ON THE REMAINDER TERM OF THE WEYL LAW FOR CONGRUENCE SUBGROUPS OF CHEVALLEY GROUPS

TOBIAS FINIS AND EREZ LAPID

Abstract. Let $X$ be a locally symmetric space defined by a simple Chevalley group $G$ and a congruence subgroup of $G(\mathbb{Q})$. In this generality, the Weyl law for $X$ was proved by Lindenstrauss–Venkatesh. In the case where $G$ is simply connected, we sharpen their result by giving a power saving estimate for the remainder term.

Contents

1. Introduction 1
2. Review of Arthur’s trace formula 5
3. A non-archimedean separation lemma 9
4. Archimedean test functions 14
5. Local bounds and Weyl’s law for the cuspidal spectrum 17
References 26

1. Introduction

1.1. The Weyl law, in its basic form, states that the number $N_X(T)$ of eigenfunctions of the Laplacian on a compact $d$-dimensional Riemannian manifold $X$ with eigenvalues $\leq T^2$ satisfies the asymptotic

$$\lim_{T \to \infty} \frac{N_X(T)}{T^d} = \frac{\text{vol}(X)}{(4\pi)^{d/2} \Gamma(\frac{d}{2} + 1)}.$$  

The definitive result for compact Riemannian manifolds is due to Hörmander [18]. His work implies in particular that

$$N_X(T) = \frac{\text{vol}(X)}{(4\pi)^{d/2} \Gamma(\frac{d}{2} + 1)} T^d + O(T^{d-1})$$

in the compact case. The order of the remainder term is optimal without further assumptions on $X$.

The problem becomes more difficult when $X$ is not compact (but still of finite volume). An interesting class to consider is the locally symmetric spaces of non-compact type, namely $X = \Gamma \backslash G/K$, where $\Gamma$ is a lattice in a semisimple Lie group $G$ with a maximal compact subgroup $K$. One of Selberg’s main motivations for developing the trace formula was to obtain information on the eigenvalues of the Laplacian. Already the existence of non-zero
eigenvalues is non-evident. For $G = \text{SL}(2, \mathbb{R})$, the case of hyperbolic surfaces, Selberg \[38\] (see also \[37, §39\]) showed that

$$N_\Gamma(T) + M_\Gamma(T) = \frac{\text{vol}(X)T^2}{4\pi} - \frac{\kappa}{\pi} T \log \frac{2T}{e} + O(T/ \log T),$$

where $N_\Gamma(T) := N_{\Gamma \backslash G/K}(T)$, $\kappa$ is the number of cusps of $\Gamma \backslash G/K$, and

$$M_\Gamma(T) = -\frac{1}{2\pi} \int_0^T \frac{\phi'}{\phi} \left(\frac{1}{2} + it\right)dt$$

is the winding number of the determinant $\phi(s)$ of the scattering matrix, a meromorphic function of a complex variable $s$ that is holomorphic and has absolute value 1 on the line $\text{Re } s = \frac{1}{2}$. The summand $M_\Gamma(T)$ can be interpreted as the contribution of the part of the continuous spectrum with Laplace eigenvalue $\leq \frac{1}{4} + T^2$. It is not difficult to see that the difference between $M_\Gamma(T)$ and the number of poles of $\phi$ with imaginary part between 0 and $T$, counted with their multiplicities, is $O(T)$ (and Selberg refined this to $cT + O(T/ \log T)$ for a constant $c \geq 0$). Selberg also showed that

$$\liminf_{T \to \infty} \frac{M_\Gamma(T)}{T \log T} \geq \frac{\kappa}{\pi}.$$  

If $\Gamma$ is a congruence subgroup of the modular group $\text{SL}(2, \mathbb{Z})$, then the scattering determinant $\phi(s)$ is given in terms of Dirichlet $L$-functions, and the classical results of Riemann and von Mangoldt on the number of zeros of such an $L$-function in a strip yield the asymptotic

$$M_\Gamma(T) = \frac{\kappa}{\pi} T \log T + O(T).$$

Therefore, in this case we have

$$N_\Gamma(T) = \frac{\text{vol}(X)T^2}{4\pi} - \frac{2\kappa}{\pi} T \log T + O(T),$$

and it is possible to refine the remainder term $O(T)$ here to $c'T + O(T/ \log T)$ for a suitable constant $c'$ depending on $\Gamma$.

This result led Selberg to speculate whether in general $M_\Gamma(T) = o(T^2)$, which would imply the Weyl law for the discrete spectrum for an arbitrary lattice $\Gamma$ in $\text{SL}(2, \mathbb{R})$. However, the subsequent work of Phillips and Sarnak on the dissolution of cusp forms under deformation of lattices led Sarnak to conjecture that the opposite extreme holds, namely that except for the Teichmüller space of the once punctured torus, the discrete spectrum of a generic non-uniform lattice in $\text{SL}(2, \mathbb{R})$ is finite (see \[35\] and the references therein).

It is well known that irreducible lattices in all other semisimple Lie groups do not form continuous families. It was conjectured by Sarnak \[33\] that the Weyl law holds in complete generality in the case of congruence subgroups. The appropriate form of the trace formula for congruence subgroups in arbitrary rank was developed by Arthur in the adelic language. It is technically much more complicated than in the rank one case.
1.2. An important part of the discrete spectrum is the cuspidal spectrum defined by the vanishing of all constant terms with respect to proper parabolic subgroups. In rank one, all but a finite part of the discrete spectrum is cuspidal, but this is not true in general. The general upper bound
\[
\limsup_{T \to \infty} \frac{N_{X,cusp}(T)}{T^d} \leq \frac{\text{vol}(X)}{(4\pi)^{d/2}\Gamma(\frac{d}{2} + 1)}
\]
on the cuspidal spectrum was proven by Donnelly [7] (it can also be obtained from a basic analysis of Arthur’s trace formula). In [25], Lindenstrauss and Venkatesh made a breakthrough by showing that the cuspidal spectrum for congruence subgroups obeys the Weyl law\(^1\) (Some special cases had been proved earlier – see [28,31,32].) Their argument uses arithmeticity in an essential way. It is based on the crucial fact that the spectral parameters of Eisenstein series at different places are not independent of one another, but satisfy certain simple relations. Taking this into account, Lindenstrauss and Venkatesh constructed from any single non-trivial Hecke operator a family of adelic test functions that act trivially on all Eisenstein series, and hence effectively only see the cuspidal spectrum. In contrast to previous instances of the simple trace formula these test functions are not factorizable, and as a family, do not entail a loss of a positive proportion of the cuspidal spectrum.

The result of [25] does not provide a bound on the remainder term in the Weyl law for the cuspidal spectrum. In fact, it seems that as it stands, the method is short of giving a good error term, since it uses a single Hecke operator. The purpose of the current paper is to push the basic idea of Lindenstrauss–Venkatesh further and to bound the error term in the cuspidal Weyl law by \(O(T^{d-\delta})\) for some \(\delta > 0\). For simplicity, we work in the setting of simple Chevalley groups \(G\) defined over \(\mathbb{Q}\), mainly because the necessary estimates for the geometric side of the trace formula have been worked out only in this case. Then, we have the following (see Theorem 5.11 and Corollary 5.12)\(^2\).

**Theorem 1.1.** Let \(G\) be a simply connected, simple Chevalley group. Then there exists \(\delta > 0\) such that for any congruence subgroup \(\Gamma\) of \(G(\mathbb{Z})\) we have
\[
N_{X,cusp}(T) = \frac{\text{vol}(X)}{(4\pi)^{d/2}\Gamma(\frac{d}{2} + 1)}T^d + O_T(T^{d-\delta}), \quad T \geq 1
\]
where \(X = \Gamma \backslash G(\mathbb{R})/K\).

Our method uses Hecke operators as well, but in a slightly different way. Namely, instead of nulling the contribution of the non-cuspidal spectrum, we amplify it in order to show its negligibility (by a factor of \(O(T^{-\delta})\)). This is somewhat analogous to the situation in Selberg’s sieve (e.g., [6]) and requires the use of \(T^c\) many Hecke operators for a suitable \(c > 0\). The argument crucially relies on a simple positivity property of Arthur’s trace formula. Apart from the basic work of Arthur’s first papers on the trace formula, we use

---

\(^1\)They state their result for quotients \(\Gamma \backslash G(\mathbb{R})/K\), where \(G\) is a split adjoint semisimple group over \(\mathbb{Q}\) and \(\Gamma\) a congruence subgroup of \(G(\mathbb{Q})\), but their method is completely general.

\(^2\)In the non-simply connected case, our result is for manifolds which may be non-connected.
the estimates on its geometric side of \cite{13}, which are based on \cite{10,11}. The exact power saving that we get can in principle be computed, but since it results from an application of the Stone–Weierstrass theorem, we expect it to be quite poor. It would be interesting to analyze this question more carefully. The final result (Theorem 5.11) is actually more general than the Weyl law for the Laplacian, since we treat the entire commutative algebra of all invariant differential operators on $G(\mathbb{R})/K$ simultaneously, following the method of \cite{8}.

More generally, our proof gives a main term for the trace of an arbitrary Hecke operator $\tau$ on the cuspidal spectrum, with a remainder term of order $O_\tau(T^{d-\delta})$ where the implicit constant is the $L^1$-norm of $\tau$ times a logarithmic factor depending on the support of $\tau$ (see Theorem 5.13 for the precise statement). As in \cite{36}, these estimates have applications to the conjectures of Katz–Sarnak on low-lying zeros of $L$-functions for the family of cuspidal automorphic representations of $G$ of a bounded level that are spherical at infinity.

1.3. An alterative strategy (see \cite{9,13,23}) is to use Arthur’s fine spectral expansion and to analyze the analytic properties of intertwining operators (i.e., in this case, of their global normalizing factors) in more depth.

A first step in this direction is to establish polynomial upper bounds for the discrete spectrum. This is the trace-class conjecture, which was solved by Werner Müller some time ago \cite{30} (see also \cite{19,20}). A refinement of this statement is the absolute convergence of the spectral side of Arthur’s trace formula, which was established in \cite{12}.

In \cite{9}, we formulated a precise analytic conjecture on intertwining operators, which would imply the Weyl law for the discrete spectrum with an error term of $O(T^{d-1})$ as in the cocompact case (with an extra logarithmic factor in the case of groups of type $A_1$ and $A_2$). It also implies that the non-cuspidal discrete spectrum is bounded by $O(T^{d-2})$ \cite{13}. Using the work of Langlands on the relation between intertwining operators and automorphic $L$-functions \cite{21}, the conjecture can be formulated in terms of the latter. This conjecture is known to hold for the general linear groups, where the pertinent $L$-functions are the Rankin–Selberg convolutions, whose analytic properties are well-understood by the work of Jacquet–Piatetski-Shapiro–Shalika and others \cite{23}. The conjecture is also known for quasi-split classical groups using Arthur’s work and for the split exceptional group $G_2$, by Shahidi’s work on the symmetric cube $L$-function for $GL_2$ (see \cite{9}). In general however, the required information on the behavior of the automorphic $L$-functions is not available.

In contrast, the method of the current paper does not give an upper bound on the non-cuspidal discrete spectrum, which remains an interesting open problem in general.

1.4. To close this introduction, we give a quick summary of the individual sections of this paper. In \cite{2} we give a summary of the first stage of the spectral expansion of Arthur’s trace formula. In \cite{3} we construct for each prime $p$, Hecke operators at $p$ suitable for the task of emphasizing the non-cuspidal contribution. The main result of this section is Proposition 3.4, which is the technical heart of the paper. In \cite{4} we collect some mostly standard facts about spherical archimedean test functions, the associated Paley–Wiener theorem and the spherical Plancherel measure. Finally, in \cite{5} we prove our main results Theorem 5.11 and Theorem 5.13. Proposition 5.5 contains the key part of the argument.
2. Review of Arthur’s trace formula

In this section we recall the relevant facts from the basic theory of Arthur’s (non-invariant) trace formula and set some notation which will be used throughout. We will freely use standard results from the textbook [29] (and by extension, [22]) as well as from Arthur’s fundamental papers [11, 12, 13] (see also the first part of [14]). On the other hand, we will not go into Arthur’s fine spectral and geometric expansions (let alone his more advanced theory of the trace formula) since we will not use them in the sequel. In fact, on the geometric side, we will only use the estimates of [10, 11, 13] (see also the first part of [15] below).

For this section, let \( G \) be an arbitrary reductive group defined over \( \mathbb{Q} \).

Here and henceforward, \( X \ll Y \) means that there is a constant \( C \) (implicitly depending on the group \( G \)) such that \( X \leq CY \). If \( C \) depends on additional parameters, we will emphasize it by writing \( X \ll_{a,b} Y \).

2.1. As usual, we denote by \( \mathbb{A} = \mathbb{R} \times \mathbb{A}_{\text{fin}} \) the ring of adeles. Let \( \mathfrak{z} \) be the center of the universal enveloping algebra of the (complexified) Lie algebra of \( G_{\infty} := G(\mathbb{R}) \). Fix a maximal compact subgroup \( K = \prod_{p<\infty} K_p \) of \( G(\mathbb{A}) \) such that the factors \( K_p \) are special for all \( p \) and hyperspecial for almost all \( p \).

