DENSITY THEOREMS FOR $\text{GL}(n)$

VALENTIN BLOMER

Abstract. Strong bounds – going beyond Sarnak’s density hypothesis – are obtained for the number of automorphic forms for the group $\Gamma_0(q) \subseteq \text{SL}(n, \mathbb{Z})$ violating the Ramanujan conjecture at any given unramified place. The proof is based on a relative trace formula of Kuznetsov type and best-possible bounds for certain Kloosterman sums for $\text{GL}(n)$. Further applications are given.

1. Introduction

The concept of families is a very fruitful one in number theory and in particular in the context of automorphic forms. It allows us to study asymptotic properties and has recently been put on some formal ground in [SST, Sa3]. On the conceptual side it damps irregularities of individual members (that may exist or whose non-existence we are unable to prove) and allows statistical concepts and deformation techniques to investigate properties within an ensemble. On the methodological it enables us to use strong analytic tools such as various types of trace formulae.

One of the key conjectures in the field of automorphic forms is the Ramanujan conjecture: cuspidal automorphic representations of the group $\text{GL}(n)$ over a number field $F$ are tempered (see [BB] for a survey). Even for $n = 2$ this appears to be far out of reach, and as a substitute one considers two types of approximations. On the one hand one can measure the worst case scenario, i.e. the largest distance from the tempered spectrum of an individual member in a family. On the other hand one can try to bound the number of members in a family violating the conjecture relative to the amount by which they violate the conjecture. This is a density result which is a familiar concept from the theory Dirichlet $L$-functions: although the Riemann hypothesis is far out of reach, we have good bounds for the number $N(\sigma, T, Q)$ of zeros with real part $\geq \sigma$ and height $\leq T$ of Dirichlet $L$-functions with conductor $q \leq Q$ (see e.g. [IK, Section 10]). The arithmetic reformulation of this is the Bombieri-Vinogradov theorem which roughly states that primes $\leq x$ are equidistributed in “almost all” residue classes modulo $q \leq x^{1/2+o(1)}$ (similarly, “almost all” short intervals contain primes). In many applications this serves as a good substitute for the Riemann hypothesis.

In this note we want to consider the automorphic analogue for the family of automorphic forms for the group $\Gamma_0(q) \subseteq \text{SL}_n(\mathbb{Z})$ of matrices whose lowest row is congruent to $(0, \ldots, 0, \ast)$ modulo $q$. This is a very natural family as it contains precisely the automorphic forms of conductor dividing $q$ [JPSS]. Let us fix a place $v$ of $\mathbb{Q}$, and for an automorphic form $\pi$ let us denote by $\mu_\pi(v) = (\mu_\pi(v, 1), \ldots, \mu_\pi(v, n))$ its local spectral parameter (each entry viewed...
modulo $\frac{2\pi i}{\log p} \mathbb{Z}$ if $v = p$ is a prime). Write
\begin{equation}
\sigma_\pi(v) = \max_j |\Re \mu_\pi(v, j)|.
\end{equation}

The representation $\pi$ is tempered at $v$ if $\sigma_\pi(v) = 0$, and the size of $\sigma_\pi(v)$ measures how far $\pi$ is from being tempered at $v$. An example of a non-tempered representation is the trivial representation which satisfies $\sigma_{\text{triv}}(v) = (n-1)/2$ for every $v$. For a finite family $\mathcal{F}$ of automorphic representations for $GL(n)$ and $\sigma \geq 0$ we define
\[ N_\nu(\sigma, \mathcal{F}) = |\{ \pi \in \mathcal{F} \mid \sigma_\pi(v) \geq \sigma \}|. \]

We have trivially $N_\nu(0, \mathcal{F}) = |\mathcal{F}|$, and if the trivial representation is contained in $\mathcal{F}$, we have $N_\nu((n-1)/2, \mathcal{F}) \geq 1$. One may hope to be able to interpolate linearly between these two extreme cases:
\begin{equation}
N_\nu(\sigma, \mathcal{F}) \ll_{v, \epsilon} |\mathcal{F}|^{1-\frac{2}{n}+\epsilon}
\end{equation}
for arbitrarily small $\epsilon > 0$ with
\begin{equation}
\sigma = (n-1)/2.
\end{equation}

