $G(\mathbb{A}_f)$. Finally, we fix the maximal compact subgroup $K_{\mathbb{A}} := K_\infty K_f$ of $G(\mathbb{A})$ and assume that it is admissible relative to M_0, so that for every $P \in \mathcal{P}(M_0)$, an Iwasawa decomposition $G(\mathbb{A}) = P(\mathbb{A}) K_{\mathbb{A}}$ holds.

Given $M \in \mathcal{M}$ and $P \in \mathcal{P}(M)$, let $H_P : G(\mathbb{A}) \to \mathfrak{a}_P$ denote the extension of H_M to a left $N_P(\mathbb{A})$ and right $K_{\mathbb{A}}$-invariant map.

3. Global preliminaries

3.1. Automorphic forms

Given $M \in \mathcal{L}$, let

$$L^2_{\mathrm{disc}}(M(F) \backslash M(\mathbb{A})^1) \cong L^2_{\mathrm{disc}}(M(F) A_M \backslash M(\mathbb{A}))$$

denote the discrete part of the spectrum of $L^2(M(F)A_M\backslash M(\mathbb{A}))$, that is the Hilbert sum of all irreducible subrepresentations of $M(\mathbb{A})$ on the space.

Let $\Pi_2(M)$ denote[4] the countable set of equivalence classes of irreducible unitary representation of $M(\mathbb{A})$ appearing in $L^2_{\text{disc}}(M(F)\backslash M(\mathbb{A})^1)$. Each element $\pi \in \Pi_2(M)$ is an abstract unitary representation of $M(\mathbb{A})$. By abuse of notations, we consider each $\pi \in \Pi_2(M)$ also as a representation of $M(\mathbb{A})^1$. By Flath's tensor product theorem (see [23]), each $\pi \in \Pi_2(M)$ can be considered as a tensor product of local representations $\pi = \otimes_v \pi_v$, where π_v is a representation of $M(F_v)$. We denote the finite part of the representation $\otimes_{v<\infty}\pi_v$ by π_f, and the infinite part of the representation $\otimes_{v|\infty}\pi_v$ by π_∞.

Let $P \in \mathcal{P}(M)$ and $\mathcal{A}^2(P)$ denote the set of automorphic forms φ on $N_P(\mathbb{A})M(F)\backslash G(\mathbb{A})$ that are square-integrable, that is φ satisfies

$$\delta_P^{-1/2}\varphi(\cdot k) \in L^2_{\text{disc}}(A_M M(F)\backslash M(\mathbb{A})), \quad \forall k \in K_{\mathbb{A}},$$

where δ_P is the modular character attached to P (in particular, $\varphi(amk) = \delta_P^{1/2}(a)\varphi(mk)$ for $a \in A_M$, $m \in M(\mathbb{A})^1$ and $k \in K_{\mathbb{A}}$). Then $\mathcal{A}^2(P)$ is a pre-Hilbert space, with an inner product

$$\langle \varphi_1, \varphi_2 \rangle_{\mathcal{A}^2(P)} := \int\limits_{A_M M(F) N_P(\mathbb{A})\backslash G(\mathbb{A})} \varphi_1(g)\overline{\varphi_2(g)}\,\mathrm{d}g.$$

Sometimes, we shorthand $\|\cdot\|_{\mathcal{A}^2(G)}$ as $\|\cdot\|_2$. Let $\overline{\mathcal{A}^2(P)}$ denote the Hilbert space completion of $\mathcal{A}^2(P)$. For $\pi \in \Pi_2(M)$, we will also denote $\mathcal{A}^2_\pi(P) \subset \mathcal{A}^2(P)$ to be the subspace of φ, such that $\delta_P^{-1/2}\varphi(\cdot k)$ lies in the π-isotypic subspace of $L^2_{\text{disc},\pi}(A_M M(F)\backslash M(\mathbb{A})) \subset L^2_{\text{disc}}(A_M M(F)\backslash M(\mathbb{A}))$ for every $k \in K_{\mathbb{A}}$.

Given $\lambda \in i(\mathfrak{a}_P^G)^*$, let $\rho(P, \lambda, \cdot)$ denote the unitary representation of $G(\mathbb{A})$ on $\overline{\mathcal{A}^2(P)}$, given by

$$\rho(P, \lambda, y)(\varphi)(x) := \varphi(xy)e^{\langle \lambda, H_P(xy)-H_P(x)\rangle}, \quad y \in G(\mathbb{A}).$$

The representation $\rho(P, \lambda, \cdot)$ is isomorphic to

$$\text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})}\left(L^2_{\text{disc}}(A_M M(F)\backslash M(\mathbb{A})) \otimes e^{\langle \lambda, H_M(\cdot)\rangle}\right)$$

(see [22, Section 4.2]). We remark that the action of every compact subgroup of $G(\mathbb{A})$ on $\mathcal{A}^2(P)$ does not depend on λ.

Recall that a *cuspidal datum* χ is a $G(F)$-conjugacy class of pairs (L, σ), where L is a Levi subgroup of G defined over F and σ is a cuspidal representation of $L(\mathbb{A})$ (see [6, Section 12] for a detailed discussion). We have a coarse spectral decomposition

$$L^2_{\text{disc}}(A_M M(F)\backslash M(\mathbb{A})) = \widehat{\bigoplus_\chi} L^2_{\text{disc},\chi}(A_M M(F)\backslash M(\mathbb{A}))$$

(see [6, (12.5)]). Here and elsewhere, $\widehat{\oplus}$ denotes a Hilbert space direct sum. We have a similar coarse spectral decomposition of $\mathcal{A}^2(P)$ (see [6, p.66–67]). Let $\mathcal{A}^2_\chi(P)$ denote the χ-part of this spectral decomposition. Finally, we also write

$$L^2_{\text{disc},\chi,\pi}(A_M M(F)\backslash M(\mathbb{A})) := L^2_{\text{disc},\chi}(A_M M(F)\backslash M(\mathbb{A})) \cap L^2_{\text{disc},\pi}(A_M M(F)\backslash M(\mathbb{A}))$$

and $\mathcal{A}^2_{\chi,\pi}(P) := \mathcal{A}^2_\chi(P) \cap \mathcal{A}^2_\pi(P)$.

---

[4]In [21], the authors denote it by $\Pi_{\text{disc}}(M(\mathbb{A}))$.

3.2. Some discussion of the coarse spectral expansion

Since this topic is less standard, for the convenience of the reader, we add here a discussion of the coarse spectral decomposition.

We start with the group GL_n, in which case, the coarse spectral decomposition is well understood. Thanks to the multiplicity one theorem, each cuspidal representation σ of a Levi subgroup L appears with multiplicity one, that is

$$\dim \operatorname{Hom}_{L(\mathbb{A})}(\sigma, L^2(A_L L(F)\backslash L(\mathbb{A}))) \leq 1.$$

Next, let us describe some of the results of [44]. First, we recall that given a cuspidal representation σ of $GL_a(\mathbb{A})$ and $b > 0$, one may construct a Speh representation $\operatorname{Speh}(\sigma, b)$ of $GL_{ab}(\mathbb{A})$, which is the unique irreducible quotient of the induction of $\sigma^{\otimes b}$ from the Levi subgroup $GL_a(\mathbb{A})^b$ and twisted by the character $GL_a(\mathbb{A})^b \ni (g_1, g_2, \ldots, g_b) \mapsto |\det(g_1)|^{(b-1)/2} |\det(g_2)|^{(b-3)/2} \ldots |\det(g_b)|^{(1-b)/2}$. It holds that $\operatorname{Speh}(\sigma, b) \in \Pi_2(GL_{ab})$, and each representation $\pi \in \Pi_2(GL_n)$ is isomorphic to a unique Speh representation $\operatorname{Speh}(\sigma, b)$ for some $b \mid n$ and cuspidal representation σ of $GL_{n/b}(\mathbb{A})$. Note that the $b = 1$ case corresponds to the cuspidal representations in $\Pi_2(GL_n)$.

Now, if M is a Levi subgroup of GL_n of a parabolic P, for each $\pi \in \Pi_2(M)$, there is a unique cuspidal datum χ, such that $\mathcal{A}^2_{\chi,\pi}(P) \neq \{0\}$, and the multiplicity of π in the space $L^2_{\text{disc}}(A_M M(F)\backslash M(\mathbb{A}))$ is equal to 1. More precisely, if $M \cong GL_{n_1} \times \cdots \times GL_{n_k}$ is the standard Levi subgroup, and $\pi = \pi_1 \otimes \cdots \otimes \pi_k$, then one can write (uniquely) $n_i = a_i b_i$, find a cuspidal representation $\sigma_i \in \Pi_2(GL_{a_i})$ and write $\pi_i = \operatorname{Speh}(\sigma_i, b_i)$. In this case, let $L := GL_{a_1}^{b_1} \times \cdots \times GL_{a_k}^{b_k}$ and $\sigma := \sigma_1^{\otimes b_1} \otimes \cdots \otimes \sigma_k^{\otimes b_k} \in \Pi_2(L)$ be cuspidal. If χ is the cuspidal datum corresponding to (L, σ), then $\mathcal{A}^2_{\chi,\pi}(P) \neq \{0\}$.

The procedure can be reversed, but one should be a bit careful about the fact that cuspidal datum is only defined up to equivalence. In general, given a Levi L and $\sigma \in \Pi_2(L)$ cuspidal, let $\chi := (L, \sigma)$ be the cuspidal datum and P be a standard parabolic containing L as a Levi subgroup. Then there are at most $n_P := |\sum_{P' \sim P} W(\mathfrak{a}_P, \mathfrak{a}_{P'})| \leq |W_0| = n!$ other pairs (L', σ') of Levi subgroups $L' \supset M_0$ and cuspidal representations $\sigma' \in \Pi_2(L')$ in χ (see [6, Section 12]). Given such a pair (L, σ) and a Levi subgroup $L \subset M$, we get at most one residual representation in $L^2_{\text{disc},\chi}(A_M M(F)\backslash M(\mathbb{A}))$. All in all, there are at most $n!$ pairs of standard parabolics P and $\pi \in \Pi_2(M_P)$, such that $\mathcal{A}^2_{\chi,\pi}(P) \neq \{0\}$.

For groups that are different than GL_n, the situation becomes much more involved. An interesting case is the group $G = G_2$. By [45, Appendix III], for $\chi = (M_0, \text{triv})$ (the torus and the trivial representation of it), there are infinitely many different $\pi \in \Pi_2(G)$, such that

$$L^2_{\text{disc},\chi,\pi}(A_G G(F)\backslash G(\mathbb{A})) \neq \{0\}$$

(see also [35, 55] for complete results about the residual spectrum in this case). Moreover, by [24], for some $\pi \in \Pi_2(G)$, there are two different cuspidal data χ_1, χ_2, such that

$$L^2_{\text{disc},\chi_i,\pi}(A_G G(F)\backslash G(\mathbb{A})) \neq \{0\},$$

with χ_1 associated with the subgroup G and χ_2 associated with a proper parabolic subgroup. Finally, the multiplicity of π in $L^2_{\text{disc},\chi_1,\pi}(A_G G(F)\backslash G(\mathbb{A}))$ can be arbitrarily large.

3.3. Eisenstein series

Given $\varphi \in \mathcal{A}^2(P)$ and $\lambda \in (\mathfrak{a}_{P,\mathbb{C}}^G)^*$, we define a corresponding Eisenstein series, as

$$\operatorname{Eis}(\varphi, \lambda)(g) := \sum_{\gamma \in P(F)\backslash G(F)} \varphi(\gamma g) e^{\langle \lambda, H_P(\gamma g) \rangle}.$$

The above converges absolutely if λ is sufficiently dominant[5]. One can then, via Langlands' deep work (see, e.g. [36]), meromorphically continue $\text{Eis}(\varphi, \lambda)$ to all $\lambda \in (\mathfrak{a}_{P,\mathbb{C}}^G)^*$.

It holds that the Eisenstein series intertwines the representation $\rho(P, \lambda, \cdot)$. Namely, it holds that

$$\text{Eis}(\varphi, \lambda)(gy) = \text{Eis}(\rho(P, \lambda, y)\varphi, \lambda)(g),$$

for $y \in G(\mathbb{A})$.

3.4. Truncation

Recall the minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$. Given $T \in \mathfrak{a}_{M_0}^G$, let

$$d(T) := \min_{\alpha \in \Delta_{P_0}} \{\alpha(T)\}.$$

Given T with $d(T)$ sufficiently large, Arthur (following Langlands; see [6, Section 13]) defined certain *truncation operator* Λ^T, which acts on locally integrable functions on $G(F)\backslash G(\mathbb{A})^1$. We do not need the exact definition of Λ^T but just the following properties of it.

Let f be a locally integrable function on $G(F)\backslash G(\mathbb{A})^1$.

1. $\Lambda^T f$ is rapidly decaying. In particular, if $\lambda \in (\mathfrak{a}_{P,\mathbb{C}}^G)^*$ is away from any pole of $\text{Eis}(\varphi, \lambda)$, then $\Lambda^T \text{Eis}(\varphi, \lambda)$ is square-integrable.
2. Given any compact $\Omega \subset G(F)\backslash G(\mathbb{A})^1$, there is a $C(\Omega) > 0$, such that for every $T \in \mathfrak{a}_{M_0}^G$ with $d(T) > C(\Omega)$, it holds that $\Lambda^T f \mid_\Omega = f \mid_\Omega$ (see [40, Lemma 6.2]).
3. When $\alpha(T) \leq \alpha(T')$ for every $\alpha \in \Delta_{P_0}$, it holds that $\Lambda^{T'}\Lambda^T = \Lambda^T$ and $\|\Lambda^{T'} f\|_2 \geq \|\Lambda^T f\|_2$ (see [20, Section 3.3]).

3.5. Intertwining operators

We follow [22, Section 4.2] to define intertwining operators and to record their relevant properties.

For $P, Q \in \mathcal{P}(M)$ and $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$, we define the standard *intertwining operator*

$$M_{Q|P}(\lambda) \colon \mathcal{A}^2(P) \to \mathcal{A}^2(Q)$$

by

$$(M_{Q|P}(\lambda)\varphi)(x) := \int_{N_Q(\mathbb{A}) \cap N_P(\mathbb{A}) \backslash N_Q(\mathbb{A})} \varphi(nx) e^{\langle \lambda, H_P(nx) - H_Q(x)\rangle} \, \mathrm{d}n. \tag{3.1}$$

The above integral converges absolutely for sufficiently dominant λ. It can then be meromorphically continued for all $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$. It can be checked that $M_{Q|P}(\lambda)^{-1} = M_{P|Q}(\lambda)$ and for $\lambda \in i(\mathfrak{a}_M^G)^*$ the operator $M_{Q|P}(\lambda)$ is unitary.

For each $\pi \in \Pi_2(M)$, we denote the restriction of $M_{Q|P}(\lambda)$ on $\mathcal{A}_\pi^2(P)$ by $M_{Q|P}(\pi, \lambda) : \mathcal{A}_\pi^2(P) \to \mathcal{A}_\pi^2(Q)$. On the other hand, we can define an intertwiner from

$$M_{Q|P}^I(\lambda, \pi) : \text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi \to \text{Ind}_{Q(\mathbb{A})}^{G(\mathbb{A})} \pi$$

that can be densely defined on K-finite and $Z(\text{Lie}(G(F_\infty)) \otimes_\mathbb{R} \mathbb{C})$-finite subspace of $\text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi$ as in (3.1) (first, for sufficiently dominant λ, then by meromorphic continuation). Then we have a unique canonical isomorphism of $G(\mathbb{A}_f) \times (\text{Lie}(G(F_\infty)) \otimes_\mathbb{R} \mathbb{C}, K_\infty)$-modules (see [22, Section 4.2])

---

[5]We call λ to be *dominant* (respectively, *sufficiently dominant*) if $\langle \lambda, \alpha \rangle \geq 0$ (respectively, $\langle \lambda, \alpha \rangle$ are sufficiently positive) for all positive root α

$$j_P : \text{Hom}_{M(\mathbb{A})}\Big(\pi, L^2(A_M M(F)\backslash M(\mathbb{A}))\Big) \otimes \text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})}\pi \to \mathcal{A}_\pi^2(P)$$

(similarly, j_Q) that can be characterized by

$$j_Q \circ \Big[\text{Id} \otimes M_{Q|P}^I(\lambda, \pi)\Big] = M_{Q|P}(\lambda, \pi) \circ j_P$$

(see [48, (2.3)] and discussion around it).

Endowing $\text{Hom}_{M(\mathbb{A})}\big(\pi, L^2(A_M M(F)\backslash M(\mathbb{A}))\big)$ with the inner product $\langle f_1, f_2 \rangle = f_2^* f_1$, where $*$ denotes adjoint (notice that by Schur's lemma, $f_2^* f_1$ is a scalar), we note that j_P is an isometry.