Fix a Haar measure on \( G(\mathbb{A}) \). Let \( C_c^\infty(G(\mathbb{A})^1) \) be the \(*\)-algebra (under convolution, with \( f^*(g) = \overline{f(g^{-1})} \)) of smooth, complex-valued, compactly supported functions on \( G(\mathbb{A})^1 \). (By definition, smoothness implies that the function is right-invariant under a suitable open subgroup of \( G(\mathbb{A}_{\text{fin}}) \).) The right regular representation gives rise to a \(*\)-representation \( R \) of \( C_c^\infty(G(\mathbb{A})^1) \) on the Hilbert space \( L^2(G(\mathbb{Q}) \backslash G(\mathbb{A})^1) \). Explicitly, for any \( f \in C_c^\infty(G(\mathbb{A})^1) \), \( R(f) \) is the operator

\[
R(f) \varphi(x) = \int_{G(\mathbb{A})^1} f(g) \varphi(xg) \, dg, \quad \varphi \in L^2(G(\mathbb{Q}) \backslash G(\mathbb{A})^1), \quad x \in G(\mathbb{Q}) \backslash G(\mathbb{A})^1
\]

which is an integral operator with kernel

\[
K_f(x,y) = \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1} \gamma y), \quad x, y \in G(\mathbb{Q}) \backslash G(\mathbb{A})^1.
\]

Let \( P = M \rtimes U \) be a parabolic subgroup of \( G \) defined over \( \mathbb{Q} \) with unipotent radical \( U \) and Levi subgroup \( M \) defined over \( \mathbb{Q} \). Denote by \( \delta_P \) the modulus function of \( P(\mathbb{A}) \) and by \( A_M \) the identity connected component of \( T_M(\mathbb{R}) \), where \( T_M \) is the split part of the center of \( M \). Thus, \( A_MM(\mathbb{Q}) \backslash M(\mathbb{A}) \) is of finite volume. Denote by \( L^2_{\text{disc}}(A_MM(\mathbb{Q}) \backslash M(\mathbb{A})) \) the discrete part of \( L^2(A_MM(\mathbb{Q}) \backslash M(\mathbb{A})) \). The space of square-integrable automorphic forms on \( A_MM(\mathbb{Q}) \backslash M(\mathbb{A}) \) is dense in \( L^2_{\text{disc}}(A_MM(\mathbb{Q}) \backslash M(\mathbb{A})) \). Let

\[
A_P^2 = \{ \varphi : U(\mathbb{A})M(\mathbb{Q}) \backslash G(\mathbb{A}) \rightarrow \mathbb{C} \text{ smooth and } \mathfrak{z}\text{-finite :} \}
\]

\[
\varphi(ag) = \delta_P(a)^{\frac{1}{2}} \varphi(g) \quad \forall a \in A_M, \ g \in G(\mathbb{A}), \int_{A_MM(\mathbb{Q})U(\mathbb{A}) \backslash G(\mathbb{A})} |\varphi(g)|^2 \, dg < \infty
\]

and let \( L^2_P \) be the Hilbert completion of \( A_P^2 \). We may identify \( L^2_P \) with the (normalized) parabolic induction \( \text{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})} L^2_{\text{disc}}(A_MM(\mathbb{Q}) \backslash M(\mathbb{A})) \). Let \( \Pi_2(M) \) be the set of equivalence
classes of irreducible representations of $M(\mathbb{A})$ that occur discretely in $L^2(A_M M(\mathbb{Q}) \backslash M(\mathbb{A}))$. (The central character of any $\pi \in \Pi_2(M)$ is trivial on $A_M$.) For any $\pi \in \Pi_2(M)$ let $\mathcal{A}_{P,\pi}^2$ be subspace of $\mathcal{A}_P^2$ consisting of the functions such that for all $x \in G(\mathbb{A})$ the function $m \mapsto \delta_\pi(m)^{-\frac{1}{2}} \varphi(mx)$ belongs to the $\pi$-isotypic component of $L^2_{disc}(A_M M(\mathbb{Q}) \backslash M(\mathbb{A}))$. Let $L^2_{P,\pi}$ be the closure of $\mathcal{A}_{P,\pi}^2$ in $L_P^2$. Thus,

$$\mathcal{A}_{P}^2 = \oplus_{\pi \in \Pi_2(M)} \mathcal{A}_{P,\pi}^2$$

and

$$L^2_{P} = \oplus_{\pi \in \Pi_2(M)} L^2_{P,\pi}.$$ 

For any $\pi \in \Pi_2(M)$ we fix an orthonormal basis $\mathcal{OB}_P(\pi)$ of $\mathcal{A}_{P,\pi}^2$.

Let $X^*(M)$ be the lattice of rational characters of $M$ and let $a_M^* = X^*(M) \otimes \mathbb{R}$ be the real vector space generated by $X^*(M)$. We also write $a_{M,\mathbb{C}}^* = X^*(M) \otimes \mathbb{C}$. The restriction map $X^*(M) \to X^*(T_M)$ identifies $a_M^*$ with $X^*(T_M) \otimes \mathbb{R}$. The dual space of $a_M^*$ will be denoted by $a_M$. Thus, $a_M = X_s(T_M) \otimes \mathbb{R}$ where $X_s(T_M)$ is the lattice of co-characters of $T_M$. We denote the canonical pairing on $a_M^* \times a_M$ by $\langle \cdot, \cdot \rangle$. Let $H_M : M(\mathbb{A}) \to A_M$ be the homomorphism characterized by

$$e^{\langle x, H_M(m) \rangle} = |\chi(m)|, \quad \chi \in X^*(M).$$

The kernel of $H_M$ is $M(\mathbb{A})^1$ and the restriction of $H_M$ to $A_M$ is an isomorphism.

Fix a maximal $\mathbb{Q}$-split torus $T_0$ of $G$ that is in a good position with respect to $K$, i.e., for every $p$ the group $K_p$ is the stabilizer of a special point in the apartment associated to $T_0$. Denote by $W$ the Weyl group $N_{G(\mathbb{Q})}(T_0)/T_0$. It acts on $a_0 := a_{T_0}$ and $a_0^*$. We may identify $a_0$ with the Lie algebra of $T_0(\mathbb{R})$. We fix a $W$-invariant Euclidean structure on $a_0$. (Of course, if $G$ is semisimple, then the Killing form is a $W$-invariant inner product on $a_0$.)

Let $\mathcal{P}$ be the finite set of parabolic subgroups defined over $\mathbb{Q}$ and containing $T_0$. Each $P \in \mathcal{P}$ admits a unique Levi decomposition $P = M \times U$ with $M \supset T_0$ (necessarily defined over $\mathbb{Q}$). We write $\mathcal{L}$ for the set of all such Levi subgroups as we vary $P \in \mathcal{P}$. In other words, $\mathcal{L}$ is the set of centralizers of subtori of $T_0$ in $G$. We say that $P, Q \in \mathcal{P}$ are associate, denoted $P \sim Q$, if their Levi parts are conjugate in $G(\mathbb{Q})$. For every $M \in \mathcal{L}$ we identify $a_M$ as a subspace of $a_0$. We have orthogonal decompositions

$$a_0 = a_0^M \oplus a_M, \quad a_0^* = (a_0^M)^* \oplus a_M^*$$

where

$$a_0^M = X_s(T_0 \cap M^{der}) \otimes \mathbb{R} = X_s(T_0/T_M) \otimes \mathbb{R}$$

and

$$(a_0^M)^* = X^*(T_0 \cap M^{der}) \otimes \mathbb{R} = X^*(T_0/T_M) \otimes \mathbb{R}. $$

Here, $M^{der}$ is the derived group of $M$. For any $M \in \mathcal{L}$ we write $\mathcal{P}(M)$ for the set of $P \in \mathcal{P}$ with Levi part $M$ and let $W(M) = N_{G(\mathbb{F})}(M)/M$, which can be identified with a subgroup of $W$. For any $P \in \mathcal{P}(M)$ let $\Delta_P \subset X^*(T_M) \subset a_M^*$ be the corresponding set of simple roots. (These are the non-zero projections to $a_M^*$ of the simple roots with respect to any minimal parabolic subgroup in $\mathcal{P}$ contained in $P$.) We extend $H_M$ to a left $U_P(\mathbb{A})$ and
right $K$-invariant map $H_P : P(A) \rightarrow a_M$. For any $\lambda \in a_{M,C}^*$ and a function $\varphi$ on $G(A)$ let $\varphi_\lambda$ be the function

$$\varphi_\lambda(x) = \varphi(x)e^{(\lambda, H_P(x))}.$$  

This gives rise to the family of representations $I_P(\lambda)$ of $G(A)$ on $L^2_P$ given by

$$(I_P(g, \lambda)\varphi)_\lambda(x) = \varphi_\lambda(xg), \quad x, g \in G(A).$$

For any $f \in C_c^\infty(G(A)^1)$ we write

$$I_P(f, \lambda) = \int_{G(A)^1} f(g)I_P(g, \lambda) \, dg,$$

a bounded operator on $L^2_P$.

Define a height $\|\cdot\|$ on $G(A)$ and heights $\|\cdot\|_p$ on $G(Q_p)$, $p \leq \infty$, as in [4]. We have $\|x\| = \Pi_{p \leq \infty}\|x_p\|_p$.

2.2. For any $\varphi \in A^2_P$ the Eisenstein series

$$E_P(\varphi, \lambda) = \sum_{\gamma \in \mathcal{P}(Q) \setminus \mathcal{G}(Q)} \varphi_{\lambda}(\gamma g), \quad g \in G(A),$$

which converges for $\text{Re}(\lambda, \alpha^\vee) \gg 1$ for all $\alpha \in \Delta_P$, admits a meromorphic continuation to $a_{M,C}^*$ and is analytic for $\lambda \in i a_M^*$ (see [21] for the non-$K$-finite case).

By the theory of Eisenstein series we have a spectral expansion

$$K_f(x, y) = \sum_{P \in \mathcal{P}/\sim} K_{f,P}(x, y)$$

over associate classes of parabolic subgroups, where for any $P \in \mathcal{P}(M)$

$$K_{f,P}(x, y) = |W(M)|^{-1} \sum_{\pi \in \Pi_2(M)} \int_{a_{M(C)}^*} \sum_{\varphi \in \mathcal{O}_{B_P}(\pi)} E_P(x, I_P(f, \pi, \lambda)\varphi, \lambda)\overline{E_P(y, \varphi, \lambda)} \, d\lambda,$$

and $I_P(f, \pi, \lambda)$ is the restriction of $I_P(f, \lambda)$ to $L^2_{P,\pi}$.

We note that for any $f_1, f_2 \in C_c^\infty(G(A)^1)$ we have

$$K_{f_1*f_2,P}(x, y) = |W(M)|^{-1} \sum_{\pi \in \Pi_2(M)} \int_{a_{M(C)}^*} \sum_{\varphi \in \mathcal{O}_{B_P}(\pi)} E_P(x, I_P(f_1, \pi, \lambda)\varphi, \lambda)\overline{E_P(y, I_P(f_2, \pi, \lambda)\varphi, \lambda)} \, d\lambda.$$

2.3. Fix a minimal parabolic subgroup $P_0 = T_0 \ltimes U_0 \in \mathcal{P}(T_0)$ and let $\Delta_0 \subset X^*(T_0)$ be the corresponding set of simple roots. For any $T \in a_0$ let $d(T) = \min_{\alpha \in \Delta_0} \langle \alpha, T \rangle$ and let $\Lambda^T$ be Arthur’s truncation operator [1] with respect to $P_0$. It takes functions of uniform moderate growth to rapidly decreasing functions provided that $d(T) > C_0$ for some constant $C_0$ depending only on $G$. Under this condition, $\Lambda^T$ also defines an orthogonal projection on $L^2(G(Q) \setminus G(A)^1)$. We will write

$$a_0^{++} = \{T \in a_0 : d(T) > C_0\}.$$
Let $\leq$ be the partial order on $\mathfrak{a}_0^{++}$ defined by $T_1 \leq T_2$ if $T_2 - T_1$ is a linear combination of simple co-roots with non-negative coefficients. Suppose that $T_1 \leq T_2$. Then by [1, Lemma 1.1] we have $\Lambda^{T_2} \Lambda^{T_1} = \Lambda^{T_1}$. Thus, for any $\varphi$ of uniform moderate growth we have

$$\|\Lambda^{T_1} \varphi\|^2 = (\Lambda^{T_1} \varphi, \varphi) = (\Lambda^{T_2} \Lambda^{T_1} \varphi, \varphi) = (\Lambda^{T_1} \varphi, \Lambda^{T_2} \varphi) \leq \|\Lambda^{T_1} \varphi\| \|\Lambda^{T_2} \varphi\|.$$  

Hence,

\begin{equation}
\|\Lambda^{T_1} \varphi\| \leq \|\Lambda^{T_2} \varphi\|.  
\end{equation}

Let $T \in \mathfrak{a}_0^{++}$ and consider the operator $\Lambda^T \circ R(f) \circ \Lambda^T$ on $L^2(G(\mathbb{Q}) \backslash G(\mathbb{A}))$. This is a trace class integral operator whose kernel is given by

$$K^T_f(x, y) = \Lambda^{(T, T)}_{G \times G} K_f(x, y),$$

where $\Lambda^{(T, T)}_{G \times G}$ denotes truncation in both variables. Let

\begin{equation}
J^T(f) = \text{tr}(\Lambda^T \circ R(f) \circ \Lambda^T) = \int_{G(\mathbb{Q}) \backslash G(\mathbb{A})} K^T_f(x, x) \, dx.  
\end{equation}

Note that

$$J^T(f) = \text{tr}(R(f) \circ \Lambda^T) = \int_{G(\mathbb{Q}) \backslash G(\mathbb{A})} \Lambda^T_x K_f(x, y)|_{y=x} \, dx,$$

which is how this distribution is defined in [1].