This is precisely Sarnak's density hypothesis [Sa2, p. 465] stated there in the context of groups $G$ of real rank 1, the principal congruence subgroup $\Gamma(q) = \{ \gamma \in G(\mathbb{Z}) \mid \gamma \equiv \text{id (mod q)} \}$ and $v = \infty$. For families of large level, Sarnak's density hypothesis has recently attracted interest in the context of lifting matrices modulo $q$ [Sa4] and the almost diameter of Ramanujan complexes, and for families with growing infinitesimal character in the context of Golden Gates and quantum computing [Sa5, PS]. In each of these cases it is not a spectral gap that is needed, but a certain kind of density result.

The shape of the bound (1.2) – (1.3) bears a certain similarity to the convexity bound for $L$-functions in the Selberg class in the critical strip. Unlike the convexity bound for $L$-functions, (1.2) – (1.3) is in general a very deep result that is completely open for general groups and families. On the other hand, it is a priori not impossible to even obtain “sub-convexity”, i.e. a proof of (1.2) with a constant $a < (n-1)/2$, if the trivial representation is not in $\mathcal{F}$. The Arthur-Selberg trace formula is usually not sensitive to whether the trivial representation is counted or not, but the Kuznetsov formula can be a versatile tool if no residual spectrum is involved.

For the group $GL(2)$ there exist strong density results for many automorphic families, for instance by Sarnak [Sa1], Iwaniec [Iw], Huxley [Hu], Blomer-Buttcane-Raulf [BBR], also in number field versions [BM1, BM2] and for general real rank 1 groups [SX, HK]. Various results are also available for $GL(3)$, see e.g. [Bl, BBM, BBR]. For higher rank groups the deep analysis of the Arthur-Selberg trace formula of Matz-Templier [MT] and Finis-Matz [FM] provides as by-products some density results for the family of Maaß forms of Laplace eigenvalue up to height $T$ and fixed level. The value of $a$ is however much larger than (1.3) for $n > 2$ (at least quadratic in $n$), so that even the “convexity bound” cannot be obtained.

In the present paper we consider the family $\mathcal{F}_I(q)$ of cuspidal automorphic representations generated by Maaß forms for the group $\Gamma_0(q) \subseteq SL_n(\mathbb{Z})$ for a large prime $q$ and Laplace eigenvalue $\lambda$ in a fixed interval $I$. If $I$ is not too small, we have $|\mathcal{F}_I(q)| \asymp I q^{n-1}$. For this family and any place $v \neq q$ of $\mathbb{Q}$, we go beyond the density hypothesis and obtain strong “subconvexity” with a value of
\[ a = (n-1)/4, \]
which is halfway between (1.3) and the Ramanujan conjecture.
Theorem 1. Let $n \geq 3$, $q$ a prime, $v$ be a place of $\mathbb{Q}$ different from $q$, $I \subseteq [0, \infty)$ a fixed interval, $\varepsilon > 0$, and $\sigma \geq 0$. Then

$$N_v(\sigma, \mathcal{F}_I(q)) \ll_{I,v,n,\varepsilon} q^{n-1-4\sigma+\varepsilon}.$$ 

Of course, by [LRS] we know that $N_v(\sigma, \mathcal{F}_I(q)) = 0$ for $\sigma \geq 1/2 - 1/(n^2 + 1)$, but for $0 < \sigma < 1/2 - 1/(n^2 + 1)$ we obtain a substantial power saving. In fact, even the rather generic (though highly non-trivial) Jacquet-Shalika bounds $N_v(\sigma, \mathcal{F}_I(q)) = 0$ for $\sigma > 1/2$ [JS] show for $n \geq 3$ that cuspidal representations are always fairly far away from the trivial representation. Unfortunately it is not clear how to combine the Luo-Rudnick-Sarnak approach with the present techniques.