Now, in the rest of this subsection, let $P|^\alpha Q$ for some $\alpha \in \Sigma_M$. In this case, $M_{Q|P}(\lambda)$ depends only upon $\langle \lambda, \alpha^\vee \rangle$. Let $\varpi \in \mathfrak{a}_M^*$ be such that $\langle \varpi, \alpha^\vee \rangle = 1$. We have

$$M_{Q|P}(s\varpi) \big|_{\mathcal{A}_\pi^2(P)} \circ j_P = n_\alpha(\pi, s) \cdot j_Q \circ \big(\text{Id} \otimes R_{Q|P}(\pi, s)\big), \tag{3.2}$$

for any $s \in \mathbb{C}$ avoiding poles of the operators on both sides of above. Here, $n_\alpha(\pi, s)$, a *global normalizing factor*, is a meromorphic function in s. Moreover, if $s \in i\mathbb{R}$, then $|n_\alpha(\pi, s)| = 1$. The operator

$$R_{Q|P}(\pi, s) : \text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})}\pi \to \text{Ind}_{Q(\mathbb{A})}^{G(\mathbb{A})}\pi$$

is a factorizable *normalized intertwining operator*, so that

$$R_{Q|P}(\pi, s) = \prod_v R_{Q|P}(\pi_v, s), \quad \pi = \otimes_v \pi_v.$$

Here, $R_{Q|P}(\pi_v, s)\mathbb{1}_{K_v} = \mathbb{1}_{K_v}$ for almost all v. Moreover, if $s \in i\mathbb{R}$, then $R_{Q|P}(\pi, s)$ is unitary (see [6, Theorem 21.4] for details).

3.6. *Langlands inner product formula*

For $Q \in \mathcal{P}(M)$, we let

$$\theta_Q(\Lambda) := v_{\Delta_Q}^{-1} \prod_{\beta \in \Delta_Q} \langle \Lambda, \beta^\vee \rangle, \quad \Lambda \in \mathfrak{a}_{M_0,\mathbb{C}}^*,$$

where v_{Δ_Q} is the covolume of the lattice in $\mathfrak{a}_Q^G$ spanned by $\{\beta^\vee \mid \beta \in \Delta_Q\}$.

For $T \in \mathfrak{a}_{M_0}^G$, we also define $Y_Q(T)$ as the projection to $\mathfrak{a}_M^G$ of $t^{-1}(T - T_0) + T_0$, where $t \in W_0$ is an element, such that $P_0 \subset tQ$, and $T_0 \in a_M^G$ is the element defined by Arthur in [2, Lemma 1.1] (also see the discussion around [6, (9.4)]).

Let $\chi = (M, \pi)$ be a cuspidal datum and $\varphi, \varphi' \in \mathcal{A}_{\chi,\pi}^2(P)$, namely, φ and φ' are cusp forms. In this case, Langlands, generalizing the classical Maass–Selberg relations, obtained a formula for the inner product of two truncated Eisenstein series attached to φ and φ' (see [6, Proposition 15.3]). The formula involves sums over various Weyl elements and parabolic subgroups of quantities involving (twisted) intertwining operators, θ_Q, and Y_Q. There is also an approximate version of the formula, due to Arthur in [5], when φ or φ' are not cuspidal. However, we do not need any of these formulæ here, but only a special case that we describe below.

For any two parabolic subgroups $P, Q \supset P_0$, let $W(\mathfrak{a}_P, \mathfrak{a}_Q)$ be the set of distinct linear isomorphisms from $\mathfrak{a}_P \subset \mathfrak{a}_{P_0}$ to $\mathfrak{a}_Q \subset \mathfrak{a}_{P_0}$, induced by elements of W_0. For any $Q \in \mathcal{P}(M)$ and $P_0 \subset P$, there is a natural action of $W(\mathfrak{a}_P, \mathfrak{a}_Q)$ on $\Pi_2(M)$. Following [6, Section 15], we call χ *unramified* if, for every $(M, \pi) \in \chi$, the stabilizer of π in $W(\mathfrak{a}_P, \mathfrak{a}_P)$ is trivial. We warn again that this notion should not be confused with the unramifiedness of representations of reductive groups over local fields.

In the unramified case, we record the Langlands inner product formula.

Lemma 3.1. *Let $\chi := (M, \pi)$ be any unramified cuspidal datum and $\varphi, \varphi' \in \mathcal{A}^2_{\chi, \pi}(P)$. Then we have*

$$\langle \Lambda^T \operatorname{Eis}(\varphi, \lambda), \Lambda^T \operatorname{Eis}(\varphi', \lambda) \rangle_{\mathcal{A}^2(G)} = \lim_{\Lambda \to 0} \sum_{Q \in \mathcal{P}(M)} e^{\langle \Lambda, Y_Q(T) \rangle} \frac{\langle M_{Q|P}(\lambda + \Lambda)\varphi, M_{Q|P}(\lambda)\varphi' \rangle_{\mathcal{A}^2(Q)}}{\theta_Q(\Lambda)}.$$

Proof. We refer to [21, pp.185] and note that $M_{Q|P}(1, \lambda)$ in this reference is the same as $M_{Q|P}(\lambda)$. □

4. Local preliminaries

4.1. Parameters associated with a representation

Let $M \in \mathcal{L}$. The goal of this subsection is to associate various parameters measuring the complexity of a certain representation $\pi \in \Pi_2(M)$.

4.1.1. Non-Archimedean complexity
First, following [22, Section 5.1], for any open-compact subgroup $K \subset K_f$, we recall the notion $\operatorname{level}(K)$. Fix $\iota : G \to \operatorname{GL}(V)$ a faithful F-rational representation and an $\mathfrak{o}$-lattice $\Lambda \subset V$ so that the stabilizer of $\hat{\mathfrak{o}} \otimes \Lambda \subset \mathbb{A}_f \otimes V$ in $G(\mathbb{A}_f)$ is K_f. Then, for any ideal $\mathfrak{q} \subset \mathfrak{o}$, we define the *principal congruence subgroup of level* $\mathfrak{q}$ by

$$K(\mathfrak{q}) := \{k \in K_f \mid \iota(k)v \equiv v \mod (\mathfrak{q}(\hat{\mathfrak{o}} \otimes \Lambda)), v \in \hat{\mathfrak{o}} \otimes \Lambda\}.$$

Now, we define $\operatorname{level}(K)$ to be the index $[\mathfrak{o} : \mathfrak{q}_K]$, where $\mathfrak{q}_K \subset \mathfrak{o}$ is the largest ideal $\mathfrak{q}$, such that $K \supset K(\mathfrak{q})$. On the other hand, given a representation π_f of $G(\mathbb{A}_f)$, we define $\operatorname{level}(\pi_f)$ as the minimum over $\operatorname{level}(K)$, such that π_f has a nontrivial K-invariant vector.[6]

Given a representation π_f of $M(\mathbb{A}_f)$ and $\alpha \in \Sigma_M$, we recall the quantity $\operatorname{level}_M(\pi_f, \hat{M}^+_\alpha)$, as defined in [22, Sections 5.1–5.2], which measures the complexity of π_f. In this paper, we do not need to know $\operatorname{level}_M(\pi_f, \hat{M}^+_\alpha)$ explicitly, but only an upper bound that we describe below.

Lemma 4.1. *Let $K \subset K_f$ be an open-compact subgroup. Assume that $\pi \in \Pi_2(M)$ is such that $\mathcal{A}^2_\pi(P)^K \neq \{0\}$ for some $P \in \mathcal{P}(M)$, that is the parabolic induction of π_f to $G(\mathbb{A}_f)$ has a K-invariant vector. Then we have*

$$\operatorname{level}_M(\pi_f, \hat{M}^+_\alpha) \leq \operatorname{level}(K).$$

Proof. It follows from [22, Lemma 5.12] that

$$\operatorname{level}_M(\pi_f, \hat{M}^+_\alpha) \leq \operatorname{level}(K; G^+_M),$$

where $G^+_M \subset G$ is a certain subgroup and $\operatorname{level}(K; G^+_M)$ denotes the relative level of K in respect to G^+_M as defined in [22, Section 5.1]. The lemma follows from the fact that $\operatorname{level}(K; G^+_M) \leq \operatorname{level}(K)$ (see [22, Section 5.1]). □

4.1.2. Archimedean complexity
Let π_∞ be a smooth representation of $M(F_\infty)$. Let $\alpha \in \Sigma_M$. We recall the quantity $\Lambda_M(\pi_\infty; \hat{M}_\alpha)$, as defined in [22, Sections 5.1–5.2], which measures the complexity of π_∞ as a representation of $M(F_\infty)$. In this paper, we do not need to know $\Lambda_M(\pi_\infty; \hat{M}_\alpha)$ explicitly, but only an upper bound that we describe below.

Let $\nu(\pi_\infty)$ be the eigenvalue for π_∞ of the Casimir operator of $M(F_\infty)$. Similarly, given $\tau \in \widehat{K_\infty}$, let $\nu(\tau)$ be the Casimir eigenvalue of τ.

---

[6]We remark that when $G = \operatorname{GL}_n$, the above definition of the level a representation of is different than the *usual* definition of the level, which is defined with respect to the *Hecke congruence subgroup* $K_0(\mathfrak{q})$ (see [31], also [22, Lemma 5.6]).

Lemma 4.2. *Let $\tau \in \widehat{K_\infty}$. Assume that $\pi \in \Pi_2(M)$ is such that $\mathcal{A}^2_\pi(P)^\tau \neq \{0\}$ for some $P \in \mathcal{P}(M)$, that is, the parabolic induction of π_∞ to $G(F_\infty)$ has a nonzero vector whose K_∞-type is τ. Then we have*

$$\Lambda_M(\pi_\infty; \hat{M}_\alpha) \ll 1 + \nu(\pi_\infty)^2 + \nu(\tau)^2.$$

Proof. It follows from [22, (10)–(11)] that

$$\Lambda_M(\pi_\infty; \hat{M}_\alpha) \ll \nu(\pi_\infty)^2 + \min_{\tau'} \nu(\tau')^2,$$

where $\tau' \in \widehat{K_\infty}$ runs over the lowest K_∞-types of $\mathrm{Ind}_{P(F_\infty)}^{G(F_\infty)} \pi_\infty$. We conclude by noting that the lowest K_∞-types have the smallest Casimir eigenvalues (equivalently, the smallest infinitesimal characters). □

4.2. The spherical function

In the rest of this section, we only work over the Archimedean places, that is F_∞ (see [15, Section 3] for references to the material below).

Let $G_\infty := A_G \backslash G(F_\infty)$ and K_∞ be a maximal compact subgroup in G_∞. Choose a maximal $\mathbb{R}$-split torus $T_\infty \subset G_\infty$, which we may assume to contain $T_0(F_\infty)$ and contained in $M_0(F_\infty)$. Let $W_\infty := N_{G_\infty}(T_\infty)/T_\infty$ be the Weyl group of G_∞ relative to T_∞. Let T_∞^0 be the connected component of the identity of T_∞ and $\mathfrak{a}_\infty$ be the Lie algebra of T_∞^0.

We remark that for any Levi $M \supset M_0$, the embedding $T_\infty \to M(F_\infty)$ defines a projection $\mathfrak{a}_\infty \twoheadrightarrow \mathfrak{a}_M^G$ and an embedding $(\mathfrak{a}_M^G)^* \hookrightarrow \mathfrak{a}_\infty^*$ (see [6, pp.118]).

We let $M_{0,\infty}$ be the Levi subgroup, which is the normalizer of T_∞ in G_∞. Let B_∞ be a Borel subgroup containing $M_{0,\infty}$ as a Levi subgroup and also containing $P_0(F_\infty)$. Let N_∞ be the unipotent radical of B_∞. We have the Iwasawa decomposition $G_\infty = B_\infty K_\infty = N_\infty T_\infty^0 K_\infty$.

Using the exponential map from $\mathfrak{a}_\infty$ to T_∞^0, each element $\mu \in \mathfrak{a}_{\infty,\mathbb{C}}^* := (\mathfrak{a}_\infty \otimes_\mathbb{R} \mathbb{C})^*$ defines a character χ_μ of T_∞, which is trivial on T_∞/T_∞^0. It extends, via N_∞-invariance, to a character of B_∞. Let $\delta_\infty : B_\infty \to \mathbb{R}_{>0}$ be the modular character of B_∞, which is attached to the character $\chi_{2\rho_\infty}$, where $2\rho_\infty$ is the sum of all the roots of the action of T_∞ on the Lie algebra of N_∞, weighted by the dimensions of the corresponding root spaces.

Given $g \in G_\infty$, we let $a(g)$ be its T_∞^0 part according to the Iwasawa decomposition $G_\infty = N_\infty T_\infty^0 K_\infty$. We define the *spherical function* $\eta_\mu : G_\infty \to \mathbb{C}$ corresponding to $\mu \in \mathfrak{a}_{\infty,\mathbb{C}}^*$ by

$$\eta_\mu(g) := \int_{K_\infty} \chi_{\mu+\rho_\infty}(a(kg)) \, dk.$$

Thus, η_μ is bi K_∞-invariant. Here, dk denotes the probability Haar measure on K_∞.

4.3. Spherical representation

We call an irreducible admissible representation (π, V) of G_∞ *spherical* if π has a nonzero K_∞-invariant vector. We can construct all irreducible admissible spherical representations of G_∞ from the unitarily induced principal series representations. Let $\mu \in \mathfrak{a}_{\infty,\mathbb{C}}^*$ and $\mathrm{Ind}_{B_\infty}^{G_\infty} \chi_\mu$ denotes the normalized parabolic induction of χ_μ from B_∞ to G_∞. It is an admissible representation and has a unique irreducible spherical subquotient. Conversely, for any irreducible spherical representation (π, V), we can find a (unique up to W_∞-action) $\mu_\pi \in \mathfrak{a}_{\infty,\mathbb{C}}^*$, such that π appears as the unique spherical subquotient of $\mathrm{Ind}_{B_\infty}^{G_\infty} \chi_{\mu_\pi}$. In this case, we call μ_π the *Langlands parameter* of π. When π is also unitary, it holds that $\|\Re(\mu_\pi)\| \leq \|\rho_\infty\|$ and $1 + \|\mu_\pi\|^2 \asymp 1 + \nu(\pi)$ (more precisely, $\nu(\pi) = \|\rho_\infty\|^2 - \|\Re(\mu_\pi)\|^2 + \|\Im(\mu_\pi)\|^2$) (see [15, (3.17)]).

If π is spherical, then the K_∞-invariant vector of π is unique up to multiplication by a scalar. Let $\mathbb{1}_{K_\infty}$ be the characteristic function on K_∞. Then $\pi(\mathbb{1}_{K_\infty})$ is well-defined and acts as a projection onto

the K_∞-invariant subspace of V. Let $0 \neq v \in \pi$ be a K_∞-invariant vector. It follows from the definition of the spherical function that $\pi(\mathbb{1}_{K_\infty})\pi(g)v = \eta_{\mu_\pi}(g)v$.

Let M_∞ be any Levi subgroup of G_∞, having T_∞ as a maximal split torus. We can associate to each irreducible admissible spherical (containing a nonzero $K_\infty \cap M_\infty$-invariant vector) representation π of M_∞ a Langlands parameter $\mu_\pi \in \mathfrak{a}_{\infty,\mathbb{C}}^*$, defined up to $W_{M,\infty} = N_{M_\infty}(T_\infty)/T_\infty \subset W_\infty$. Moreover, if Q_∞ is a parabolic subgroup of G_∞ with M_∞ as its Levi subgroup, $\mathrm{Ind}_{Q_\infty}^{G_\infty}\pi$ has a unique spherical subrepresentation. On the other hand, if $\mathrm{Ind}_{Q_\infty}^{G_\infty}\pi$ has a K_∞-invariant vector, then π necessarily has a $K_\infty \cap M_\infty$-invariant vector. In any case, by the transitivity of induction, μ_π is also the Langlands parameter of the spherical subquotient of $\mathrm{Ind}_{Q_\infty}^{G_\infty}\pi$ (however, the new parameter is defined up to W_∞).

