RESEARCH ARTICLE

A density theorem for $\text{Sp}(4)$

Siu Hang Man

Mathematisches Institut, Universität Bonn, Bonn, Germany

Correspondence
Siu Hang Man, Mathematisches Institut, Universität Bonn, Endenicher Allee 60, D-53115 Bonn, Germany.
Email: shman@math.uni-bonn.de

Funding information
DAAD; German Research Foundation, Grant/Award Number: EXC-2047/1 - 390685813

Abstract
Strong bounds are obtained for the number of automorphic forms for the group $\Gamma_0(q) \subseteq \text{Sp}(4, \mathbb{Z})$ violating the Ramanujan conjecture at any given unramified place, which go beyond Sarnak’s density hypothesis. The proof is based on a relative trace formula of Kuznetsov type, and explicit evaluation for certain Kloosterman sums for $\text{Sp}(4)$.

MSC (2020)
11F72, 11L05 (primary)

1 INTRODUCTION

In number theory, we often study families of objects, and this is especially the case in the context of automorphic forms, and this methodology has been described formally in [22, 23]. On the one hand, this can smooth out irregularities of individual members (which may exist, or whose proof of non-existence is beyond our reach) and allows the use of statistical concepts and deformation techniques to investigate properties within a given family. On the other hand, strong analytic tools, such as various types of trace formulae, are available, to give non-trivial statements to the families as a whole.

One of the key conjectures in the theory of automorphic forms is the Ramanujan conjecture, which states that cuspidal automorphic representations of the group $\text{GL}(n)$ over a number field $F$ are tempered. While the conjecture is still far out of reach even for the $\text{GL}(2)$ case, one can think of generalisations to automorphic forms of other reductive groups, such as $\text{Sp}(2n)$. It is well known that the naive generalisation of the Ramanujan conjecture is false for the group $\text{Sp}(4)$, because of the presence of Saito–Kurokawa lifts, which are not tempered. This is not the end of the investigation, however, because it is still an open question whether generic cuspidal automorphic representations of $\text{Sp}(4)$ are tempered. While this is still far out of reach, we can consider approximations to the conjecture as a substitute, and try to bound the number of members in a family violating the conjecture relative to the amount by which they violate the conjecture. This
gives a density result, which is analogous to the ones given in [2] for the GL(n) case. This does not prove the conjecture, but such density results often suffice in applications.

In this paper we consider the family of generic cuspidal automorphic representations for the group $\Gamma_0(q) \subseteq \text{Sp}(4, \mathbb{Z})$ of matrices whose lower left $2 \times 2$ block is divisible by $q$. Fix a place $v$ of $\mathbb{Q}$. For an automorphic representation $\pi = \bigotimes_v \pi_v$, we denote by $\mu_\pi(v) = (\mu_\pi(v, 1), \mu_\pi(v, 2))$ its local Langlands spectral parameter (see (2.2)), each entry viewed modulo $\frac{2\pi i}{\log p} \mathbb{Z}$ if $v = p$ is a prime. We write

$$\sigma_\pi(v) = \max \left\{ \left| \Re \mu_\pi(v, 1) \right|, \left| \Re \mu_\pi(v, 2) \right| \right\}. \quad (1.1)$$

The representation $\pi$ is tempered at $v$ if $\sigma_\pi(v) = 0$, and the size of $\sigma_\pi(v)$ gives a measure on how far $\pi$ is from being tempered at $v$. An example of a non-tempered representation is the trivial representation, which satisfies $\sigma_{\text{triv}}(v) = \frac{3}{2}$ for all places $v$.

For a finite family $\mathcal{F}$ of automorphic representations of $\text{Sp}(4)$ and $\sigma \geq 0$ we define

$$N_v(\sigma, \mathcal{F}) = |\{ \pi \in \mathcal{F} \mid \sigma_\pi(v) \geq \sigma \}|.$$

Trivially, we have $N_v(0, \mathcal{F}) = |\mathcal{F}|$, and if $\mathcal{F}$ contains the trivial representation, then we have $N_v(3/2, \mathcal{F}) \geq 1$. One may hope to interpolate linearly between the two extreme cases, and obtain a bound of the form

$$N_v(\sigma, \mathcal{F}) \ll_v |\mathcal{F}|^{1 - \frac{\sigma}{a} + \varepsilon} \quad (1.2)$$

with $a = 3/2$. In the context of groups $G$ of real rank 1, for the principal congruence subgroup $\Gamma(q) = \{ \gamma \in G(\mathbb{Z}) \mid \gamma = \text{id} \pmod{q} \}$ and $v = \infty$, this is known as Sarnak’s density hypothesis [21, p. 465].

Density theorems have attracted much attention in the history, and many strong density results are known for various automorphic families on GL(2) with different settings [4–6, 12, 14, 20]. Via Kuznetsov-type trace formulae on GL(3), strong density results on GL(3) were obtained in [1, 3, 4]. Blomer [2] further generalised the technique to obtain results in GL(n) beyond Sarnak’s density hypothesis. However, relatively little is known for general reductive groups of rank $r$. Finis–Matz [9] gives as by-products some density results for the family of Maass forms of Laplace eigenvalue up to a height $T$ and fixed level. However, the value of $a$ is large, and is at least quadratic in $r$, so that even the equivalent of Sarnak’s density hypothesis cannot be obtained.

More concretely, in this paper we consider the family $\mathcal{F}_I(q)$ of generic cuspidal automorphic representations for the group

$$\Gamma_0(q) = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \text{Sp}(4, \mathbb{Z}) \bigg| C \equiv 0 \pmod{q} \right\} \subseteq \text{Sp}(4, \mathbb{Z})$$

for a large prime $q$, and Laplace eigenvalue $\lambda$ in a fixed interval $I$. A simple application of Weyl’s law shows that $|\mathcal{F}_I(q)| \asymp q^3$ when the size of $I$ is sufficiently large, noting that the contribution from the continuous spectrum has size $O(q)$. For this family and any place $v \neq q$ of $\mathbb{Q}$, we go beyond the density hypothesis and obtain $a = 3/4$, which is halfway between the density hypothesis and the Ramanujan conjecture.
Theorem 1.1. Let $q$ be a prime, and $v$ a place of $\mathbb{Q}$ different from $q$, $I \subseteq [0, \infty)$ a fixed interval, $\varepsilon > 0$, and $\sigma \geq 0$. Then

$$N_v(\sigma, F_I(q)) \ll_{I, v, n, \varepsilon} q^{3-4\sigma+\varepsilon}.$$

The proof is based on a careful analysis on the arithmetic side of the Kuznetsov formula, and on the spectral side through a relation of Fourier coefficients of automorphic forms and Hecke eigenvalues. Let $\lambda(m, \pi)$ be the Hecke eigenvalue of $\pi \in F_I(q)$ for the $m$th standard Hecke operator $T(m)$. It is convenient to adopt the normalisation $\lambda'(m, \pi) := m^{-3/2}\lambda(m, \pi)$.

Theorem 1.2. Keep the notations as above. Let $m \in \mathbb{N}$ be coprime to $q$ and $Z \geq 1$. Then

$$\sum_{\pi \in F_I(q)} |\lambda'(m, \pi)|^2 Z^{2\sigma_{\pi}(\infty)} \ll_{I, \varepsilon} q^{3+\varepsilon}$$

uniformly in $mZ \ll q^2$ for a sufficiently small implied constant depending on $I$.

Let us roughly sketch the proof of Theorem 1.2. We denote by $\{\pi\}$ the isomorphism classes of generic automorphic representations for $\Gamma_0(q)$, and we denote by $\int_{(q)} d\pi$ the integral over the generic automorphic representations with respect to the automorphic Plancherel measure. Very roughly, the Kuznetsov formula takes the form

$$\int_{(q)} |A_{\pi}(M)|^2 Z^{2\sigma_{\pi}(\infty)} \delta_{\pi \in I} d\pi \approx 1 + \sum_{id \neq w \in W} \sum_{c_1, c_2} \frac{KL_{q,w}(c, M, M)}{c_1c_2},$$

where $M = (1, m) \in \mathbb{Z}^2$, $A_{\pi}(M)$ is the $M$th Fourier coefficient of $\pi$, defined in (5.1), $W$ is the Weyl group of Sp(4), and $KL_{q,w}(c, M, M)$ is a generalised Kloosterman sum, defined in (4.2) below, associated with the Weyl element $w$, and moduli $c = (c_1, c_2)$. Note that the Kuznetsov formula only extracts the generic spectrum.

However, the situation here is very different from GL(n) case found in [2]. In the symplectic case, there are no simple relations between the Fourier coefficients $A_{\pi}(M)$ of a cuspidal newform $\pi$ and Hecke eigenvalues $\lambda'(m, \pi)$ of the corresponding automorphic representation (i.e. $\pi \in V_{\pi}$). This is in stark contrast with the GL(n) case, where the Fourier coefficients and Hecke eigenvalues are proportional [11, Theorem 9.3.11]. It is because of this obstacle that the Kuznetsov formula is not yet a standard tool for the group GSp(4), and the present paper seems to be the first application of the Kuznetsov formula that is seen in action for a group other than GL(n).

While the Fourier coefficients in principle contain the information on Hecke eigenvalues, it is not obvious how to extract it. A detailed analysis of the relations between them is found in Section 6. In Theorem 6.2 we establish a recursive formula of $\lambda(p^r, \pi)$ in terms of Fourier coefficients. We also outline an algorithm for computing arbitrary Fourier coefficients of a cuspidal form in terms of its Hecke eigenvalues in the Appendix. While this is not needed for the proof of the theorems, such results serve an independent interest in number theory, in laying the groundwork for further applications of the Kuznetsov formula on Sp(4), as well as Fourier analysis of automorphic forms on Sp(4) in general.

Using Theorem 6.2, we deduce from Lemma 6.4 that for a prime $p \nmid q$ and $r \in \mathbb{N}$, the size of Fourier coefficients $A_{\pi}(1, p^r)$ of an $L^2$-normalised generic cuspidal form $\pi$ is often as big as
$q^{-3/2-\varepsilon} p^{r\sigma_{\star}(p)}$. Through this relation, we are able to use the Kuznetsov formula to derive information on $\sigma_{\pi}(p)$ from an analysis of the Kloosterman sums. Meanwhile, the factor $Z^{2\sigma_{\pi}(\infty)}$ deals with the infinite place, so the test function $|A_{\alpha}(M)| Z^{2\sigma_{\pi}(\infty)}$ treats the finite places and the infinite place essentially on the same footing.

When $mZ \ll q$, the Kloosterman sums associated to non-trivial Weyl elements are empty; hence, the off-diagonal terms vanish completely. We will use this observation to prove Theorem 1.3 below. To obtain stronger density results, we have to deal with the Kloosterman sums appearing in the off-diagonal term, and improve the trivial bound $|S_{q,\nu}(c, M, N)| \leq c_1c_2$. Obtaining such bounds for general groups remains a major open problem.

Finally, we give an application of Theorem 1.2, for a large sieve inequality. This is analogous to large sieve inequalities for $GL(n)$, which are known in the level aspect [2] and in the analytic conductor aspect [16].

**Theorem 1.3.** Let $q$ be prime and $\{\alpha(m)\}_{m \in \mathbb{N}}$ any sequence of complex numbers. Then

$$\sum_{\pi \in \mathcal{F}_1(q)} \left| \sum_{m \leq x} \alpha(m) \lambda'(m, \pi) \right|^2 \ll_{f, \varepsilon} q^{3+\varepsilon} \sum_{m \leq x} |\alpha(m)|^2$$

uniformly in $x \ll q$ for a sufficiently small implied constant depending on $I$.

And as a corollary, we establish a bound for the second moment of spinor L-functions on the critical line. Precisely, let $L(s, \pi)$ be the spinor L-function associated to $\pi$, normalised such that its critical strip is $0 < \Re s < 1$.

**Corollary 1.4.** For $q$ prime and $t \in \mathbb{R}$, we have

$$\sum_{\pi \in \mathcal{F}_1(q)} |L(1/2 + it, \pi)|^2 \ll_{f, \varepsilon} q^{3+\varepsilon}.$$

### 2 PRELIMINARIES

Let $U \subseteq Sp(4)$ be the standard unipotent subgroup

$$U = \left\{ \begin{pmatrix} 1 & x_{12} & x_{13} & x_{14} \\ 1 & 1 & x_{23} & x_{24} \\ x_{43} & 1 & 1 & \end{pmatrix} \Bigg| x_{ij} \in \mathbb{G}_a, \right. \begin{array}{l} x_{12} = -x_{43}, \\ x_{14} = x_{23} + x_{12}x_{24} \end{array} \right\}.$$

Let $V \subseteq Sp(4)$ be the group of diagonal matrices with entries $\pm 1$. Let $W$ be the Weyl group of $Sp(4)$, which is generated by the matrices

$$s_{\alpha} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad s_{\beta} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$
as the elements in $\text{Sp}(4)/V$. We denote the long element of the Weyl group by $w_0 := s_\alpha s_\beta s_\alpha s_\beta$. For $w \in W$, we define $U_w = w^{-1} U^T w \cap U$. For $N \in \mathbb{R}^2$ we define a character $\vartheta_N : U(\mathbb{R}) \to S^1$ by

$$\vartheta_N(x) = e(\mathbb{R}_+ x_1^2 + N_2 x_2^2),$$

where $e(x) = e^{2\pi i x}$ is the standard additive character. Note that if $N \in \mathbb{Z}^2$, this defines a character $U(\mathbb{Z}) \to S^1$. If $N = (1, 1)$, we drop it from the notation of the character.