The theorem remains true for $n = 2$ (by a slightly different proof), but it is known in this case (see [Iw] for $v = \infty$, and the proof for finite $v$ is similar) and recovers Selberg’s 3/16 bound for exceptional eigenvalues. For $n = 3$ and $v = \infty$ this is [BBM, Theorem 4].

As mentioned above, for larger $n$ Theorem 1 is completely new. As we shall outline below it appears to be the limit of what is available by any trace formula approach, even in the case $n = 2$ nothing better is known.

The proof is based on a careful analysis of the arithmetic side of the Kuznetsov formula with a test function on the spectral side that blows up on exceptional Langlands parameters at $v$ (and therefore increases the complexity on the arithmetic side). We denote by $\lambda_\pi(m)$ the $m$-th Hecke eigenvalue of $\pi \in \mathcal{F}_I(q)$.

Theorem 2. Keep the assumptions and notation of Theorem 1. Let $m \in \mathbb{N}$ be coprime to $q$ and $Z \geq 1$. Then

$$\sum_{\pi \in \mathcal{F}_I(q)} |\lambda_\pi(m)|^2 Z^{2\sigma_{\pi}(\infty)} \ll_{I,v,n,\varepsilon} q^{n-1+\varepsilon}$$

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

We shall see in Lemma 4 below that $|\lambda_\pi(p^\nu)|^2$ is often as big as $p^{2\nu\sigma_{\pi}(p)}$ for a prime $p$ and $\nu \in \mathbb{N}$, so that the “test function” $|\lambda_\pi(m)|^2 Z^{2\sigma_{\pi}(\infty)}$ treats finite places and the infinite place essentially on equal footing.

Let us roughly sketch how one may hope to arrive at Theorem 2. Since the Laplacian eigenvalue is fixed, the Whittaker transforms in the Kuznetsov formula play no major role, and the battle is decided on the level of Kloosterman sums. Very roughly, the Kuznetsov formula takes the shape

$$\frac{1}{|\mathcal{F}_I(q)|} \sum_{\pi \in \mathcal{F}_I(q)} |\lambda_\pi(m)|^2 Z^{2\sigma_{\pi}(\infty)} \approx 1 + \sum_{id \neq w \in W} \sum_{q|c_1, \ldots, q|c_{n-1}, c_1, \ldots, c_{n-1} \ll mZ} S_{q,w}(M, M, c)$$

where $W$ is the Weyl group of permutation matrices, $M = (m, 1, \ldots, 1) \in Z^{n-1}$ and $S_{q,w}(M, M, c)$ is a certain generalized Kloosterman sum, defined in (4.2) below, associated with the Weyl element $w$ and moduli $c = (c_1, \ldots, c_{n-1})$. If $mZ \ll q$, then the off-diagonal term vanishes completely and we are done. We will use this observation in Theorem 4 below. This range of $mZ$ recovers the “convexity bound” with the value (1.3). For larger values of $mZ$ and stronger density results we must deal with the Kloosterman sums appearing in the off-diagonal term and improve on the trivial bound $|S_{q,w}(\ast, \ast, c)| \ll (c_1 \cdots c_{n-1})^{1+\varepsilon}$, see (4.7) below. To obtain such bounds for general groups is a famous open problem. In an ideal world we would have complete Weil-type square root cancellation $|S_{q,w}(\ast, \ast, c)| \ll (c_1 \cdots c_{n-1})^{1/2+\varepsilon}$ (at least under certain coprimality assumptions) which allows us to take
as large as $q^2$. This square-root cancellation implies the statement of Theorem 2 and a density hypothesis halfway between the trivial representation and the Ramanujan conjecture. (Additional square root cancellation in the $c_1, \ldots, c_{n-1}$ sum would give the full Ramanujan conjecture, but this is of course not in the cards).