4.4. The spherical transform

We call the convolution algebra on $C_c^\infty(K_\infty \backslash G_\infty / K_\infty)$ the *spherical Hecke algebra* of G_∞. For $h \in C_c^\infty(K_\infty \backslash G_\infty / K_\infty)$, we let $\tilde{h} \colon \mathfrak{a}_{\infty,\mathbb{C}}^* \to \mathbb{C}$ be the *spherical transform* of h, defined by

$$\tilde{h}(\mu) := \int_{G_\infty} h(g)\eta_\mu(g)\,\mathrm{d}g.$$

If π is spherical and $0 \neq v \in \pi$ is a K_∞-invariant vector, then it holds that

$$\pi(h)v = \int_{G_\infty} h(g)\pi(g)v\,\mathrm{d}g = \tilde{h}(\mu_\pi)v.$$

We define the *Abel–Satake transform* (also known as *Harish-Chandra transform*) to be the map $\mathcal{S} : C_c(K_\infty \backslash G_\infty / K_\infty) \to C_c(T_\infty)$ defined by

$$h \mapsto \mathcal{S}h : a \mapsto \delta_\infty(a)^{1/2} \int_{N_\infty} h(an)\,\mathrm{d}n,$$

where $\mathrm{d}n$ is a Haar measure on N_∞. Since $\mathcal{S}h$ is left-$K_\infty \cap T_\infty$-invariant, it is actually a map on T_∞^0. We have the exponential map $\exp\colon \mathfrak{a}_\infty \to T_\infty^0$, which gives an identification of $\mathfrak{a}_\infty$ with T_∞^0. So we may as well realize $\mathcal{S}h \in C_c(\mathfrak{a}_\infty)$ after precomposing with the exp map. It holds that $\mathcal{S}f$ is W_∞-invariant. In fact, Gangolli showed that

$$\mathcal{S}\colon C_c^\infty(K_\infty \backslash G_\infty / K_\infty) \to C_c^\infty(\mathfrak{a}_\infty)^{W_\infty}$$

is an isomorphism of topological algebras (see [15, eq.(3.21)] (also see [25]). Let us denote the Fourier–Laplace transform $C_c(\mathfrak{a}_\infty) \to C(\mathfrak{a}_{\infty,\mathbb{C}}^*)$ by the map

$$h \mapsto \hat{h} : \mu \mapsto \int_{\mathfrak{a}_\infty} h(\alpha)e^{\langle \mu, \alpha \rangle}\,\mathrm{d}\alpha,$$

Harish-Chandra showed that the spherical transform $\tilde{h} = \widehat{\mathcal{S}(h)}$. Here, $\mathrm{d}\alpha$ denotes the Lebesgue measure inherited from the $\mathbb{R}$-vector space structure of $\mathfrak{a}_\infty$.

4.5. The Paley–Wiener theorem

Recall the Abel–Satake transform $\mathcal{S}$ above. Gangolli (see [15, eq.(3.22)]) proved that if $h \in C_c^\infty(\mathfrak{a}_\infty)^{W_\infty}$ is such that $\mathrm{supp}(h) \subset \{\alpha \in \mathfrak{a}_\infty \mid \|\alpha\| \le b\}$, then

$$\mathrm{supp}(\mathcal{S}^{-1}h) \subset K_\infty \{\exp \alpha \mid \alpha \in \mathfrak{a}_\infty, \|\alpha\| \le b\} K_\infty. \tag{4.1}$$

We also record the classical Paley–Wiener theorem, which implies that for $h \in C_c^\infty(\mathfrak{a}_\infty)^{W_\infty}$ with support in $\{\alpha \in \mathfrak{a}_\infty \mid \|\alpha\| \le b\}$, we have

$$|\hat{h}(\mu)| \ll_{N,h} \exp(b\|\mathfrak{R}(\mu)\|)(1 + \|\mu\|)^{-N} \tag{4.2}$$

for every $N \ge 0$.

5. A bound on an intertwining operator

In this section, we prove Theorem 2 and provide one of the main ingredients of the proof of Theorem 3.

First, we record the required estimates of the logarithmic derivatives of the normalizing factors and the normalized intertwining operators, denoted by properties (TWN+) and (BD), respectively.

5.1. Property (TWN+)

In [18, Definition 3.3][7], the authors defined the *(TWN+) property, strong version*, of a reductive group. Property (TWN+) is a global property that concerns the size of the logarithmic derivative of $n_\alpha(\pi, s)$. Below, we record an implication of property (TWN+) which we will use in this article.

Lemma 5.1. *Let G satisfy property (TWN+). Let $M \in \mathcal{L}$ be proper, $\pi \in \Pi_2(M)$ and $\alpha \in \Sigma_M$. Let $K \subset K_f$ be open-compact and $\tau \in \widehat{K_\infty}$ so that $\mathcal{A}_\pi^2(P)^{\tau,K} \ne \{0\}$ for some $P \in \mathcal{P}(M)$. Then it holds that*

$$\int_{t_0}^{t_0+1} \left| \frac{n_\alpha'(\pi, it)}{n_\alpha(\pi, it)} \right| dt \ll \log(1 + |t_0| + \nu(\pi_\infty) + \nu(\tau) + \mathrm{level}(K)),$$

for any $t_0 \in \mathbb{R}$.

Proof. Note that $|n_\alpha(\pi, it)| = 1$. Now the proof follows immediately by applying Lemmas 4.1 and 4.2 to the inequality in [18, Definition 3.3]. □

Finally, we keep a record that property (TWN+) is not vacuous.

Proposition 5.2. *[18, Theorem 3.11]. Property (TWN+) holds for G being GL_n and its inner form, a quasi-split classical group[8], and G_2. Moreover, it holds for any group whose derived group coincides with the derived group of any group mentioned above.*

5.2. Property (BD)

In [19, Definitions 1–2], the authors defined the *(BD) property* of a reductive group. Property (BD) is a local property that concerns the size of the logarithmic derivative of $R_{Q|P}(\pi_v, s)$. Below, we record an implication of property (BD) which we will use in this article.

---

[7]The quantities $\Lambda(\pi_\infty; p^{\mathrm{sc}})$ and $\mathrm{level}(\pi, p^{\mathrm{sc}})$ are denoted by $\Lambda(\pi_\infty; \hat{M}_\alpha)$ and $\mathrm{level}(\pi_f; \hat{M}_\alpha^+)$, respectively, in this article, following [22].

[8]The proof of [18] for the classical groups crucially depends on the endoscopic classification of automorphic representations, due to Arthur [7] and Mok [46].

Lemma 5.3. *Let G satisfy property (BD) at every place. Let $M \in \mathcal{L}$ be proper, $\pi \in \Pi_2(M)$ and $P, Q \in \mathcal{P}(M)$ be adjacent. Let $K \subset K_f$ be open-compact and $\tau \in \widehat{K_\infty}$ so that $\mathcal{A}_\pi^2(P)^{\tau,K} \neq \{0\}$. Then it holds that*

$$\int_{t_0}^{t_0+1} \left\| R_{Q|P}(\pi, it)^{-1} R'_{Q|P}(\pi, it) \big|_{\left(\mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi\right)^{\tau,K}} \right\|_{\mathrm{op}} dt \ll \log(1 + \nu(\tau) + \mathrm{level}(K)),$$

for any $t_0 \in \mathbb{R}$.

To prove Lemma 5.3, we use a different formulation of property (BD), namely, [22, Definition 5.9]. This formulation is implied by the formulation in [19, Definitions 1–2] (see [19, Remark 3]).

Proof. We closely follow the proof of [22, Proposition 5.16]. As $R_{Q|P}(\pi, s) = \prod_v R_{Q|P}(\pi_v, s)$ and as $R_{Q|P}(\pi_v, s)\big|_{\left(\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v\right)^{K_v}}$ is independent of s for almost all v (see [22, Remark 5.13]), it is enough to show that

$$\int_{t_0}^{t_0+1} \left\| R_{Q|P}(\pi_\infty, it)^{-1} R'_{Q|P}(\pi_\infty, it) \big|_{\left(\mathrm{Ind}_{P(F_\infty)}^{G(F_\infty)} \pi_\infty\right)^{\tau}} \right\|_{\mathrm{op}} dt \ll \log(1 + \nu(\tau)),$$

and for each non-Archimedean v

$$\int_{t_0}^{t_0+1} \left\| R_{Q|P}(\pi_v, it)^{-1} R'_{Q|P}(\pi_v, it) \big|_{\left(\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v\right)^{K_v}} \right\|_{\mathrm{op}} dt \ll \log(1 + \mathrm{level}(K_v)),$$

where $K = \prod_{v < \infty} K_v$ and $\mathrm{level}(K_v)$ is the v-adic factor of $\mathrm{level}(K)$ and the implied constant is uniform in v (i.e. uniform in the order q_v of the corresponding residue fields).

To prove the non-Archimedean part, we only modify [22, (14)]. Keeping the same notations (in particular, denoting $R_{Q|P}(\pi_v, s)\big|_{\left(\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v\right)^{K_v}} = A_v(q_v^{-s})$) and assumptions as in [22], we bound

$$\int_{t_0}^{t_0+1} \left\| R_{Q|P}(\pi_v, it)^{-1} R'_{Q|P}(\pi_v, it) \big|_{\left(\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v\right)^{K_v}} \right\|_{\mathrm{op}} dt$$

$$= \log q_v \int_{t_0}^{t_0+1} \|A'_v(q_v^{-it})\|_{\mathrm{op}} dt \ll \log q_v \int_{S^1} \|A'_v(z)\|_{\mathrm{op}} |dz|,$$

where the last estimate follows by changing variable $q_v^{-it} \mapsto z$. In the first equality above, the factor $\log q_v$ comes from

$$R'_{Q|P}(\pi_v, s) = -\log q_v A'_v(q_v^{-s}).$$

In the second estimate above, we have changed variable

$$q_v^{-it} \mapsto z, \quad \frac{dz}{z} = -i \log q_v \, dt.$$

On the other hand, as t moves along $[t_0, t_0 + 1] \subset \mathbb{R}$, the variable z winds S^1 asymptotically $\log q_v$ times, uniformly in t_0. From here, we follow the discussion after [22, (14)] and conclude.

To prove the Archimedean part, we first modify [22, Lemma 5.19] and its proof. Keeping the same notations and assumptions as in [22], we claim that

$$\int\limits_{t_0}^{t_0+1} \|A'(it)\|_{\mathrm{op}} \, dt \ll \sum_{1 \leq j \leq m: \, u_j \neq 0} \min\left(1, \frac{1}{|u_j|}\right).$$

Here, $A(s) = A_\infty(s) = R_{Q|P}(\pi_\infty, s)\big|_{\left(\mathrm{Ind}_{P(F_\infty)}^{G(F_\infty)} \pi_\infty\right)^\tau}$. Following the proof of [22, Lemma 5.19], we see that the above claim follows from the following claim

$$\int\limits_{t_0}^{t_0+1} |\phi'_w(it)| \, dt \ll \min(1, 1/|u|),$$

where $w = u + iv$ with $u, v \in \mathbb{R}$,

$$\phi_w(z) := \frac{z + \overline{w}}{z - w}, \quad |\phi'_w(it)| = \frac{2|u|}{u^2 + (t - v)^2}.$$

If $u \neq 0$, we clearly have

$$\int\limits_{t_0}^{t_0+1} |\phi'_w(it)| \, dt \ll \int\limits_{t_0}^{t_0+1} \frac{1}{|u|} \, dt \ll \frac{1}{|u|}.$$

We also have

$$\int\limits_{t_0}^{t_0+1} |\phi'_w(it)| \, dt \leq \int\limits_{\mathbb{R}} \frac{2|u|}{u^2 + (t - v)^2} \, dt = 2 \arctan\left(\frac{t - v}{|u|}\right)\Big|_{-\infty}^{\infty} = 2\pi.$$

Thus, the claim follows.

Now, it follows from the discussion and the proof in [22, pp.620–621] that

$$\int\limits_{t_0}^{t_0+1} \left\| R_{Q|P}(\pi_\infty, it)^{-1} R'_{Q|P}(\pi_\infty, it)\big|_{\left(\mathrm{Ind}_{P(F_\infty)}^{G(F_\infty)} \pi_\infty\right)^\tau} \right\|_{\mathrm{op}} \, dt \ll \sum_{j=1}^{r} \sum_{k=1}^{m} \min\left(1, \frac{1}{|u_j - ck|}\right),$$

where $|u_j|, r \ll_G 1$,

$$m \ll 1 + \|\tau\| \asymp 1 + \nu(\tau)^2,$$

and $c > 0$ depends only on M. We conclude the proof in the Archimedean case by noting that the double sum above is bounded by $\ll_{G,M} \log m$. $\square$

Finally, we keep a record that the property (BD) is not vacuous.

Proposition 5.4. *[19, Corollary 1]. Property (BD) holds at every place if G is GL_n or its inner form, or a split group of rank 2. It also holds at every non 2-adic place if G is special orthogonal, or symplectic.*

Fix $T \in \mathfrak{a}_{M_0}^G$. Let $M \in \mathcal{L}$, and fix $P \in \mathcal{P}(M)$. For $Q \in \mathcal{P}(M)$ and $\lambda, \Lambda \in i(\mathfrak{a}_M^G)^*$, we define an operator $\mathcal{M}_Q^T(\lambda, \Lambda, P)$ acting on $\mathcal{A}^2(P)$ by

$$\mathcal{M}_Q^T(\lambda, \Lambda, P) := e^{\langle \Lambda, Y_Q(T) \rangle} M_{Q|P}(\lambda)^{-1} M_{Q|P}(\lambda + \Lambda). \tag{5.1}$$

Similarly, for $L \in \mathcal{L}(M)$, $Q \in \mathcal{P}(L)$ and $\lambda, \Lambda \in i(\mathfrak{a}_L^G)^* \subset i(\mathfrak{a}_M^G)^*$, we define

$$\mathcal{M}_Q^T(\lambda, \Lambda, P) := \mathcal{M}_{Q_1}^T(\lambda, \Lambda, P),$$

for arbitrary $Q_1 \subset Q$, $Q_1 \in \mathcal{P}(M)$. The choice of Q_1 does not matter here. Here, in the left-hand side, λ, Λ are realized as restrictions on $\mathfrak{a}_L$ (see [6, pp.133]).

Following definitions in [6, Section 21], the set $\{\mathcal{M}_Q^T(\lambda, \Lambda, P)\}_{Q \in \mathcal{P}(L)}$ is an example of a '(G, L)-family' (see [6, Section 17] for a general discussion on (G, M)-family). For $L \in \mathcal{L}(M)$ and $\lambda \in i(\mathfrak{a}_L^G)^*$, we define the operator $\mathcal{M}_L^T(\lambda, P)$ as the value at $\Lambda = 0$ of

$$\sum_{Q \in \mathcal{P}(L)} \frac{\mathcal{M}_Q^T(\lambda, \Lambda, P)}{\theta_Q(\Lambda)}.$$

Below, we prove the main estimate on the operator norm of $\mathcal{M}_L^T(\lambda, P)$.

Proposition 5.5. *Assume that G satisfies properties (TWN+) and (BD). Fix $M \in \mathcal{L}$, $P \in \mathcal{P}(M)$ and $L \in \mathcal{L}(M)$. Let $\tau \in \widehat{K_\infty}$, $K \subset K_f$ be an open-compact and $\pi \in \Pi_2(M)$. Then, for every $\lambda_0 \in i(\mathfrak{a}_L^G)^*$,*

$$\int\limits_{\substack{\lambda \in i(\mathfrak{a}_L^G)^* \\ \|\lambda - \lambda_0\| \leq 1}} \left\| \mathcal{M}_L^T(\lambda, P)|_{\mathcal{A}_\pi^2(P)^{\tau, K}} \right\|_{\mathrm{op}} d\lambda \ll (\|T\| \log(1 + \nu(\tau) + \nu(\pi_\infty) + \|\lambda_0\| + \mathrm{level}(K)))^{\dim \mathfrak{a}_L^G},$$

where $T \in \mathfrak{a}_{M_0}^G$ with sufficiently large $d(T)$.

Before proving Proposition 5.5, we give a proof of Theorem 2.

Proof of Theorem 2. Using Lemma 3.1 and the fact that $M_{Q|P}(\lambda)$ unitary for $\lambda \in i(\mathfrak{a}_M^G)^*$, and recalling Equation (5.1), we can write

$$\|\Lambda^T \mathrm{Eis}(\varphi, \lambda)\|_2^2 = \langle \mathcal{M}_M^T(\lambda, P)\varphi, \varphi \rangle_{\mathcal{A}^2(P)}.$$

We conclude using Proposition 5.5 and that φ is a unit. □

The rest of this section is devoted to the proof of Proposition 5.5.

Let $P|^\alpha Q$ for some $\alpha \in \Sigma_M$. Let $\varpi \in \mathfrak{a}_M^*$ be such that $\langle \varpi, \alpha^\vee \rangle = 1$. We define

$$\delta_{Q|P}(\lambda) := M_{Q|P}^{-1}(\lambda) D_\varpi M_{Q|P}(\lambda) \tag{5.2}$$

as in [22, Section 4.3] as an operator from $\mathcal{A}^2(P)$ to $\mathcal{A}^2(P)$. Here $D_\varpi M_{Q|P}(\lambda)$ is the derivative of $M_{Q|P}(\lambda)$ in the direction ϖ. The definition is independent of the choice of ϖ; see [22, pp.608].