Let $T \subseteq \text{Sp}(4)$ be the diagonal torus. The standard minimal parabolic subgroup is given by $P_0 = TU$. We embed $y = (y_1, y_2) \in \mathbb{R}_+^2$ into $T(\mathbb{R})$ via the map $\iota(y) = (y_1 y_2^{1/2}, y_2^{1/2}, 1/y_1^{1/2} y_2^{1/2}, 1/y_1^{1/2} y_2^{1/2})$. We denote the image of $\mathbb{R}_+^2$ in $T(\mathbb{R})$ by $T(\mathbb{R}_+)$. An element $g \in \text{Sp}(4, \mathbb{R})$ admits Iwasawa decomposition $g = x \tilde{y} k$, with $x = U(\mathbb{R})$, $\tilde{y} \in T(\mathbb{R}_+)$ and $k \in K$, where $K = \text{SO}(4, \mathbb{R}) \cap \text{Sp}(4, \mathbb{R})$ is the maximal compact subgroup of $\text{Sp}(4, \mathbb{R})$. We denote by $y(g) = \iota^{-1}(\tilde{y})$ the Iwasawa $y$-coordinates of $g$. Any $\tilde{y} \in T(\mathbb{R}_+)$ can be uniquely identified with $y := y(\tilde{y}) \in \mathbb{R}_+^2$, so we abuse notation and denote $\tilde{y} \in T(\mathbb{R}_+)$ also by $y$. For $w \in W$, $y \in \mathbb{R}_+^2$ we write $w y = y(\iota^{-1}(w) w^{-1})$.

For $\alpha \in \mathbb{C}_2$, $y \in \mathbb{R}_+^2$, we write $y^\alpha = y_1^{\alpha_1} y_2^{\alpha_2}$. Let $\eta = (2, 3/2)$. We define measures

$$dx = dx_1 dx_2 dx_3 dx_4, \quad d^* y = y^{-2\eta} \frac{dy_1}{y_1} \frac{dy_2}{y_2}$$

on $U(\mathbb{R})$ and $\mathbb{R}_+^2$, respectively. We denote the pushforward of $d^* y$ to $T(\mathbb{R}_+)$ by $\iota$ also by $d^* y$. Then $dx$ is the Haar measure on $U(\mathbb{R})$, and $dx d^* y$ is a left $\text{Sp}(4, \mathbb{R})$-invariant measure on $\text{Sp}(4, \mathbb{R})/K$.

We define another embedding of $\mathbb{R}_+^2$ into $T(\mathbb{R}_+)$ by

$$c = (c_1, c_2) \mapsto c^* = \text{diag}(1/c_1, c_1/c_2, c_1, c_2/c_1).$$

A simple calculation shows that $y(c^*) = (c_1 c_2)^{-1}$.

Let $\pi = \otimes \pi_v$ be a globally generic irreducible spherical automorphic representation of $\text{GSp}(4)$ with trivial central character. Using notations in [19], $\pi_v$ is induced from the character $\chi_1 \times \chi_2 \rtimes \sigma$, given by

$$\text{diag}(t_1, t_2, t_1^{-1} v, t_2^{-1} v) \mapsto \chi_1(t_1) \chi_2(t_2) \sigma(v).$$

As $\pi_v$ is right $K_v$-invariant, we may assume that $\chi_1, \chi_2, \sigma$ are unramified, and we may write $\chi_1 = | \cdot |^{\alpha_1}, \chi_2 = | \cdot |^{\alpha_2}$ and $\sigma = | \cdot |^{\beta}$. As $\pi_v$ has trivial central character, we have $\alpha_1 + \alpha_2 + 2\beta = 0$. So the $L$-parameter is given by

$$(\chi_1 \chi_2 \sigma, \chi_1 \sigma, \chi_2 \sigma, \sigma) = \left( \begin{array}{c} \alpha_1 + \alpha_2, \alpha_1 - \alpha_2, -\alpha_1 + \alpha_2, -\alpha_1 - \alpha_2 \end{array} \right).$$

We then take

$$\mu_\pi(v) = \left( \begin{array}{c} \alpha_1 + \alpha_2, \alpha_1 - \alpha_2 \end{array} \right),$$

so the $L$-parameter becomes $(\mu_\pi(v, 1), \mu_\pi(v, 2), -\mu_\pi(v, 1), -\mu_\pi(v, 2))$. When $\pi_v$ is lifted to a self-dual representation of $\text{GL}(4)$, this is precisely the natural Langlands parameter of the lift.
\section{Auxiliary Results}

Since the Iwasawa decomposition $\text{Sp}(4, \mathbb{R}) = U(\mathbb{R}) T(\mathbb{R}_+) K$ is actually the Gram–Schmidt orthogonalisation of rows, we can compute $y(g)$ explicitly. Let $\Delta_1$ be the norm of the third row of $g$, and $\Delta_2$ be the area of the parallelogram spanned by the bottom two rows of $g$. Then we have

\[
  g \equiv \begin{pmatrix}
  1/\Delta_1 & * & * & *
  \\
  \Delta_1/\Delta_2 & * & * & *
  \\
  \Delta_1 & * & \Delta_2/\Delta_1 & *
\end{pmatrix} \pmod{K}.
\]

In particular, we have $y(g) = (\Delta_2/\Delta_1^2, \Delta_1^2/\Delta_2^2)$. Conversely, if $y(g) = (Y_1, Y_2)$, then $\Delta_1(g) = Y_1^{-1} Y_2^{-1/2}$ and $\Delta_2(g) = Y_1^{-1} Y_2^{-1}$.

**Lemma 3.1.** Let $w \in W, x \in U_w(\mathbb{R})$, and $y, c, B \in \mathbb{R}_+^2$. Write $y(\iota(B)c^* wxi(y)) = Y \in \mathbb{R}_+^2$ and $A = \iota(B)c^*$. Let $w, y, Y$ be fixed, and let $x, c, B$ vary. Then we have

\[
  c_1 \ll_{y,Y} B_1 B_2^{1/2}, \quad c_2 \ll_{y,Y} B_1 B_2,
\]

and

\[
  1 \leq \Delta_1(wx) \ll_{y,Y} y(A)_1 y(A)_2^{1/2}, \quad 1 \leq \Delta_2(wx) \ll_{y,Y} y(A)_1 y(A)_2.
\]

**Proof.** Note that $\Delta_i(wx) \geq 1$ as one of its minors is always 1. For the first statement, we compute

\[
  \Delta_1(\iota(B)c^* wxi(y)) = \frac{c_1}{B_1 B_2^{1/2}} \Delta_1(wxi(y)), \quad \Delta_2(\iota(B)c^* wxi(y)) = \frac{c_2}{B_1 B_2} \Delta_2(wxi(y)).
\]

Then we obtain

\[
  c_1 \leq c_1 \Delta_1(wx) \ll_{y} c_1 \Delta_1(wxi(y)) = \Delta_1(\iota(B)c^* wxi(y)) B_1 B_2^{1/2} \ll_{y} B_1 B_2^{1/2},
\]

\[
  c_2 \leq c_2 \Delta_2(wx) \ll_{y} c_2 \Delta_2(wxi(y)) = \Delta_2(\iota(B)c^* wxi(y)) B_1 B_2 \ll_{y} B_1 B_2.
\]

For the second statement, we observe that

\[
  c_1^{-1} = y(c^*)_1 y(c^*)_2^{1/2}, \quad c_2^{-1} = y(c^*)_1 y(c^*)_2.
\]

Hence

\[
  \Delta_1(wx) \ll_{y,Y} B_1 B_2^{1/2} c_1^{-1} = (B_1 y(c^*)_1)(B_2 y(c^*)_2)^{1/2} = y(A)_1 y(A)_2^{1/2},
\]

\[
  \Delta_2(wx) \ll_{y,Y} B_1 B_2 c_2^{-1} = (B_1 y(c^*)_1)(B_2 y(c^*)_2) = y(A)_1 y(A)_2,
\]

finishing the proof. \qed
Lemma 3.2. Let $N \in \mathbb{N}^2$ and $w \in W$. For $x \in U_w(\mathbb{R})$, define $x' = \iota(N)x\iota(N)^{-1}$. Then

$$dx' = (wN)^n N^n dx.$$

Proof. This is just direct computation. For $w = w_0$, we have

$$x = \begin{pmatrix} 1 & x_{12} & x_{13} & x_{14} \\ 1 & x_{23} & x_{24} & 1 \\ 1 & -x_{12} & 1 \end{pmatrix} \leadsto x' = \begin{pmatrix} 1 & N_1x_{12} & N_1^2N_2x_{13} & N_1N_2x_{14} \\ 1 & N_1N_2x_{23} & N_2x_{24} & 1 \\ -N_1x_{12} & 1 \end{pmatrix}.$$ 

It follows that $dx' = N_1^4N_2^3 dx = (w_0N)^n N^n dx$. The proof is similar for other Weyl elements. \qed

Lemma 3.3. Let $B \in \mathbb{R}^2_{+}$, and $w = s_\rho s_\alpha s_\beta$. Then

$$\text{vol} \left\{ x \in U_w(\mathbb{R}) \mid \Delta_j(wx) \leq B_j, j = 1, 2 \right\} \ll (B_1B_2)^{1+\varepsilon}$$

for any $\varepsilon > 0$.

Proof. We can assume without loss of generality that $B_j \geq 1$; otherwise, we deduce from Lemma 3.1 that the volume is 0. We have

$$x = \begin{pmatrix} 1 & x_{13} & x_{23} \\ 1 & x_{23} & x_{24} \\ 1 & -x_{12} & 1 \end{pmatrix} \in U_w(\mathbb{R}) \leadsto wx = \begin{pmatrix} 1 & -1 \\ 1 & x_{23} & x_{24} \\ -1 & -x_{13} & -x_{23} \end{pmatrix}.$$ 

Then we obtain bounds

$$|x_{23}| \leq B_1, |x_{13}x_{24} - x_{23}^2| \leq B_2.$$ 

We also have $|x_{13}|, |x_{24}| \leq b := 1 + \max\{B_1, B_2\}$. If $I \subseteq \mathbb{R}$ is any interval of length $|I| \geq 1$, then

$$\text{vol} \left\{ (x, y) \in [-b, b]^2 \mid xy \in I \right\} \leq \int_{|y| \leq b} \min \left\{ \frac{|I|}{|y|}, 2b \right\} dy \leq 4|I|(1 + \log b).$$

Hence, if $|x_{23}| \leq B_1$ is fixed, then

$$\text{vol} \left\{ (x_{13}, x_{24}) \in [-b, b]^2 \mid x_{13}x_{24} - x_{23}^2 \leq B_2 \right\} = O(B_2 \log b).$$

This establishes the bound. \qed
Properties of Kloosterman sums for $\text{Sp}(4, \mathbb{Z})$ were given in [18]. They generalise in a natural way to the congruence subgroup $\Gamma_0(q)$. The Bruhat decomposition gives $\text{Sp}(4, \mathbb{Q}) = \bigcup_{w \in W} G_w(\mathbb{Q})$ with $G_w = U_w T U_w$ as a disjoint union. Let $M, N, c \in \mathbb{N}^2$, $w \in W$. Then if

$$\vartheta_M(\omega c^* x(c^*)^{-1} w^{-1}) = \vartheta_N(x)$$  \hfill (4.1)

for all $x \in U \cap w^{-1} U w$, the Kloosterman sum

$$\text{Kl}_{q,w}(c, M, N) := \sum_{x w c^* x' \in U(\mathbb{Z}) \setminus G_w(\mathbb{Q}) \cap \Gamma_0(q) / U_w(\mathbb{Z})} \vartheta_M(x) \vartheta_N(x')$$  \hfill (4.2)

is well defined [18, Proposition 5.1]. If (4.1) does not hold, we set $\text{Kl}_{q,w}(c, M, N) = 0$. The Kloosterman sum $\text{Kl}_{q,w}(c, M, N)$ is non-zero only if $w = id, s_\alpha s_\beta s_\alpha, s_\beta s_\alpha s_\beta, w_0$ [18].