Interestingly, this heuristic sketch turns out to be quite far from the truth. Square root cancellation for the size of Kloosterman sums may fail badly, so that we have to arrive at Theorem 2 by a rather different analysis. The key lemma is the following, which seems to be the first explicit analysis of general $GL(n)$ Kloosterman sums beyond hyper-Kloosterman sums $[Fr]$ associated to the Weyl element $w = (I_{n-1}^{-1})$, where $I_n$ denotes the $n$-by-$n$ identity matrix.

**Theorem 3.** Let $q$ be a prime and let $M, N \in \mathbb{Z}^{n-1}$ with entries coprime to $q$ (in particular non-zero). Let $n \geq 3$ and let $w \in W$. Then $S_{q,w}(M, N, (q, \ldots, q)) = 0$ unless

$$w = w_* := \left( I_{n-2}^{-1} \right)$$

in which case $S_{q,w}(M, N, (q, \ldots, q)) = q^{n-2}$.

Note that this is in sharp contrast to the case $n = 2$, where a Kloosterman sum to prime modulus $q$ has no closed evaluation. The key point is that by multiplicativity the Kloosterman sums in (1.4) contain $S_{q,w}(\ast, \ast, (q, \ldots, q))$ as a large chunk. The critical case is the term corresponding to $w = w_*$ where the Kloosterman sum is much bigger than the product of the square root of the moduli (if $n > 3$). Luckily in this case the remaining piece with moduli $(c_1/q, \ldots, c_{n-1}/q)$ comes with additional savings since the Weyl element $w_*$ imposes certain relations among the $c_j$. That the critical Weyl element is not the long Weyl element, but rather the permutation $1 \leftrightarrow n$ that is relatively “close” to the identity, may also be quite surprising in this context. This analysis is sensitive to $q$ being prime. While the general technique can be applied in a rather broad context (see e.g. $[AB, Ma]$), the estimation in Theorem 3 is particularly designed to the particular setup of $\Gamma_0(q)$, $q$ prime.

Theorem 2 and variations of it have other applications of which we mention here only one, namely a large sieve inequality.

**Theorem 4.** Let $q$ be prime and $(\alpha(m))$ any sequence of complex numbers. Then

$$\sum_{\pi \in \mathcal{F}_1(q)} \left| \sum_{\substack{m \leq x \\ (m,q)=1}} \alpha(m) \lambda_\pi(m) \right|^2 \ll_{I,n,\varepsilon} q^{n-1+\varepsilon} \sum_{\substack{m \leq x \\ (m,q)=1}} |\alpha(m)|^2$$

uniformly in $x \ll q$ for a sufficiently small implied constant (in terms of $I$ and $n$).

This result holds (with literally the same proof) for all $q \in \mathbb{N}$. For comparison, Venkatesh [Ve, Theorem 1] obtained this with $x \ll q^{1/(2n-2)}$. A simple corollary is the following best-possible bound for a second moment of $L$-functions on the critical line:

**Corollary 5.** For $q$ prime and $t \in \mathbb{R}$ we have

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

The author would like to thank Farrell Brumley for encouragement and numerous discussions on the subject.
2. Basic notation

Let $U \subseteq \text{GL}_n$ be the subgroup of unipotent upper triangular matrices. The Haar measure on $U(\mathbb{R})$ is given by $dx = \prod_{1 \leq i < j \leq n} dx_{ij}$. As before let $W$ be the Weyl group of permutation matrices; we identify a permutation matrix $w = (w_{ij}) \in W$ with the permutation $i \mapsto j$ for $w_{ij} = 1$. For $w \in W$ we define

$$U_w = w^{-1} U^T w \cap U$$

and we continue to write $dx$ for the induced measure on the subgroup $U_w(\mathbb{R})$. (Here $^T$ denotes the transpose, so that $U^T$ is the set of unipotent lower triangular matrices.)

As

$$w^{-1}(x_{ij})w = (x_{w^{-1}(i),w^{-1}(j)}),$$

the group $U_w$ has entries at $(i,j)$ with $i < j$ exactly when $w^{-1}(i) > w^{-1}(j)$ (since $U^T$ consists of lower triangular matrices). Equivalently, $U_w$ has entries at

$$(w(i),w(j))$$

for $1 \leq j < i \leq n$ whenever $w(i) < w(j)$.