Lemma 5.6. *Keep the notations as above and assumptions as in Proposition 5.5. We have*

$$\int\limits_{t_0}^{t_0+1} \left\| \delta_{Q|P}(it\varpi)|_{\mathcal{A}_\pi^2(P)^{\tau, K}} \right\|_{\mathrm{op}} dt \ll \log(1 + \nu(\tau) + \nu(\pi_\infty) + |t_0| + \mathrm{level}(K)),$$

for any $t_0 \in \mathbb{R}$.

Proof. We take the inverse on both sides of (3.2) and compose it with the derivative of (3.2) and obtain

$$j_P^{-1} \circ \delta_{Q|P}(s\varpi) \circ j_P = \frac{n_\alpha'(\pi, s)}{n_\alpha(\pi, s)} \mathrm{Id} + R_{Q|P}^{-1}(\pi, s) R_{Q|P}'(\pi, s).$$

Recall that j_P can be chosen to be isometric, that is $\|j_P\| = 1$. We conclude the proof by applying Lemmas 5.1 and 5.3.   □

Given $L \in \mathcal{L}(M)$, we let $\mathfrak{B}_{P,L}$ be the set of tuples $\underline{\beta} = (\beta_1^\vee, ..., \beta_m^\vee)$ of $\Sigma_P^\vee$ with $m := \dim \mathfrak{a}_L^G$, such that their projection to $\mathfrak{a}_L^G$ form a basis of $\mathfrak{a}_L^G$. We denote $\mathrm{vol}(\underline{\beta})$ to be the covolume of the lattice generated by the projection of $\beta_1^\vee, ..., \beta_m^\vee$ in $\mathfrak{a}_L^G$. Let

$$\Xi_L(\underline{\beta}) := \{(P_1, ..., P_m, P_1', ..., P_m') \in \mathcal{P}(M)^{2m} \mid P_i, P_i' \in \mathcal{P}(M), P_i|^{\beta_i} P_i'\}.$$

In [21, pp.179–180], (also see [22, pp.608–609]) the authors define a map[9]

$$\mathfrak{B}_{P,L} \ni \underline{\beta} \mapsto \xi(\underline{\beta}) \in \Xi_L(\underline{\beta}).$$

For $P|^\beta P'$, we find $\alpha_{P,P'} \in \mathfrak{a}_M^*$, such that

$$Y_P(T) - Y_{P'}(T) = \langle \alpha_{P,P'}, T - T_0 \rangle \beta^\vee.$$

Such an $\alpha_{P,P'}$ exists (see [21, pp.186]). Finally, for $\xi \in \Xi_L(\underline{\beta})$, we define the operator

$$\Delta_\xi^T(P, \lambda) := \frac{\mathrm{vol}(\underline{\beta})}{m!} M_{P|P_1'}(\lambda) \left(\delta_{P_1|P_1'}(\lambda) + \langle \alpha_{P_1,P_1'}, T - T_0 \rangle \mathrm{Id} \right) M_{P_1'|P_2'}(\lambda)$$

$$\dots \left(\delta_{P_m|P_m'}(\lambda) + \langle \alpha_{P_m,P_m'}, T - T_0 \rangle \mathrm{Id} \right) M_{P_m'|P}(\lambda). \quad (5.3)$$

Note that the definition of $\Delta_\xi^T(P, \lambda)$, here, looks a little different from [21, pp.186], although they are the same. This is because the definition of $\delta_{P_i|P_i'}$ in [21, pp.179] is different than ours (also compare with the definition of $\Delta_\xi^T(P, \lambda)$ in [22, pp.609]).

Following, Proposition 5.7 expresses the operator $\mathcal{M}_L^T$ in terms of Δ_ξ^T. A proof of this is essentially contained in [21, Theorem 4].

Proposition 5.7. *Keep the same notations as in Proposition 5.5. We have*

$$\mathcal{M}_L^T(P, \lambda) = \sum_{\underline{\beta} \in \mathfrak{B}_{P,L}} \Delta_{\xi(\underline{\beta})}^T(P, \lambda)$$

as operators on $\mathcal{A}_\pi^2(P)^{\tau,K}$.

Proof. We closely follow the proof of [21, Theorem 4]. The Taylor expansion $\mathcal{A}_Q$ of

$$M_{Q|P}(\lambda)^{-1} M_{Q|P}(\lambda + \Lambda), \quad \text{at } \Lambda = 0$$

form a compatible family (in the sense of [21]), when restricted to the finite-dimensional subspace $\mathcal{A}_\pi^2(P)^{\tau,K}$. This follows from the argument in [21, pp.185] for $s = 1$. On the other hand, the Taylor series c_Q of $e^{\langle \Lambda, Y_Q(T) \rangle}$ at $\Lambda = 0$ form a scalar-valued compatible family. Thus, following the argument in [21, pp.186], we conclude that $c_Q \mathcal{A}_Q$ form a compatible family when restricted to the finite-dimensional subspace $\mathcal{A}_\pi^2(P)^{\tau,K}$. Hence, as in [21], we compute the limit of (5.1) at $\Lambda = 0$ using [17, Theorem 8.1], which concludes the proof.   □

Finally, we can prove Proposition 5.5.

---

[9]The definition of ξ also depends on L and on a certain $\mu \in \mathfrak{a}_M^*$, but we do not need the precise definition of the map.

Proof of Proposition 5.5. Using Proposition 5.7, we see that it is enough to bound for each $\beta \in \mathfrak{B}_{P,L}$,

$$\int_{\substack{\lambda \in i(\mathfrak{a}_L^G)^* \\ \|\lambda - \lambda_0\| \leq 1}} \left\| \Delta^T_{\xi(\underline{\beta})}(P,\lambda)|_{\mathcal{A}^2_\pi(P)^{\tau,K}} \right\|_{\mathrm{op}} \mathrm{d}\lambda \ll (\|T\| \log(1 + \nu(\tau) + \nu(\pi_\infty) + \|\lambda_0\| + \mathrm{level}(K)))^m.$$

Let $\vartheta_i \in (\mathfrak{a}_L^G)^*$ be the dual basis to $\beta_j^\vee$. Using this basis, the above integral can be majorized by

$$\ll \int_{\substack{\lambda = i \sum_{j=1}^m t_j \vartheta_j \\ t_j \in \mathbb{R}, \, |t_j - t_{0,j}| \leq C}} \left\| \Delta^T_{\xi(\underline{\beta})}(P,\lambda)|_{\mathcal{A}^2_\pi(P)^{\tau,K}} \right\|_{\mathrm{op}} \mathrm{d}\lambda,$$

where $it_{0,j} := \langle \lambda_0, \beta_j^\vee \rangle$ and C is a constant, depending only on the root system. Now we recall (5.3) and use the fact that $M_{Q|P}(\lambda)$ is unitary for purely imaginary λ to see that the above integral can be bounded by

$$\prod_{j=1}^m \int_{|t_j - t_{0,j}| \leq C} \left(\left\| \delta_{P_j|P_j'}(i\vartheta_j t_j)|_{\mathcal{A}^2_\pi(P_j')^{\tau,K}} \right\|_{\mathrm{op}} + \|T\| \right) \mathrm{d}t_j.$$

We conclude the proof by dividing the above interval in the j'th integral in $O(1)$ number of length one subinterval centring the points of size $O(t_{0,j})$ and applying Lemma 5.6 for the integral over each subinterval. $\qquad\square$

6. The spectral side of the trace formula

For any $M \in \mathcal{L}$ and $\pi \in \Pi_2(M)$, we denote $\mathcal{B}_{\chi,\pi}(P)$ to be an orthonormal basis of $\mathcal{A}^2_{\chi,\pi}(P)$. Let $h \in C_c^\infty(G(\mathbb{A})^1)$ be a $K_\infty K_f$-finite function and $T \in \mathfrak{a}^G_{M_0}$ with sufficiently large $d(T)$.

Let χ be cuspidal data. In [1, Section 3], Arthur defined a distribution $J^T_\chi(h)$ via a *geometric* expansion (for $d(T)$ sufficiently large) and showed that $J^T_\chi(h)$ is a polynomial in T. We do not need to know the exact definition of the distribution $J^T_\chi(h)$ but will record *spectral* formulæ for it that will be useful for us.

We recall certain norm $\|g\|$ on $G(\mathbb{A}) \ni g$ as defined in [3, pp.1252–1253]. For $P \supset P_0$, we let

$$n_P := \sum_{P_0 \subset P' \sim P} |W(\mathfrak{a}_P, \mathfrak{a}_{P'})|,$$

where $\sim$ denotes association.

Proposition 6.1. *[3, Proposition 2.2] There is a constant $C_0 > 0$, such that if*

$$d(T) \geq C_0(1 + \sup\{\log \|g\| \mid g \in \mathrm{supp}(h)\}),$$

then

$$J^T_\chi(h) = \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_{\chi,\pi}(P)} \int_{i(\mathfrak{a}_P^G)^*} \langle \Lambda^T \mathrm{Eis}(\rho(P,\lambda,h)\varphi, \lambda), \Lambda^T \mathrm{Eis}(\varphi, \lambda) \rangle_{\mathcal{A}^2(G)} \, \mathrm{d}\lambda,$$

and moreover, is a polynomial in T.

The π-sum in Proposition 6.1 is actually finite (see [3, pp.1245]), although we will not use this fact in this paper.

The polynomial expression of $J_\chi^T(h)$ can be explicitly calculated. We record the expression in the following Proposition 6.2.

For $L \in \mathcal{L}(M)$, let $W^L(M_P)$ be the set of $w \in W(M)$, such that

$$\{H \in \mathfrak{a}_M \mid wH = H\} = \mathfrak{a}_L.$$

We record the following important proposition. For each $w \in W^L(M_P)$, we consider the intertwining operator $M(P, w)$ defined in [4, pp.1309]. We do not need an explicit definition (which is a bit convoluted) of this operator but only need some of its properties, which are described below.

Proposition 6.2. *The quantity* $J_\chi^T(h)$ *is equal to*

$$\sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{L \in \mathcal{L}(M_P)} \sum_{w \in W^L(M_P)} \iota_w \int_{i(\mathfrak{a}_L^G)^*} \mathrm{tr}\Big(\mathcal{M}_L^T(P, \lambda) M(P, w) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\Big) \mathrm{d}\lambda,$$

where $\iota_w := \left|\det\left((w - 1)\big|_{\mathfrak{a}_M^L}\right)\right|.$

Arthur in [3, 4] proved Proposition 6.2 only for $T = T_0$, but below, we show that the same arguments can also be made to work for general T.

Proof. We fix a Weyl group invariant test function $B \in C_c^\infty(i(\mathfrak{a}_{M_0}^G)^*)^{W_0}$ with $B(0) = 1$. As in [3, pp.1273], for $\epsilon > 0$ and $\pi \in \Pi_2(M)$, we define

$$B_\pi^\epsilon(\lambda) := B(\epsilon(i\Im(\mu_\pi) + \lambda)), \quad \lambda \in i(\mathfrak{a}_L^G)^*,$$

where $\mu_\pi \in \mathfrak{a}_{M_0,\mathbb{C}}^*$ is the infinitesimal character of π. Combining [3, Theorem 6.3(ii)] and [4, Theorem 4.1], we obtain that $J_\chi^T(h)$ is equal to the limit of

$$\sum_{P \supset P_0} \frac{1}{n_P} \sum_{L \in \mathcal{L}(M_P)} \sum_{w \in W^L(M_P)} \iota_w \sum_{\pi \in \Pi_2(M_P)} \int_{i(\mathfrak{a}_L^G)^*} \mathrm{tr}\Big(\mathcal{M}_L^T(P, \lambda) M(P, w) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\Big) B_\pi^\epsilon(\lambda) \, \mathrm{d}\lambda$$

as $\epsilon \to 0$. Now, if we show that for each $P \in \mathcal{P}$, $L \in \mathcal{L}(M_P)$, and $w \in W^L(M_P)$

$$\sum_{\pi \in \Pi_2(M_P)} \int_{i(\mathfrak{a}_L^G)^*} \left|\mathrm{tr}\Big(\mathcal{M}_L^T(P, \lambda) M(P, w) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\Big)\right| \mathrm{d}\lambda < \infty,$$

then by dominated convergence, we can pass the $\epsilon \to 0$ limit inside and conclude.

Now we work as in [21, Section 5.1] to show the above absolute convergence. As the intertwiner $M(P, w)$ commutes with the group action ρ and is unitary (see [21, pp.184]), we have

$$\mathrm{tr}\Big(\mathcal{M}_L^T(P, \lambda) M(P, w) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\Big) \leq \left\|\mathcal{M}_L^T(P, \lambda) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\right\|_1.$$

Here, on the right-hand side above and elsewhere, $\|\cdot\|_1$ denotes the trace norm of the corresponding operator. Moreover, using Proposition 5.7, it suffices to show that

$$\sum_{\pi \in \Pi_2(M_P)} \int_{i(\mathfrak{a}_L^G)^*} \left\|\Delta_\xi^T(P, \lambda) \rho(P, \lambda, h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\right\|_1 \mathrm{d}\lambda < \infty,$$

for any $\beta \in \mathcal{B}_{P,L}$ and $\xi \in \Xi_L(\beta)$. As h is K_f-finite, there exists an open-compact $K_h \subset K_f$, such that h is right $\overline{K}_h$-invariant. Now decomposing into K_∞-types, it suffices to show that

$$\sum_{\tau \in \overline{K_\infty}} \sum_{\pi \in \Pi_2(M_P)} \int_{i(\mathfrak{a}_L^G)^*} \left\| \Delta_\xi^T(P,\lambda)\rho(P,\lambda,h)|_{\mathcal{A}_{\chi,\pi}^2(P)^{\tau,K_h}} \right\|_1 \, \mathrm{d}\lambda < \infty.$$

Working, as in [21, Section 5.1], with the Laplacian on $G(F_\infty)$, noting that h is compactly supported and using $\|A|_V\|_1 \le \dim(V)\|A\|_{\mathrm{op}}$, we obtain the above integral is bounded by

$$\ll_{h,N} \dim\!\left(\mathcal{A}_{\chi,\pi}^2(P)^{\tau,K_h}\right) \int_{i(\mathfrak{a}_L^G)^*} (1+\|\lambda\|+\nu(\pi_\infty)+\nu(\tau))^{-N} \left\| \Delta_\xi^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{\tau,K_h}} \right\|_{\mathrm{op}} \mathrm{d}\lambda,$$

for any $N > 0$. Now we use (5.3) and work as in the proof of Proposition 5.5 to bound

$$\left\| \Delta_\xi^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{\tau,K_h}} \right\|_{\mathrm{op}} \ll_T \prod_j \left(1 + \left\| \delta_{P_j|P_j'}(i\vartheta_j t_j)|_{\mathcal{A}_\pi^2(P_j')^{\tau,K_h}} \right\|_{\mathrm{op}} \right),$$

where, following the notation in the proof of Proposition 5.5, above, we have written $\lambda = i\sum_j t_j \vartheta_j$ with $\vartheta_j \in (\mathfrak{a}_L^G)^*$. Finally, from here on, proceeding as in [21, Section 5.1], we conclude the proof. □

We denote $\mathcal{A}_\pi^2(P) := \bigoplus_\chi \mathcal{A}_{\chi,\pi}^2(P)$ and $\mathcal{B}_\pi(P) := \bigcup_\chi \mathcal{B}_{\chi,\pi}(P)$, which is an orthonormal basis of $\mathcal{A}_\pi^2(P)$. On the other hand, the sum

$$J^T(h) := \sum_\chi J_\chi^T(h), \tag{6.1}$$

where χ runs over all equivalence classes of cuspidal data, is absolutely convergent (see [6, pp.80]).

Proposition 6.3. *Let T be as in Proposition 6.1. Then*

$$J^T(h) = \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_\pi(P)} \int_{i(\mathfrak{a}_P^G)^*} \langle \Lambda^T \mathrm{Eis}(\rho(P,\lambda,h)\varphi,\lambda), \Lambda^T \mathrm{Eis}(\varphi,\lambda)\rangle_{\mathcal{A}^2(G)} \, \mathrm{d}\lambda$$

$$= \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{L \in \mathcal{L}(M_P)} \sum_{w \in W^L(M_P)} \iota_w \int_{i(\mathfrak{a}_L^G)^*} \mathrm{tr}\!\left(\mathcal{M}_L^T(P,\lambda)M(P,w)\rho(P,\lambda,h)|_{\mathcal{A}_\pi^2(P)}\right) \mathrm{d}\lambda,$$

where ι_w and $M(P,w)$ are as in Proposition 6.2.

Proof. The proofs are similar (almost verbatim) to those of Propositions 6.1 and 6.2, after replacing $\mathcal{A}_{\chi,\pi}^2$ and $\mathcal{B}_{\chi,\pi}$ with $\mathcal{A}_\pi^2$ and $\mathcal{B}_\pi$, respectively. □

7. Bounds for a specific maximal compact

In this section, we prove Theorem 3. First, we develop some necessary local ingredients that go in the proof.