Now suppose that the entries of $M = (M_1, M_2)$ and $N = (N_1, N_2)$ are coprime to $q$. Considering the Bruhat decomposition of $\Gamma_0(q)$, we deduce that the Kloosterman sum $\text{Kl}_{q,w}(c, M, N)$ is non-empty only if

$$q | c_1 \text{ for } w = s_\alpha s_\beta s_\alpha, \quad q | c_1 \text{ and } q^2 | c_2 \text{ for } w = s_\beta s_\alpha s_\beta, w_0.$$  \hfill (4.3)

Meanwhile, the well-definedness condition (4.1) says that the Kloosterman sums are well defined precisely if

$$N_2 = M_2 \frac{c_2^2}{c_1} \text{ if } w = s_\alpha s_\beta s_\alpha, \quad N_1 = M_1 \frac{c_2}{c_1} \text{ if } w = s_\beta s_\alpha s_\beta.$$

Hence the Kloosterman sums are well defined only if

$$v_q(c_1) = v_q(c_2) \text{ if } w = s_\alpha s_\beta s_\alpha, \quad v_q(c_2) = 2v_q(c_1) \text{ if } w = s_\beta s_\alpha s_\beta.$$  \hfill (4.4)

From the abstract definition [18, 26], the Kloosterman sums $\text{Kl}_{q,w}(c, M, N)$ also enjoy certain multiplicativity in the moduli. We state one particular case. Let $q$ be prime. For $c = (c_1, c_2) \in \mathbb{N}^2$, let $c_j' = q^{-v_q(c_j)} c_j$, $j = 1, 2$, and $c' = (c_1', c_2')$. Then we have

$$\text{Kl}_{q,w}(c, M, N) = \text{Kl}_{q,w} \left( (q^{v_q(c_1)}, q^{v_q(c_2)}), M', N' \right) \text{Kl}_{1,w}(c', M'', N'')$$  \hfill (4.5)

for some $M', N', M'', N'' \in \mathbb{N}^2$. Moreover, if the entries of $M, N$ are coprime to $q$, then so are $M', N'$. From [8], we have a trivial bound

$$\text{Kl}_{1,w}(c', M'', N'') \leq |U(\mathbb{Z}) \setminus G_w(\mathbb{Q}) \cap \text{Sp}(4, \mathbb{Z}) / U_w(\mathbb{Z})| \leq c_1' c_2'.$$  \hfill (4.6)
4.1 Evaluations of Kloosterman sums

For the proof of the theorems in Section 8, we compute the following Kloosterman sums:

\[
\begin{align*}
\text{Kl}_{q,s_{a}q_{s_{a}}}, \quad & \text{Kl}_{q,s_{b}q_{s_{b}}}, \\
\text{Kl}_{q,w_{0}}, \quad & \text{Kl}_{q,w_{0}}.
\end{align*}
\]

(i) Consider the Bruhat decomposition for summands in \(\text{Kl}_{q,s_{a}q_{s_{a}}}, ((q, q), M, N)\):

\[
\gamma = \begin{pmatrix}
1 & \beta_1 & \beta_2 & \beta_3 \\
1 & \beta_4 & \beta_5 \\
-\beta_1 & 1
\end{pmatrix}
\begin{pmatrix}
-q^{-1} & q^{-1} & 1 \\
1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
v_2/q & v_3/q & v_4/q \\
v_3/q & v_4/q & 1
\end{pmatrix}
\in \Gamma_0(q),
\]

with \(v_2, v_3, v_4 \pmod{q}\) chosen such that \((v_3, v_4, (q, v_2)) = 1\), and \(((q, v_2)^2, qv_3 + v_2v_4) = q\). As \(\gamma \in \Gamma_0(q)\), by considering the lower left block, we deduce that \(v_2 = 0\), and solve \(\beta_1 \equiv 0 \pmod{1}\). The conditions on \(v_3, v_4\) then simplify as \((q, v_3) = 1\). Considering the second row, we solve

\[
\beta_4 \equiv -\frac{v_3v_4}{q} \pmod{1}, \quad \beta_5 \equiv \frac{v_2^2}{q} \pmod{1}.
\]

So the Kloosterman sum is given by

\[
\text{Kl}_{q,s_{a}q_{s_{a}}}, ((q, q), M, N) = \sum_{v_3 \pmod{q}} \sum_{v_4 \pmod{q}} e\left(\frac{M_2v_3v_4^2}{q}\right) = 0.
\]

(ii) Consider the Bruhat decomposition for summands in \(\text{Kl}_{q,s_{b}q_{s_{b}}}, ((q, q^2), M, N)\):

\[
\gamma = \begin{pmatrix}
1 & \beta_1 & \beta_2 & \beta_3 \\
1 & \beta_4 & \beta_5 \\
-\beta_1 & 1
\end{pmatrix}
\begin{pmatrix}
q^{-1} & q^{-1} & 1 \\
1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
-v_{23}/q^2 & v_{13}/q^2 \\
v_{13}/q^2 & v_{14}/q^2 & 1
\end{pmatrix}
\in \Gamma_0(q),
\]

\[
\gamma = \begin{pmatrix}
-\beta_3q & \beta_2q & \beta_2v_{13}/q + \beta_1/q + \beta_3v_{23}/q & \beta_2v_{14}/q - \beta_3v_{13}/q - 1/q \\
-\beta_5q & \beta_4q & \beta_4v_{13}/q + \beta_5v_{23}/q + 1/q & \beta_4v_{14}/q - \beta_5v_{13}/q \\
0 & q & v_{13}/q & v_{14}/q \\
-q & -\beta_1q & -\beta_1v_{13}/q + v_{23}/q & -\beta_1v_{14}/q - v_{13}/q
\end{pmatrix}
\in \Gamma_0(q),
\]
with \( v_{13}, v_{14}, v_{23} \) (mod \( q^2 \)) chosen such that \( (q^2, v_{13}, v_{14}) = q \), and \( (q, v_{23}, v_{34}) = 1 \), where \( v_{34} = -\frac{v_{13}^2 + v_{14}v_{23}}{q^2} \). As \( \gamma \in \Gamma_0(q) \), by considering the lower left block, we solve \( \beta_1 \equiv 0 \) (mod 1). Then, \(-\beta_1 v_{13}/q + v_{23}/q\) being an integer implies \( q \mid v_{23} \), so \( (q, v_{34}) = 1 \). Write \( v_{13} = q v'_{13} \), \( v_{14} = q v'_{14} \), and \( \beta_4 = \beta'_4/q \), \( \beta_5 = \beta'_5/q \) for some \( \beta'_4, \beta'_5 \in \mathbb{Z} \). By considering the second row, we deduce that
\[
\beta'_4 v'_{13} + \beta'_5 v_{23}/q + 1, \beta'_4 v'_{14} - \beta'_5 v'_{13} \in q \mathbb{Z},
\]
from which we deduce \( \beta'_5 \equiv v'_{14} v_{34} \) (mod \( q \)), and \( \beta_5 \equiv \frac{v'_{14} v_{34}}{q} \) (mod 1). Writing \( v_{23} = q v'_{23} \), the Kloosterman sum is given by
\[
\text{Kl}_{q, s} \left( (q, q^2), M, N \right) = \sum_{v'_{13}, v'_{14}, v'_{23} \pmod{q}} e \left( \frac{M v'_{14} v_{34} + N v'_{14}}{q} \right),
\]
where \( v_{34} = -(v'_{13}^2 + v'_{14} v'_{23}) \). We evaluate
\[
\sum_{v'_{13}, v'_{23} \pmod{q}} 1 + \sum_{v'_{13}, v'_{14}, v'_{23} \pmod{q}} e \left( \frac{M v'_{14} v_{34} + N v'_{14}}{q} \right)
\]
\[
= q(q - 1) - \sum_{v'_{13}, v'_{14} \pmod{q}} e \left( \frac{N v'_{14}}{q} \right) = q^2.
\]

(iii) Consider the Bruhat decomposition for summands in \( \text{Kl}_{q, w_0} \left( (q, q^2), M, N \right) \):
\[
\gamma = \begin{pmatrix}
1 & \beta_1 & \beta_2 & \beta_3 \\
1 & \beta_4 & \beta_5 & 1 \\
-\beta_1 & 1 & 1 & \qquad q^{-1} & -q^{-1} & q & q^{-1} & 1 & 1 & -v_2/q & 1 & 1 & -v_2/q & 1 & -v_2/q & 1 & \v_3/q & \v_4/q & \v_13/q^2 & \v_14/q^2 \\
\beta_2 q & \beta_2 v_2 + \beta_3 q & \beta_2 v_3 + \beta_3 v_{13}/q + \beta_1 v_2/q^2 - 1/q & \beta_2 v_4 - \beta_1/q + \beta_3 v_{14}/q & q \\
\beta_4 q & \beta_4 v_2 + \beta_5 q & \beta_4 v_3 + \beta_5 v_{13}/q + \beta_2 v_2/q^2 & \beta_4 v_4 + \beta_5 v_{14}/q - 1/q & -\beta_1 q & -\beta_1 v_2 + q & -\beta_1 v_2 + \v_13/q & -\beta_1 v_4 + \v_14/q & 1
\end{pmatrix} \in \Gamma_0(q),
\]
with \( v_2, v_3, v_4 \) (mod \( q \)), \( v_{13}, v_{14} \) (mod \( q^2 \)) chosen such that \( v_{13} q + v_2 v_{14} - v_4 q^2 = 0 \), \( (q, v_2, v_3, v_4) = 1 \), and \( (q^2, v_{13}, v_{14}, v_{23}, v_{34}) = 1 \), where \( v_{23} = \frac{v_2 v_{13} - v_4 v_{14}}{q} \) and \( v_{34} = \frac{v_3 v_{14} - v_4 v_{13}}{q} \). As \( \gamma \in \Gamma_0(q) \), by considering the lower left block, we deduce that \( v_2 = 0 \), and solve \( \beta_1 \equiv 0 \) (mod 1). The last row being integers implies that \( q \mid v_{13}, v_{14} \). Write \( v_{13} = q v'_{13}, v_{14} = q v'_{14} \).
The relation \( v_{13}q + v_2v_{14} - v_4q^2 = 0 \) says \( v'_{13} = v_4 \). We check that \( q \mid v_{23} \) as well, so \((q, v_{34}) = 1\). Write \( \beta_4 = \beta'_4/q, \beta_5 = \beta'_5/q \) for some \( \beta'_4, \beta'_5 \in \mathbb{Z} \). By considering the second row, we deduce that

\[
\beta'_4 v_3 + \beta'_5 v'_{13}, \beta'_4 v_4 + \beta'_5 v'_{14} - 1 \in q \mathbb{Z},
\]

from which we deduce \( \beta'_5 \equiv v_3 v_{34} \pmod{q} \), and \( \beta'_5 = \frac{v_3 v_{34}}{q} \pmod{1} \). The Kloosterman sum is given by

\[
\text{Kl}_{q, w_0} ((q, q^2), M, N) = \sum_{v_3, v_4, v'_{14} \pmod{q}} e \left( \frac{M_2 v_3 u_{34} + N_2 v'_{14}}{q} \right),
\]

where \( u_{34} = v_3 v'_{14} - v_4^2 \).

Fix \( v_4, v'_{14} \neq 0 \). As \( v_3 \neq 0 \) varies, \( \overline{v_3} v_{34} = v'_{14} - v_4^2 \overline{v_3} \) runs through non-zero residues except \( v'_{14} \) modulo \( q \); hence, as \( v_3 \) varies, \( v_3 v_{34} \) runs through all residues except \( v'_{14} \) modulo \( q \). Hence

\[
\sum_{v_3, v_4, v'_{14} \pmod{q}} e \left( \frac{M_2 v_3 u_{34} + N_2 v'_{14}}{q} \right) = - \sum_{v_4, v'_{14} \pmod{q}} e \left( \frac{M_2 v'_{14} + N_2 v'_{14}}{q} \right)
\]

\[
= -(q - 1) S(M_2, N_2; q).
\]

If \( v_4 \neq 0 \) and \( v'_{14} = 0 \), then \( v_{34} = -v_4^2 \). The corresponding part of the sum becomes

\[
\sum_{v_3, v_4 \pmod{q}} e \left( \frac{-M_2 v_3 \overline{v_4}^2}{q} \right) = 0.
\]

Meanwhile, for \( v_4 = 0 \), we have \( v'_{14} \neq 0 \), and \( v_{34} = v_3 v'_{14} \), so \( \overline{v_3 v_{34}} = \overline{v'_{14}} \). Hence this part of the sum is

\[
\sum_{v_3, v'_{14} \pmod{q}} e \left( \frac{M_2 v'_{14} + N_2 v'_{14}}{q} \right) = (q - 1) S(M_2, N_2; q).
\]

Combining the parts above, we conclude that \( \text{Kl}_{q, w_0} ((q, q^2), M, N) = 0 \).
(iv) Consider the Bruhat decomposition for summands in $K_{l(q, w_0)}((q, q^3), M, N)$:

$$\gamma = \begin{pmatrix}
\beta_1 & \beta_2 & \beta_3 \\
1 & \beta_4 & \beta_5 \\
-\beta_1 & 1 & \beta_2
\end{pmatrix} \begin{pmatrix}
-q^{-1} \\
-q^{-2} \\
q^2
\end{pmatrix} \begin{pmatrix}
v_2/q & v_3/q & v_4/q \\
v_3/q & v_4/q & 1 \\
v_2/q & 1
\end{pmatrix}
$$

$$\in \Gamma_0(q),$$

with $v_2, v_3, v_4 \pmod{q}$, $v_{13}, v_{14} \pmod{q^2}$ chosen such that $v_{13}q + v_2v_{14} - v_4q^3 = 0$, $(q, v_2, v_3, v_4) = 1$, and $(q^2, v_{13}, v_{14}, v_{23}, v_{34}) = 1$, where $v_{23} = v_{2}
\frac{v_{13}q}{q^2}$ and $v_{34} = v_3\frac{v_{14}-v_4v_{13}}{q^3}$.