Let $V \subseteq \text{GL}_n$ be the group of diagonal matrices with entries $\pm 1$.

For $N \in \mathbb{Z}^{n-1}$ we define a character $\theta_N : U(\mathbb{R})/U(\mathbb{Z}) \to S^1$ by

$$(2.3) \quad \theta_N(x) = e(N_{n-1}x_{12} + \ldots + N_1 x_{n-1,n}).$$

For $v \in V$ we write $\theta_N(v^{-1} x v)$ (note that $v^{-1} U v = U$). If $N = (1,\ldots,1)$, we drop it from the notation of the character.

Let $T \subseteq \text{GL}_n$ be the diagonal torus. We embed $y = (y_1,\ldots,y_n) \in \mathbb{G}_m^{n-1}$ into $T$ as

$$(2.4) \quad \iota(y) = \text{diag}(y_{n-1} \cdots y_1, y_2 y_1, y_1, 1).$$

We multiply two elements in $y, y' \in \mathbb{G}_m^{n-1}$ componentwise, written $y \cdot y'$, so that $\iota$ is a homomorphism. We denote the image of $\mathbb{R}_{>0}^{n-1}$ in $T$ by $\hat{T}(\mathbb{R})$. Then $H = U(\mathbb{R})\hat{T}(\mathbb{R})$ is the generalized upper half plane in the sense of [Go, Chapter 1]. We identify $H$ with $\text{GL}_n(\mathbb{R})/O_n(\mathbb{R})\mathbb{Z}^+$ where $\mathbb{Z}^+ \cong \mathbb{R}_{>0}$ is the subgroup of diagonal scalar matrices with positive entries. For $g = xyk\alpha \in U(\mathbb{R})$ with $x \in U(\mathbb{R}), y \in \hat{T}(\mathbb{R}), k \in O_n(\mathbb{R}), \alpha \in \mathbb{Z}^+$, we write $y(g) = \iota^{-1}y \in \mathbb{R}_{>0}^{n-1}$ for $(n-1)$-tuple of Iwasawa $y$-coordinates. In particular, for $g = \text{diag}(y_1,\ldots,y_n)$ with positive $y_j$ we have

$$(2.5) \quad y(g) = \left(\frac{y_{n-1}}{y_n}, \ldots, \frac{y_1}{y_2}\right) \in \mathbb{R}_{>0}^{n-1}.$$ 

For $w \in W, y \in \mathbb{R}_{>0}^{n-1}$ we write

$$y(w) = \left(\frac{y_{n-1} \cdots y_n w(n-j+1)}{y_1 \cdots y_n w(n-j)}\right)_{1 \leq j \leq n-1}$$

for the Iwasawa $y$-coordinates of $w y (n-j-1)w^{-1}$. Explicitly, combining (2.1) with $w^{-1}$ in place of $w$, (2.4) and (2.5), we have

$$(2.6) \quad w y = \left(\frac{y_1 \cdots y_{n-w(n-j+1)}}{y_1 \cdots y_{n-w(n-j)}}\right)_{1 \leq j \leq n-1}.$$ 

For $\alpha \in \mathbb{C}^{n-1}, y \in \mathbb{R}_{>0}^{n-1}$ we write $y^\alpha = y_{n-1}^{\alpha_{n-1}} \cdots y_1^{\alpha_1} \in \mathbb{C}$. Let

$$(2.7) \quad \eta = (\eta_1,\ldots,\eta_{n-1}) = \left(\frac{1}{2} j(n-j)\right)_{1 \leq j \leq n-1}.$$
We define a measure on \( \mathbb{R}_{>0}^{n-1} \) by \( d^*y = y^{-2n} \frac{dy}{y_1} \cdots \frac{dy_{n-1}}{y_{n-1}} \) and correspondingly an inner product by

\[
\langle f, g \rangle = \int_{\mathbb{R}_{>0}^{n-1}} f(y) \overline{g}(y) d^*y.
\]

We denote the push forward of \( d^*y \) to \( \tilde{T}(\mathbb{R}) \) by \( \iota \) also by \( d^*y \). Then \( dx \, d^*y \) is a left \( \text{GL}_n(\mathbb{R}) \) invariant measure on \( \mathcal{H} \).