7.1. Choice of test function

We freely use the notations of Section 4.2. Recall that $\mathfrak{a}_\infty$ is the Lie algebra of a maximal $\mathbb{R}$-split torus of $G(F_\infty)$. Let $\delta > 0$ be a sufficiently small constant. We choose a fixed L^1-normalized nonnegative test function $f_{0,\delta}$ on $\mathfrak{a}_\infty$ supported on the ball of radius $\delta/2$ around 0. Let $f_\delta := f_{0,\delta} * f_{0,\delta}^*$, where for any

$f \in C_c(\mathfrak{a}_\infty)$, we define $f^*(\alpha) := \overline{f(-\alpha)}$. It holds that $\widehat{f^*}(\mu) = \overline{\hat{f}(-\overline{\mu})}$. Thus, $\widehat{f_\delta}$ is nonnegative on $i\mathfrak{a}_\infty^*$. Note that f_δ is also nonnegative, L^1-normalized and supported on the ball of radius δ around 0.

We fix

$$C_0 := 2\|\rho_\infty\| + 1. \tag{7.1}$$

Now, we make $\delta > 0$ small enough so that

$$\mathfrak{R}(\widehat{f_\delta}(\mu)) \geq 1/2, \quad \text{for } \|\mu\| \leq C_0. \tag{7.2}$$

This follows from continuity and the fact that $\widehat{f_\delta}(0) = 1$.

Given $\mu_0 \in i\mathfrak{a}_\infty^*$, we define $f_{\mu_0} \in C_c^\infty(\mathfrak{a}_\infty)^{W_\infty}$ by

$$f_{\mu_0}(\alpha) := \sum_{w \in W_\infty} f_\delta(w\alpha) e^{-\langle \mu_0, w\alpha \rangle},$$

so that

$$\widehat{f_{\mu_0}}(\mu) = \sum_{w \in W_\infty} \widehat{f_\delta}(w\mu - \mu_0), \quad \mu \in \mathfrak{a}_{\infty,\mathbb{C}}^*.$$

Finally, we define

$$h_{\mu_0} := \mathcal{S}^{-1}(f_{\mu_0}^* * f_{\mu_0}) \in C_c^\infty(K_\infty \backslash G(F_\infty)/K_\infty).$$

Thus, the spherical transform of h_{μ_0} is

$$\tilde{h}_{\mu_0}(\mu) = \widehat{f_{\mu_0}}(\mu) \overline{\widehat{f_{\mu_0}}(-\overline{\mu})}$$

(see Section 4.4). Then h_{μ_0} satisfies the following properties:

1. The support of h_{μ_0} is bounded independently of μ_0, which follows from (4.1).
2. $\tilde{h}_{\mu_0}(\mu) \geq 0$, if $\mu = -w\overline{\mu}$ for some $w \in W_\infty$; in particular, if $\mu \in i\mathfrak{a}_\infty^*$, or more generally, if μ is the Langlands parameter of a unitary representation. This is immediate from W_∞-invariance of $\widehat{f_{\mu_0}}$.
3. There is a constant b depending only on the support of f_δ, such that for every $N > 0$, it holds that

$$|\tilde{h}_{\mu_0}(\mu)| \ll_N \sum_{w \in W_\infty} \exp(b\|\mathfrak{R}(\mu)\|)(1 + \|w\mu - \mu_0\|)^{-N},$$

   which follows from (4.2).
4. For C_0 as in (7.1), we have that if $\|\mu - \mu_0\| \leq C_0$, then $|\tilde{h}_{\mu_0}(\mu)| \geq 1/10$, perhaps after making $\delta > 0$ smaller.

To see the last claim, it suffices to show that

$$\mathfrak{R}\left(\widehat{f_{\mu_0}}(\mu)\right) \geq 1/3, \quad \text{for } \|\mu - \mu_0\| = \|-\overline{\mu} - \mu_0\| < C_0.$$

Working similarly to the proof of [21, Lemma 3.2], below, we prove the above sufficient formulation.

We see that

$$\widehat{f_\delta}(\mu) = \int\limits_{\mathfrak{a}_\infty} f_\delta(\alpha) \Big(e^{i\langle \Im(\mu), \alpha \rangle} - e^{i\langle \Im(\mu), \alpha \rangle} + e^{\langle \mu, \alpha \rangle} \Big) \, d\alpha$$

$$= \widehat{f_\delta}(i\Im(\mu)) + \int\limits_{\mathfrak{a}_\infty} f_\delta(\alpha) e^{i\langle \Im(\mu), \alpha \rangle} \Big(e^{\langle \Re(\mu), \alpha \rangle} - 1 \Big) \, d\alpha.$$

As $\|\Re(\mu)\| \le C_0$, we have $|e^{\langle \Re(\mu), \alpha \rangle} - 1| < C_1 \delta$ for α lying in a δ-radius ball around the origin and for certain C_1 depending on C_0. Consequently, we have

$$\Re(\widehat{f_\delta}(\mu)) \ge -C_1 \delta,$$

as $\widehat{f_\delta}(i\Im(\mu)) \ge 0$. Thus, upon applying (7.2), and the fact that

$$\|\Re(w\mu - \mu_0)\| = \|\Re(\mu)\| = \|\Re(\mu - \mu_0)\| \le \|\mu - \mu_0\| \le C_0,$$

we obtain

$$\Re(\widehat{f_{\mu_0}}(\mu)) = \Re(\widehat{f_\delta}(\mu - \mu_0)) + \sum_{w \ne 1} \Re(\widehat{f_\delta}(w\mu - \mu_0)) \ge 1/2 - C_2 \delta,$$

for some C_2 depending on C_0 and the group. We conclude by making δ sufficiently small.

7.2. Langlands parameters

Let $\chi := (M_1, \pi_1)$ be a cuspidal datum, determined by a Levi subgroup M_1 and a representation $\pi_1 \in \Pi_2(M_1)$. Let $P_1 \supset P_0$ be such that $M_1 = M_{P_1}$. We assume that $\pi_{1,\infty}$ is spherical, that is, has a nonzero $K_\infty \cap M_1(F_\infty)$-invariant vector, so we may associate to it a Langlands parameter μ_{π_1}.

We first notice that χ is actually an equivalence class, whose size by [6, Section 12], is at most

$$n_{P_1} = \sum_{P_0 \subset P' \sim P_1} |W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P'})|.$$

We let $(M_1, \pi_1), ..., (M_k, \pi_k)$ be the elements in this equivalence class, with $k \le n_{P_1} \le n_{P_0}$, and $M_i = M_{P_i}$ for $P_i \supset P_0$, all associated with P_1. Notice that the Langlands parameter μ_{π_i} is in the W_∞-orbit of μ_{π_1}, so $\|\mu_{\pi_i}\| = \|\mu_{\pi_1}\|$.

Now, let $P \supset P_0$ and $\pi \in \Pi_2(M_P)$ so that $\mathcal{A}^2_{\chi,\pi}(P)^{K_\infty K} \ne \{0\}$. It follows from the Langlands construction of the residual spectrum that π is defined using iterated residues from an Eisenstein series of one of the π_i, $1 \le i \le k$, for some $P_i \sim P$. Let $\rho_\pi \in (\mathfrak{a}^M_{M_i})^* \subset (\mathfrak{a}^G_{M_0})^* \subset (\mathfrak{a}_\infty)^*$ be the point where the residual representation is defined by iterated residues (see, e.g. [5, Section 2]). Then $\mu_\pi = \mu_{\pi_i} + \rho_\pi$. The crucial point here is that ρ_π is real and universally bounded (in fact, $\|\rho_\pi\| \le \|\rho_\infty\|$, where ρ_∞ is as in Section 4.2). Therefore, $\|\mu_\pi - \mu_{\pi_i}\| \le \|\rho_\infty\|$ and

$$1 + \|\mu_\pi\| \asymp 1 + \|\mu_{\pi_i}\|. \tag{7.3}$$

The above properties of ρ_π follow from Langlands' construction of the residual spectrum (see e.g. [47, pp.1152, Proof of Lemma 4.3]).

Next, the Langlands parameter of $\mathrm{Ind}_{P(F_\infty)}^{G(F_\infty)} \Big(\pi_\infty \otimes e^{\langle \lambda, H_{M_P}(\cdot) \rangle} \Big)$ for $\lambda \in (\mathfrak{a}^G_{P,\mathbb{C}})^*$ is $\mu_{\pi,\lambda} := \mu_\pi + \lambda$.[10] If $\lambda \in i(\mathfrak{a}^G_P)^*$, then this induction is unitary. Correspondingly, it holds that

---

[10]Here, we identify $i(\mathfrak{a}^G_P)^* \subset \mathfrak{a}^*_{\infty,\mathbb{C}}$.

○ $\|\mathfrak{R}(\mu_{\pi,\lambda})\| \le \|\rho_\infty\|$, and
○ there is a $w \in W_\infty$ so that $-\overline{\mu_{\pi,\lambda}} = w\mu_{\pi,\lambda}$ (as a matter of fact, we may choose $w \in N_{M_P(F_\infty)}(T_\infty)/T_\infty \subset W_\infty$)

(see, e.g. [15, Proposition 3.4].
   We now prove Theorem 3.

Proof of Theorem 3. We use property (3) in Section 3.4 to assume that T in Theorem 3 is such that $d(T)$ is sufficiently large (in terms of G and F) but fixed.

   We assume that the cuspidal datum $\chi := (M, \pi_0)$ with the equivalence class $(M_1, \pi_1), \ldots, (M_k, \pi_k)$ as above. Recall that $\|\mu_{\pi_0}\| \asymp \|\mu_{\pi_i}\|$.

   Let $i\mathfrak{a}_\infty \ni \mu_j := i\mathfrak{I}(\mu_{\pi_j}) + \lambda_0$ for $j = 1, \ldots, k$. We construct h_{μ_j} as in Section 7.1, with the choice of C_0 given in (7.1). Now let $h_j \in C_c^\infty(G(\mathbb{A})^1)$ be of the form

$$h_j(g_\infty, g_f) := h_{\mu_j}(g_\infty) \frac{\mathbb{1}_K(g_f)}{\mathrm{vol}(K)}.$$

Note that Item 1 implies that support of h_j has bounded measure.
   Clearly, for any $P \supset P_0$ and $\lambda \in i(\mathfrak{a}_P^G)^*$, the operator $\rho(P, \lambda, h_j)$ projects on $\mathcal{A}^2(P)^{K_\infty K}$. Moreover, for any $\pi \in \Pi_2(M_P)$ and $\varphi \in \mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}$, we have

$$\rho(P, \lambda, h)\varphi = \tilde{h}_{\mu_j}(\mu_{\pi,\lambda})\varphi.$$

Recall from the discussion about the Langlands parameters above that there exists $w \in W_\infty$ so that $w\mu_{\pi,\lambda} = -\overline{\mu_{\pi,\lambda}}$. Thus, using Item 2, we obtain that

$$\tilde{h}_{\mu_j}(\mu_{\pi,\lambda}) \ge 0.$$

   On the other hand, if $\|\lambda - \lambda_0\| \le 1$, and $P \supset P_0$ and $\pi \in \Pi_2(M_P)$ are such that $\mathcal{A}_{\chi,\pi}^2(P) \ne \{0\}$, then by the discussion above, there is $1 \le j \le k$, such that

$$\|\mu_{\pi,\lambda} - \mu_j\| \le \|\mathfrak{R}(\mu_{\pi_j})\| + \|\rho_\pi\| + \|\lambda - \lambda_0\| \le 2\|\rho_\infty\| + 1 = C_0.$$

Hence, applying Item 4 it holds that

$$\tilde{h}_{\mu_j}(\mu_{\pi,\lambda}) \ge 1/10.$$

   Now, we fix $T \in \mathfrak{a}_{M_0}^G$ with sufficiently large $d(T)$ and apply Proposition 6.1 to write

$$J_\chi^T(h_j) = \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_{\chi,\pi}(P)^{K_\infty K}} \int_{i(\mathfrak{a}_P^G)^*} \tilde{h}_{\mu_j}(\mu_{\pi,\lambda}) \|\Lambda^T \mathrm{Eis}(\varphi, \lambda)\|_2^2 \, d\lambda.$$

Here, $\mathcal{B}_{\chi,\pi}(P)^{K_\infty K}$ is an orthonormal basis of $\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}$. Using the above properties of $\tilde{h}_{\mu_j}$, we obtain that

$$\sum_j J_\chi^T(h_j) \gg \sum_{P \supset P_0} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_{\chi,\pi}(P)^{K_\infty K}} \int_{\substack{\lambda \in i(\mathfrak{a}_P^G)^* \\ \|\lambda - \lambda_0\| \le 1}} \|\Lambda^T \mathrm{Eis}(\varphi, \lambda)\|_2^2 \, d\lambda.$$

Note that the right-hand side above equals to the left-hand side of the estimate in Theorem 3.
   Hence, it remains to bound $J_\chi^T(h_j)$ for $j = 1, \ldots, k$. As the proof is the same for all j without loss of generality, we will bound for $j = 1$ and write $h = h_1$.

Using Proposition 6.2, we see that

$$J_\chi^T(h) \ll \sum_{P \supset P_0} \max_{L \in \mathcal{L}(M_P)} \max_{w \in W^L(M_P)} \sum_{\pi \in \Pi_2(M_P)} \int_{i(\mathfrak{a}_L^G)^*} \mathrm{tr}\Big(\mathcal{M}_L^T(P,\lambda) M(P,w) \rho(P,\lambda,h)|_{\mathcal{A}_{\chi,\pi}^2(P)}\Big) \, d\lambda.$$

Recall that $\rho(P,\lambda,h)|_{\mathcal{A}_{\chi,\pi}^2(P)}$ projects on $\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}$ and acts there by the scalar $\tilde{h}_{\mu_1}(\mu_\pi + \lambda)$. Noting that $M(P,w)$ is unitary and commutes with $\rho(P,\lambda,h)$, we get that the integral on the right-hand side above is bounded by

$$\dim\left(\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}\right) \int_{i(\mathfrak{a}_L^G)^*} \tilde{h}_{\mu_1}(\mu_\pi + \lambda) \left\|\mathcal{M}_L^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right\|_{\mathrm{op}} d\lambda.$$

Applying Item 3, we can bound the inner integral in the above display by

$$\ll_N \max_{w \in W_\infty} \int_{i(\mathfrak{a}_L^G)^*} \left(1 + \|w\lambda + wi\Im(\mu_\pi) - \lambda_0 - i\Im(\mu_{\pi_1})\|\right)^{-N} \left\|\mathcal{M}_L^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right\|_{\mathrm{op}} d\lambda. \quad (7.4)$$

For each $w \in W_\infty$, we define

$$\lambda_1 := \lambda_{1,w,\pi,\pi_1} := w^{-1}\big(\lambda_0 - i\Im(w\mu_\pi - \mu_{\pi_1})\big) \in i(\mathfrak{a}_\infty)^*.$$

Using (7.3), we obtain

$$\|\lambda_1\| \ll \|\lambda_0\| + \|\mu_\pi\| + \|\mu_{\pi_1}\| \asymp \|\lambda_0\| + \sqrt{\nu(\pi_{1,\infty})}.$$

Thus, for each w, we bound the integral on the right-hand side of (7.4) by

$$\int_{i(\mathfrak{a}_L^G)^*} \left(1 + \|\lambda - \lambda_1\|\right)^{-N} \left\|\mathcal{M}_L^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right\|_{\mathrm{op}} d\lambda$$

$$\leq \sum_{m=1}^{\infty} \frac{1}{m^N} \int_{\substack{\lambda \in i(\mathfrak{a}_L^G)^* \\ m-1 \leq \|\lambda - \lambda_1\| \leq m}} \left\|\mathcal{M}_L^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right\|_{\mathrm{op}} d\lambda. \quad (7.5)$$

We can find $\{\eta_l\}_{l=1}^n \subset i(\mathfrak{a}_L^G)^*$ with $\|\eta_l\| \leq m$ and $n \ll m^{\dim \mathfrak{a}_L^G}$, such that

$$\{\lambda : \|\lambda - \lambda_1\| \leq m\} \subset \cup_{l=1}^n \{\lambda : \|\lambda - \lambda_1 - \eta_l\| \leq 1\}.$$

Thus, the integral on the right-hand side of (7.5) is bounded by

$$\sum_{l=1}^n \int_{\substack{\lambda \in i(\mathfrak{a}_L^G)^* \\ \|\lambda - \lambda_1 - \eta_l\| \leq 1}} \left\|\mathcal{M}_L^T(P,\lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}}\right\|_{\mathrm{op}} d\lambda.$$

We apply Proposition 5.5 to estimate the above integral and obtain that

$$
\int\limits_{i(\mathfrak{a}_L^G)^*} \tilde{h}_{\mu_1}(\mu_\pi + \lambda) \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}} \right\|_{\mathrm{op}} \, d\lambda
$$

$$
\ll_N \max_{w \in W_\infty} \sum_{m=1}^\infty \frac{1}{m^N} \sum_{l=1}^n (\|T\| \log(1 + \|\lambda_1 + \eta_l\| + \nu(\pi_\infty) + \mathrm{level}(K)))^{\dim \mathfrak{a}_L^G}.
$$

Using the bounds of λ_1, η_l, and n, the right-hand side above is bounded by

$$
\ll_N \sum_{m=1}^\infty \frac{1}{m^{N-\dim \mathfrak{a}_L^G}} \left(\|T\| \log\left(1 + m + \|\lambda_0\| + \nu(\pi_{1,\infty}) + \mathrm{level}(K)\right)\right)^{\dim \mathfrak{a}_L^G}
$$

$$
\ll \left(\|T\| \log\left(1 + \|\lambda_0\| + \nu(\pi_{1,\infty}) + \mathrm{level}(K)\right)\right)^{\dim \mathfrak{a}_L^G},
$$

by making N large enough. Thus, we obtain

$$
J_\chi^T(h) \ll \sum_{P \supset P_0} \max_{L \in \mathcal{L}(M_P)} \sum_{\pi \in \Pi_2(M_P)} \dim \left(\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K}\right)
$$

$$
\times \left(\|T\| \log\left(1 + \|\lambda_0\| + \nu(\pi_{1,\infty}) + \mathrm{level}(K)\right)\right)^{\dim \mathfrak{a}_L^G}.
$$

We conclude noting that $\max_{L \in \mathcal{L}(M_P)} \dim \mathfrak{a}_L^G = \dim \mathfrak{a}_M^G$. □

We now prove Corollary 1.6 and Theorem 1.