As $\gamma \in \Gamma_0(q)$, by considering the lower left block, we deduce that $v_2 = 0$, and solve $\beta_1 = 0 \pmod{1}$. The last row being integers implies that $q \nmid v_{13}, v_{14}$. Write $v_{13} = qv_3', v_{14} = qv_4'$. The relation $v_{13}q + v_2v_{14} - v_4q^3 = 0$ says $v_3' = qv_4'$. We check that $q \nmid v_{23}$ as well, so $(q, v_{34}) = 1$. Write $\beta_4 = \beta_4'/q$, $\beta_5 = \beta_5'/q^2$ for some $\beta_4', \beta_5' \in \mathbb{Z}$. By considering the second row, we deduce that

$$\beta_4'v_3'q + \beta_5'v_{13}', \beta_4'v_4'q + \beta_5'v_{14}' - 1 \in q^2\mathbb{Z},$$

from which we deduce $\beta_5' \equiv v_3'v_{34} \pmod{q^2}$, and $\beta_5 = \frac{v_3'v_{34}}{q^2} \pmod{1}$. The Kloosterman sum is given by

$$K_{l(q, w_0)}((q, q^3), M, N) = \sum_{v_3, v_4 \pmod{q}, (q, v_3) = 1, (q, v_4) = 1} e\left(\frac{M_2v_3v_{34} + N_2v_{14}'}{q^2}\right),$$

where $v_{34} = v_3'v_{14}' - v_4'^2q$.

Fix $v_4' \neq 0$. Then from $(q, v_{34}) = 1$ we have $(q, v_{14}') = 1$, and $v_3 \neq 0$. For a fixed $v_{14}'$, we see that as $v_3$ varies, $v_3v_{34} = v_{14}' - q^2v_3q$ runs through non-zero residues modulo $q^2$ that are congruent to $v_{14}' \pmod{q}$, except $v_{14}'$; hence, as $v_3$ varies, $v_3\bar{v}_{34}$ runs through all residues modulo $q^2$ that are congruent to $v_{14}' \pmod{q}$, except $\bar{v}_{14}'$. Hence

$$\sum_{v_3, v_{14}' \pmod{q}, (q, v_3) = 1, (q, v_{14}') = 1} e\left(\frac{M_2v_3v_{34} + N_2v_{14}'}{q^2}\right) = -\sum_{v_4' \pmod{q^2}, (q, v_{14}') = 1} e\left(\frac{M_2v_{14}' + N_2v_{14}'}{q^2}\right)$$

$$= -(q - 1)S(M_2, N_2; q^2).$$
Meanwhile, for \( v_4 = 0 \), we have \((q, v'_{14}) = 1\), and \( v_{34} = v_3 v'_{14}\), so \( v_3 \overline{v_{34}} = \overline{v'_{14}}\). Hence this part of the sum is

\[
\sum_{\substack{v_3 \pmod{q} \\ v'_{14} \pmod{q^2} \\ (q, v_3) = 1, (q, v'_{14}) = 1}} e\left(\frac{M_2 \overline{v'_{14}} + N_2 v'_{14}}{q^2}\right) = (q - 1)S(M_2, N_2; q).
\]

Combining the parts above, we conclude that \( K_{lq, w_0}((q, q^3), M, N) = 0 \).

# 5 Automorphic Forms and Whittaker Functions

We denote by \( \{\pi\} \) an orthonormal basis of right \( K \)-invariant automorphic forms for \( \Gamma_0(q) \), cuspidal or Eisenstein series. The space \( L^2(\Gamma_0(q) \backslash \text{Sp}(4, \mathbb{R})/K) \) is equipped with the standard inner product

\[
\langle f, g \rangle = \int_{\Gamma_0(q) \backslash \text{Sp}(4, \mathbb{R})/K} f(xy) \overline{g(xy)} dx \, d^*y.
\]

An integral over the complete spectrum of \( L^2(\Gamma_0(q) \backslash \text{Sp}(4, \mathbb{R})/K) \) is denoted by \( \int_{(q)} d\varpi \). All the automorphic forms \( \varpi \) belong to representations \( \pi \) of level \( q' \mid q \), and we assume that \( \{\varpi\} \) contains all cuspidal newvectors of level \( q' \mid q \). For simplicity in notations, we denote the local archimedean spectral parameter \( \mu_\pi(\infty) \) by \( \mu = (\mu_1, \mu_2) \).

Let \( \varpi \) be an automorphic form for \( \Gamma_0(q) \), with spectral parameter \( \mu \). We suppose that \( \varpi \) is generic throughout the section. For \( M = (M_1, M_2) \in \mathbb{Z}^2 \), the \( M \)th Fourier coefficient of \( \varpi \) is given by

\[
\varpi_M(g) = \int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} \varpi(xg) \overline{\varpi_M(x)} dx.
\]

The Fourier coefficients \( \varpi_M(g) \) are actually Whittaker functions. For \( g = x y k \in \text{Sp}(4, \mathbb{R}) \), we have

\[
\varpi_M(g) = \frac{A_{\varpi}(M)}{M^\eta} \varpi_M(x) \cdot W_\mu((M)y),
\]

where \( A_{\varpi}(M) \in \mathbb{C} \) is a constant, called the \( M \)th Fourier coefficient of \( \varpi \), \( W_\mu : \mathbb{R}^2_+ \to \mathbb{C} \) is the standard Whittaker function on \( \text{Sp}(4, \mathbb{R}) \), with detailed descriptions found in [13]. As in the \( GL(n) \) case, the size of \( \sigma_\pi(\infty) \) captures the growth of \( W_\mu \) near the origin. Precisely, for a function \( E \) on \( \mathbb{R}^2_+ \) and \( X \in \mathbb{R}^2_+ \), we define

\[
E^{(X)}(y_1, y_2) = E(X_1 y_1, X_2 y_2).
\]

For \( \beta \in \mathbb{C} \), let \( D_\beta = -y \partial_y + \beta \). This is a commutative family of differential operators, which correspond to multiplication by \( s + \beta \) under Mellin transform. In the proof of Lemma 5.2, we need the following technical lemma, found in [2].
Lemma 5.1. Let $\alpha \geq 0$, and $\beta \in \mathbb{C}$ such that $\text{Re} \beta \leq \alpha$. Let $I = [a, b] \subset (0, 1)$ be an interval with $a < b < 2a$, and $w : I \to \mathbb{C}$ a smooth function satisfying

$$|D_{\beta} w(y)| \geq c_1 y^{-\alpha}, \quad |\partial_y (D_{\beta} w)(y)| \leq c_2 \|D_{\beta} w\|^{-1} \tag{5.3}$$

for $y \in I$ and some $c_1, c_2 > 0$, where $\|w\|$ denotes the sup-norm of $w$. Then there exists constants $a', b', c_1', c_2' > 0$, depending only on $c_1, c_2, \alpha, \beta$ (not on $a, b$) such that $a \leq a' < b \leq b$, and

$$|w(y)| \geq c_1' y^{-\alpha}, \quad |w'(y)| \leq c_2' \|w\|^{-1} \tag{5.4}$$

for $y \in [a', b']$.

Proof. Let $\tilde{w}(y) = w(y) y^{-\beta}$. Then $\tilde{w}' = -y^{-\beta-1} D_{\beta} w$. We deduce that (5.3) implies

$$|y \tilde{w}'(y)| \geq c_1 y^{-\tilde{\alpha}}, \quad \|\tilde{w}''\| \leq \tilde{c}_2 \|\tilde{w}'\|^{-1}$$

for some constant $\tilde{c}_2 > c_2$, and $\tilde{\alpha} = \alpha + \text{Re} \beta$. Let $y_0 = \max_{y \in I} |\tilde{w}'(y)|$. Changing $\tilde{w}$ by a fourth root of unity if necessary, we may assume that

$$\text{Re} \tilde{w}'(y_0) \geq \frac{1}{\sqrt{2}} \max \{c_1 y_0^{-\alpha-1}, \|\tilde{w}'\|\}.$$

Meanwhile, the condition $|\tilde{w}''(y)| \leq c_2 \|\tilde{w}'\|^{-1}$ implies that the following inequality

$$\text{Re} \tilde{w}'(y) \geq \frac{1}{2\sqrt{2}} \max \{c_1 y_0^{-\tilde{\alpha}-1}, \|\tilde{w}'\|\} \geq \text{Re} \tilde{w}'(y_0)$$

holds for $y \in I_0 = [a', b']$, for some $a' < b'$ such that $y_0 \in I_0$. Now we show that $y|\tilde{w}'(y)| \ll |\tilde{w}||_{[a', b']}$ for $y \in [a', b']$. Let $c_3 > 0$ be a sufficiently small constant. We distinguish two cases:

(i) Suppose $\text{Re} \tilde{w}(a') \leq -c_3 y_0 \text{Re} \tilde{w}'(y_0)$. Then $|\tilde{w}||_{[a', b']} \geq c_3 y_0 \text{Re} \tilde{w}'(y_0) \gg y|\tilde{w}'(y)|$ for $y \in [a', b']$.

(ii) Suppose $\text{Re} \tilde{w}(a') > -c_3 y_0 \text{Re} \tilde{w}'(y_0)$. When $c_3 > 0$ is sufficiently small, we have $\text{Re} \tilde{w}(b') \gg y_0 \text{Re} \tilde{w}'(y_0)$, and hence $|\tilde{w}||_{[a', b']} \gg y \text{Re} \tilde{w}'(y_0) \gg y|\tilde{w}'(y)|$ for $y \in [a', b']$.

From the bound $y|\tilde{w}'(y)| \ll |\tilde{w}|$, it follows immediately that $|\tilde{w}(y)| \gg y^{-\tilde{\alpha}}$ on $[a', b']$. Reverting back to $w$ yields (5.4). \qed

Lemma 5.2. Assume that $\mu = (\mu_1, \mu_2)$ varies in some compact set $\Omega$, and let $Z \geq 1$. There exists $r \in \mathbb{N}$ and a compact set $S \subseteq \mathbb{R}^2_+$ depending only on $\Omega$ (independent of $Z$), and a finite collection of functions $E_1, \ldots, E_r : \mathbb{R}^2_+ \to \mathbb{R}$ depending on $\Omega$ and $Z$ that are uniformly bounded and supported in a compact subset of $S$ such that

$$\sum_{j=1}^r \left| \left\langle E_j^{(1,Z)}, W_{\mu} \right\rangle_{T(\mathbb{R}_+)} \right|^2 \gg \Omega Z^{2\eta_2 + 2\sigma_\mu(\infty)} = Z^{3 + 2\sigma_\mu(\infty)},$$

where $\langle \cdot, \cdot \rangle_{T(\mathbb{R}_+)}$ is the inner product with respect to the measure $d^* y$. 

Proof. For $Z \ll 1$, we follow the argument from [2, 3], which was stated for $\text{GL}(n)$, but also works for $\text{Sp}(4)$. For each $\mu \in \Omega$, choose an open set $S_\mu \subseteq \mathbb{R}^2_+$ such that $\Re W_\mu(y) \neq 0$ for all $y \in S_\mu$ or $3W_\mu(y) \neq 0$ for all $y \in S_\mu$. Now choose open neighbourhoods $U_\mu$ about $\mu$ such that $\Re W_\mu(y) \neq 0$ for all $y \in S_\mu$ and $\mu^* \in U_\mu$, or $3W_\mu(y) \neq 0$ for all $y \in S_\mu$ and $\mu^* \in U_\mu$. By compactness, $\Omega$ is covered by a finite collection of neighbourhoods $U_{\mu_1}, ..., U_{\mu_r}$, and we may pick corresponding $E_j$ to be real-valued functions with supports on $S_{\mu_j}$ and non-vanishing on the interior $S_{\mu_j}^o$.

Now suppose that $Z \gg 1$ is sufficiently large. Consider the following renormalisation of the Whittaker function:

$$W^*_\mu(y) := y^{-\eta}W_\mu(y).$$

The Mellin transform $M^*_\mu(s) = \int_{\mathbb{R}^2_+} W^*_\mu(y)y^s \frac{dy_1}{y_1} \frac{dy_2}{y_2}$ is given by [13] (where $\nu_1, \nu_2$ in [13] are $\mu_1 + \mu_2$ and $-\mu_1 + \mu_2$ in our notation)

$$M^*_\mu(s) = 2^{-4} \Gamma\left(\frac{s_1 + \mu_1 + \mu_2}{2}\right) \Gamma\left(\frac{s_1 + \mu_1 - \mu_2}{2}\right) \Gamma\left(\frac{s_1 - \mu_1 + \mu_2}{2}\right) \Gamma\left(\frac{s_1 - \mu_1 - \mu_2}{2}\right)$$

$$\Gamma\left(\frac{s_2 + \mu_1}{2}\right) \Gamma\left(\frac{s_2 - \mu_1}{2}\right) \Gamma\left(\frac{s_2 + \mu_2}{2}\right) \Gamma\left(\frac{s_2 - \mu_2}{2}\right)$$

$$\left\{ \Gamma\left(\frac{s_1 + s_2 + \mu_2}{2}\right) \Gamma\left(\frac{s_1 + s_2 - \mu_2}{2}\right) \right\}^{-1} \mathbf{F}_2\left(\frac{s_1}{s_1 + s_2 + \mu_2}, \frac{s_2 - \mu_1}{s_1 + s_2 - \mu_2}; \frac{s_1}{s_1 + s_2 + \mu_2}, \frac{s_2 - \mu_1}{s_1 + s_2 - \mu_2}; 1\right).$$