On the other hand, Arthur introduced in [4] the modified kernel $k^T_f$ and its integral

\begin{equation}
\mathcal{J}^T(f) = \int_{G(\mathbb{Q}) \backslash G(\mathbb{A})} k^T_f(x, x) \, dx,  
\end{equation}

which by [11, Theorem 5.1] is absolutely convergent and a polynomial function of the parameter $T$ for all $T \in \mathfrak{a}_0^{++}$.

By a basic result of Arthur [2, Proposition 2.2] there exists a constant $C_1 > 0$ depending only on $G$ we have

\begin{equation}
J^T(f) = \mathcal{J}^T(f) \quad \text{for} \quad d(T) > C_1 \quad \text{and} \quad \text{supp} \, f \subset \{x \in G(\mathbb{A}) : \log \|x\| < C_1^{-1} d(T)\}.
\end{equation}

Spectrally expanding $K^T_f$, we may write

$$K^T_f(x, y) = \sum_{P \in \mathcal{P} / \sim} \Lambda^{(T, T)}_{G \times G} K_{f, P}(x, y),$$

where

$$\Lambda^{(T, T)}_{G \times G} K_{f, P}(x, y) = \frac{1}{|W(M)|^{-1}} \sum_{\pi \in \Pi_2(M)} \int_{i(a_0^+)^*} \sum_{\varphi \in \mathcal{O}_{B_P}(\pi)} \Lambda^T E_P(x, I_P(f, \pi, \varphi, \lambda)) \Lambda^T E_P(y, \varphi, \lambda) \, d\lambda$$

for any $P \in \mathcal{P}(M)$. Thus,

$$J^T(f) = \sum_{P \in \mathcal{P} / \sim} J^T_P(f).$$
with
\[ J_P^T(f) = \int_{G(\mathbb{Q}) \backslash G(\mathbb{A})^1} \Lambda_{G \times G}^{(T,T)} K_{f,P}(x,x) \, dx = \]
\[ |W(M)|^{-1} \sum_{\pi \in \Pi_2(M)} \int_{i(\mathfrak{a}_M^0)^*} \sum_{\varphi \in \mathcal{OB}_P(\pi)} \int_{G(\mathbb{Q}) \backslash G(\mathbb{A})^1} \Lambda^T E_P(x, I_P(f,\pi,\lambda) \varphi, \lambda) \Lambda^T E_P(x, \varphi, \lambda) \, dx \, d\lambda. \]

In particular, we have the following crucial positivity property.
\[ J_P^T(f \ast f^*) = |W(M)|^{-1} \sum_{\pi \in \Pi_2(M)} \int_{i(\mathfrak{a}_M^0)^*} \sum_{\varphi \in \mathcal{OB}_P(\pi)} \| \Lambda^T E_P(\cdot, I_P(f,\lambda) \varphi, \lambda) \|_2^2 \, d\lambda \geq 0. \]

Note that \( J_P^T(f) \) is not necessarily a polynomial in \( T \), even if \( P = G \) and \( d(T) \) is large. (In general, \( J_P^T(f) \) approximates a polynomial in \( T \) as \( d(T) \to \infty \) but we will not use this fact.)

Let \( L^2_{\text{cusp}}(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \) be the cuspidal part of \( L^2(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \) and let \( L^2_{\text{res}}(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \) be the orthogonal complement of \( L^2_{\text{cusp}}(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \) in \( L^2(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \). Denote by \( \Pi_{\text{cusp}}(G) \) (resp., \( \Pi_{\text{res}}(G) \)) the set of equivalence classes of irreducible representations that occur in \( L^2_{\text{cusp}}(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \) (resp., \( L^2_{\text{res}}(A_G G(\mathbb{Q}) \backslash G(\mathbb{A})) \)). Note that in general \( \Pi_{\text{cusp}}(G) \) and \( \Pi_{\text{res}}(G) \) are not disjoint. For any \( f \in C_c^\infty(G(\mathbb{A})^1) \) let \( R_{\text{cusp}}(f) \) denote the restriction of \( R(f) \) to \( L^2_{\text{cusp}}(G(\mathbb{Q}) \backslash G(\mathbb{A})) \). We will need the following result due to Wallach.

**Lemma 2.1** ([10]). The local components \( \pi_p \) of any \( \pi \in \Pi_{\text{res}}(G) \) are non-tempered.

Although in [ibid.] this is technically only stated for the archimedean components, the same proof, but easier, applies to the non-archimedean components as well. Namely, the cuspidal exponents of any square-integrable automorphic form \( \varphi \) lie in the negative obtuse Weyl chamber [29, I.4.11]. On the other hand, if \( \varphi \) occurs in the space of \( \pi \in \Pi_2(G) \), then every cuspidal exponent of \( \varphi \) is an exponents of \( \pi_p \) for all \( p \). Hence, \( \pi_p \) cannot be tempered unless \( \varphi \) is cuspidal.

### 3. A non-archimedean separation lemma

In this section we construct the Hecke operators (one for each prime) that will be used to amplify the non-cuspidal part of the trace formula. The construction is elementary, using the Stone–Weierstrass theorem as the main tool.

From now on we assume that \( G \) is a Chevalley group over \( \mathbb{Q} \) of rank \( r \) (in this section it would be sufficient to assume that \( G \) is split reductive over \( \mathbb{Q} \)).

3.1. Let us first recall some basic facts and set some notation pertaining to unramified representations and the Satake isomorphism. See [15, 20] for standard references. Let \( \hat{T}_0 \) be the torus dual to \( T_0 \), considered as a torus over \( \mathbb{C} \). We denote by \( \nu \mapsto \nu^\vee \) the resulting isomorphism \( X^*(\hat{T}_0) \to X_*(T_0) \) between the lattices of rational characters of \( \hat{T}_0 \) and the co-characters of \( T_0 \).
For every prime \( p \) let \( G_p = G(\mathbb{Q}_p) \), \( K_p = G(\mathbb{Z}_p) \) and denote by \( \widehat{G}_p^{unr} \) the unramified (admissible) spectrum of \( G_p \). For any \( \lambda \in \widehat{T}_0(\mathbb{C}) \) let \( \chi_\lambda \) be the unramified character of \( T_0(\mathbb{Q}_p) \) such that \( \chi_\lambda(\nu'(p)) = \nu(\lambda) \) for any \( \nu \in X^*(\widehat{T}_0) \). The map \( \lambda \mapsto \chi_\lambda \) defines an isomorphism of topological groups between \( \widehat{T}_0(\mathbb{C}) \) and the group of unramified characters of \( T_0(\mathbb{Q}_p) \). (Note that the Levi part of \( P_0 \) is \( T_0 \) since \( G \) is split.) For any \( \lambda \in \widehat{T}_0(\mathbb{C}) \), the induced representation \( \text{Ind}^{G(\mathbb{Q}_p)}_{P_0(\mathbb{Q}_p)} \chi_\lambda \) admits a unique unramified irreducible subquotient \( \pi_\lambda \), and the isomorphism class of \( \pi_\lambda \) depends only on \( W\lambda \). The map \( \lambda \mapsto \pi_\lambda \) gives rise to a homeomorphism

\[
\widehat{T}_0(\mathbb{C})/W \to \widehat{G}_p^{unr}.
\]

For any \( \pi \in \widehat{G}_p^{unr} \) we will write \( \lambda_\pi \in \widehat{T}_0(\mathbb{C})/W \) for the Frobenius–Hecke parameter of \( \pi \). For instance, if \( \pi \) is the identity representation, then \( \lambda_\pi \) is the \( W \)-orbit of the element \( \rho_p \in \widehat{T}_0(\mathbb{C}) \) that corresponds to the square-root of the modulus character of \( P_0(\mathbb{Q}_p) \), an unramified character of \( T_0(\mathbb{Q}_p) \).

Let \( \widehat{T}_0(\mathbb{C})^{hm} \) be the hermitian part of \( \widehat{T}_0(\mathbb{C}) \), namely

\[
\widehat{T}_0(\mathbb{C})^{hm} = \bigcup_{w \in W} \{ t \in \widehat{T}_0(\mathbb{C}) : w(t) = t^{-1} \}
\]

where \( \bar{t} \) is the complex conjugate of \( t \). Clearly \( \widehat{T}_0(\mathbb{C})^{hm} \) is closed in \( \widehat{T}_0(\mathbb{C}) \). Let

\[
\widehat{T}_0(\mathbb{C})^1 = \{ t \in \widehat{T}_0(\mathbb{C}) : \bar{t} = t^{-1} \} \subset \widehat{T}_0(\mathbb{C})^{hm}
\]

be the maximal compact subgroup of \( \widehat{T}_0(\mathbb{C}) \). Finally, set

\[
\widehat{T}_0(\mathbb{C})^{nt} = \widehat{T}_0(\mathbb{C})^{hm} \setminus \widehat{T}_0(\mathbb{C})^1.
\]

The sets \( \widehat{T}_0(\mathbb{C})^{hm} \), \( \widehat{T}_0(\mathbb{C})^1 \) and \( \widehat{T}_0(\mathbb{C})^{nt} \) are \( W \)-invariant. Their quotients by \( W \) parameterize the sets of hermitian, tempered and non-tempered (unramified, irreducible) representations of \( G_p \), respectively. In particular, the set of unitarizable, unramified, irreducible representations of \( G_p \) corresponds to a compact, \( W \)-invariant subset \( \widehat{T}_0(\mathbb{C})^{nt} \) of \( \widehat{T}_0(\mathbb{C})^{hm} \) (which of course depends on \( p \)). Note that \( \widehat{T}_0(\mathbb{C})^{nt} \) strictly contains \( \widehat{T}_0(\mathbb{C})^1 \) (unless \( G \) is trivial).

For any commutative \( * \)-algebra \( A \) we denote by \( A_\mathbb{R} \) the real subalgebra of self-adjoint elements of \( A \) and by \( A_{\geq 0} \subset A_\mathbb{R} \) the convex cone generated by \( x^*x \), \( x \in A \).

Let \( \mathbb{C}[[\widehat{T}_0]]^W \) be the commutative \( * \)-algebra of \( W \)-invariant, regular functions on \( \widehat{T}_0 \) where

\[
h^*(t^{-1}) = \overline{h(t)} = h(\bar{t}).
\]

As a vector space, \( \mathbb{C}[[\widehat{T}_0]]^W \) has the basis \( \lambda \chi = \frac{1}{w} \sum_{w \in W} w\lambda \lambda \), \( \lambda \in \widehat{T}_0(\mathbb{C})/W \). Note that \( h^*(t) = \overline{h(t)} \) for any \( h \in \mathbb{C}[[\widehat{T}_0]]^W \) and \( t \in \widehat{T}_0(\mathbb{C})^{hm} \). Thus,

\[
h(t) \in \mathbb{R} \quad (\text{resp.}, \ h(t) \geq 0) \quad \text{for all} \quad h \in \mathbb{C}[[\widehat{T}_0]]^W_{\geq 0} \quad (\text{resp.}, \ h \in \mathbb{C}[[\widehat{T}_0]]^W_{\geq 0} \text{ and } t \in \widehat{T}_0(\mathbb{C})^{hm}).
\]

For any \( \chi \in X^*(T_0) \) with corresponding co-character \( \chi^\vee \in X_*(\widehat{T}_0) \) let \( p^\chi = \chi^\vee(p) \in \widehat{T}_0(\mathbb{C}) \). We extend the homomorphism \( \chi \in X^*(T_0) \mapsto p^\chi \in \widehat{T}_0(\mathbb{C}) \) to a surjective homomorphism \( A_{0,\mathbb{C}}^* \to \widehat{T}_0(\mathbb{C}) \) characterized by \( p^{i\chi} = \chi^\vee(p^z) \) for any \( \chi \in X^*(T_0) \) and \( z \in \mathbb{C} \).
3.2. Fix the Haar measure on $G_p$ such that $\text{vol} K_p = 1$. Let $\mathcal{H}_p$ be the Hecke algebra of bi-$K_p$-invariant, compactly supported functions on $G_p$ with respect to convolution, with identity element $e_{K_p}$. For any $f \in \mathcal{H}_p$ we write $f^*(x) = \overline{f(x^{-1})}$. We denote by

$$S = S_p : \mathcal{H}_p \to \mathbb{C}[\hat{T}_0]^W$$

the Satake transform. It is an isomorphism of commutative $*$-algebras which is characterized by the property

$$\hat{f}(\pi) = (S_p f)(p^\lambda)$$

for any $\pi = \pi_\lambda \in \hat{G}_p^\text{unr}$, where $\hat{f}(\pi)$ is the scalar by which $f$ acts on the one-dimensional space $\pi^{K_p}$.