Proof of Corollary 1.6. By symmetries of the Eisenstein series, we may assume that $P \supset P_0$. Let $\chi = (M, \pi_0)$ be as in the statement of the corollary. Then, by (7.3), it holds that $1 + \nu(\pi_\infty) \asymp 1 + \nu(\pi_{0,\infty})$ and the claim follows from Theorem 3. □

Proof of Theorem 1. We will prove the slightly stronger statement given in Remark 1.2.

The classification of the residual spectrum for GL_n due to Mœglin and Waldspurger [45] (see Section 3.2) implies that for every χ and $P \supset P_0$ it holds that

$$
\#\{\pi \in \Pi_2(M_P) \mid \mathcal{A}_{\chi,\pi}^2(P) \neq \{0\}\} \ll_n 1.
$$

The classification also implies that the multiplicity of π in $L^2_{\mathrm{disc},\chi}(A_{M_P} M_P(F) \backslash M(\mathbb{A}))$ is at most 1. So the representation $\rho(P, 0, \cdot)$ on $\mathcal{A}_{\chi,\pi}^2(P)$ is isomorphic to $\mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi$, and, in particular

$$
\mathcal{A}_{\chi,\pi}^2(P)^{K_\infty K} = (\mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi)^{K_\infty K}.
$$

On the other hand, we may write $\pi = \otimes_v' \pi_v$ and obtain

$$
(\mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} \pi)^{K_\infty K} = \bigotimes_v {}' (\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v)^{K_v},
$$

where $K_\infty K = \prod_v K_v$.

Now using uniform admissibility due to Bernstein [10], which says that

$$
\dim(\mathrm{Ind}_{P(F_v)}^{G(F_v)} \pi_v)^{K_v} \ll_{K_v} 1,
$$

we deduce that

$$F(\chi; \text{triv}, K) = \sum_{P \supset P_0} \sum_{\pi \in \Pi_2(M_P)} \dim\left(\mathcal{A}^2_{\chi,\pi}(P)^{K_\infty K}\right) \ll_{n,K} 1.$$

Let χ_0 be a cuspidal datum and $\pi_0 \in \Pi_2(M_P)$, such that $\varphi_0 \in \mathcal{A}^2_{\chi_0,\pi_0}(P)^{K_\infty K}$. Then we have $\nu(\varphi_0) \asymp \nu(\pi_{0,\infty})$. Now the result follows from Corollary 1.6 along with Propositions 5.2 and 5.4. □

8. Optimal lifting after Assing–Blomer

In this section, we prove Theorems 4 and 5. Aligning with our usual convention of suppressing the dependence on G in Vinogradov's notations, we also suppress the dependence on n in this section.

We start with a discussion of the link between the adelic setting of the group GL_n and the classical setting of $SL_n(\mathbb{R})$. A good place to follow is [39, Section 3].

From now on, let $G := GL_n$ over $\mathbb{Q}$. We identify A_G as the positive scalar matrices in $G(\mathbb{R})$. Note that $G(\mathbb{R})/A_G \cong SL_n(\mathbb{R}) \times \{\pm 1\}$. We can therefore identify

$$G(\mathbb{A})^1 = (G(\mathbb{R})/A_G) \times \prod_p{}' G(\mathbb{Q}_p),$$

where $'$ denotes the restricted product relative to $G(\mathbb{Z}_p)$.

Given an integer $q > 0$ with prime factorization $q = \prod_{p<\infty} p^{r_p}$ with $r_p \in \mathbb{Z}_{\geq 0}$, let $K(q) \subset G(\mathbb{A}_f)$ be the principal congruence subgroup of level q. We have $K(q) = \prod_{p<\infty} K_p(q)$, where

$$K_p(q) := \{g \in G(\mathbb{Z}_p) \mid g \equiv I \mod p^{r_p}\}.$$

The group $SL_n(\mathbb{R})$ acts from the right on the space $G(\mathbb{Q})A_G\backslash G(\mathbb{A})/K(q)$. The space decomposes into $\varphi(q)$ (φ being Euler's totient function) connected components, determined by the determinant map into

$$\mathbb{R}_{>0}\mathbb{Q}^\times\backslash\mathbb{A}^\times/\prod_{p|q}\{a \in \mathbb{Z}_p^\times \mid a \equiv 1 \mod p^{r_p}\mathbb{Z}_p\} \cong (\mathbb{Z}/q\mathbb{Z})^\times.$$

Each orbit is isomorphic to $\Gamma(q)\backslash SL_n(\mathbb{R})$, where $\Gamma(q)$ is the principal congruence subgroup of level q of $\Gamma(1) := SL_n(\mathbb{Z})$.

Let $K_\infty^0 = SO(n)$, which is a maximal compact subgroup of $SL_n(\mathbb{R})$, and an index 2 subgroup of the maximal compact subgroup $K_\infty := O(n)$ of $G(\mathbb{R})$. Let $\|\cdot\|$ denote the Frobenius norm on $SL_n(\mathbb{R})$, given by $g \mapsto \sqrt{\text{tr}(g^\top g)}$. We note that the norm is both left and right K_∞^0-invariant. It holds that

$$\|g_1 g_2\| \leq \|g_1\|\|g_2\|.$$

For $R > n$, we define

$$B_R := \{g \in SL_n(\mathbb{R}) \mid \|g\| \leq R\}.$$

Note that

$$\text{vol}(B_R) = c_n R^{n(n-1)}(1 + o(1)), \tag{8.1}$$

see [16, Appendix 1], which also calculates the precise coefficient c_n.

We now construct a test function in the Archimedean Hecke algebra following [8, Section 2.7]. Let $h_0 \colon \mathbb{R}_{\geq 0} \to [0, 1]$ be a smooth nonnegative function with $h_0(x) = 1$ for $x \in [0, n]$ and $h_0(x) = 0$ for

$x > 2n$. For $R > 0$, we define

$$h_R \in C_c^\infty(K_\infty^0 \backslash \mathrm{SL}_n(\mathbb{R})/K_\infty^0), \quad g \mapsto \frac{1}{\|h_0(\frac{\|\cdot\|}{R})\|_{L^1}} h_0\left(\frac{\|g\|}{R}\right).$$

A direct consequence of [8, (2.34)–(2.35)] is that the spherical transform of h_R satisfies

$$|\tilde{h}_R(\mu)| \ll_{N,\epsilon} (1 + \|\mu\|)^{-N} R^{-\frac{n(n-1)}{2} + n\|\Re(\mu)\|_\infty + \epsilon} \tag{8.2}$$

for $\mu \in \mathfrak{a}_{0,\mathbb{C}}^*$, such that $w\mu = -\overline{\mu}$ for some $w \in W_\infty$ (and, in particular, for any $\mu \in \mathfrak{a}_{0,\mathbb{C}}^*$ corresponding to the Langlands parameters of a unitary representation). Here, $\|\mu\|_\infty$ denotes the ℓ^∞-norm of μ when $\mathfrak{a}_{0,\mathbb{C}}^*$ is realized as a complex n-tuple.

For the non-Archimedean part, we let $h_q \in C_c^\infty(G(\mathbb{A}_f))$ be the normalized characteristic function of $K(q)$, that is

$$h_q := \frac{1}{\mathrm{vol}(K(q))} \mathbb{1}_{K(q)}.$$

Let $h_{R,q,1} \in C_c^\infty(G(\mathbb{A})^1)$ be given by

$$h_{R,q,1}(g_\infty, g_f) := h_R(g_\infty) h_q(g_f).$$

Finally, let $h_{R,q} = h_{R,q,1} * h_{R,q,1}^*$, where $h^*(g) := \overline{h(g^{-1})}$. In other words, we have

$$h_{R,q}(g_\infty, g_f) = (h_R * h_R^*)(g_\infty) h_q(g_f).$$

Clearly, $h_{R,q} \in C_c^\infty(G(\mathbb{A})^1)$ and is bi-$K_\infty^0 K_f$-finite.

The following is the main technical result of this section and will be used to prove Theorems 4 and 5. First, we recall the definition of $J^T(h)$ from (6.1).

Proposition 8.1. *For any q square-free and any $R > 1$, it holds that for every $\epsilon > 0$*

$$J^T(h_{R,q}) \ll_\epsilon \|T\|^{n-1}(Rq)^\epsilon (q + R^{-n(n-1)} q^{n^2}),$$

for any $T \in \mathfrak{a}_{M_0}$.

Remark 8.2. As will be evident from the proof, the summand q comes from the $\varphi(q)$ possible Dirichlet characters modulo q. The factor $R^{-n(n-1)} q^{n^2}$ comes from the contribution of cusp forms. The number of such forms with bounded Laplacian eigenvalues is $\asymp [K(1) : K(q)] \ll q^{n^2}$, and the contribution of each such form to $J^T(h_{R,q})$ *assuming the (GRC)* is $R^{-n(n-1)}$. The results of Assing and Blomer [8, Theorem 1.1] on the *density hypothesis* allow us to prove the proposition unconditionally of the GRC. However, a significant difficulty arises to control the continuous spectrum, as also experienced in [8], managing that is the main contribution of Proposition 8.1.

We will assume Proposition 8.1 for now and prove Theorems 4 and 5. In the next subsection, we prove Proposition 8.6 using Proposition 8.1 and then deduce these two theorems from it in the following two subsections. At the end of this section, we prove Proposition 8.1.

8.1. Initial reductions

Let us fix $R_0 > n$ and a compact domain $\Omega \subset G(\mathbb{Q})\backslash G(\mathbb{A})^1$ of the form $\Omega := G(\mathbb{Q})\backslash B_{R_0} \times K(1)$. The implied constants in various estimates below are allowed to depend on R_0, however, we will not mention

it in the notations. We define the adelic kernel $K_{h_{R,q}}(x,y)$ by

$$K_{h_{R,q}}(x,y) := \sum_{\gamma \in G(\mathbb{Q})} h_{R,q}(x^{-1}\gamma y),$$

for $x, y \in G(\mathbb{A})^1$.

Proposition 8.3. *There is a constant C (possibly depending on R_0), such that if $d(T) \geq C(1+\log(R))$, then it holds that*

$$\int_\Omega K_{h_{R,q}}(x,x)\,\mathrm{d}x \leq J^T(h_{R,q}),$$

for any R and q.

Proof. We start with the spectral decomposition of $K_{h_{R,q}}(x,x)$ (see, e.g. [6, (7.6)]) which implies that

$$\int_\Omega K_{h_{R,q}}(x,x)\,\mathrm{d}x$$
$$= \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_\pi(P)} \int_{i(\mathfrak{a}_P^G)^*} \int_\Omega \mathrm{Eis}(\rho(P,\lambda,h_{R,q})\varphi,\lambda)(x)\overline{\mathrm{Eis}(\varphi,\lambda)(x)}\,\mathrm{d}x\,\mathrm{d}\lambda.$$

Note that absolute convergence of the spectral side [21] allows us to pull the Ω-integral inside. Here, we choose a basis $\mathcal{B}_\pi(P)$, such that each of its elements is either $K_\infty^0 K(q)$-invariant or orthogonal to the $K_\infty^0 K(q)$-invariant subspace.

From the definition of $h_{R,q}$, we note that the operator $\rho(P,\lambda,h_{R,q})$ projects $\mathcal{A}_\pi^2(P)$ onto $\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}$, and acts on it by the scalar $(\tilde{h}_R \cdot \tilde{h}_R^*)(\mu_\pi + \lambda) = |\tilde{h}_R(\mu_\pi + \lambda)|^2$. Focusing on the innermost integral above and assuming that φ is $K_\infty^0 K(q)$-invariant, we have

$$\int_\Omega \mathrm{Eis}(\rho(P,\lambda,h_{R,q})\varphi,\lambda)(x)\overline{\mathrm{Eis}(\varphi,\lambda))(x)}\,\mathrm{d}x = |\tilde{h}_R(\mu_\pi + \lambda)|^2 \int_\Omega |\mathrm{Eis}(\varphi,\lambda)(x)|^2\,\mathrm{d}x.$$

On the other hand, by property (2) of Section 3.4, we have

$$\int_\Omega |\mathrm{Eis}(\varphi,\lambda)(x)|^2\,\mathrm{d}x \leq \int_{G(\mathbb{Q})\backslash G(\mathbb{A})} |\Lambda^T \mathrm{Eis}(\varphi,\lambda)(x)|^2\,\mathrm{d}x$$

for $d(T)$ sufficiently large. Plugging this into the above integral and reverse engineering the above manipulation with $h_{R,q}$, we deduce that

$$\int_\Omega K_f(x,x)\,\mathrm{d}x$$
$$\leq \sum_{P \supset P_0} \frac{1}{n_P} \sum_{\pi \in \Pi_2(M_P)} \sum_{\varphi \in \mathcal{B}_\pi(P)} \int_{i(\mathfrak{a}_P^G)^*} \langle \Lambda^T \mathrm{Eis}(\rho(P,\lambda,h_{R,q})\varphi,\lambda), \Lambda^T \mathrm{Eis}(\varphi,\lambda) \rangle_{\mathcal{A}^2(G)}\,\mathrm{d}\lambda.$$

Now, the statement of the proposition follows from Proposition 6.3. □

Corollary 8.4. *Let q be square-free. There is an $x \in \Omega$, such that*

$$K_{h_{R,q}}(x,x) \ll_\epsilon (Rq)^\epsilon q(1 + R^{-n(n-1)}q^{n^2-1}),$$

for every $\epsilon > 0$.

Proof. The corollary is immediate from Propositions 8.1 and 8.3. □

We move the discussion to the classical language. We let $K^q_{h_R * h_R^*}(x, y)$ be the classical automorphic kernel of $h_R * h_R^*$, that is

$$K^q_{h_R * h_R^*}(x, y) = \sum_{\gamma \in \Gamma(q)} (h_R * h_R^*)(x^{-1} \gamma y),$$

where $x, y \in \Gamma(q) \backslash \mathrm{SL}_n(\mathbb{R})$.

Lemma 8.5. *Let q be square-free, there is an $x \in \Gamma(q) \backslash B_{R_0} \subset \Gamma(q) \backslash \mathrm{SL}_n(\mathbb{R})$, such that*

$$K^q_{h_R * h_R^*}(x, x) \ll_\epsilon (Rq)^\epsilon (q^{-(n^2-1)} + R^{-n(n-1)}),$$

for every $\epsilon > 0$.

Proof. This follows from the classical-adelic dictionary. Let $x_0 \in B_{R_0} \times K(1) \subset G(\mathbb{A})^1$ be a representative point from Corollary 8.4, and let $x := x_{0,\infty} \in B_{R_0} \subset \mathrm{SL}_n(\mathbb{R})$ be its ∞-coordinate.

We claim that

$$K_{h_{R,q}}(x_0, x_0) = \mathrm{vol}(K(q))^{-1} K^q_{h_R * h_R^*}(x, x),$$

which implies the claim together with Corollary 8.4 and the fact that $\mathrm{vol}(K(q)) \ll_\epsilon q^{-n^2+\epsilon}$.