By Weyl group symmetry, we may assume without loss of generality that $\sigma_{\eta}(\infty) = \Re \mu_2$. For $\Re(s_1)$ sufficiently large, $M^*_\mu(s)$ is holomorphic for $\Re(s_2) > \sigma_{\pi}(\infty)$, as poles can only occur at $s_2 = \pm \mu_1 - k, \pm \mu_2 - k$ for $k \in \mathbb{N}_0$. Hence, for $\Re(s_1)$ sufficiently large, the function

$$M^+_\mu(s) := M^*_\mu(s)(s_2 + \mu_1)(s_2 - \mu_1)(s_2 + \mu_2)(s_2 - \mu_2)$$

is holomorphic for $\Re(s_2) > \sigma_{\pi}(\infty) - 1$. Now let

$$\hat{M}_\mu(s) := M^+_\mu(s) = M^*_\mu(s)(s_2 + \mu_2)(s_2 + \mu_1)(s_2 - \mu_2).$$

The inverse Mellin transform of $\hat{M}_\mu(s)$ is then given by

$$\hat{W}_\mu(y) := D_{\mu_2} D_{\mu_1} D_{-\mu_2} W^*_\mu(y),$$

where the differential operators are applied to $y_2$. On the other hand, we compute the inverse Mellin transform directly. By shifting the contour to $\Re(s_2) = \sigma_{\pi}(\infty) - \frac{1}{2}$, we get

$$\hat{W}_\mu(y) = \frac{1}{2\pi i} \int_{(c)} \text{Res}_{s_2 = \mu_2} (\hat{M}_\mu(s)y^{-s}) ds_1 + \frac{1}{(2\pi i)^2} \int_{(c)} \int_{(\sigma_{\pi}(\infty) - \frac{1}{2})} \hat{M}_\mu(s)y^{-s} ds_2 ds_1. \quad (5.6)$$
As \( y_2 \to 0 \), the second integral has size \( O_{y_1, \mu}(y_2^{-\sigma_\pi(\infty) + \frac{1}{2}}) \). We evaluate the first integral:

\[
\frac{1}{2\pi i} \int_{(c)} \text{Res}_{s_2 = \mu_2} (\hat{M}_\mu(s)y^{-s})ds_1
\]

\[
= y_2^{-\mu_2} \Gamma(\mu_2 + 1) \Gamma\left(\frac{\mu_1 + \mu_2}{2} + 1\right) \Gamma\left(\frac{-\mu_1 + \mu_2}{2} + 1\right) \frac{1}{2\pi i} \times \int_{(c)} \Gamma\left(\frac{s_1 + \mu_1 + \mu_2}{2}\right) \Gamma\left(\frac{s_1 + \mu_1 - \mu_2}{2}\right) \Gamma\left(\frac{s_1 - \mu_1 + \mu_2}{2}\right) \Gamma\left(\frac{s_1 - \mu_1 - \mu_2}{2}\right)
\]

\[
\times_2 \text{F}_1\left(\frac{\mu_1 + \mu_2}{2}, \frac{-\mu_1 + \mu_2}{2}; \frac{s_1 + 2\mu_2}{2}; 1\right) y_1^{-s_1} ds_1
\]

\[
= y_2^{-\mu_2} \Gamma(\mu_2 + 1) \Gamma\left(\frac{\mu_1 + \mu_2}{2} + 1\right) \Gamma\left(\frac{-\mu_1 + \mu_2}{2} + 1\right) \frac{1}{2\pi i} \int_{(c)} \Gamma\left(\frac{s_1 + \mu_1 - \mu_2}{2}\right) \Gamma\left(\frac{s_1 - \mu_1 - \mu_2}{2}\right) y_1^{-s_1} ds_1
\]

\[
= y_2^{-\mu_2} \Gamma(\mu_2 + 1) \Gamma\left(\frac{\mu_1 + \mu_2}{2} + 1\right) \Gamma\left(\frac{-\mu_1 + \mu_2}{2} + 1\right) W^*_\mu(\mu_1 y_1) y_1^{-\mu_2},
\]

where \( W^*_\mu(y_1) = y_1^{-1/2} W_{\mu_1}(y_1) \) denotes the standard normalised GL(2)-Whittaker function. Here we used the well-known identity

\[
_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)}, \quad \Re(c) > \Re(a + b).
\]

Now we set

\[
W^{**}_\mu(y_1) := \Gamma(\mu_2 + 1) \Gamma\left(\frac{\mu_1 + \mu_2}{2} + 1\right) \Gamma\left(\frac{-\mu_1 + \mu_2}{2} + 1\right) W^*_\mu(\mu_1 y_1) y_1^{-\mu_2},
\]

and conclude that \( \hat{W}_\mu(y) = y_2^{-\mu_2} W^{**}_\mu(y_1) + O_{y_1, \mu}(y_2^{-\sigma_\pi(\infty) + \frac{1}{2}}) \) as \( y_2 \to 0 \).

The rest of the proof follows the argument in [2]. Applying Lemma 5.1 repeatedly, we obtain constants \( \frac{1}{2} < \gamma_1 < \gamma_2 < 1 \) such that the bound

\[
||| W^*_\mu(y) ||| \gg y_2^{-\sigma_\pi(\infty)} ||| W^{**}_\mu(y_1) |||
\]

holds for \( y_2 \in [\gamma_1/Z, \gamma_2/Z] \), when \( y_1 \) and \( \mu \) vary in some fixed compact domain. Now choose functions \( E^{**}_j : \mathbb{R}^+ \to \mathbb{C} \), depending on \( \Omega \) but not \( Z \), such that \( \sum_j |\langle E^{**}_j, W^{**}_\mu \rangle_{\mathbb{R}^+}|^2 \gg 1 \) for \( \mu \in \Omega \), where \( \langle -, - \rangle_{\mathbb{R}^+} \) denotes the inner product with respect to the Haar measure on \( \mathbb{R}^+ \). Now define \( E^*_j(y_1, y_2) = \delta_{y_1 < y_2 < y_2} E^{**}_j(y_1) \). This choice depends on \( Z \), but the support of \( E^*_j \) varies inside some interval depending only on \( \Omega \). We then obtain

\[
\sum_j \left| \left\langle (E^*_j)^{(1, Z)}, W^*_\mu \right\rangle_{\mathbb{R}^2} \right|^2 = \sum_j \int_{\mathbb{R}^2} E^*_j(y_1, Zy_2) W^*_\mu(y) \frac{dy_1}{y_1} \frac{dy_2}{y_2} \gg Z^{2\sigma_\pi(\infty)}.
\]
Using the relation (5.5), we obtain

\[ Z^{2\sigma_\pi(\infty)} \ll \sum_j \left| \int_{\mathbb{R}^2^+} y^{-\eta} E_j^\pi(y_1, Zy_2) \frac{dy_1}{y_1} \frac{dy_2}{y_2} \right|^2 = Z^{-2\eta_2} \sum_j \left| \left\langle E_j^{(1, Z)}, W_\mu \right\rangle_{T(\mathbb{R}^+)} \right|^2 \]

upon defining \( E_j(y_1, y_2) = y^{2\eta} E_j^\pi(y_1, y_2) \). This finishes the proof of the lemma. \( \square \)

## 6 HECKE EIGENVALUES AND FOURIER COEFFICIENTS

Let \( \mathcal{M} \) be a set of matrices in \( \text{GSp}(4, \mathbb{Q})^+ \) that is left and right invariant under \( \Gamma = \text{Sp}(4, \mathbb{Z}) \) and is a finite union \( \bigcup_j \mathcal{M}_j \) of left cosets. Then \( \mathcal{M} \) defines a Hecke operator \( T_{\mathcal{M}} \) on the space of automorphic forms by

\[ T_{\mathcal{M}} \varphi(g) = \sum_j \varphi(\mathcal{M}_j g). \]

For a matrix \( g \in \text{GSp}(4, \mathbb{Q})^+ \), we denote by \( T_g \) the Hecke operator \( T_{\Gamma(g)} \). For \( m \in \mathbb{N} \), let

\[ S(m) := \{ M \in \text{GSp}(4, \mathbb{Z})^+ \mid M^\top J M = mJ \}, \quad J = \begin{pmatrix} -I_2 & I_2 \end{pmatrix}. \]

The \( m \)th standard Hecke operator is then given by \( T(m) = T_{S(m)} \). The set of matrices

\[ \mathcal{H}(m) = \left\{ \begin{pmatrix} A & m^{-1}BD \\ D \end{pmatrix} \in S(m) \mid A = \begin{pmatrix} a_1 & a_2 \\ a_2 & \end{pmatrix}, B = \begin{pmatrix} b_1 & b_2 & b_3 \\ b_2 & b_3 \end{pmatrix}, 0 \leq a_1, a_2 < m, AD^\top = mI_2, BD \equiv 0 \pmod{m} \right\} \]

(6.1)

gives a complete system of left coset representatives for \( \Gamma \setminus S(m) \) [25]. For \( r \in \mathbb{N}_0, 0 \leq a \leq b \leq r/2 \) and any prime \( p \), define

\[ T^{(r)}_{a,b}(p) := T_{\text{diag}(p^a, p^b, p^{r-a}, p^{r-b})}. \]

When the context is clear, we suppress \( p \) from the notation, and write \( T^{r}_{a,b} \) instead. Then \( T(p^r) \) admits a decomposition

\[ T(p^r) = \sum_{0 \leq a \leq b \leq r/2} T^{(r)}_{a,b}(p). \]

It is well known that the Hecke algebra \( \mathcal{H} \) of \( \text{Sp}(4, \mathbb{R}) \) is generated by \( T(p) = T^{(1)}_{0,0}(p) \) and \( T^{(2)}_{0,1}(p) \) for primes \( p \), along with the identity.
We also define involutions $T_{\varepsilon_1}, T_{\varepsilon_2}$ on the space of cuspidal automorphic forms by

$$T_{\varepsilon_1} \varphi(g) = \varphi(\varepsilon_1 g), \quad T_{\varepsilon_2} \varphi \left( \begin{array}{c} Y \\ X \\ (Y^{-1})^T \end{array} \right) = \varphi \left( \begin{array}{c} Y \\ -X \\ (Y^{-1})^T \end{array} \right),$$

where $\varepsilon_1 = \text{diag}(-1, 1, -1, 1)$. It is clear that

$$(T_{\varepsilon_1} \varphi)(M_1, M_2)(g) = \varphi(-M_1, M_2)(g), \quad (T_{\varepsilon_2} \varphi)(M_1, M_2)(g) = \varphi(M_1, -M_2)(g). \quad (6.2)$$

It is also straightforward to check that $T_{\varepsilon_1}, T_{\varepsilon_2}$ commute with the Hecke operators and the invariant differential operators. So we may assume that an automorphic form $\varphi$ is also an eigenfunction of $T_{\varepsilon_1}, T_{\varepsilon_2}$.

Let $\pi$ be the irreducible automorphic representation corresponding to $\varphi$. We write $\lambda(m, \pi)$ and $\lambda^{(r)}_{a,b}(p, \pi)$ to denote the eigenvalue of $\varphi$ with respect to $T(m)$ and $T^{(r)}_{a,b}$, respectively, and write $\lambda'(m, \pi) := m^{-3/2}\lambda(m, \pi)$. Again, we omit $\pi$ from the notation when the context is clear.

It is known that if $\varphi$ is generic cuspidal and $L^2$-normalised, then by [7, Theorem 1.1] and [17, Theorem 3], we have

$$|A_{\varphi}(1, 1)|^2 \asymp \frac{1}{[\text{Sp}(4, \mathbb{Z}) : \Gamma_0(q)]L(1, \pi, \text{Ad})} \gg q^{-3-\varepsilon}. \quad (6.3)$$

In particular, $A_{\varphi}(1, 1) \neq 0$.

**Notation.** Let $\varphi$ be an $L^2$-normalised generic cuspidal newform. For the rest of the section, it is, however, instructive to have an alternative normalisation, such that the $(1,1)$st Fourier coefficient is 1. To avoid confusion, we always denote by $\varphi$ an $L^2$-normalised form, and by $\varphi_1$ a scalar multiple of $\varphi$ such that $A_{\varphi_1}(1, 1) = 1$. From (6.3), we see that $\varphi_1 = k \varphi$ for some $|k| \ll q^{3/2+\varepsilon}$. When $\varphi$ is an Eisenstein series, there is no $L^2$-normalisation; nevertheless, we can pick a normalisation $\varphi_1$ of $\varphi$ such that $A_{\varphi_1}(1, 1) = 1$.