For any $\mu \in X^*(\hat{T}_0)$ in the positive Weyl chamber, the image under $S$ of the space of functions in $\mathcal{H}_p$ supported in $\cup \lambda \leq \mu K_p \lambda^\vee(p) K_p$ is the span of $e_{\lambda}$, $\lambda \leq \mu$. Here, $\mu - \lambda$ is a sum of simple roots with non-negative coefficients. It follows that

(5) for any $h \in \mathbb{C}[\hat{T}_0]^W$ there exists $a = a(h) > 0$ such that

$$\text{supp} S_p^{-1}(h) \subset \{ x \in G_p : \|x\|_p \leq p^a \}$$

for all $p$.

Denote by $\mu_{\text{pl},p}$ the Plancherel measure on $\hat{T}_0(\mathbb{C})^\text{unt}$ with respect to $G_p$. It is the probability measure characterized by the property

$$f(e) = \mu_{\text{pl},p}(S_p f), \ f \in \mathcal{H}_p,$$

or equivalently

$$\|f\|_2^2 = \mu_{\text{pl},p}(\|S_p f\|^2), \ f \in \mathcal{H}_p.$$  

It is well known that the support of $\mu_{\text{pl},p}$ is $\hat{T}_0(\mathbb{C})^1$. We will need the following fact.

**Lemma 3.1.** The support of the weak-$*$ limit of any convergent subsequence of $\mu_{\text{pl},p}$ is $\hat{T}_0(\mathbb{C})^1$. In other words, $\inf_p \mu_{\text{pl},p}(U) > 0$ for any open subset $U \neq \emptyset$ of $\hat{T}_0(\mathbb{C})^1$.

**Proof.** The measure $\mu_{\text{pl},p}$ is absolutely continuous with respect to the Haar measure of $\hat{T}_0(\mathbb{C})^1$. The density function is given by Macdonald’s formula [26, Chap. 5]. From this, it easily follows that the weak-$*$ limit of $\mu_{\text{pl},p}$ as $p \to \infty$ exists and is also absolutely continuous, with an explicit density function whose support is $\hat{T}_0(\mathbb{C})^1$. \qed

**Remark 3.2.** In fact, the weak-$*$ limit $\mu_{\text{ST}}$ of $\mu_{\text{pl},p}$ as $p \to \infty$, which is often called the Sato–Tate measure, can be described as follows [39, Proposition 5.3] (an observation going back at least to [31]). Let $\hat{K}$ be a maximal compact subgroup of $\hat{G}(\mathbb{C})$. The conjugacy classes of $\hat{K}$ are parameterized by $\hat{T}_0(\mathbb{C})^1/W$. Let $c : \hat{K} \to \hat{T}_0(\mathbb{C})^1/W$ be the corresponding conjugation invariant map. Then $\mu_{\text{ST}}$ is the pushforward of the normalized Haar measure on $\hat{K}$ under $c$, considered as a $W$-invariant measure on $\hat{T}_0(\mathbb{C})^1$. 


3.3. For any $M \in \mathcal{L}$ we may identify the dual torus of $T_0/T_M$ with a subtorus $\hat{T}_0^M$ of $\hat{T}_0$. Thus,

$$\hat{T}_0^M(\mathbb{C}) = \{ \lambda \in \hat{T}_0(\mathbb{C}) : \chi \lambda |_{\tau_M(\mathbb{Q}_p)} = 1 \}.$$  

Let $\hat{T}_0^M(\mathbb{C})^1 = \hat{T}_0^M(\mathbb{C}) \cap \hat{T}_0(\mathbb{C})^1$. We will use the following elementary result.

**Lemma 3.3.** For any $P \in \mathcal{P}(M)$, $M \neq G$, there exist two disjoint, $W$-invariant, open subsets $U_1, U_2 \neq \emptyset$ of $\hat{T}_0(\mathbb{C})^1$ and an open neighborhood $U$ of the identity in $\hat{T}_0(\mathbb{C})^1$ with the following property: for every $\xi \in \hat{T}_0(\mathbb{C})^1$ the set $\xi U \hat{T}_0^M(\mathbb{C})^1$ is disjoint from $U_1$ or $U_2$.

**Proof.** Fix two $W$-orbits $O_1, O_2$ in $\hat{T}_0(\mathbb{C})^1$ such that $O_1 \hat{T}_0^M(\mathbb{C})^1 \cap O_2 \hat{T}_0^M(\mathbb{C})^1 = \emptyset$. Then there exist open, $W$-invariant neighborhoods $U_i$ of $O_i$ in $\hat{T}_0(\mathbb{C})^1$ such that $U_1 \hat{T}_0^M(\mathbb{C})^1$ and $U_2 \hat{T}_0^M(\mathbb{C})^1$ are disjoint, where $A^\dagger$ denotes the closure of $A$ in $\hat{T}_0(\mathbb{C})^1$. Therefore, every $\xi \in \hat{T}_0(\mathbb{C})^1$ has an open neighborhood that is disjoint from $U_1 \hat{T}_0^M(\mathbb{C})^1$ or from $U_2 \hat{T}_0^M(\mathbb{C})^1$. By compactness, there exists an open neighborhood $U$ of the identity in $\hat{T}_0(\mathbb{C})^1$ such that for any $\xi \in \hat{T}_0(\mathbb{C})^1$ the neighborhood $\xi U$ of $\xi$ is disjoint from $U_1 \hat{T}_0^M(\mathbb{C})^1$ or from $U_2 \hat{T}_0^M(\mathbb{C})^1$, which is the assertion of the lemma. □

We now state the main technical result of this section.

**Proposition 3.4.** Let $U \neq \emptyset$ be an open, $W$-invariant subset of $\hat{T}_0(\mathbb{C})^1$. Then, there exist constants $A, B, a > 0$ and for every $p$, an element $\tau_{U,p} \in (\mathcal{H}_p)_\mathbb{R}$ such that

1. $\tau_{U,p}(e) = 0$.
2. $\|\tau_{U,p}\|_1 \leq Bp^A$.
3. $\|\tau_{U,p}\|_2 \leq B$.
4. $S_p \tau_{U,p}(x) \geq 1$ for all $x \in \hat{T}_0(\mathbb{C})^\text{hm} \setminus U$.
5. $\text{supp } \tau_{U,p} \subset \{ x \in G_p : \|x\|_p \leq p^a \}$.

**Remark 3.5.** As can be seen from the application in Proposition [5.5] later, it is mainly the constant $A$ that influences the quality of our estimates. It would be interesting to find an explicit value for $A$ (for subsets $U_1$ and $U_2$ satisfying the conclusion of Lemma 3.3).

Before proving the proposition, we need a few auxiliary results. The following lemma is necessary to take care of the non-tempered spectrum. For $G = \text{PGL}(n)$ and $G = \text{GL}(n)$ it is contained in [27, Lemma 3.1].

**Lemma 3.6.** There exists $h \in \mathbb{C}[\hat{T}_0]^{W_{\geq 0}}$ such that $\max_{\hat{T}_0(\mathbb{C})^1} h = 1$ and the restriction of $h$ to $\hat{T}_0(\mathbb{C})^\text{hm}$ defines a proper map

$$h|_{\hat{T}_0(\mathbb{C})^\text{hm}} : \hat{T}_0(\mathbb{C})^\text{hm} \to \mathbb{R}_{\geq 0}.$$  

**Proof.** The variety $\hat{T}_0/W$ is affine and hence by the Noether normalization theorem, it admits a finite (and in particular, proper) map

$$f = (p_1, \ldots, p_r) : \hat{T}_0/W \to \mathbb{A}^r.$$
In fact, if $G$ is adjoint then by [5, Théorème VI.3.1], $f$ may be taken to be an isomorphism and in general, $f$ can be easily constructed from this case.

Take $h = \sum_{i=1}^{r} p_i p_i^t$. Then $h|_{\hat{T}_0(C)_{hm}} = \sum_{i=1}^{r} |p_i|^2$ and therefore, it is a proper map to $\mathbb{R}_{\geq 0}$. Finally, we may normalize $h$ so that $\max_{\hat{T}_0(C)_{1}} h = 1$. □

**Lemma 3.7.** Let $C$ be a compact, $W$-invariant subset of $\hat{T}_0(C)_{hm}$ and let $C_1$, $C_2$ be two disjoint, closed, $W$-invariant subsets of $C$. Then, for every $\epsilon > 0$ there exists $f \in \mathbb{C}[\hat{T}_0]_{\geq 0}$ such that

1. $f|_C \leq 1$.
2. $f|_{C_1} \leq \epsilon$.
3. $f|_{C_2} \geq 1 - \epsilon$.

**Proof.** By the Stone-Weierstrass theorem, $\mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W$ is dense in the space of continuous, $W$-invariant, real-valued functions on $C$. Thus, there exists $h \in \mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W$ such that $h(C) \subset [0, 1]$, $h(C_1) \subset [0, \epsilon]$ and $h(C_2) \subset [1 - \frac{1}{2} \epsilon, 1]$. We then take $f = h^2$. □

**Lemma 3.8.** For any $h \in \mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W$ there exist $A, B > 0$ such that $\|S_p^{-1}h\|_1 \leq Bp^A$ for all $p$.

**Proof.** Clearly, it is enough to prove the lemma for elements $h$ that form a basis for $\mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W$. Let $\hat{G}$ be the dual group of $G$ (with maximal torus $\hat{T}_0$). The restriction to $\hat{T}_0$ defines an isomorphism of algebras $\mathbb{C}[\hat{G}]^{\hat{G}} \rightarrow \mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W$. We take the basis $\{h_\lambda : \lambda \in \Lambda_+\}$ formed by the traces of the irreducible rational representations of $\hat{G}$, indexed by their highest weight. By the Lusztig–Kato formula (see e.g. [16]) if $h_\lambda = S_p f_{\lambda,p}$, then $f_{\lambda,p} \geq 0$ as a function on $G_p$. (Up to a power of $p$, the values of $f_{\lambda,p}$ are given by the values of certain Kazhdan–Lusztig polynomials at $p$.) Hence, $\|f_{\lambda,p}\|_1 = h_\lambda(p) \leq B(\lambda)p^{(\lambda,p)}$. The lemma follows. □

**Remark 3.9.** The proof yields the value $A = \max_{\lambda \in \Lambda^+: c_\lambda \neq 0} \langle \lambda, \rho \rangle$ for $h = \sum_{\lambda \in \Lambda^+} c_\lambda h_\lambda$.

**Proof of Proposition 3.4.** We may assume without loss of generality that $U$ consists of regular elements, so that $U$ is open in $\hat{T}_0(C)_{hm}$. Let $V := \emptyset$ be a $W$-invariant, open subset of $\hat{T}_0(C)$ such $V^{cl} \subset U$. By Lemma 3.1 $\delta := \inf_{p} \mu_{pl,p}(V) > 0$. Let $X > 0$ be a parameter, to be determined below. Let $h$ be as in Lemma 3.6 and let $C$ be a compact, $W$-invariant subset of $\hat{T}_0(C)_{hm}$, containing $\hat{T}_0(C)$, such that $h > X + 2$ outside $C$. Let $f$ be as in Lemma 3.7 with $C_1 = V^{cl}$, $C_2 = C \setminus U$ and $\epsilon = \frac{3}{4} \delta$. Clearly,

$$
\mu_{pl,p}(f) \leq 1 - (1 - \epsilon) \mu_{pl,p}(V) \leq 1 - \frac{3}{4} \delta.
$$

Set

$$
g_p = X(f - \mu_{pl,p}(f)) + h - \mu_{pl,p}(h) \in \mathbb{C}[\hat{T}_0]_{\mathbb{R}}^W
$$

and $\tau_p = S_p^{-1}(g_p) \in (\mathcal{H}_p)_{\mathbb{R}}$. Clearly, $\tau_p(\epsilon) = \mu_{pl,p}(g_p) = 0$. On the other hand, for any $x \in \hat{T}_0(C)_{hm} \setminus U$ we have

$$
g_p(x) \geq \begin{cases} 
X(\frac{3}{4} \delta - \epsilon) - 1 = \frac{1}{2} X \delta - 1, & \text{if } x \in C, \\
-X + X + 2 - 1 = 1, & \text{otherwise}.
\end{cases}
$$
Thus, taking $X = 4\delta^{-1}$ we get $g_p \geq 1$ on $\hat{T}_0(\mathbb{C})^\text{hm} \setminus U$. It is also clear that

$$\|\tau_p\|_2^2 = \mu_{pt,p}(g_p^2) \leq \max_{\hat{T}_0(\mathbb{C})^\text{hm}} g_p^2 \leq (X + 1)^2.$$Moreover, since $\|\tau_p\|_1 \leq \|S_p^{-1}(Xf + h)\|_1 + X + 1$ we may apply Lemma 3.8 to infer the existence of constants $A, B$ such that $\|\tau_p\|_1 \leq Bp^A$ for all $p$. Finally, the support condition on $\tau_p$ follows immediately from [5].

\section{Archimedean test functions}

Next, we recall the Paley–Wiener theorem for spherical functions, following Harish-Chandra. See [14, 17] for standard references. Recall that $G$ denotes a Chevalley group defined over $\mathbb{Q}$ of rank $r$. Let $\Phi$ be the set of roots of $(G, T_0)$ and $\Phi^+$ a fixed subset of positive roots. Let $d = \dim G - \dim K_\infty$ be the dimension of the symmetric space $G(\mathbb{R})/K_\infty$. The difference $d - r$ is the dimension of a maximal unipotent subgroup, i.e., the number of positive roots of $G$. The Killing form defines an inner product on $a_0$, and hence on $a_0^\star$. We endow $a_0$ with the Lebesgue measure with respect to the Euclidean structure.