Indeed,

$$K_{h_{R,q}}(x_0, x_0) = \sum_{\gamma \in G(\mathbb{Q})} h_{R,q}(x_0^{-1} \gamma x_0) = \sum_{\gamma \in G(\mathbb{Q})} (h_R * h_R)^*(x_{0,\infty}^{-1} \gamma x_{0,\infty}) h_q(x_{0,f}^{-1} \gamma x_{0,f}).$$

An element $\gamma \in G(\mathbb{Q})$ that has nonzero contribution to the above sum must satisfy $x_{0,\infty}^{-1} \gamma x_{0,\infty} \in \mathrm{SL}_n(\mathbb{R})$, so $\gamma \in \mathrm{SL}_n(\mathbb{R})$. Also, it must hold that $x_{0,f}^{-1} \gamma x_{0,f} \in K(q)$. In addition, since $x_{0,f} \in K(1)$ and $K(q)$ is normal in $K(1)$, it holds that $\gamma \in K(q)$. Therefore, $\gamma \in \Gamma(q) \subset \mathrm{SL}_n(\mathbb{Z})$, and $h_q(x_{0,f}^{-1} \gamma x_{0,f}) = \frac{1}{\mathrm{vol}(K(q))}$. Therefore,

$$K_{h_{R,q}}(x_0, x_0) = \mathrm{vol}(K(q))^{-1} \sum_{\gamma \in \Gamma(q)} (h_R * h_R)^*(x^{-1} \gamma x) = \mathrm{vol}(K(q))^{-1} K^q_{h_R * h_R^*}(x, x),$$

and the proof is complete. □

We now claim that we may replace $x \in B_{R_0}$ with $x = e$ in Lemma 8.5.

Proposition 8.6. *Let q be square-free. It holds that*

$$K^q_{h_R * h_R^*}(e, e) \ll_\epsilon (Rq)^\epsilon (q^{-(n^2-1)} + R^{-n(n-1)}),$$

for every $\epsilon > 0$.

Proof. Let x be the point from Lemma 8.5. We choose for x a representative in B_{R_0}. This implies that for $C > 0$ sufficiently large and for every $g \in \mathrm{SL}_n(\mathbb{R})$, we have

$$\|x^{-1} g\| \le C \|g\|, \quad \|g x\| \le C \|g\|.$$

From the definition of h_R, the above implies that for $c > 0$ sufficiently small (e.g. $c = C^{-1}/2$) and every $g \in \mathrm{SL}_n(\mathbb{R})$

$$h_{cR}(g) \ll h_R(x^{-1} g), \quad h_{cR}^*(g) \ll h_R^*(g x),$$

and therefore, for every $\gamma \in \Gamma(q)$,

$$(h_{cR} * h_{cR}^*)(\gamma) \ll (h_R * h_R^*)(x^{-1}\gamma x).$$

We deduce using Lemma 8.5 that

$$K_{h_{cR}*h_{cR}^*}^q(e,e) \ll K_{h_R*h_R^*}^q(x,x) \ll_\epsilon (Rq)^\epsilon (q^{-(n^2-1)} + R^{-n(n-1)}).$$

After changing R to cR, the proposition follows. $\qquad\qquad\square$

8.2. Proof of Theorem 4

We adopt the notations developed for the classical spectral decomposition of $L^2(\Gamma(q)\backslash\mathrm{SL}_n(\mathbb{R})/K_\infty^0)$ in [8, Section 2.5]. We write the spectral decomposition of the kernel of $h \in C_c^\infty(K_\infty^0\backslash\mathrm{SL}_n(\mathbb{R})/K_\infty^0)$ as

$$K_h^q(x,y) = \int_{\Gamma(q)} \tilde{h}(\mu_\varpi)\varpi(x)\overline{\varpi(y)}\,\mathrm{d}\varpi, \qquad (8.3)$$

where $\int_{\Gamma(q)} V_\varpi\,\mathrm{d}\varpi$ is a spectral decomposition of $L^2(\Gamma(q)\backslash\mathrm{SL}_n(\mathbb{R})/K_\infty^0)$, respecting the unramified Hecke algebra (see [8, Section 2.5] for more details). This classical spectral decomposition may be deduced from the more precise adelic spectral decomposition (e.g. [6, (7.6)]). The value μ_ϖ is the Langlands parameter of the K_∞^0-invariant representation V_ϖ.

Lemma 8.7. *We have*

$$K_{h*h}^q(e,e) \le K_{h*h^*}^q(e,e)$$

for any $h \in C_c^\infty(K_\infty^0\backslash\mathrm{SL}_n(\mathbb{R})/K_\infty^0)$.

Proof. We have

$$K_{h*h}^q(e,e) = \int_{\Gamma(q)} \tilde{h}(\mu_\varpi)^2|\varpi(e)|^2\,\mathrm{d}\varpi \le \int_{\Gamma(q)} |\tilde{h}(\mu_\varpi)|^2|\varpi(e)|^2\,\mathrm{d}\varpi = K_{h*h^*}^q(e,e),$$

where the first and last equalities follow from the spectral decompositions of the corresponding kernels as in (8.3). $\qquad\qquad\square$

Proof of Theorem 4. Let $h_R^{(1)} := \frac{1}{\mathrm{vol}(B_R)}\mathbb{1}_{B_R}$ be the normalized characteristic function of the ball B_R. It holds for every $g \in \mathrm{SL}_n(\mathbb{R})$ that $h_R^{(1)}(g) \ll h_R(g)$, which, upon applying Lemma 8.7 and Proposition 8.6, implies that

$$K_{h_R^{(1)}*h_R^{(1)}}^q(e,e) \ll K_{h_R*h_R}^q(e,e) \le K_{h_R*h_R^*}^q(e,e) \ll_\epsilon (Rq)^\epsilon (q^{-(n^2-1)} + R^{-n(n-1)}). \qquad (8.4)$$

Notice that the kernel $K_{h_R^{(1)}*h_R^{(1)}}^q(e,e)$ is well-defined even though $h_R^{(1)}$ is not smooth. The same will be true for other functions below.

Next, we claim that for $c > 0$ sufficiently small and R sufficiently large, it holds that for every $g \in \mathrm{SL}_n(\mathbb{R})$

$$h_{cR^2}^{(1)}(g) \ll (h_R^{(1)} * h_R^{(1)})(g). \qquad (8.5)$$

To prove this claim, notice that

$$\mathbb{1}_{B_R} * \mathbb{1}_{B_R}(g) = \mathrm{vol}(B_R \cap gB_R^{-1}).$$

For $c > 0$ small enough and R large enough, the above is $\gg 1$ as long as $\|g\| \leq cR^2$. Indeed, we may assume that g is diagonal with positive values on the diagonal. Then $B_R \cap gB_R^{-1}$ contains all the element of the form $g^{1/2}x$, for $\|x\| \leq C$ (for $c > 0$ sufficiently small relative to C). Therefore, for every $g \in \mathrm{SL}_n(\mathbb{R})$,

$$\mathbb{1}_{B_{cR^2}}(g) \ll (\mathbb{1}_R * \mathbb{1}_R)(g).$$

Using (8.1), we deduce that (8.5) holds. Thus, using (8.4), we obtain

$$\frac{1}{\mathrm{vol}(B_{R^2})} \sum_{\gamma \in \Gamma(q)} \mathbb{1}_{B_{R^2}}(\gamma) = K^q_{h^{(1)}_{R^2}}(e, e) \ll (Rq)^\epsilon (q^{-(n^2-1)} + R^{-n(n-1)}).$$

Using (8.1) once again and changing R to $\sqrt{R}$, we complete the proof. □

8.3. Proof Theorem 5

To prove the theorem, we first prove a couple of lemmata. Fix $\eta > 0$. We consider the function $m_R := h_R * h_{R^\eta}$.

Lemma 8.8. *There is a $\delta > 0$ depending only on η, such that for every q square-free and $R \geq q^{1+1/n}$*

$$K^q_{m_R * m_R^*}(e, e) - \frac{1}{\mathrm{vol}(\Gamma(q) \backslash \mathrm{SL}_n(\mathbb{R}))} \ll R^{-\delta} q^{-(n^2-1)}.$$

Proof. We apply the spectral decomposition as in (8.3) to obtain

$$K^q_{m_R * m_R^*}(e, e) = \int_{\Gamma(q)} |\tilde{m}_R(\mu_\varpi)|^2 |\varpi(e)|^2 \, \mathrm{d}\varpi.$$

Among the automorphic forms ϖ appearing in the spectral decomposition, we have the L^2-normalized constant function ϖ_{triv}, for which it holds that

$$|\varpi_{\mathrm{triv}}(e)|^2 = \frac{1}{\mathrm{vol}(\Gamma(q) \backslash \mathrm{SL}_n(\mathbb{R}))}.$$

Also, using the facts that $\tilde{m}_R(\lambda) = \tilde{h}_R(\lambda) \tilde{h}_{R^\eta}(\lambda)$ and $\tilde{h}_R(\mu_{\varpi_{\mathrm{triv}}}) = \|h_R\|_1 = 1$, it follows that $\tilde{m}_R(\mu_{\varpi_{\mathrm{triv}}}) = 1 = \|m_R\|_1$.

On the other hand, there is a $\delta_1 > 0$, such that for every $\varpi \neq \varpi_{\mathrm{triv}}$, it holds that $\|\Re(\mu_\varpi)\|_\infty \leq (n-1)/2 - \delta_1$. This follows from the explicit property (T) for $n \geq 3$ (where we may actually take $\delta_1 = 1/2$), and from Selberg's spectral gap theorem for $n = 2$ (see [50]).

Using (8.2), we deduce that there is a $\delta_2 > 0$ depending only on η (and n), such that for $\varpi \neq \varpi_{\mathrm{triv}}$

$$|\tilde{m}_R(\mu_\varpi)| \leq R^{-\delta_2} |\tilde{h}_R(\mu_\varpi)|.$$

Combining all the information above, we obtain

$$K^q_{m_R * m^*_R}(e, e) - \frac{1}{\mathrm{vol}(\Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R}))} = \int\limits_{\varpi \neq \varpi_{\mathrm{triv}}} |\tilde{m}_R(\mu_\varpi)|^2 |\varpi(e)|^2 \, d\varpi$$

$$\leq R^{-2\delta_2} \int\limits_{\varpi \neq \varpi_{\mathrm{triv}}} |\tilde{h}_R(\mu_\varpi)|^2 |\varpi(e)|^2 \, d\varpi \leq R^{-2\delta_2} K^q_{h_R * h^*_R}(e, e).$$

The claim now follows from Proposition 8.6. □

Lemma 8.9. *Let* $M^q_R \colon \Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R}) \to \mathbb{R}$ *be the function* $M^q_R(x) := \sum_{\gamma \in \Gamma(q)} m_R(\gamma x)$. *Then*

$$K^q_{m_R * m^*_R}(e, e) = \|M^q_R\|^2_2.$$

Proof. We see that

$$K^q_{m_R * m^*_R}(e, e) = \sum_{\gamma \in \Gamma(q)} (m_R * m^*_R)(\gamma) = \sum_{\gamma \in \Gamma(q)} \int\limits_{x \in \mathrm{SL}_n(\mathbb{R})} m_R(x) m_R(\gamma^{-1} x) \, dx.$$

Folding the integral over $\Gamma(q)$, we obtain the above equals

$$\int\limits_{x \in \Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R})} \sum_{\gamma_1 \in \Gamma(q)} m_R(\gamma_1 x) \sum_{\gamma \in \Gamma(q)} m_R(\gamma^{-1}\gamma_1 x) \, dx = \int\limits_{x \in \Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R})} M^q_R(x) M^q_R(x) \, dx.$$

We conclude that the above is $\|M^q_R\|^2_2$ as m_R is real-valued. □

Proof of Theorem 5. Let $U_q \in L^2(\Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R}))$ be the L^1-normalized constant function on $\Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R})$. Then by Lemma 8.9, the fact that

$$\int\limits_{\Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R})} M^q_R(x) \, dx = \int\limits_{\mathrm{SL}_n(\mathbb{R})} m_R(x) \, dx = 1,$$

and Lemma 8.8, we have that there is a $\delta > 0$ depending only on $\eta > 0$, such that for $R \geq q^{1+1/n}$

$$\|M^q_R - U_q\|^2_2 = \|M^q_R\|^2_2 - \frac{1}{\mathrm{vol}(\Gamma(q)\backslash \mathrm{SL}_n(\mathbb{R}))} \ll R^{-\delta} q^{-(n^2-1)}. \tag{8.6}$$

We identify $\overline{g} \in \Gamma(q)\backslash \Gamma(1) \cong \mathrm{SL}_n(\mathbb{Z}/q\mathbb{Z})$ with some representative of it in $\Gamma(1) := \mathrm{SL}_n(\mathbb{Z}) \subset \mathrm{SL}_n(\mathbb{R})$. We notice that if $\overline{g}$ has no lift to $\Gamma(1)$ of size bounded by $C_1 R^{1+\eta}$ for certain sufficiently large but fixed $C_1 > 0$, then for all $\gamma \in \Gamma(q)$, we have

$$\|\gamma \overline{g} w\| \geq \|\gamma \overline{g}\| \|w^{-1}\|^{-1} \geq C_3 R^{1+\eta},$$

for certain $C_3 > 0$ sufficiently large, where $w \in \Gamma(1)\backslash \mathrm{SL}_n(\mathbb{R})$ with $\|w\| \leq C_2$ for certain $C_2 \geq n$. Consequently, noting that the support of m_R is on $g \in \mathrm{SL}_n(\mathbb{R})$, such that $\|g\| \leq C_4 R^{1+\eta}$ for certain fixed $C_4 > 0$, we conclude that

$$M^q_R(\overline{g} w) = 0.$$

Thus, for such $\overline{g}$, we have

$$\int\limits_{\substack{w \in \mathcal{F} \\ \|w\| \leq C_2}} \left| M_R^q(\overline{g}w) - \frac{1}{\text{vol}(\Gamma(q) \backslash \text{SL}_n(\mathbb{R}))} \right|^2 dw \gg \frac{1}{\text{vol}(\Gamma(q) \backslash \text{SL}_n(\mathbb{R}))^2} \gg \frac{1}{q^{2(n^2-1)}},$$

where $\mathcal{F} \subseteq \text{SL}_n(\mathbb{R})$ is the standard (i.e. containing identity) fundamental domain of $\Gamma(1) \backslash \text{SL}_n(\mathbb{R})$. Summing the above over such $\overline{g}$, we obtain

$$\frac{N(q, C_1 R^{1+\eta})}{q^{2(n^2-1)}} \ll \sum_{\overline{g} \in \Gamma(q) \backslash \Gamma(1)} \int\limits_{\mathcal{F}} \left| M_R^q(\overline{g}w) - \frac{1}{\text{vol}(\Gamma(q) \backslash \text{SL}_n(\mathbb{R}))} \right|^2 dw,$$

where $N(q, R)$ is the number of $\overline{g} \in \text{SL}_n(\mathbb{Z}/q\mathbb{Z})$ without a lift to $\Gamma(1)$ of norm bounded by R. Note that

$$\sum_{\overline{g} \in \Gamma(q) \backslash \Gamma(1)} \left| M_R^q(\overline{g}w) - \frac{1}{\text{vol}(\Gamma(q) \backslash \text{SL}_n(\mathbb{R}))} \right|^2$$

as a function of w is left-invariant by $\Gamma(1)$. Thus, we can write the w-integral of the above over $\mathcal{F}$ as

$$\sum_{\overline{g} \in \Gamma(q) \backslash \Gamma(1)} \int\limits_{\Gamma(1) \backslash \text{SL}_n(\mathbb{R})} \left| M_R^q(\overline{g}w) - \frac{1}{\text{vol}(\Gamma(q) \backslash \text{SL}_n(\mathbb{R}))} \right|^2 dw = \|M_R^q - U_q\|_2^2.$$

On the other hand, for $R \geq q^{1+1/n}$, we have

$$\|M_R^q - U_q\|_2^2 \ll R^{-\delta} q^{-(n^2-1)},$$

which follows from (8.6). Finally, plugging in $R = q^{1+1/n}$, we get that for $C > 0$ sufficiently large

$$N(q, C_1 q^{1+1/n+\eta(1+1/n)}) \ll q^{n^2-1-\delta}.$$

Modifying η, we conclude the proof. $\square$

8.4. Proof of Proposition 8.1

We start by applying the second formulation of $J^T(h)$ as in Proposition 6.3. Note that $\rho(P, \lambda, h_{R,q})|_{\mathcal{A}_\pi^2(P)}$ projects onto $\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}$. Thus, working as in the proof of Theorem 3, it suffices to show that

$$\sum_{P \supset P_0} \max_{L \in \mathcal{L}(M_P)} \sum_{\pi \in \Pi_2(M_P)} \dim \left(\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)} \right)$$

$$\times \int\limits_{i(\mathfrak{a}_L^G)^*} |\tilde{h}_R(\mu_\pi + \lambda)|^2 \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}} \right\|_{\text{op}} d\lambda \ll_\epsilon \|T\|^{n-1}(Rq)^\epsilon (q + R^{-n(n-1)}q^{n^2}). \quad (8.7)$$