Now fix a prime $p \nmid q$. Let $M = (M_1, M_2)$, and $0 \leq c, d \leq r$ such that $p^{d-c} \mid M_1$ and $p^{r-2d} \mid M_2$. Write

$$\Gamma \text{diag}(p^a, p^b, p^{r-a}, p^{r-b}) \Gamma = \bigcup_i \Gamma h_i$$

as a finite union of left cosets. We can assume that $h_i \in U(\mathbb{Q})T(\mathbb{Q}^+)$. Consider the decomposition $h_i = \hat{y}_i \hat{x}_i$, with $\hat{y}_i \in T(\mathbb{Q}^+), \hat{x}_i \in U(\mathbb{Q}^+)$. We define exponential sums

$$\mathcal{E}_{a,b,M}^{(r)}(c, d) := \sum_{\Gamma h_i \subseteq \Gamma \text{diag}(p^a, p^b, p^{r-a}, p^{r-b}) \Gamma} \vartheta_M(\hat{x}_i),$$

and

$$\mathcal{E}^{(r)}(c, d) := \sum_{0 \leq a, b \leq r/2} \mathcal{E}_{a,b,(1,1)}^{(r)}(c, d) = \sum_{\Gamma h_i \subseteq S(p^r)} \vartheta(\hat{x}_i).$$
Proposition 6.1. We have
\[ \lambda_{a,b}^{(r)}(p)A_{\varpi}(M) = \sum_{0 \leq c, d \leq r} \zeta_{a,b,M}^{(r)}(c, d) p^{2c + d - \frac{3r}{2}} A_{\varpi}(M_1 p^{d-c}, M_2 p^{r-2d}). \]

Proof. We compute the Fourier coefficient of \( T_{a,b}^{(r)} \varpi \) in two ways. On the one hand, we have
\[
\int_{U(\mathbb{Z}) \setminus U(\mathbb{R})} T_h \varpi(xy) \overline{\vartheta_M(x)} \, dx = \lambda_{a,b}^{(r)}(p) \frac{A_{\varpi}(M)}{M^r} \mathcal{W}_\mu(t(M)y). \tag{6.4}
\]
On the other hand, we expand the Hecke operator
\[
\int_{U(\mathbb{Z}) \setminus U(\mathbb{R})} T_h \varpi(xy) \overline{\vartheta_M(x)} \, dx = \sum_{\Gamma h_1} \int_{U(\mathbb{Z}) \setminus U(\mathbb{R})} \varpi(h_1 xy) \overline{\vartheta_M(x)} \, dx
\]
\[= p^{-4r} \sum_{\Gamma h_1} \int_{U(p'^r \mathbb{Z}) \setminus U(\mathbb{R})} \varpi(h_1 xy) \overline{\vartheta_M(x)} \, dx. \]
Write \( h_1 x = x' \tilde{y}_i \), with \( x' \in U(\mathbb{R}) \), and \( \tilde{y}_i = \text{diag}(c_1, \ldots, c_4) \). A simple calculation shows that
\[ x'_{kl} = c_l \sum_j (h_i)_{kj} x_j. \]
In particular, we have
\[ x_{12} = \frac{c_2}{c_1} x'_{12} - (h_i)_{12} \frac{x_{12}}{c_1}, \quad x_{24} = \frac{c_4}{c_2} x'_{24} - (h_i)_{24} \frac{x_{24}}{c_2}. \]
Making this substitution, the expression becomes
\[ p^{-4r} \sum_{\Gamma h_1} \prod_{k,l} \int_{0}^{c_k p^r} \varpi(x' \tilde{y}_i y) e \left( \frac{c_2}{c_1} M_1 x_{12} + \frac{c_4}{c_2} M_2 x_{24} \right) \frac{c_l}{c_k} dx'_{kl}, \]
where \( (k, l) \) runs through the indices \((1, 2), (1, 3), (2, 3), (2, 4)\). By periodicity, we shift the integral and get
\[ p^{-4r} \sum_{\Gamma h_1} \prod_{k,l} \int_{0}^{c_k p^r} \varpi(x' \tilde{y}_i y) e \left( \frac{c_2}{c_1} M_1 x_{12} + \frac{c_4}{c_2} M_2 x_{24} \right) \frac{c_l}{c_k} dx'_{kl}. \]
Since \( \varpi(x' \tilde{y}_i y) \) is 1-periodic with respect to \( x'_{kl} \), this integral vanishes unless \( c_1 \mid c_2 M_1 \) and \( c_2 \mid c_4 M_2 \). We sum over the terms with the same \( \tilde{y}_i = (p^r, p^d, p^{r-d}, p^{r-2d}) \) and get
\[ \sum_{0 \leq c, d \leq r} \zeta_{a,b,M}^{(r)}(c, d) \int_{U(\mathbb{Z}) \setminus U(\mathbb{R})} \varpi(x' \tilde{y}_i y) e \left( -p^{d-c} M_1 x_{12} - p^{r-2d} M_2 x_{24} \right) dx'. \]
Evaluating the integral gives
\[
\sum_{0 \leq c, d \leq r} \mathcal{G}^{(r)}_{a, b, M}(c, d) p^{2c + d - \frac{3r}{2}} \frac{A_\varpi(p^{d-c}M_1, p^{r-2d}M_2)}{M^r} W_\nu(i(M)y). \tag{6.5}
\]
Comparing (6.4) and (6.5) gives the result. \hfill \square

**Theorem 6.2.** Let \( \varpi_1 \in V_\pi \) be a cuspidal newform such that \( A_\varpi_1(1, 1) = 1 \), and \( p \nmid q \) a prime. The Hecke eigenvalues \( \lambda(p^r, \pi) \) of \( \pi \) with respect to \( T(p^r) \) are given by
\[
\lambda(p, \pi) = p^{3/2} A_{\varpi_1}(1, p),
\]
\[
\lambda(p^r, \pi) = p^{3r/2} \left( A_{\varpi_1}(1, p^r) - p^{-1} A_{\varpi_1}(1, p^{r-2}) \right), \quad r \geq 2.
\]

**Proof.** Plugging in \( M = (1, 1) \) to Proposition 6.1 gives
\[
\lambda(p^r) A_{\varpi_1}(1, 1) = \sum_{0 \leq a, b \leq r/2} \mathcal{G}^{(r)}_{a, b, M}(1, 1) = \sum_{0 \leq c, d \leq r/2} \mathcal{G}^{(r)}_{c, d, M}(p^{d-c}, p^{r-2d}). \tag{6.6}
\]
We evaluate \( \mathcal{G}^{(r)}(c, d) \) explicitly. We set \( A_a := \left( \begin{array}{cc} p^c & a \\ p^d & b \end{array} \right) \), and partition the sum
\[
\mathcal{G}^{(r)}(c, d) = \sum_{0 \leq a < p^d} \mathcal{G}^{(r)}(c, d; a), \quad \text{where} \quad \mathcal{G}^{(r)}(c, d; a) := \sum_{\Gamma h_i \subseteq \mathcal{S}(p^r)} \theta(\hat{x}_i),
\]
and \( A(h_i) \) denotes the top left \( 2 \times 2 \) block of \( h_i \). Using representatives in (6.1), we rewrite
\[
\mathcal{G}^{(r)}(c, d; a) = \sum_{\Gamma h_i \subseteq \mathcal{S}(p^r)} \theta(\hat{x}_i) = \sum_{\Gamma h_i \subseteq \mathcal{S}(p^r)} e\left( \frac{a}{p^c} + \frac{b_3}{p^{2d}} \right) e\left( \frac{a}{p^c} \right) \sum_{\Gamma h_i \subseteq \mathcal{S}(p^r)} e\left( \frac{b_3}{p^{2d}} \right).
\]
The condition \( BD \equiv 0 \pmod{p^r} \) in (6.1) says
\[
p^{-r}BD = \left( \begin{array}{cc} b_1 p^{-c} - ab_3 p^{-d-c} & b_2 p^{-d} \\ b_2 p^{-c} - ab_3 p^{-d-c} & b_3 p^{-d} \end{array} \right) \in M_2(\mathbb{Z}). \tag{6.7}
\]
Note that the summation over \( B \) depends only on \( v_p(a) \). We partition the sum with respect to \( v_p(a) \). For \( v_p(a) \leq c - 2 \), we have
\[
\sum_{0 \leq a < p^d \atop v_p(a) \leq c - 2} \mathcal{G}^{(r)}(c, d; a) = \sum_{0 \leq a < p^{d-c} \atop a' \geq 0} \sum_{0 \leq d-v \atop (a', p) = 1} e\left( \frac{a'}{p^{d-u}} \right) \sum_{\Gamma h_i \subseteq \mathcal{S}(p^r)} \theta(\hat{x}_i) = 0.
\]
For \( v_p(a) = c - 1 \), we have \( d \geq c \geq 1 \), and

\[
\sum_{0 \leq a < p^d \atop \nu_p(a) \leq c-1} \mathbb{G}^{(r)}(c, d; a) = \sum_{0 < a' < p^{d+1}} e \left( \frac{a'}{p} \right) \sum_{(a', p) = 1 \atop \Gamma_{h_i} \subseteq S(r') \atop A(h_i) = A_{p-1}} e \left( \frac{b_3}{p^{2d}} \right) = -p^{d-c} \sum_{\Gamma_{h_i} \subseteq S(r') \atop A(h_i) = A_{p-1}} e \left( \frac{b_3}{p^{2d}} \right).
\]

The integrality conditions in (6.7) forces \( p^{d+1} \mid b_3, p^d \mid b_2, \) and \( p^{d+1} \mid b_1^d p^{d+c-1} + b_2 \). Hence

\[
\sum_{0 \leq a < p^d \atop \nu_p(a) \leq c-1} \mathbb{G}^{(r)}(c, d; a) = -p^{d-c} \sum_{0 \leq b_1, b_2, b_3 < p^r \atop \Gamma_{h_i} \subseteq S(r') \atop p^{d+1} \mid b_3, p^d \mid b_2} e \left( \frac{b_3}{p^{2d}} \right) = \begin{cases} -p^{3r-c-2d-1} & \text{if } d = 1, \\ 0 & \text{otherwise}. \end{cases}
\]

For \( \nu_p(a) \geq c \), the integrality condition in (6.7) forces \( p^d \mid b_2, b_3 \), and \( p^e \mid b_1 \). Hence

\[
\sum_{0 \leq a < p^d \atop \nu_p(a) \geq c} \mathbb{G}^{(r)}(c, d; a) = p^{d-c} \sum_{0 \leq b_1, b_2, b_3 < p^r \atop \Gamma_{h_i} \subseteq S(r') \atop p^d \mid b_2, b_3 \atop p^e \mid b_1} e \left( \frac{b_3}{p^{2d}} \right) = \begin{cases} p^{3r-c-2d} & \text{if } d = 0, \\ 0 & \text{otherwise}. \end{cases}
\]

Hence we conclude

\[
\mathbb{G}^{(r)}(c, d) = \begin{cases} p^{3r} & \text{if } (c, d) = (0, 0), \\ -p^{3r-4} & \text{if } (c, d) = (1, 1), \\ 0 & \text{otherwise}. \end{cases}
\]

Putting this back into (6.6) gives the statement. \( \square \)

Hecke eigenvalues can also be expressed in terms of local Satake parameters \( \alpha_p, \beta_p \) associated to \( \pi \). Without loss of generality, assume \( |\alpha_p| \geq |\beta_p| \geq 1 \). Then up to some ordering we have \( p^{\mu\pi(p, 1)} = \alpha_p, \ p^{\mu\pi(p, 2)} = \beta_p \), and \( \sigma_\pi(p) = \mu_\pi(p, 1) \). By an identity of Shimura [24, Theorem 2], we have

\[
\sum_{r=0}^{\infty} \lambda(p^r)x^r = (1 - p^2x^2)(1 - p^{3/2}\alpha_p x)^{-1}(1 - p^{3/2}\alpha_p^{-1} x)^{-1}(1 - p^{3/2}\beta_p x)^{-1}(1 - p^{3/2}\beta_p^{-1} x)^{-1}.
\]

(6.8)

For convenience, we define \( \sigma_\pi^+(p) = \frac{3}{2} + \sigma_\pi(p) \).

**Lemma 6.3.** For a prime \( p \nmid q \) and \( r \geq 3 \), the following inequality

\[
\left| \lambda(p^{r-j}) \right| \geq \frac{p^{(r-j)\sigma_\pi^+(p)}}{16}
\]

holds for some \( j \in \{0, 1, 2, 3\} \).
Proof. We derive from (6.8) that
\[
(1 - p^{3/2} \alpha_p x)(1 - p^{3/2} \beta_p x)(1 - p^{3/2} \beta^{-1}_p x) \sum_{r=0}^{\infty} \lambda(p^r) x^r = (1 - p^2 x^2) \sum_{r=0}^{\infty} \left( p^{3/2} \alpha_p \right)^r x^r.
\]
Comparing coefficients gives
\[
\lambda(p^r) - \lambda(p^{r-1}) p^{3/2} \left( \alpha_p^{-1} + \beta_p^{-1} \right) + \lambda(p^{r-2}) p^3 \left( \alpha_p^{-1} \beta_p + \alpha_p^{-1} \beta^{-1}_p + 1 \right) + \lambda(p^{r-3}) p^{9/2} \alpha_p = p^{3r/2} \left( \alpha_p^r - p^{-1} \alpha_p^{r-2} \right).
\]
Assume the contrary. Then the left-hand side is bounded by
\[
\frac{p^{r+1}(p)}{2} \leq p^{r+1} - \frac{p^{2+(r-2)\sigma_2}(p)}{2} \leq p^{3r/2} \left| \alpha_p^r - p^{-1} \alpha_p^{r-2} \right|,
\]
a contradiction.

Lemma 6.4. Let \( \varpi_1 \in V_\pi \) be a cuspidal newform such that \( A_{\varpi_1}(1,1) = 1 \), \( p \nmid q \) a prime, and \( r_0 \in \mathbb{N}_0 \). Then the inequality
\[
\left| A_{\varpi_1}(1, p^r) \right| \geq \frac{p^{r+\sigma_2}(p)}{32}
\]
holds for some \( r_0 \leq r \leq r_0 + 5 \).

Proof. By Lemma 6.3, we have
\[
\left| \lambda(p^r) \right| \geq \frac{p^{r+\sigma_2}(p)}{16}
\]
for some \( r_0 + 2 \leq r \leq r_0 + 5 \). By Theorem 6.2, we have
\[
p^{3r/2} \left( \left| A_{\varpi_1}(1, p^r) \right| + p^{-1} \left| A_{\varpi_1}(1, p^{r-2}) \right| \right) \geq \frac{p^{r+\sigma_2}(p)}{16},
\]
and the statement follows.