\subsection{The space of $W$-invariant Paley–Wiener functions}

Let $\mathcal{P}(a_{0,C}^\star)^W$ denote the space of $W$-invariant Paley–Wiener functions on $a_{0,C}^\star$. Thus,

$$\mathcal{P}(a_{0,C}^\star)^W = \cup_{R>0} \mathcal{P}_R(a_{0,C}^\star)^W$$

where $\mathcal{P}_R(a_{0,C}^\star)^W$ is the Fréchet space of $W$-invariant entire functions $f$ on $a_{0,C}^\star$ such that

$$\sup_{\lambda \in a_{0,C}^\star} (1 + \|\lambda\|^2)^n e^{-R|\text{Re} \lambda|} |f(\lambda)| < \infty$$

for all $n > 0$. The space $\mathcal{P}(a_{0,C}^\star)^W$ is a commutative $\ast$-algebra under pointwise multiplication and the involution $h^\ast(\lambda) = \overline{h(-\lambda)}$. Moreover, the subspaces $\mathcal{P}_R(a_{0,C}^\star)^W$, $R > 0$ are invariant under $\ast$ and $\mathcal{P}_R_1(a_{0,C}^\star)^W \mathcal{P}_R_2(a_{0,C}^\star)^W \subset \mathcal{P}_{R_1+R_2}(a_{0,C}^\star)^W$ for all $R_1, R_2 > 0$.

For any $\lambda \in a_0$, the induced representation $\text{Ind}^{G_\infty}_{G_0} e^{(\lambda,H_0(\cdot))}$ admits a unique unramified irreducible subquotient $\pi_\lambda$, which up to equivalence depends only on $W\lambda$. The map $\lambda \mapsto \pi_\lambda$ defines a homeomorphism between $a_0^\ast/W$ and the unramified dual of $G_\infty$. We denote by $\lambda_\pi$ the $W$-orbit (or a representative thereof) corresponding to an irreducible unramified representation $\pi$ of $G_\infty$. Define

$$a_{0,\text{hm}}^\ast = \cup_{w \in W} \{ \lambda \in a_{0,C}^\ast : w\lambda = -\overline{\lambda} \},$$

$$i a_{0,\text{ms}}^\ast \subset a_{0,\text{int}}^\ast = \{ \lambda \in a_{0,C}^\ast : \pi_\lambda \text{ is unitarizable} \} \subset a_{0,\text{hm}}^\ast \cap \{ \lambda \in a_{0,C}^\ast : \|\text{Re} \lambda\| \leq \|\rho\| \}.$$We have $h(\lambda) \in \mathbb{R}$ (resp., $h(\lambda) \geq 0$) for all $h \in \mathcal{P}(a_{0,C}^\star)^W$ (resp., $h \in \mathcal{P}(a_{0,C}^\star)^W_{\geq 0}$) and $\lambda \in a_{0,\text{hm}}^\ast$.

Denote by $C_c^\infty(a_0)^W$ the $\ast$-algebra of space of smooth, $W$-invariant, compactly supported functions on $a_0$ under convolution, with $f^\ast(X) = \overline{f(-X)}$. For any $R > 0$ denote by $C_R^\infty(a_0)^W$ the subspace of $C_c^\infty(a_0)^W$ consisting of functions supported on the ball $B_R$ of
radius $R$ around 0. It is a Fréchet space with respect to the usual topology. The Fourier–Laplace transform
\[ f \mapsto \int_{a_0} f(X) e^{\langle \lambda, X \rangle} \, dX \]
defines for every $R > 0$ an isomorphism of Fréchet spaces $C_R^\infty(a_0)^W \to \mathcal{P}_R(a_0^\ast, \mathbb{C})^W$, as well as a $\ast$-algebra isomorphism $C_c^\infty(a_0)^W \to \mathcal{P}(a_0^\ast, \mathbb{C})^W$.

On the other hand, let $\exp : a_0 \to T_0(\mathbb{R})$ be the exponential map. Thus, the image of $\exp$ is the identity component of $T_0(\mathbb{R})$ and $\exp x \chi = \chi(e^x)$ for all $x \in \mathbb{R}$, $\chi \in X_*(T_0)$. Fix the Haar measure on $U_0(\mathbb{R})$ as in [8, p. 37] and take the Haar measure on $G_\infty$ such that
\[ \int_{G_\infty} f(g) \, dg = \int_{K_\infty} \int_{U_0(\mathbb{R})} \int_{a_0} f(\exp Xu) \, dX \, du \, dk \]
where the measure on $K_\infty$ is normalized by $\text{vol}(K_\infty) = 1$. Let $C_c^\infty(G_\infty//K_\infty)$ be the $\ast$-algebra of smooth, compactly supported, bi-$K_\infty$-invariant functions on $G_\infty$ with respect to convolution and with $f^\ast(g) = \overline{f(g^{-1})}$. For each $R > 0$ let $C_R^\infty(G_\infty//K_\infty)$ be the subspace of $C_c^\infty(G_\infty//K_\infty)$ consisting of those functions that are supported in $K_\infty \exp B_R K_\infty$. The Harish-Chandra transform
\[ f \mapsto \delta^\frac{1}{2}_0(\exp X) \int_{U_0(\mathbb{R})} f(\exp Xu) \, du \]
defines a $\ast$-algebra isomorphism $C_c^\infty(G_\infty//K_\infty) \to C_c^\infty(a_0)^W$ which for any $R > 0$, restricts to an isomorphism of Fréchet spaces $C_R^\infty(G_\infty//K_\infty) \to C_R^\infty(a_0)^W$. Composing this with the Fourier–Laplace transform, we get a $\ast$-algebra isomorphism
\[ S_\infty : C_c^\infty(G_\infty//K_\infty) \to \mathcal{P}(a_0^\ast, \mathbb{C})^W \]
which restricts to isomorphisms of Fréchet spaces $C_R^\infty(G_\infty//K_\infty) \to \mathcal{P}_R(a_0^\ast, \mathbb{C})^W$ for all $R > 0$.

Let $\beta$ be the Plancherel density. It is given by
\[ \beta(\lambda) = |c(\lambda)c(\rho)^{-1}|^{-2}, \quad \lambda \in i\mathfrak{a}_0^\ast, \]
where $c$ is Harish-Chandra’s $c$-function. More precisely, if
\[ \phi(s) = \frac{\Gamma_R(s)}{\Gamma_R(s + 1)}, \quad \Gamma_R(s) = \pi^{-s/2} \Gamma\left(\frac{s}{2}\right), \]
then
\[ c(\lambda) = \prod_{\alpha \in \Phi^+} \phi(\langle \lambda, \alpha^\vee \rangle). \]

Therefore we have the explicit formula
\[ \beta(\lambda) = \prod_{\alpha \in \Phi^+} \left[ \frac{\Gamma\left(\frac{\langle \lambda, \alpha^\vee \rangle}{2}\right)}{\Gamma\left(\frac{\langle \lambda, \alpha^\vee \rangle + 1}{2}\right)} \right]^2 \prod_{\alpha \in \Phi^+} \frac{\langle \lambda, \alpha^\vee \rangle}{2} \tanh\left(\frac{\pi \langle \lambda, \alpha^\vee \rangle}{2}\right). \]
For any \( f \in C^\infty_c(G_\infty//K_\infty) \) we have
\[
    f(e) = \int_{i\mathfrak{a}_0^*} S_\infty f(\lambda) \beta(\lambda) \, d\lambda
\]
where the measure on \( i\mathfrak{a}_0^* \) is the dual to the one on \( \mathfrak{a}_0 \).

Following Duistermaat–Kolk–Varadarajan [8], it is useful to introduce the \( W \)-invariant function
\[
    \tilde{\beta}(\lambda) = \prod_{\alpha \in \Phi^+} (1 + |\langle \lambda, \alpha^\vee \rangle|), \quad \lambda \in i\mathfrak{a}_0^*.
\]
We have
\[
    (7) \quad \beta(\lambda) \ll \tilde{\beta}(\lambda) \ll (1 + \|\lambda\|)^{d-r}, \quad \lambda \in i\mathfrak{a}_0^*
\]
and
\[
    (8) \quad \tilde{\beta}(\lambda_1 + \lambda_2) \leq \tilde{\beta}(\lambda_1)\tilde{\beta}(\lambda_2), \quad \lambda_1, \lambda_2 \in i\mathfrak{a}_0^*.
\]

4.2. Again following [8], for any \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \) and \( \mu \in i\mathfrak{a}_0^* \) define \( g_\mu \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \) by
\[
    g_\mu(\lambda) = \frac{1}{|W|} \sum_{w \in W} g(\lambda - w\mu).
\]
Clearly, \((g_\mu)^* = (g^*)_\mu\) and if \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \) then \( g_\mu \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \). In particular, if \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \) then \( g_\mu \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \), although it is not true in general that \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)_{\geq 0} \) implies \( g_\mu \in \mathcal{P}(\mathfrak{a}_{0,C}^*)_{\geq 0} \).

The main feature of \( g_\mu \) is that it localizes near \( W\mu \). More precisely, for every \( n, R > 0 \) we have
\[
    (9) \quad |g_\mu(\lambda)| \ll_n g, R (1 + \min_{w \in W} \|\lambda - w\mu\|)^{-n} \quad \forall \lambda \in \mathfrak{a}_{0,C}^* \text{ such that } \|\Re \lambda\| \leq R.
\]
If \( S_\infty f = g \) then we define \( f_\mu = S_\infty^{-1}(g_\mu) \in C^\infty_c(G_\infty//K_\infty) \).

**Definition 4.1.**

1. We say that a function \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)^W \) has a small derivative if
\[
    \sup_{\|\Re \lambda\| \leq \|\rho\|} \|\nabla g(\lambda)\| = \sup_{\|\Re \lambda\| \leq \|\rho\|} \sup_{z \in \mathfrak{a}_{0,C}^* : \|z\| = 1} |\partial_z g(\lambda)| \leq \frac{1}{\sqrt{\|\rho\|^2 + 1}}.
\]

2. We say that \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)_{\geq 0}^W \) (and likewise, \( S_\infty^{-1} g \in C^\infty_c(G_\infty//K_\infty_{\geq 0}) \)) is special if \( g(0) = 2 |W| \) and \( g \) has a small derivative.

It is easy to see that special Paley–Wiener functions exist. Simply take any \( g \in \mathcal{P}(\mathfrak{a}_{0,C}^*)_{\geq 0}^W \) such that \( g(0) = 2 |W| \) and consider the function \( g(t\cdot) \) for \( t > 0 \) sufficiently small.

**Lemma 4.2.**

1. If \( g \) has a small derivative, then the same is true for \( g_\mu \) for any \( \mu \in i\mathfrak{a}_0^* \).

2. Let \( g \) be special. Then
\[
    (10) \quad |g_\mu(\lambda)| \geq 1 \quad \forall \lambda \in \mathfrak{a}_{0,C}^* \text{ such that } \|\Re \lambda\| \leq \|\rho\| \text{ and } \|\Im(\lambda - \mu)\| \leq 1.
\]
Proof. The first part is clear. Suppose that $g$ is special. Then $g_\mu(\mu) \geq \frac{q(0)}{|W|} \geq 2$ and by the mean value theorem, for any $\lambda \in a_{\ast, c}^*$ such that $\|\text{Re} \lambda\| \leq \|\rho\|$ we have

$$|g_\mu(\lambda) - g_\mu(\mu)| \leq \sqrt{\|\text{Re} \lambda\|^2 + \|\text{Im} (\lambda - \mu)\|^2} \sup_{\|\text{Re} \lambda\| \leq \|\rho\|} \|\nabla g_\mu(\lambda)\|$$

$$\leq \sqrt{\frac{\|\text{Re} \lambda\|^2 + \|\text{Im} (\lambda - \mu)\|^2}{\|\rho\|^2 + 1}}$$

since $g$ has a small derivative. The second part follows.

□

5. Local bounds and Weyl’s law for the cuspidal spectrum

In this section we will prove the main result of the paper, namely, the Weyl law with remainder for the cuspidal spectrum. Roughly speaking, the Weyl law is essentially the statement that for suitable test functions, the main term on the spectral (resp., geometric) side of the trace formula is the contribution of the cuspidal spectrum (resp., of the central elements). These properties have to be formulated precisely and quantitatively.

We go back to the general setup of Arthur’s trace formula in §2 except for our running assumption that $G$ is a simple Chevalley group defined over $\mathbb{Q}$. We take the product Haar measure on $G(\mathbb{A})$, where the Haar measures on $G_p$ and $G_\infty$ are as in §3.2 and §4.1, respectively. For the remainder of the paper we fix a compact open subgroup $K$ of $G(\mathbb{A}_{\text{fin}})$ and let $$v_K = |Z(\mathbb{Q}) \cap K| \frac{\text{vol}(G(\mathbb{Q}) \backslash G(\mathbb{A}))}{\text{vol} K}.$$ Let $S = S_K$ be the finite set of (finite) primes $p$ such that $K_p \not\subset K$. Thus, $K = K_S K^S$ where $K_S = K \cap G(\mathbb{Q}_S)$. Let $e_K$ be the idempotent corresponding to $K$, i.e., the characteristic function of $K$ normalized by $(\text{vol} K)^{-1}$, viewed as a smooth function on $G(\mathbb{Q}_S)$.