We apply (8.2) to bound the inner integral on the left-hand side of the above by

$$\ll_{N,\epsilon} R^{-n(n-1)+2n\|\Re(\mu_\pi)\|_\infty+\epsilon} \int\limits_{i(\mathfrak{a}_L^G)^*} (1 + \|\mu_\pi + \lambda\|)^{-N} \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}} \right\|_{\text{op}} d\lambda.$$

As π is a representation of $M(\mathbb{A})^1$ and $\lambda \in i(\mathfrak{a}_M^G)^*$, we have $\langle \mu_\pi, \lambda \rangle = 0$, consequently, we have $\|\mu_\pi + \lambda\|^2 \asymp \|\mu_\pi\|^2 + \|\lambda\|^2$. Thus, the above integral can be bounded by

$$(1 + \|\mu_\pi\|)^{-N/2} \int_{i(\mathfrak{a}_L^G)^*} (1 + \|\lambda\|)^{-N/2} \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}} \right\|_{\mathrm{op}} \, d\lambda.$$

To estimate the above integral, we proceed exactly as in the proof Theorem 3. We first write it as the sum of the subintegrals over $m - 1 \le \|\lambda\| \le m$ where m varies over $\mathbb{N}$. We further majorize each subintegral by a sum of integrals over $\|\lambda - \eta\| \le 1$ with $\eta \in i(\mathfrak{a}_L^G)^*$ and $\|\eta\| \ll m$. Finally, we apply Proposition 5.5 to bound each integral after making N large enough. This gives us

$$\int_{i(\mathfrak{a}_L^G)^*} (1 + \|\lambda\|)^{-N/2} \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}} \right\|_{\mathrm{op}} \, d\lambda \ll (\|T\| \log(1 + \|\mu_\pi\| + \mathrm{level}(K(q)))^{\dim \mathfrak{a}_L^G}.$$

Hence, we obtain

$$\max_{L \in \mathcal{L}(M_P)} \int_{i(\mathfrak{a}_L^G)^*} |\tilde{h}_R(\mu_\pi + \lambda)|^2 \left\| \mathcal{M}_L^T(P, \lambda)|_{\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}} \right\|_{\mathrm{op}} \, d\lambda$$

$$\ll_{\epsilon, N} \|T\|^{n-1} (Rq)^\epsilon R^{-n(n-1) + 2n\|\mathfrak{R}(\mu_\pi)\|_\infty} (1 + \|\mu_\pi\|)^{-N}, \quad (8.8)$$

for all sufficiently large $N > 0$.

Now, we move on to estimate the π-sum in (8.7). First, let us estimate the contribution for $\|\mu_\pi\| \ge (Rq)^\epsilon$. Dividing into dyadic intervals, and using (8.8) and the fact that $\|\mathfrak{R}(\mu_\pi)\|_\infty \ll 1$, we estimate this contribution by

$$\ll (Rq)^{O(1)} \sum_{m \ge (Rq)^\epsilon} \sum_{\substack{\pi \in \Pi_2(M_P) \\ \|\mu_\pi\| \asymp m}} \dim\left(\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)} \right) \|\mu_\pi\|^{-N}.$$

There is an absolute K, such that the inner sum is bounded by $\ll m^{-N} (mq)^K$, which follows from [8, (7.2)], classification of discrete series [44] and local Weyl law [49] (also see the discussion in [8, pp.35]). Making N large enough, we can bound the above double sum by $\ll_N (Rq)^{-N}$. Thus, we may restrict the π-sum in (8.7) with $\|\mu_\pi\| \le (Rq)^\epsilon$. We conclude the proof by applying Lemma 8.10 below.

Lemma 8.10. *We have*

$$\sum_{P \supset P_0} \sum_{\substack{\pi \in \Pi_2(M_P) \\ \|\mu_\pi\| \le (Rq)^\epsilon}} \dim\left(\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)} \right) R^{2n\|\mathfrak{R}(\mu_\pi)\|_\infty} \ll_\epsilon (Rq)^\epsilon q(R^{n(n-1)} + q^{n^2 - 1}),$$

with R, q as in Proposition 8.1.

The statement and the proof below are quite similar to those of [8, Theorem 7.1]. To make the similarities comprehensible, we make the dictionary that T and M in [8] are R^n and $(Rq)^\epsilon$, respectively, in our case.

Proof. First, trivially, we have

$$\sum_{P \supset P_0} \sum_{\substack{\pi \in \Pi_2(M_P) \\ \|\mu_\pi\| \le (Rq)^\epsilon}} \dim\left(\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}\right) R^{2n\|\Re(\mu_\pi)\|_\infty}$$

$$\le \sum_{P \supset P_0} \sum_{\substack{\pi \in \Pi_2(M_P) \\ \|\mu_\pi\| \le (Rq)^\epsilon}} \dim\left(\mathcal{A}_\pi^2(P)^{K_\infty^0 K(q)}\right) R^{2n\|\Re(\mu_\pi)\|_\infty} \int_{\substack{\lambda \in i(\mathfrak{a}_P^G)^* \\ \|\lambda\| \le (Rq)^\epsilon}} \mathrm{d}\lambda.$$

In the classical language as in [8, Theorem 7.1], the right-hand side above is bounded by

$$q \int_{\substack{\Gamma(q) \\ \|\mu_\varpi\| \le 2(Rq)^\epsilon}} R^{2n\|\Re(\mu_\varpi)\|_\infty} \, \mathrm{d}\varpi. \tag{8.9}$$

For the relevant notations, we refer to [8, Section 2]. The factor q comes from the $\varphi(q) \le q$ possible twists by Dirichlet characters when moving from the adelic to the classical language (see [8, Section 2.6.2] and also [39, Section 3.4]).

If $R \le R_0 := q^{1+1/n}(Rq)^{-\epsilon}$ for sufficiently small $\epsilon > 0$, we directly apply [8, Theorem 7.1] to bound the integral in (8.9) by $\ll_\epsilon (Rq)^\epsilon q^{n^2-1}$.

If $R > R_0$, then we work as in [8, Section 8.2]. Note that $\|\Re(\mu_\varpi)\|_\infty \le \frac{n-1}{2}$ (which is attained when ϖ is one-dimensional). In this case, we write the integral in (8.9) as

$$(R/R_0)^{n(n-1)} \int_{\substack{\Gamma(q) \\ \|\mu_\varpi\| \le 2(Rq)^\epsilon}} R_0^{2n\|\Re(\mu_\varpi)\|_\infty} \, \mathrm{d}\varpi \ll_\epsilon (Rq)^\epsilon R^{n(n-1)}.$$

This completes the proof. □

References

[1] J. Arthur, 'A trace formula for reductive groups. II. Applications of a truncation operator', *Compositio Math.* **40**(1) (1980), 87–121.

[2] J. Arthur, 'The trace formula in invariant form', *Ann. of Math. (2)* **114**(1) (1981), 1–74.

[3] J. Arthur, 'On a family of distributions obtained from Eisenstein series. I. Application of the Paley-Wiener theorem', *Amer. J. Math.* **104**(6) (1982), 1243–1288.

[4] J. Arthur, 'On a family of distributions obtained from Eisenstein series. II. Explicit formulas', *Amer. J. Math.* **104**(6) (1982), 1289–1336.

[5] J. Arthur, 'On the inner product of truncated Eisenstein series', *Duke Math. J.* **49**(1) (1982), 35–70.

[6] J. Arthur, 'An introduction to the trace formula', in Harmonic Analysis, the Trace Formula, and Shimura Varieties, *Clay Mathematics Proceedings* vol. 4 (American Mathematical Society, Providence, RI, 2005), 1–263.

[7] J. Arthur, 'The endoscopic classification of representations, in *American Mathematical Society Colloquium Publications* vol. 61 (American Mathematical Society, Providence, RI, 2013), vii–589. Orthogonal and symplectic groups.

[8] E. Assing and V. Blomer, 'The density conjecture for principal congruence subgroups', *Duke Math. J . To appear*, Preprint, 2022, arXiv:2204.08868.

[9] E. Assing, V. Blomer and P. D. Nelson, 'Local analysis of the Kuznetsov formula and the density conjecture', Preprint, 2024, arXiv:2404.05561.

[10] I. N. Bernšteĭn, 'All reductive 𝔭-adic groups are of type I', *Funkcional. Anal. i Priložen.* **8**(2) (1974), 3–6.

[11] V. Blomer, 'Epstein zeta-functions, subconvexity, and the purity conjecture', *J. Inst. Math. Jussieu* **19**(2) (2020), 581–596.

[12] V. Blomer and C. Lutsko, 'Hyperbolic lattice point counting in unbounded rank', *J. Reine Angew. Math . To appear*, Preprint, 2023, arXiv:2309.00522.

[13] V. Blomer and P. Maga, 'Subconvexity for sup-norms of cusp forms on $PGL(n)$', *Selecta Math. (N.S.)* **22**(3) (2016), 1269–1287.

[14] F. Brumley and N. Templier, 'Large values of cusp forms on GL_n', *Selecta Math. (N.S.)* **26**(4) (2020), Paper No. 63, 71.

[15] J. J. Duistermaat, J. A. C. Kolk and V. S. Varadarajan, 'Spectra of compact locally symmetric manifolds of negative curvature', *Invent. Math.* **52**(1) (1979), 27–93.

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

[17] T. Finis and E. Lapid, 'On the spectral side of Arthur's trace formula—combinatorial setup', *Ann. of Math. (2)* **174**(1) (2011), 197–223.

[18] T. Finis and E. Lapid, 'On the analytic properties of intertwining operators I: Global normalizing factors', *Bull. Iranian Math. Soc.* **43**(4) (2017), 235–277.

[19] T. Finis and E. Lapid, 'On the analytic properties of intertwining operators II: Local degree bounds and limit multiplicities', *Israel J. Math.* **230**(2) (2019), 771–793.

[20] T. Finis and E. Lapid, 'On the remainder term of the Weyl law for congruence subgroups of Chevalley groups', *Duke Math. J.* **170**(4) (2021), 653–695.

[21] T. Finis, E. Lapid and W. Müller, 'On the spectral side of Arthur's trace formula—Absolute convergence', *Ann. of Math. (2)* **174**(1) (2011), 173–195.

[22] T. Finis, E. Lapid and W. Müller, 'Limit multiplicities for principal congruence subgroups of $GL(n)$ and $SL(n)$', *J. Inst. Math. Jussieu* **14**(3) (2015), 589–638.

[23] D. Flath, 'Decomposition of representations into tensor products', in Automorphic Forms, Representations and L-functions (Proceedings of Symposia in Pure Mathematics, Oregon State Univ., Corvallis, Ore., 1977), Part 1 (1979), *Proceedings of Symposia in Pure Mathematics*, XXXIII (American Mathematical Society, Providence, R.I., 1979), 179–183.

[24] W. T. Gan, N. Gurevich and D. Jiang, 'Cubic unipotent Arthur parameters and multiplicities of square integrable automorphic forms', *Invent. Math.* **149**(2) (2002), 225–265.

[25] R. Gangolli, 'On the Plancherel formula and the Paley-Wiener theorem for spherical functions on semisimple Lie groups', *Ann. of Math. (2)* **93** (1971), 150–165.

[26] A. Ghosh, A. Gorodnik and A. Nevo, 'Best possible rates of distribution of dense lattice orbits in homogeneous spaces', *J. Reine Angew. Math.* **745** (2018), 155–188.

[27] K. Golubev and A. Kamber, 'On Sarnak's density conjecture and its applications', *Forum Math. Sigma* **11** (2023), Paper No. e48, 51.

[28] A. Good, 'The square mean of Dirichlet series associated with cusp forms', *Mathematika* **29**(2) (1982), 278–295.

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

[30] A. Gorodnik and A. Nevo, 'Quantitative ergodic theorems and their number-theoretic applications', *Bull. Amer. Math. Soc. (N.S.)* **52**(1) (2015), 65–113.

[31] H. Jacquet, I. I. Piatetski-Shapiro and J. Shalika, 'Conducteur des représentations du groupe linéaire', *Math. Ann.* **256**(2) (1981), 199–214.

[32] S. Jana and A. Kamber, 'Optimal Diophantine exponents for SL(n)', *Adv. Math.* **443** (2024), Paper No. 109613.

[33] A. Kamber and H. Lavner, 'Optimal lifting for the projective action of $SL_3(\mathbb{Z})$', *Algebra Number Theory* **17**(3) (2023), 749–774.

[34] A. Kamber and P. Varjú, 'Lifting all elements in $SL_n(\mathbb{Z}/q\mathbb{Z})$', Preprint, 2023, arXiv:2310.10269.

[35] H. H. Kim, 'The residual spectrum of G_2', *Canad. J. Math.* **48**(6) (1996), 1245–1272.

[36] R. P. Langlands, 'Eisenstein series', in *Algebraic Groups and Discontinuous Subgroups (Proceedings of Symposia in Pure Mathematics, Boulder, Colo., 1965)* (1966), (American Mathematical Society, Providence, R.I., 1966), 235–252.

[37] R. P. Langlands, 'Eisenstein series, the trace formula, and the modern theory of automorphic forms', in *Number Theory, Trace Formulas and Discrete Groups (Oslo, 1987)* (Academic Press, Boston, MA, 1989), 125–155.

[38] E. Lapid, 'Some perspectives on Eisenstein series', Preprint, 2022, arxiv.2204.02914.

[39] E. Lapid and W. Müller, 'Spectral asymptotics for arithmetic quotients of $SL(n, \mathbb{R})/SO(n)$', *Duke Math. J.* **149**(1) (2009), 117–155.

[40] E. M. Lapid, 'On the fine spectral expansion of Jacquet's relative trace formula', *J. Inst. Math. Jussieu* **5**(2) (2006), 263–308.

[41] E. Lindenstrauss and A. Venkatesh, 'Existence and Weyl's law for spherical cusp forms', *Geom. Funct. Anal.* **17**(1) (2007), 220–251.

[42] S. Marshall, 'Upper bounds for Maass forms on semisimple groups', Preprint, 2014, arXiv:1405.7033.

[43] S. D. Miller, 'On the existence and temperedness of cusp forms for $SL_3(\mathbb{Z})$', *J. Reine Angew. Math.* **533** (2001), 127–169.

[44] C. Mœglin and J.-L. Waldspurger, 'Le spectre résiduel de GL(n)', *Ann. Sci. École Norm. Sup. (4)* **22**(4) (1989), 605–674.

[45] C. Mœglin and J.-L. Waldspurger, 'Spectral decomposition and Eisenstein series', in *Cambridge Tracts in Mathematics* vol. 113 (Cambridge University Press, Cambridge, 1995), i–297. Une paraphrase de l'Écriture [A paraphrase of Scripture]

[46] C. P. Mok, 'Endoscopic classification of representations of quasi-split unitary groups', *Mem. Amer. Math. Soc*. **235**(1108) (2015), vi+248.

[47] W. Müller, 'On the singularities of residual intertwining operators', *Geom. Funct. Anal.* **10**(5) (2000), 1118–1170.

[48] W. Müller, 'On the spectral side of the Arthur trace formula', *Geom. Funct. Anal.* **12**(4) (2002), 669–722.

[49] W. Müller, 'Weyl's law for the cuspidal spectrum of SL_n', *Ann. of Math. (2)* **165**(1) (2007), 275–333.

[50] P. Sarnak, 'Notes on the generalized Ramanujan conjectures', in Harmonic Analysis, the Trace Formula, and Shimura Varieties, *Clay Mathematics Proceedings* vol. 4 (American Mathematical Society, Providence, RI, 2005), 659–685.

[51] P. Sarnak, 'Letter to Stephen D. Miller and Naser Talebizadeh Sardari on optimal strong approximation by integral points on quadrics', available at http://publications.ias.edu/sarnak, 2015.

[52] P. Sarnak and X. X. Xue, 'Bounds for multiplicities of automorphic representations', *Duke Math. J.* **64**(1) (1991), 207–227.

[53] P. C. Sarnak, 'Diophantine problems and linear groups', in *Proceedings of the International Congress of Mathematicians, Vols. I, II (Kyoto, 1990)* (1991), (The Mathematical Society of Japan, Tokyo, 1991), 459–471.

[54] C. D. Sogge, 'Concerning the L^p norm of spectral clusters for second-order elliptic operators on compact manifolds', *J. Funct. Anal*. **77**(1) (1988), 123–138.

[55] S. Žampera, 'The residual spectrum of the group of type G_2', *J. Math. Pures Appl. (9)* **76**(9) (1997), 805–835.

[56] L. Zhang, 'Quantum unique ergodicity of degenerate Eisenstein series on $GL(n)$', *Comm. Math. Phys*. **369**(1) (2019), 1–48.