7 POINCARÉ SERIES AND THE KUZNETSOV FORMULA

Let \( E : \mathbb{R}_+^2 \to \mathbb{C} \) be a fixed function with compact support, \( X \in \mathbb{R}_+^2 \) a ‘parameter’. We define
\[
E^{(X)}(y_1, y_2) = E(X_1 y_1, X_2 y_2),
\]
and a right \( K \)-invariant function \( F^{(X)} : \text{Sp}(4, \mathbb{R}) \to \mathbb{C} \) by
\[
F^{(X)}(x y) = \vartheta(x) E^{(X)}(y y) \quad (7.1)
\]
for \( x \in U(\mathbb{R}) \) and \( y \in T(\mathbb{R}_+) \), where \( \theta = \theta_{(1,1)} \) is as in (2.1). For \( N \in \mathbb{N}^2 \), we define the Poincaré series of level \( q \) to be

\[
P_N^{(X)}(xy) = \sum_{\gamma \in U(\mathbb{Z}) \backslash \Gamma_0(q)} F^{(X)}(\iota(N)xy)\gamma xy.
\]

Note that \( F^{(X)}(\iota(N)xy) = \theta_N(x)E^{(X)}(NY(y)) = \theta_N(x)E(XNY(y)) \). For \( w \in W \), let \( G_w = UwTU \), and \( \Gamma_w := U(\mathbb{Z}) \cap w^{-1}U(\mathbb{Z})^\perp w \). Let \( R_w(q) \) be a complete system of coset representatives for \( P_0 \cap \Gamma_0(q) \backslash \Gamma_0(q) \cap G_w / \Gamma_w \).

We compute the Fourier coefficients of the Poincaré series:

\[
\int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} P_M^{(X)}(xy)\overline{\theta_N(x)}dx
\]

\[
= \sum_{\gamma \in P_0 \cap U_0(q) \backslash \Gamma_0(q)} \int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} F^{(X)}(\iota(M)\gamma xy)\overline{\theta_N(x)}dx
\]

\[
= \sum_{w \in W} \sum_{\gamma \in R_w(q)} \sum_{\ell \in \Gamma_w} \int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} F^{(X)}(\iota(M)\ell xy)\overline{\theta_N(x)}dx
\]

For fixed \( y \), it follows from Lemma 3.1 and \( E \) having compact support that the \( c \)-sum runs over a finite set, and \( U_w(\mathbb{R}) \) runs over a compact domain. In particular, the right-hand side is absolutely convergent. Now let \( \varpi \) be an automorphic form, not necessarily cuspidal, in the spectrum of \( L^2(\Gamma_0(q) \backslash \text{Sp}(4, \mathbb{R}) / K) \). By unfolding, (5.1) and a change of variables \( \iota(N)y \mapsto y \), we obtain

\[
\left\langle \varpi, P_N^{(X)} \right\rangle = \int_{T(\mathbb{R}_+)} \int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} \varpi(xy)\overline{\theta_N(-x)E^{(X)}(N \cdot y(y))}dx d^*y = N^\eta A_{\varpi}(N) \left\langle W_{\mu^*}, E^{(X)} \right\rangle_{T(\mathbb{R}_+)}. \]

Since \( E \) has compact support, the integral above is absolutely convergent even when \( \varpi \) is not cuspidal, and the inner product is always well defined. Note that taking inner product with \( P_N^{(X)} \) kills the non-generic spectrum. By Parseval, we obtain

\[
\left\langle P_M^{(X)}, P_N^{(X)} \right\rangle = M^\eta N^\eta \int_{(q)} A_{\varpi}(M)A_{\varpi}(N) \left\langle W_{\mu^*}, E^{(X)} \right\rangle_{T(\mathbb{R}_+)}^2 d\varpi.
\]

Meanwhile, unfolding the inner product directly gives

\[
\left\langle P_M^{(X)}, P_N^{(X)} \right\rangle = \int_{T(\mathbb{R}_+)} \int_{U(\mathbb{Z}) \backslash U(\mathbb{R})} P_M^{(X)}(xy)\overline{\theta_N(-x)E^{(X)}(N \cdot y(y))}dx d^*y
\]

\[
= \sum_{w \in W} \sum_{c \in \mathbb{N}^2} \text{Kl}_{q,w}(c, M, N) \int_{U_w(\mathbb{R})} F^{(X)}(\iota(M)c^* wxy)\overline{\theta_N(-x)E(XN \cdot y(y))}dx d^*y.
\]
Now define
\[
A = t(XM)c^*w(XN)^{-1}w^{-1} = t((XM) \cdot w(XN))c^* \in T(\mathbb{R}_+). \tag{7.2}
\]

Then \(y(A)c_1c_2 = ((XM) \cdot w(XN))^\eta\). By change of variables \(t(XN)y \mapsto y, t(XN)x(XN)^{-1} \mapsto x\), we can express \(\langle P_M^{(X)}, P_N^{(X)} \rangle\) as
\[
\sum_{w \in W} \sum_{c \in \mathbb{N}^2} Kl_{q,w}(c, M, N)(XM)^\eta(XN)^\eta c_1c_2y(A)^\eta \int_{T(\mathbb{R}_+)} \int_{U_w(\mathbb{R})} F^{(X)}(t(X)^{-1}Awx)y \partial_{X^{-1}}(-x)E(y(y))dx d^*y.
\]

We then conclude a Kuznetsov-type pre-trace formula.

**Lemma 7.1.** Let \(M, N \in \mathbb{N}^2, X \in \mathbb{R}_+^2\), \(E\) a function on \(\mathbb{R}_+^2\) with compact support, and define \(F^{(X)}\) as in (7.1). Then
\[
\int_{(q)} \left| A_{\nu}(M)A_{\nu}(N) \left( W_{\mu}, E^{(X)} \right)_{T(\mathbb{R}_+)} \right|^2 d\nu
\]
\[
= \sum_{w \in W} \sum_{c \in \mathbb{N}^2} Kl_{q,w}(c, M, N) X^{2\eta} c_1c_2y(A)^\eta \int_{T(\mathbb{R}_+)} \int_{U_w(\mathbb{R})} F^{(X)}(t(X)^{-1}Awx)y \partial_{X^{-1}}(-x)E(y(y))dx d^*y,
\]
with \(A\) as in (7.2).

**8 PROOF OF THEOREMS**

We establish the following proposition, from which the other theorems are proved.

**Proposition 8.1.** Keep the notations as above. Let \(m \in \mathbb{N}\) be coprime to \(q\) and \(Z \geq 1\). Then
\[
\int_{(q)} \left| A_{\nu}(1, m) \right|^2 Z^{2\sigma_\infty(\infty)} \delta_{\lambda = \in I} d\nu \ll_I \varepsilon q^\varepsilon
\]
uniformly in \(mZ \ll q^2\) for a sufficiently small implied constant depending on \(I\).

**Proof.** We take \(X = (1, Z), M = N = (1, m)\), and apply Lemma 7.1. By Lemma 5.2, there is a finite set of compactly supported functions \(E_j\) such that
\[
Z^{2\eta_2 + 2\sigma_\infty(\infty)} \delta_{\lambda = \in I} \ll_I \sum_j \left| \left( W_{\mu}, E_j^{(X)} \right)_{T(\mathbb{R}_+)} \right|^2. \tag{8.1}
\]

Now we consider the arithmetic side of the Kuznetsov formula for a fixed \(E^{(X)} = E_j^{(X)}\). It suffices to consider the Weyl elements \(w \in W\) for which the Kloosterman sum \(Kl_{q,w}(c, M, N)\) does not vanish, namely \(w = \text{id}, s_x, s_\beta s_x, s_\beta s_x s_\beta, w_0\).

For \(w = \text{id}\), we have \(c_1 = c_2 = 1\), and hence the contribution is \(O(Z^{2\eta_2}) = O(Z^3)\).
Now let \( w \in \{ s_\alpha s_\beta s_\alpha, s_\beta s_\alpha s_\beta, w_0 \} \). Apply Lemma 3.1 with \( B = (XM) \cdot w(XN) \). Concretely, we set

\[
(B_1, B_2) = \begin{cases} 
(mZ, 1) & \text{if } w = s_\alpha s_\beta s_\alpha, \\
(1, (mZ)^2) & \text{if } w = s_\beta s_\alpha s_\beta, w_0.
\end{cases}
\]

Then we obtain

\[
c_1 \ll B_1^{1/2} = mZ, \quad c_2 \ll B_1 B_2 = \begin{cases} 
mZ & \text{if } w = s_\alpha s_\beta s_\alpha, \\
(mZ)^2 & \text{if } w = s_\beta s_\alpha s_\beta, w_0.
\end{cases}
\]

We assume \( mZ \ll q^2 \) with a sufficiently small implied constant, such that

\[
c_1, c_2 < q^2 \text{ for } w = s_\alpha s_\beta s_\alpha, \quad \text{and } c_1 < q^2, c_2 < q^4 \text{ for } w = s_\beta s_\alpha s_\beta, w_0 \quad (8.2)
\]

always hold. Now we consider the Kloosterman sums

\[
K_{q, w}(c, M, N) = \sum_{xcw'x' \in U(Z) \cap \Gamma_0(q) / U_w(Z)} \theta_M(x') \theta_N(x'),
\]

where the entries of \( M = (M_1, M_2) \) and \( N = (N_1, N_2) \) are coprime to \( q \). The Kloosterman sums are non-zero only when (4.3) and (4.4) are satisfied. By (4.5), (4.6) and (8.2), the problem reduces to computing the Kloosterman sums

\[
K_{q, s_\alpha s_\beta s_\alpha} ((q, q), M, N), \quad K_{q, s_\alpha s_\beta s_\beta} ((q, q^2), M, N),
\]

\[
K_{q, w_0} ((q, q^2), M, N), \quad K_{q, w_0} ((q, q^3), M, N).
\]

From Section 4.1, we see that only \( K_{q, s_\beta s_\alpha s_\beta} ((q, q^2), M, N) \) does not vanish. So only \( w = s_\beta s_\alpha s_\beta \) contributes.

The next step is to estimate for \( w = s_\beta s_\alpha s_\beta \) the integral

\[
\int_{T(\mathbb{R}^+)} \int_{U_w(\mathbb{R})} F(X)(\iota(X)^{-1} Awxy) \partial_X^{-1}(-x) E(y(y)) dx dy,
\]

which can be considered as the archimedean analogue of Kloosterman sums. We first observe that the integral is bounded by

\[
\int_{T(\mathbb{R}^+)} \int_{U_w(\mathbb{R})} |E(y(Awxy))| E(y(y))|,
\]

and this integral is bounded by the size of the set of \( x \in U_w(\mathbb{R}) \) such that \( y(Awxy) \) lies in the support of \( E \). Using Lemma 3.1 and Lemma 3.3, we deduce that

\[
\left| \int_{T(\mathbb{R}^+)} \int_{U_w(\mathbb{R})} F(X)(\iota(X)^{-1} Awxy) \partial_X^{-1}(-x) E(y(y)) dx dy \right| \ll E \text{ vol } \left\{ x \in U_w(\mathbb{R}) \left| \Delta_1(wx) \ll_E y(A)_1 y(A)_2^{1/2}, \Delta_2(wx) \ll_E y(A)_1 y(A)_2 \right\} \ll_E y(A)^{\eta(1+\varepsilon)}.
\]
So the contribution from \( w = s_\beta s_\alpha s_\beta \) is given by

\[
\sum_{c \in \mathbb{N}^2} \text{KL}_{q,w}(c,M,N) \frac{X^{2\eta}}{c_1c_2 y(A)^{\eta}} \int_{T(\mathbb{R}^+)} \int_{U_w(\mathbb{R})} F^{(x)}(t(x)^{-1} A w x y) \theta_{X^{-1}}(-x) E(y(y)) dx dy \]

\[
\ll_E \sum_{c' = c_1/q, c_2 \leq mZ/q} \frac{Z^{2\eta_2} y(A)^{\xi}}{q} \ll Z^{3} q^{\xi}.
\]

Combining the estimates with (8.1), we obtain

\[
\int_{(q)} |A_\pi(1,m)|^2 Z^{2+2\sigma_\pi(\infty)} \delta_{\lambda_\pi \in I} d\omega \ll \int_{(q)} |A_\pi(1,m)|^2 \left| \left< W_\mu, E^{(x)} \right>_{T(\mathbb{R}^+)} \right|^2 d\omega \ll \varepsilon Z^{3} q^{\xi}.
\]

Dividing both sides by \( Z^3 \) yields the theorem.