Consider $H^S = \otimes_{p \in S} H_p$. For any $f \in H^S$ and an admissible representation $\pi^S$ of $G(\mathbb{A}^S)$ such that $\pi^{K^S}$ is one-dimensional, we denote by $\hat{f}(\pi^S)$ the scalar by which $f$ acts on $\pi^{K^S}$. We set for any $h \in H^S$

$$m_s(h) = \sup_{x \in G(\mathbb{A}^S)/h(x) \neq 0} \log \|x\|,$$

interpreted as 0 if $h = 0$.

Let $f \in C^\infty_c(G_\infty/K_\infty)$, $h \in H^S$ and $T \in a_{\ast, c}^+$ and consider $J^T(f \otimes e_K \otimes h)$ as introduced in (2). By (4), we have

(12) $J^T(f \otimes e_K \otimes h) = J^T(f \otimes e_K \otimes h)$ for $d(T) > C_1(1 + \max_{x \in K} \log \|x\| + R + m_s(h))$. 
On the geometric side we have the following estimate for the polynomial function $\mathcal{J}^T(f \otimes e_K \otimes h)$ \cite{FinisMatz} \footnote{In \cite{ibid.}, the theorem is stated only for the constant term of the polynomial $\mathcal{J}^T(f \otimes e_K \otimes h)$. However, the proof yields the full statement of Theorem 5.1.} As in \cite{FinisMatz} (2.2), (3.1) set

$$D(\lambda) = \min_{M \in \mathcal{L}, M \neq G} \prod_{\alpha \in \Phi \setminus \Phi^{M,+}} (1 + |\langle \lambda, \alpha^\vee \rangle|)^{\frac{1}{2}},$$

for classical $G$, and

$$D(\lambda) = \frac{1}{\log(2 + \|\lambda\|)} \min_{M \in \mathcal{L}, S \subset \Phi^{+} \setminus \Phi^{M,+}} \max_{|S| = 2r} \prod_{\alpha \in S} (1 + |\langle \lambda, \alpha^\vee \rangle|)^{\frac{1}{2}}$$

for exceptional $G$. For a qualitative result it is only important that

$$|\mathcal{J}^T(f \otimes e_K \otimes h) - \tilde{v}_K h(e) f(e)| \ll_{K,R} \|h\|_1 (1 + \|T\|)^r \int_{i\mathbb{R}} |S_\infty f(\lambda)| \frac{\beta(\lambda)}{D(\lambda)} \log(2 + \|\lambda\|)^r d\lambda.$$

By (7), (8) and (9) we conclude (cf. \cite{ibid.}, Corollary 3.8):}

**Corollary 5.2.** For any $f \in C^\infty_c(G_\infty//K_\infty)$ and either $F = f_\mu$ or $F = f_\mu * f_\mu^*$ we have

$$|\mathcal{J}^T(F \otimes e_K \otimes h) - \tilde{v}_K h(e) F(e)| \ll_{f,K} \|h\|_1 (1 + \|T\|)^r \log(2 + \|\mu\|)^r \frac{\beta(\mu)}{D(\mu)}$$

for all $h \in \mathcal{H}^S$, $\mu \in i\mathbb{R}$ and $T \in a_0^{++}$.}

Turning to the spectral side, for any $M \in \mathcal{L}$ denote by $\Pi_2(M)^{K_\infty K}$ the subset of $\Pi_2(M)$ consisting of the representations such that $I_P(\pi, 0)$ admits a non-zero $K_\infty K$-invariant vector for any (or equivalently, all) $P \in \mathcal{P}(M)$. In particular, if $\pi = \otimes_{p \leq \infty} \pi_p \in \Pi_2(M)^{K_\infty K}$, then $\pi_p$ is unramified for all $p \notin S$ (including $p = \infty$). As in \cite{ibid.} for any $f \in C^\infty_c(G_\infty//K_\infty)$, $h \in \mathcal{H}^S$ and $T \in a_0^{++}$ we have

$$J^T(f \otimes e_K \otimes h) = \sum_{P \in \mathcal{P}/\sim} J^T_P(f \otimes e_K \otimes h),$$

where

$$J^T_P(f \otimes e_K \otimes h) = |W(M)|^{-1} \sum_{\pi \in \Pi_2(M)^{K_\infty K}} \int_{i\mathbb{R}_M} S_\infty f(\lambda_{\pi_\infty} + \lambda) \hat{h}(I_P(\pi^S, \lambda)) \sum_{\varphi \in OB_P(\pi)^{K_\infty K}} \|A^T E_P(\cdot, \varphi, \lambda)\|^2 d\lambda$$

\footnote{In \cite{ibid.}, the theorem is stated only for the constant term of the polynomial $\mathcal{J}^T(f \otimes e_K \otimes h)$. However, the proof yields the full statement of Theorem 5.1.}
for $P \in \mathcal{P}(M)$. Here $\lambda_{\pi, \infty} \in (a_0^M)^*/W_M$ is the archimedean parameter of $\pi$ and $\mathcal{O}B_P(\pi)^{K_{\infty}K}$ is an orthonormal basis of the finite-dimensional space $(A_{P, \pi}^{(2)})^{K_{\infty}K}$ (the $K_{\infty}K$-invariant part of $A_{P, \pi}^{(2)}$).

Let $\nu_{P}^{T,K}$ be the Radon measure on $i a_0^* / W$ given by

$$\nu_{P}^{T,K}(f) = |W(M)|^{-1} \int_{\pi \in \Pi_{G}(M)^{K_{\infty}K}} f(\text{Im} \lambda_{\pi, \infty} + \lambda) \sum_{\varphi \in \mathcal{O}B_P(\pi)^{K_{\infty}K}} \|\Lambda_T E_P(\cdot, \varphi, \lambda)\|_2^2 d\lambda, \quad f \in C_c(i a_0^*)^W,$$

and let

$$\nu^{T,K} = \sum_{P \in \mathcal{P}/\sim} \nu_{P}^{T,K}.$$

It is clear from (1) that (15) $\nu_{T_1,K} \leq \nu_{T_2,K}$ if $T_1 \leq T_2$.

We also introduce the Radon measures

$$\nu_{\text{cusp}}^{K} = \sum_{\pi \in \Pi_{\text{cusp}}(G)} \dim((A_{G_{\text{cusp}}}^{(2)})^{K_{\infty}K}) \delta_{\text{Im} \lambda_{\pi, \infty}},$$

$$\nu_{\text{res}}^{T,K} = \nu_{G}^{T,K} - \nu_{\text{cusp}}^{T,K},$$

$$\nu_{\text{ncusp}}^{T,K} = \nu_{\text{res}}^{T,K} - \nu_{\text{cusp}}^{T,K} = \sum_{P \in \mathcal{P}, P \not \sim G/\sim} \nu_{P}^{T,K} + \nu_{\text{res}}^{T,K}.$$

Let

$$m_{\text{cusp}}^{K}(A) = \sum_{\pi \in \Pi_{\text{cusp}}(G)^{K_{\infty}K}, \lambda_{\pi, \infty} \in A} \dim(A_{G_{\text{cusp}}}^{(2)})^{K_{\infty}K}$$

for any bounded, $W$-invariant subset $A$ of $a_0^* \subset \mathbb{C}$, and note that

$$\nu_{\text{cusp}}^{K}(A) = m_{\text{cusp}}^{K}(A + a_0^*), \quad A \subset i a_0^*.$$

Finally, we decompose the Radon measure $\nu_{\text{cusp}}^{K}$ as the sum of two Radon measures

$$\nu_{\text{cusp}}^{K} = \nu_{\text{cusp, temp}}^{K} + \nu_{\text{cusp, nt}}^{K}$$

(the tempered and non-tempered part of $\nu_{\text{cusp}}^{K}$) where

$$\nu_{\text{cusp, temp}}^{K}(A) = m_{\text{cusp}}^{K}(A), \quad A \subset i a_0^*.$$

For any $\mu \in i a_0^*$ and $R \geq 0$ denote by $B_R(\mu)$ the ball of radius $R$ around $\mu$ in $i a_0^*$.

The case $h = 1$ of Corollary 5.2 is already sufficient to give the following qualitative local bounds. Let

$$\mathfrak{d}(\mu) = \min_{1 \neq w \in W} \|w\mu - \mu\| = \min_{\alpha > 0} |\langle \mu, \alpha^\vee \rangle| \|\alpha\|.$$


Lemma 5.3. For any $\mu \in ia_0^+$ and $T \in a_0^{++}$ we have

$$\nu^{T,K}(WB_1(\mu)) \ll_K (1 + \|T\|)^r \tilde{\beta}(\mu).$$

In particular,

$$\nu_{cusp}^{K}(WB_1(\mu)) \ll_K \tilde{\beta}(\mu)$$

and

$$\nu_{cusp,nt}^{K}(WB_1(\mu)) \ll_K \begin{cases} (1 + \|\mu\|)^{d-r-1}, & \text{if } \delta(\mu) \leq 2, \\ 0, & \text{otherwise}. \end{cases}$$

Proof. Fix $f \in C_c^\infty(G_\infty/\K_\infty)_{\geq 0}$ special (see Definition 4.1). By (10) and (14) we have

$$J^T(f_\mu \ast f_\mu \otimes e_\K) \geq \nu^{T,K}(WB_1(\mu))$$

for any $T \in a_0^{++}$. On the other hand, by Corollary 5.2 (with $h = 1$) and (12) we have

$$|J^T(f_\mu \ast f_\mu \otimes e_\K)| \ll_{f,K} (1 + \|T\|)^r \tilde{\beta}(\mu)$$

for all $T \in a_0^{++}$ with $d(T)$ sufficiently large (depending on the support of $f$ and on $K$). By (13) this holds for all $T \in a_0^{++}$.

The last estimate is a consequence of the second one together with the fact that $\tilde{\beta}(\mu) \ll (1 + \|\mu\|)^{d-r-1}$ if $\delta(\mu) \leq 2$ and $WB_1(\mu) \cap a_0^{++,\text{fin}} = \emptyset$ otherwise. The lemma follows. \qed

It is possible to use positivity to deduce a more precise upper bound on the cuspidal spectrum. To go beyond upper bounds, we will need the full force of Theorem 5.1 (i.e., for an arbitrary $h$). First, we make the following elementary, but crucial observation.

Lemma 5.4. For any $M \in \mathcal{L}$ there exists an integer $N$ (depending on $K$), divisible by $\prod_{p \in S} p$, such that for any $\pi \in \Pi_2(M)^K$, the central character of $\pi_p$ is trivial on $T_M(\Q_p)$ whenever $p \equiv 1 \pmod{N}$.

Proof. This is simply because if $K'$ is an open compact subgroup of $\A_0^*$, then the characters of $\R \times \Q^* \setminus \A^*$ of level $K'$ are precisely the Dirichlet characters of level $N'$ where $N'$ is determined by $K'$. \qed

The following proposition is the key assertion for the proof of the Weyl law.

Proposition 5.5. We have

$$\nu^{T,K}_{\text{ncusp}}(WB_1(\mu)) \ll_K \tilde{\beta}(\mu)(1 + \|T\|)^r \log(2 + \|\mu\|)^{3r} D(\mu)^{2A+1-1} \ll \tilde{\beta}(\mu)(1 + \|T\|)^r (1 + \|\mu\|)^{\delta}$$

for all $\mu \in ia_0^+$ and $T \in a_0^{++}$, where $A$ is the constant from Proposition 3.4 (which depends only on $G$), and $0 < \delta < (4A + 2)^{-1}$.

Proof. First note that it is enough to bound $\nu^{T,K}_P(WB_1(\mu))$ for all proper parabolic subgroups $P$, as well as $\nu^{T,K}_{\text{res}}(WB_1(\mu))$.

To deal with the first case, fix a proper parabolic subgroup $P \in \mathcal{P}(M)$ and let $N$ be as in Lemma 5.4. By Lemma 3.3 there exist two open, $W$-invariant subsets $U_1, U_2 \neq \emptyset$
of $\hat{T}_0(\mathbb{C})^1$ and a number $\delta_1 > 0$, such that for every $p$ and $\mu \in i\mathfrak{a}_0^+$ there exists an index $j_p(\mu) \in \{1, 2\}$ with
\begin{equation}
\label{eq:16}
p^{\overline{\lambda} T_0^M(\mathbb{C}) \cap U_{j_p(\mu)} = \emptyset} \quad \text{for all } \lambda \in \mathcal{B}_{\overline{\lambda} T_0^M(\mu_p^*)}.
\end{equation}

Let $\mu \in i\mathfrak{a}_0^+$ and let $X \geq 2$ be a parameter depending on $\mu$ (to be determined below). Set
\[ P_{X,N} = \{ p \leq X : p \equiv 1 \pmod{N} \}, \quad Y = \#P_{X,N}, \]
and let $R = \frac{\delta}{\log X}$. For each $p \in P_{X,N}$ let $\tau_{U_p} \in (\mathcal{H}^S_p)^\mathbb{R}$, $j = 1, 2$, be as in Proposition 3.4 and consider the following combination of Hecke operators:
\[ \theta = \theta_{X,\mu,p} = \sum_{p \in P_{X,N}} \tau_{U_{j_p(\mu)}} p \in \mathcal{H}_p^S. \]

As a consequence of Proposition 3.4, $\hat{\theta}$ satisfies the following properties.