**Proof of Theorem 1.2.** It follows easily from Proposition 8.1, Theorem 6.2 and the estimate \((6.3)\) that

\[
\sum_{\pi \in \mathcal{I}(q)} |\lambda'(m,\pi)|^2 Z^{2\sigma_\pi(\infty)} \delta_{\lambda_\pi \in I} d\omega \ll \int_{(q)} |A_\pi(1,m)|^2 \left| \left< W_\mu, E^{(x)} \right>_{T(\mathbb{R}^+)} \right|^2 d\omega \ll \varepsilon q^{3+\xi}.
\]

**Proof of Theorem 1.3.** This is just a simple variation of the proofs above. Again we have

\[
\sum_{\pi \in \mathcal{I}(q)} \left| \sum_{\substack{m \leq x \\text{any}\ \pi \in \mathcal{I}(q)\ \text{is} \text{unit}}} \alpha(m) \lambda'(m,\pi) \right|^2 \ll \varepsilon q^{3+\xi} \int_{(q)} |A_\pi(1,m)|^2 Z^{2\sigma_\pi(\infty)} \delta_{\lambda_\pi \in I} d\omega
\]

\[
= q^{3+\xi} \sum_{\substack{m_1,m_2 \leq x \\text{any}\ \pi \in \mathcal{I}(q)\ \text{is} \text{unit}}} \alpha(m_1) \overline{\alpha(m_2)} \int_{(q)} A_\pi(1,m_1) A_\pi(1,m_2) \delta_{\lambda_\pi \in I} d\omega,
\]

where \( M = (1,m), M_1 = (1,m_1), M_2 = (1,m_2) \). Now we apply Lemma 7.1 and evaluate the Kloosterman sums on the arithmetic side. For \( w \neq \text{id} \), apply Lemma 3.1 with \( B = M_1 w M_2 \). We get

\[
c_1 \ll (m_1 m_2)^{1/2} \leq x, \quad c_2 \ll m_1 \leq x \quad \text{for } w = s_\alpha s_\beta s_\alpha,
\]

\[
c_1 \ll (m_1 m_2)^{1/2} \leq x, \quad c_2 \ll m_1 m_2 \leq x^2 \quad \text{for } w = s_\beta s_\alpha s_\beta, w_0.
\]

Note that when \( x \ll q \) with a sufficiently small implied constant, the condition \( (m,q) = 1 \) is void, and we deduce from (4.3) that the Kloosterman sums \( \text{KL}_{q,w}(c,M,N) \) are empty for \( w \neq \text{id} \). Hence only the trivial Weyl element contributes, and we obtain the desired bound.

**Proof of Corollary 1.4.** Observe that for \( \pi \in \mathcal{P}(q) \) an approximate functional equation has length \( q^{1/2} \) (see [15, Section 5]). So, for all but \( O(1) \) cuspidal representations \( \pi \in \mathcal{P}(q) \) (and \( \varepsilon < 1/2 \), we have
The statement then follows from Theorem 1.3. □

Proof of Theorem 1.1. We first assume \( \nu = p \neq q \) is a finite place. We choose \( \nu_0 \) maximal such that \( p^{\nu_0} \approx q^2 \) with an implied constant that is admissible to Proposition 8.1. Then by Lemma 6.4 and the estimate (6.3), there exists \( \nu_0 - 5 \leq \nu_\pi \leq \nu_0 \) such that

\[
|A_{\varpi}(1, p^{\nu_\pi})|^2 \gg q^{-3-\varepsilon} p^{2\nu_\pi \sigma(p)}.
\]

Note that \( p^{\nu_\pi} \approx q^2 \) as well. We apply Proposition 8.1 with \( m = p^{\nu_\pi}, Z = 1 \), and conclude that

\[
N_p(\sigma, F_1(q)) \leq \sum_{\pi \in F_1(q)} \frac{p^{2\nu_\pi \sigma(p)}(p)}{p^{2\nu_\pi \sigma(p)}} \ll q^{3-4\sigma+\varepsilon} \int_{(q)} \sum_{\nu_0-5 \leq \nu_\pi \leq \nu_0} |A_{\varpi}(1, p^{\nu})|^2 \delta_{\lambda_{\varpi} \equiv 1} \ll_{\lambda, \varepsilon} q^{3-4\sigma+\varepsilon}.
\]

For \( \nu = \infty \), we use the estimate (6.3), apply Proposition 8.1 with \( m = 1, Z \ll q^2 \), and conclude that

\[
N_\infty(\sigma, F_1(q)) \leq \sum_{\pi \in F_1(q)} Z^{2\sigma(\infty)-2\sigma} \ll q^{3-4\sigma+\varepsilon} \int_{(q)} |A_{\varpi}(1, 1)|^2 Z^{2\sigma(\infty)} \ll_{\lambda, \varepsilon} q^{3-4\sigma+\varepsilon}.
\]

APPENDIX: COMPUTATION OF FOURIER COEFFICIENTS

In this appendix, we outline an algorithm for computing arbitrary Fourier coefficients of a Hecke eigenform \( \varpi_1 \in V_\pi \) with \( A_{\varpi_1}(1, 1) = 1 \). For this purpose, it suffices to compute the actions of \( T(p) \) and \( T^{(2)}_{0,1}(p) \), which generate the Hecke algebra. By Proposition 6.1, we compute

\[
\lambda(p, \pi) A_{\varpi_1}(M_1, M_2)
\]

\[
= p^{3/2} \left( A_{\varpi_1}(M_1, pM_2) + A_{\varpi_1}(p^{-1}M_1, pM_2) \right) + \left( A_{\varpi_1}(pM_1, p^{-1}M_2) + A_{\varpi_1}(M_1, p^{-1}M_2) \right)
\]

(A.1)

and if \( p \nmid M_2 \),

\[
\left( \lambda^{(2)}_{0,1}(p, \pi) + 1 \right) A_{\varpi_1}(M_1, M_2) = p^2 \left( A_{\varpi_1}(pM_1, M_2) + A_{\varpi_1}(p^{-1}M_1, p^2M_2) + A_{\varpi_1}(p^{-1}M_1, M_2) \right)
\]

(A.2)
We proceed to show how the Fourier coefficients $A_{\varpi_1}(p^{k_1}, p^{k_2})$ are obtained. Starting from $A_{\varpi_1}(1, 1) = 1$, we apply (A.1) and (A.2) with $M = (1, 1)$ and solve the coefficients

$$A_{\varpi_1}(p, 1) = p^{-2}\left(\lambda_{0,1}^{(2)}(p) + 1\right), \quad A_{\varpi_1}(1, p) = p^{-3/2}\lambda(p).$$

Inductively, suppose that the Fourier coefficients $A_{\varpi_1}(p^{k_1}, p^{k_2})$ are known for all $k_1 + k_2 \leq r$. For $0 \leq k \leq r$, applying (A.1) with $M = (p^k, p^{r-k})$ yields the coefficient $A_{\varpi_1}(p^k, p^{r-k+1})$. Then, applying (A.2) with $M = (p^r, 1)$ yields the coefficient $A_{\varpi_1}(p^{k+1}, 1)$, since the coefficient $A_{\varpi_1}(p^{k-1}, p^2)$ has already been determined. This shows that the Fourier coefficients $A_{\varpi_1}(p^{k_1}, p^{k_2})$ with $k_1 + k_2 \leq r + 1$ can be expressed in terms of $\lambda(p)$ and $\lambda_{0,1}^{(2)}(p)$, finishing the induction.

Let $X := p^{-3/2}\lambda(p, \pi)$ and $Y := p^{-2}(\lambda_{0,1}^{(2)}(p, \pi) + 1)$. In fact, the recursion implies that the generating function

$$A(u, v) : = \sum_{i,j \geq 0} A_{\varpi_1}(p^i, p^j)u^iv^j$$

is actually a rational function of a relatively simple form. Concretely, we have

$$A(u, v) = \frac{1 + u + uw - Xuv}{\left(1 + (1 - Y)u + (X^2 - 2Y)u^2 + (1 - Y)u^3 + u^4\right)\left(1 - Xv + (Y + 1)v^2 - Xv^3 + v^4\right)}.$$  

(A.3)

Meanwhile, it is evident from Proposition 6.1 that Fourier coefficients are multiplicative, that is,

$$A_{\varpi_1}(M_1N_1, M_2N_2) = A_{\varpi_1}(M_1, M_2)A_{\varpi_1}(N_1, N_2) \quad \text{if } (M_1M_2, N_1N_2) = 1. \quad (A.4)$$

Using (A.4), we are able to compute $A_{\varpi_1}(M)$ for every $M \in \mathbb{N}^2$.

Remark. From Theorem 6.2, we obtain $\lambda(p^2, \pi) = p^3(X^2 - Y - 1) - p^2$. Hence the Fourier coefficients can also be expressed in terms of eigenvalues $\lambda(p^r, \pi)$ of standard Hecke operators.

Finally, we give a useful characterisation of the generating function $A(u, v)$. Consider the Lie algebra $\mathfrak{sp}_4$, with simple roots $\alpha, \beta$, and fundamental weights $\omega_\alpha, \omega_\beta$. Let $x = e^{\omega_\alpha}$ and $y = e^{\omega_\beta - \omega_\alpha}$ be formal symbols. Let $X = x + x^{-1} + y + y^{-1}$, and $Y = (x + x^{-1})(y + y^{-1}) + 1$. Then it is shown in [10] that

$$A(u, v) = \sum_{i,j \geq 0} \text{ch} V(j\omega_\alpha + i\omega_\beta)u^iv^j$$

is the generating function of Weyl characters of irreducible representations of $\mathfrak{sp}_4$.

ACKNOWLEDGEMENTS
The author would like to thank Valentin Blomer for his guidance, and Edgar Assing for his helpful explanation on the subject. The author would also like to thank Don Zagier for his valuable comments.

The author is supported in part by DAAD Graduate School Scholarship Programme, and by the German Research Foundation under Germany’s Excellence Strategy - EXC-2047/1 - 390685813.
REFERENCES
1. V. Blomer, Applications of the Kuznetsov formula on GL(3), Invent. Math. 194 (2013), 673–729.
2. V. Blomer, Density theorems for GL(n), in preprint, https://arxiv.org/abs/1906.07459, 2019.
3. V. Blomer, J. Buttcane, and P. Maga, Applications of the Kuznetsov formula on GL(3) II: the level aspect, Math. Ann. 369 (2017), 723–759.
4. V. Blomer, J. Buttcane, and N. Raulf, A Sato-Tate law for GL(3), Comm. Math. Helv. 89 (2014), 895–919.
5. R. W. Bruggeman and R. J. Mi tello, Sum formula for SL₂ over a number field and Selberg type estimate for exceptional eigenvalues, Geom. Funct. Anal. 8 (1998), 627–655.
6. R. W. Bruggeman and R. J. Miatello, Density results for automorphic forms on Hilbert modular groups, Geom. Funct. Anal. 13 (2003), 681–719.
7. S.-Y. Chen and A. Ichino, On Petersson norms of generic cusp forms and special values of adjoint L-functions for GSp₄, https://arxiv.org/abs/1902.06429, 2019.
8. R. Dąbrowski and M. Reeder, Kloosterman sets in reductive groups, J. Number Theory 73 (1998), 228–255.
9. T. Finis and J. Matz, On the asymptotics of Hecke operators for reductive groups, https://arxiv.org/abs/1905.09078, 2019.
10. J. Fernández-Núñez, Wifredo García-Fuertes, and A. M. Perelomov, On the generating function of weight multiplicities for the representations of the Lie algebra c₂, J. Math. Phys. 56 (2015), no. 4, 041702.
11. D. Goldfeld, Automorphic forms and L-functions for the Group GL(n, ℝ), Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2006.
12. M. Huxley, Exceptional eigenvalues and congruence subgroups, The Selberg trace formula and related topics, vol. 53 of Contemp. Math., American Mathematical Society, Providence, RI, 1986, pp. 341–349.
13. T. Ishii, On principal series Whittaker functions on Sp(2, ℝ), J. Funct. Anal. 225 (2005), 1–32.
14. H. Iwaniec, Small eigenvalues of Laplacian for γ₀(n), Acta Arith. 56 (1990), 65–82.
15. H. Iwaniec and E. Kowalski, Analytic number theory, volume 53 of AMS Colloq. Publ., American Mathematical Society, Providence, RI, 2004.
16. S. Jana, Applications of analytic newvectors for GL(n), Math. Ann. 380 (2021), no. 3, 915–952.
17. X. Li, Upper bounds on L-functions at the edge of the critical strip, Int. Math. Res. Not. IMRN 2010 (2010), no. 4, 727–755.
18. S. H. Man, Symplectic Kloosterman sums and Poincaré series, https://arxiv.org/abs/2006.03036, 2020.
19. B. Roberts and R. Schmidt, Local Newforms for GSp(4), vol. 1918 of Springer Lecture Note in Mathematics, Springer-Verlag, Berlin, Heidelberg, 2007.
20. P. Sarnak, Statistical properties of eigenvalues of the Hecke operators, Analytic number theory and Diophantine problems, vol. 70 of Progr. Math., Birkhäuser, Boston, 1987, pp. 321–331.
21. P. Sarnak, Diophantine problems and linear groups, Proceedings of the ICM Kyoto, 1990, pp. 459–471.
22. P. Sarnak, Definition of families of L-functions, https://publications.ias.edu/sites/default/files/definition%20of%20family%202008_0.pdf, 2008.
23. P. Sarnak, S. W. Shin, and N. Templier, Families of L-functions and their symmetry, Families of automorphic forms and the trace formula, Simons Symposium, Springer International Publishing, Cham, 2016, pp. 531–578.
24. G. Shimura, On modular correspondences for Sp(n, ℤ) and their congruence relations, Proc. Natl. Acad. Sci USA 49 (1963), 824–828.
25. E. Spence, m-symplectic matrices, Trans. Amer. Math. Soc. 170 (1972), 447–457.
26. G. Stevens, Poincaré series on GL(r) and Kloostermann sums, Math. Ann. 277 (1987), 25–51.