1. For any $\pi \in \Pi_2(M)^K$, $\lambda \in \mathcal{B}_R(\mu_p)$ and $p \in P_{X,N}$ we have $\overline{\tau_{U_{j_p(\mu)}}}(I_p(\pi, \lambda)) \geq 1$ by (16) and Lemma 5.4. Hence, $\hat{\theta}(I_p(\pi^S, \lambda)) \geq Y$.
2. The functions $\overline{\tau_{U_{j_p(\mu)}}}$, $p \in P_{X,N}$ are pairwise orthogonal in $\mathcal{H}^S$.
3. $\|\theta\|_1 \leq \sum_{p \in P_{X,N}} \|\overline{\tau_{U_{j_p(\mu)}}}\|_1 \leq B X^4 Y$.
4. $\|\theta\|_2^2 = \sum_{p \in P_{X,N}} \|\overline{\tau_{U_{j_p(\mu)}}}\|^2_2 \leq B^2 Y$.
5. $\text{ms}(\theta) \ll \log X$.

Fix $f \in C_c^\infty(G_\infty//K_\infty)^{\geq 0}$ special and let $F_{\mu,X} = f \ast f \otimes e_K \otimes \theta \ast \theta \in C_c^\infty(G(\mathbb{A})^1)$. By positivity, the first property of $\theta$ implies that
\[ Y^2 v_p^{T,K}(W\mathcal{B}_R(\mu)) \leq J_p^T(F_{\mu,X}) \leq J^T(F_{\mu,X}) \]
for all $T \in \mathfrak{a}_0^{++}$. On the other hand, Corollary 3.2 and the remaining properties of $\theta$ yield that for $d(T) > C_2 \log X$, where $C_2$ is independent of $\mu$, we have
\[ J^T(F_{\mu,X}) = J_p^T(F_{\mu,X}) \ll K B^2 \tilde{\beta}(\mu) \left[ Y + \frac{X^2 A Y^2 (1 + \|T\|)^r \log(2 + \|\mu\|)^r}{D(\mu)} \right]. \]

Since $Y \geq c_3 X / \log X$ for all $X \geq C_3$, where $c_3 > 0$ and $C_3 \geq 2$ depend only on $K$, we can take $X = \max(D(\mu)^{(2A+1)^{-1}}, C_3)$ to get
\[ \nu_p^{T,K}(W\mathcal{B}_R(\mu)) \ll K \tilde{\beta}(\mu) (1 + \|T\|)^r \log(2 + \|\mu\|)^r D(\mu)^{(2A+1)^{-1}}. \]

Turning to the residual contribution, it follows from Lemma 2.1 and the fourth property of Proposition 3.4 that for any $\pi \in \Pi_{\text{res}}(G)$ we have
\[ \hat{\theta}(\pi^S) \geq Y. \]

Thus, by the same argument as above
\[ \nu_{\text{res}}^{T,K}(W\mathcal{B}_R(\mu)) \ll K \tilde{\beta}(\mu) (1 + \|T\|)^r \log(2 + \|\mu\|)^r D(\mu)^{(2A+1)^{-1}}. \]
Since $R \geq \frac{\delta}{\log(1+\|\mu\|)}$ for a suitable $\delta > 0$, we obtain

$$\nu_{\text{nucsp}}^{T,K}(WB_1(\mu)) \ll_K \frac{\bar{\beta}(\mu)(1 + \|T\|)^r \log(2 + \|\mu\|)^{2r}}{D(\mu)(2A+1)^{2r}}$$

at least under the condition $d(T) > C_2 \log X \gg \log(2 + \|\mu\|)$. However, using (15) we can remove this condition (at the cost of replacing the exponent $2r$ of $\log(2 + \|\mu\|)$ by $3r$). □

**Theorem 5.6.** There exists $\delta > 0$ (depending only on $G$) such that for any $f \in C^\infty_c(G_\infty//K_\infty)$ and $\mu \in ia_0^*$ we have

$$|\text{tr} R_{\text{cusp}}(f_\mu \otimes e_K) - v_K f_\mu(e)| \ll_{f,K} \frac{\bar{\beta}(\mu)}{(1 + \|\mu\|)^{\delta}}.$$

**Proof.** For any $g \in P(a_0^,(\cdot),C)$ let $Mg \in C(i a_0^*)^W$ be given by

$$Mg(\lambda) = \max_{x \in a_0^*: x \leq \|\mu\|} |g(\lambda + x)|, \quad \lambda \in ia_0^*.$$

Then, for any $f \in C^\infty_c(G_\infty//K_\infty)$ we have

$$|\text{tr} R_{\text{cusp}}(f \otimes e_K) - J^T (f \otimes e_K)| \leq \nu_{\text{nucsp}}^{K,T}(M(S_\infty f)).$$

Thus, using (12) and Corollary 5.2, in order to finish the proof it remains to prove that

$$\nu_{\text{nucsp}}^{K,T}(M(S_\infty f_\mu)) \ll_{f,K} \frac{\bar{\beta}(\mu)(1 + \|T\|)^r}{(1 + \|\mu\|)^{\delta}}$$

for all $T \in a_0^{++}$. Clearly,

$$\nu_{\text{nucsp}}^{K,T}(M(S_\infty f_\mu)) \leq \sum_{k=1}^{\infty} \nu_{\text{nucsp}}^{K,T}(WB_k(\mu)) \sup_{(WB_{k-1}(\mu))^c} M(S_\infty f_\mu).$$

By (9), for every $n > 0$ we have

$$\sup_{(WB_{k-1}(\mu))^c} M(S_\infty f_\mu) \ll_{f,n} k^{-n}.$$

Covering $B_k(\mu)$ by $O(k^r)$ balls of radius 1 with centers in $B_k(\mu)$ and using either Proposition 5.5 and (8) if $k \leq \frac{1}{2} \|\mu\|$ or Lemma 5.3 and (7) if $k > \frac{1}{2} \|\mu\|$, we get

$$\nu_{\text{nucsp}}^{K,T}(WB_k(\mu)) \ll_K k^d(1 + \|T\|)^r \cdot \begin{cases} \frac{\bar{\beta}(\mu)}{(1 + \|\mu\|)^{\delta}} & \text{if } k \leq \frac{1}{2} \|\mu\|, \\ 1 & \text{otherwise.} \end{cases}$$

Our claim follows. □

**Corollary 5.7.** For any $f \in C^\infty_c(G_\infty//K_\infty)$ and $\mu \in ia_0^*$ we have

$$|\nu^K_{\text{cusp,temp}}(S_\infty f_\mu) - v_K f_\mu(e)| \ll_{f,K} (1 + \|\mu\|)^{d-r-\delta}.$$
Proof. Indeed,
\[ \left| \text{tr} R_{\text{cusp}}(f_\mu \otimes e_K) - \nu^K_{\text{cusp, temp}}(S_\infty f_\mu) \right| \leq \nu^K_{\text{cusp, nt}}(|S_\infty f_\mu|). \]
On the other hand, by (9), for all \( n > 0 \) we have
\[ \nu^K_{\text{cusp, nt}}(|S_\infty f_\mu|) \ll f_n \sum_{k=1}^\infty \nu^K_{\text{cusp, nt}}(W B_k(\mu)) k^{-n}, \]
and by Lemma 5.3
\[ \nu^K_{\text{cusp, nt}}(W B_k(\mu)) \ll k^r (k + \|\mu\|)^{d-r-1}. \]
Hence, the corollary follows from Theorem 5.6. \( \square \)

**Definition 5.8.** For any \( A \subset \mathfrak{a}_0^* \) and \( R > 0 \) let
\[ \partial_R A = \{ x \in \mathfrak{a}_0^* : \mathcal{B}_R(x) \cap A \neq \emptyset \text{ and } \mathcal{B}_R(x) \nsubseteq A \}. \]

Note that for any \( x \in \partial_R A \) we have \( \mathcal{B}_R(x) \subset \partial_{2R} A \). Hence, by the Vitali covering lemma, \( \partial_A A \) is covered by \( O(\text{vol } \partial_1 A) \) balls of radius 1. On the other hand, it is clear that if \( x \in \partial_{R} A \), then \( d(x, \partial_A A) \leq R \) for any \( \epsilon > 0 \). (Here \( d(x, B) = \inf_{b \in B} \| x - b \| \).) It follows that for every \( R \geq 1 \),
\[ \partial_R A \text{ is covered by } O(R^d \text{vol } \partial_1 A) \text{ balls of radius 1.} \]

In particular, \( \text{vol}(\partial_R A) \ll R^d \text{vol } (\partial_1 A) \).

Let \( D \subset \mathfrak{a}_0^* \) be a \( W \)-invariant bounded measurable set. For any \( f \in C_c^\infty(G_\infty//K_\infty) \) define \( f_D \in C_c^\infty(G_\infty//K_\infty) \) by
\[ f_D(g) = \int_D f_\mu(g) \, d\mu, \quad g \in G_\infty. \]
Similarly, for any \( g \in \mathcal{P}(\mathfrak{a}_0^*//G_\infty) \) let \( g_D \in \mathcal{P}(\mathfrak{a}_0^*//G_\infty) \) given by
\[ g_D(\lambda) = \int_D g_\mu(\lambda) \, d\mu = \int_D g(\lambda - \mu) \, d\mu, \quad \lambda \in \mathfrak{a}_0^*. \]
Thus, if \( S_\infty f = g \) then \( S_\infty f_D = g_D \).

**Lemma 5.9.** Let \( D \subset \mathfrak{a}_0^* \) be a \( W \)-invariant bounded measurable set and \( \chi_D \) its characteristic function. Let \( f \in \mathcal{P}(\mathfrak{a}_0^*//G_\infty) \) with \( \int_{\mathfrak{a}_0^*} f(\lambda) \, d\lambda = 1 \). Then for any \( n > 0 \) and \( \lambda \in \mathfrak{a}_0^* \) we have
\[ |f_D(\lambda) - \chi_D(\lambda)| \ll_f n \begin{cases} (1 + d(\lambda, D^c))^{-n}, & \lambda \in D, \\ (1 + d(\lambda, D))^{-n}, & \text{otherwise}. \end{cases} \]
This follows immediately from the rapid decay of \( f \) on \( \mathfrak{a}_0^* \).

**Corollary 5.10.** Let \( \nu \) be a measure on \( \mathfrak{a}_0^*/W \) satisfying
\[ \nu(W \mathcal{B}_1(\lambda)) \ll (1 + \|\lambda\|)^m, \quad \lambda \in \mathfrak{a}_0^*. \]
for some $m \geq 0$. Then for any $W$-invariant, bounded measurable set $D$ in $\mathfrak{ia}_0^*$ and any $f \in \mathcal{P}(\mathfrak{a}_{0,\mathbb{C}}^*)^W$ such that $\int_{\mathfrak{ia}_0^*} f(\lambda) \, d\lambda = 1$ we have

$$|\nu(D) - \nu(f_D)| \ll_f \text{vol}(\partial_1 D)(1 + \|D\|)^m,$$

where $\|D\| = \sup_{\lambda \in D}\|\lambda\|$.

**Proof.** By Lemma [5.9] for every $n > 0$ we have

$$|\nu(D) - \nu(f_D)| \ll_{f,n} \sum_{k=1}^\infty \nu(\partial_k D)k^{-n}.$$

On the other hand, by (18), for any $k \geq 1$, $\partial_k D$ is covered by $O(k^r \text{vol}(\partial_1 D))$ balls of radius 1 and we may as well assume that the centers of these balls lie in $\mathcal{B}(0, \|D\| + k)$. Hence, by the assumption on $\nu$ we have

$$\sum_{k=1}^\infty \nu(\partial_k D)k^{-n} \ll \text{vol}(\partial_1 D)(1 + \|D\|)^m$$

provided that $n \geq r + m + 2$. The corollary follows. \hfill \Box

**Theorem 5.11.** There exists $\delta > 0$ such that for any open subgroup $K$ of $G(\mathfrak{a}_{\text{fin}})$ and any $W$-invariant, bounded measurable set $D$ in $\mathfrak{ia}_0^*$ we have

$$|m^{K}_{\text{cusp}}(D) - v_K\mu_{pl}(D)| \ll_K \text{vol}(\partial_1 D)(1 + \|D\|)^{d-r} + (1 + \|D\|)^{d-\delta}.$$

In particular, if the boundary of $D$ is rectifiable (or more generally, if the $(r-1)$-dimensional upper Minkowski content of $D$ is finite), then for all $t \geq 1$ we have

$$|m^{K}_t(D) - v_K\mu_{pl}(D)| \ll_{K,D} t^{d-\delta}.$$

Note that $\mu_{pl}(D) = C_D t^d + O(t^{d-1})$ for $t \to \infty$.

**Proof.** Fix $f \in \mathcal{P}(\mathfrak{a}_{0,\mathbb{C}}^*)^W$ such that $\int_{\mathfrak{ia}_0^*} f(\lambda) \, d\lambda = 1$. By Corollary 5.7

$$|\nu^{K}_{\text{cusp, temp}}(f_\mu) - v_K\mu_{pl}(f_\mu)| \ll_{f,K} (1 + \|\mu\|)^{d-r-\delta}.$$

Integrating over $\mu \in D$ we get

$$|\nu^{K}_{\text{cusp, temp}}(f_D) - v_K\mu_{pl}(f_D)| \ll_{f,K} (1 + \|D\|)^{d-\delta}.$$

On the other hand, by Corollary 5.10 we have

$$|\nu^{K}_{\text{cusp, temp}}(f_D) - m^{K}_{\text{cusp}}(D)| \ll_{f,K} \text{vol}(\partial_1 D)(1 + \|D\|)^{d-r}$$

and

$$|\mu_{pl}(f_D) - \mu_{pl}(D)| \ll_f \text{vol}(\partial_1 D)(1 + \|D\|)^{d-r}.$$

The theorem follows. \hfill \Box

For the balls $\mathcal{B}_t(0)$, $t \geq 1$, we obtain:
Corollary 5.12.

\[(19) \quad m^K_{\text{cusp}}(B_t(0)) = \frac{v_K t^d}{(4\pi)^{d/2} \Gamma(d/2 + 1)} + O_K(t^{d-\delta})\]

and

\[(20) \quad m^K_{\text{cusp}}(\{\lambda \in a^*_0 : ||\Im \lambda||^2 - ||\Re \lambda||^2 + ||\rho||^2 \leq t^2\}) = \frac{v_K t^d}{(4\pi)^{d/2} \Gamma(d/2 + 1)} + O_K(t^{d-\delta}).\]

Thus, the Weyl law with remainder holds for the cuspidal spectrum of the adelic quotient \(G(\mathbb{Q}) \backslash \mathbb{G}(\mathbb{A})/K_{\infty}K\). If in addition \(G\) is simply connected, then we obtain

\[N_{\Gamma \backslash G(\mathbb{R})/K_{\infty}}(t) = \frac{\text{vol}(\Gamma \backslash G(\mathbb{R})/K_{\infty}) t^d}{(4\pi)^{d/2} \Gamma(d/2 + 1)} + O_K(t^{d-\delta})\]

for \(\Gamma = G(\mathbb{Q}) \cap K\) by strong approximation (cf. Theorem 5.11).

Indeed, the first statement follows from the computation of \(\mu_{\text{pl}}(B_t(0))\) (cf. [8, §8]). The second statement follows from the first one since \(||\Re \lambda|| \leq ||\rho||\) for any \(\lambda \in a^*_{0,\text{unt}}\).

Finally, note that the Laplace eigenvalue corresponding to an archimedean parameter \(\lambda\) is \(||\Im \lambda||^2 - ||\Re \lambda||^2 + ||\rho||^2\).

Finally, we give a generalization of Theorem 5.11 incorporating Hecke operators \(\tau \in \mathcal{H}_S\). Set

\[m^K_{\text{cusp}}(A, \tau) = \sum_{\pi \in \Pi_{\text{cusp}}(G)^{K_{\infty}K}, \lambda \in a^*_0} \text{tr} \tau|_{(A_{\text{cusp}, \pi}^K)^K_{\infty}K} \text{tr} \tau\]

for any bounded, \(W\)-invariant subset \(A\) of \(a^*_0\).

Theorem 5.13. There exists \(\delta > 0\) such that for any open subgroup \(K\) of \(G(\mathbb{A}_{\text{fin}})\), any \(\tau \in \mathcal{H}_S\) and any \(W\)-invariant, bounded measurable set \(D\) in \(i a^*_0\) we have

\[|m^K_{\text{cusp}}(D, \tau) - v_K \tau(e) \mu_{\text{pl}}(D)| \ll_K ||\tau||_1 (\text{vol}(\partial D)(1 + ||D||)^{d-r} + (1 + ms(\tau))^r (1 + ||D||)^{d-\delta}).\]

In particular, if the \((r-1)\)-dimensional upper Minkowski content of \(D\) is finite, (e.g., if the boundary of \(D\) is rectifiable), then for all \(t \geq 1\) we have

\[|m^K_{\text{cusp}}(tD, \tau) - v_K \tau(e) \mu_{\text{pl}}(tD)| \ll_K ||\tau||_1 (1 + ms(\tau))^r t^{d-\delta}.\]

Proof. We first prove the analog of Theorem 5.6 for any \(f \in C^\infty_c(G_{\infty}/K_{\infty})\) and \(\mu \in i a^*_0\) we have

\[(21) \quad |\text{tr} R_{\text{cusp}}(f_\mu \otimes e_K \otimes \tau) - v_K \tau(e) f_\mu(e)| \ll_{f,K} ||\tau||_1 (1 + ms(\tau))^r \frac{\tilde{\beta}(\mu)}{(1 + ||\mu||)^{\delta}}.\]

As in the proof of Theorem 5.6, we start with

\[|\text{tr} R_{\text{cusp}}(f_\mu \otimes e_K \otimes \tau) - J^K_T(f_\mu \otimes e_K \otimes \tau)| \leq ||\tau||_1 v^K_{\text{ncusp}}(M(S_{\infty} f_\mu))\]

for all \(\tau \in \mathcal{H}_S\) and \(T \in a^*_{0,++}\). Moreover, by (12) we have

\[J^K_T(f_\mu \otimes e_K \otimes \tau) = \mathbb{J}^T(f_\mu \otimes e_K \otimes \tau)\]
for $d(T) \gg_{f,K} 1 + ms(\tau)$. Using the upper bound of (17) for $\nu_{\text{cusp}}^{K,T}(M(\mathcal{S}_\infty f_\mu))$ and the estimate of Corollary 5.2 for $\mathcal{J}(f_\mu \otimes e_K \otimes \tau)$ for a suitable value of $T$, one obtains (21).

Denote by $R_{\text{cusp,temp}}$ the restriction of $R_{\text{cusp}}$ to the space of cuspidal representations tempered at infinity. As in the proof of Corollary 5.7, one obtains

$$|\text{tr} R_{\text{cusp}}(f_\mu \otimes e_K \otimes \tau) - \text{tr} R_{\text{cusp,temp}}(f_\mu \otimes e_K \otimes \tau)| \ll_{f,K} \|\tau\|_1(1 + \|\mu\|)^{d-r-1},$$

and if in addition $\int_{\text{int}_0^*} \mathcal{S}_\infty f(\lambda) \, d\lambda = 1$, then by Corollary 5.10

$$|\text{tr} R_{\text{cusp,temp}}(f_D \otimes e_K \otimes \tau) - m_{\text{cusp}}^K(D, \tau)| \ll_{f,K} \|\tau\|_1 \text{vol}(\partial_1 D)(1 + \|D\|)^{d-r}.$$

From these estimates one can deduce the theorem exactly as in the proof of Theorem 5.11. □

References

[1] James Arthur, A trace formula for reductive groups. II. Applications of a truncation operator, Compositio Math. 40 (1980), no. 1, 87–121. MR558260 (81b:22018)

[2] , On a family of distributions obtained from Eisenstein series. I. Application of the Paley-Wiener theorem, Amer. J. Math. 104 (1982), no. 6, 1243–1288. MR681737 (85k:22044)

[3] , An introduction to the trace formula, Harmonic analysis, the trace formula, and Shimura varieties, 2005, pp. 1–263. MR2192011 (2007d:11058)

[4] James G. Arthur, A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbb{Q})$, Duke Math. J. 45 (1978), no. 4, 911–952. MR518111 (80d:10043)

[5] N. Bourbaki, Éléments de mathématique. Fasc. XXXIV. Groupes et algèbres de Lie. Chapitre IV: Groupes de Coxeter et systèmes de Tits. Chapitre V: Groupes engendrés par des réflexions. Chapitre VI: systèmes de racines, Actualités Scientifiques et Industrielles, No. 1337, Hermann, Paris, 1968. MR0240238 (39 #1590)

[6] Alina Carmen Cojocaru and M. Ram Murty, An introduction to sieve methods and their applications, London Mathematical Society Student Texts, vol. 66, Cambridge University Press, Cambridge, 2006. MR2200366 (2006k:11184)

[7] Harold Donnelly, On the cuspidal spectrum for finite volume symmetric spaces, J. Differential Geom. 17 (1982), no. 2, 239–253. MR664496 (83m:58079)

[8] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature, Invent. Math. 52 (1979), no. 1, 27–93. MR532745 (82a:58050a)

[9] T. Finis and E. Lapid, On the analytic properties of intertwining operators I: global normalizing factors, Bull. Iranian Math. Soc. 43 (2017), no. 4, 235–277. MR3711830

[10] Tobias Finis and Erez Lapid, On the continuity of Arthur’s trace formula: the semisimple terms, Compos. Math. 147 (2011), no. 3, 784–802. MR2801400

[11] , On the Continuity of the Geometric Side of the Trace Formula, Acta Math. Vietnam. 41 (2016), no. 3, 425–455. MR3534542

[12] Tobias Finis, Erez Lapid, and Werner Müller, On the spectral side of Arthur’s trace formula—absolute convergence, Ann. of Math. (2) 174 (2011), no. 1, 173–195. MR2811597

[13] Tobias Finis and Jasmin Matz, On the asymptotics of Hecke operators for reductive groups, 2019. arXiv:1905.09078.

[14] Ramesh Gangolli and V. S. Varadarajan, Harmonic analysis of spherical functions on real reductive groups, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], vol. 101, Springer-Verlag, Berlin, 1988. MR954385 (89m:22015)

[15] Benedict H. Gross, On the Satake isomorphism, Galois representations in arithmetic algebraic geometry (Durham, 1996), 1998, pp. 223–237. MR1696481 (2000e:22008)
[16] Thomas J. Haines, Robert E. Kottwitz, and Amritanshu Prasad, *Iwahori-Hecke algebras*, J. Ramanujan Math. Soc. 25 (2010), no. 2, 113–145. MR2642451

[17] Sigurdur Helgason, *Groups and geometric analysis*, Mathematical Surveys and Monographs, vol. 83, American Mathematical Society, Providence, RI, 2000. Integral geometry, invariant differential operators, and spherical functions, Corrected reprint of the 1984 original. MR1790156

[18] Lars Hörmander, *The spectral function of an elliptic operator*, Acta Math. 121 (1968), 193–218. MR0609014 (58 #29418)

[19] Lizhen Ji, *The trace class conjecture and the Weyl upper bound on the discrete spectrum of arithmetic groups*, Math. Res. Lett. 4 (1997), no. 4, 501–511. MR1470422 (98i:11037)

[20] Robert P. Langlands, *Euler products*, Yale University Press, New Haven, Conn., 1971. A James K. Whittemore Lecture in Mathematics given at Yale University, 1967, Yale Mathematical Monographs, 1. MR0419366 (54 #7387)

[21] Lars Hörmander, *The spectral function of an elliptic operator*, Acta Math. 121 (1968), 193–218. MR0609014 (58 #29418)

[22] Erez Lapid and Werner Müller, *Spectral asymptotics for arithmetic quotients of $\text{SL}(n, \mathbb{R})/\text{SO}(n)$*, Duke Math. J. 149 (2009), no. 1, 117–155. MR2541128

[23] Erez M. Lapid, *A remark on Eisenstein series*, Lecture Notes in Mathematics, Vol. 544, Springer-Verlag, Berlin, 1976. MR0579181 (58 #28319)

[24] Erez Lapid and Werner Müller, *Spectral asymptotics for arithmetic quotients of $\text{SL}(n, \mathbb{R})/\text{SO}(n)$*, Duke Math. J. 149 (2009), no. 1, 117–155. MR2541128

[25] I. G. Macdonald, *Spherical functions on a group of $p$-adic type*, Ramanujan Institute, Centre for Advanced Study in Mathematics, University of Madras, Madras, 1971. Publications of the Ramanujan Institute, No. 2. MR0435301

[26] Peter Sarnak, *On cusp forms. II*, Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part II (Ramat Aviv, 1989), 1990, pp. 237–250. MR1159118 (93e:11068)

[27] Werner Müller, *The trace class conjecture in the theory of automorphic forms*, Ann. of Math. (2) 130 (1989), no. 3, 473–529. MR1025165

[28] Peter Sarnak, Sug Woo Shin, and Nicolas Templier, *Families of $L$-functions and their symmetry*, Families of automorphic forms and the trace formula, 2016, pp. 531–578. MR3675175

[29] Atle Selberg, *Collected papers. Vol. I*, Springer-Verlag, Berlin, 1989. With a foreword by K. Chandrasekharan. MR1117906 (92h:01083)
[38] , Remarks on the distribution of poles of Eisenstein series, Festschrift in honor of I. I. Piatetski-
Shapiro on the occasion of his sixtieth birthday, Part II (Ramat Aviv, 1989), 1990, pp. 251–278.
MR1159119 (93c:11035)

[39] Sug Woo Shin and Nicolas Templier, Sato-Tate theorem for families and low-lying zeros of automorphic
L-functions, Invent. Math. 203 (2016), no. 1, 1–177. Appendix A by Robert Kottwitz, and Appendix
B by Raf Cluckers, Julia Gordon and Immanuel Halupczok. MR3437869

[40] N. R. Wallach, On the constant term of a square integrable automorphic form, Operator algebras and
group representations, Vol. II (Neptun, 1980), 1984, pp. 227–237. MR733320 (86i:22029)

Universität Leipzig, Mathematisches Institut, PF 10 09 20, D-04009 Leipzig, Germany

E-mail address: finis@math.uni-leipzig.de

Department of Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel

E-mail address: erez.m.lapid@gmail.com