We define a different embedding of \( \mathbb{R}_{>0}^{n-1} \) into \( T(\mathbb{R}) \) by

\[
c = (c_1, \ldots, c_{n-1}) \mapsto c^* = \text{diag}(1/c_{n-1}, c_{n-1}/c_{n-2}, \ldots, c_2/c_1, c_1).
\]

From (2.5), it is useful to observe that

\[
y(c^*) = \left( \frac{c_{j-1}c_{j+1}}{c_j^2} \right) \quad 1 \leq j \leq n-1
\]

where \( c_0 = c_n = 1 \), and a simple computation shows

\[
y(c^*)^n = (c_1 \cdots c_{n-1})^{-1}.
\]

### 3. Auxiliary results

As the Iwasawa decomposition (with the compact group on the right) is the Gram-Schmidt orthogonalization of rows starting with the last row, we can compute \( y(g) \) explicitly. For \( 1 \leq j \leq n \) let \( \Delta_j = \Delta_j(g) \) be the volume of the parallelepiped spanned by last \( j \) rows of \( g \). Then

\[
g \equiv \begin{pmatrix}
\Delta_n/\Delta_{n-1} & * & \cdots & * \\
\Delta_{n-1}/\Delta_{n-2} & \cdots & * \\
& \ddots & \ddots & \\
& & \Delta_1 & \\
\end{pmatrix} \quad \text{(mod } O_n(\mathbb{R}))
\]

so that by (2.5) we have

\[
y(g) = \left( \frac{\Delta_{j+1}(g)\Delta_{j-1}(g)}{\Delta_j(g)^2} \right) \quad 1 \leq j \leq n-1
\]

with the convention \( \Delta_0(g) = 1 \). By [FP, Corollary 4.2 and p. 11] (or by hand) we confirm the inversion formula for the following \( (n-1) \)-by-\( (n-1) \) tridiagonal Toeplitz matrix

\[
\left( \begin{array}{cccc}
-2 & 1 & \cdots & 1 \\
1 & -2 & \ddots & \ddots \\
\ddots & \ddots & \ddots & 1 \\
1 & \cdots & 1 & -2
\end{array} \right)^{-1} = \left( -s(i,j) \right)_{ij}, \quad s(i,j) = \frac{1}{n} \begin{cases} i(n-j), & i \leq j, \\
j(n-i), & i > j.
\end{cases}
\]

Therefore, given \( y(g) = (Y_1, \ldots, Y_{n-1}) \) and \( \Delta_n(g) = |\det(g)| \) we can solve (3.1) explicitly for \( \Delta_1, \ldots, \Delta_{n-1} > 0 \) getting

\[
\Delta_j(g) = |\det(g)|^{j/n} \prod_{i=1}^{n-1} Y_i^{-s(i,j)}.
\]

Our first lemma will be used to bound the moduli \( c \) on the arithmetic side of the Kuznetsov formula.
Lemma 1. Let $w \in W$, $x \in U_w(\mathbb{R})$, $y, c, B \in \mathbb{R}_{>0}^{n-1}$. Write $y(\iota(B)c^*wx_\iota(y)) = Y \in \mathbb{R}_{>0}^{n-1}$ and $A = \iota(B)c^*$. Then

$$c_j \ll_{y,Y} \prod_{i=1}^{n-1} B_i^{s(i,j)} \quad \text{and} \quad 1 \leq \Delta_j(wx) \ll_{y,Y} \prod_{i=1}^{n-1} y(A)_i^{s(i,j)}$$

for $1 \leq j \leq n-1$.

Proof. We have

$$\Delta_j(\iota(B)c^*wx_\iota(y)) = \Delta_j(wx_\iota(y)) c_j \prod_{i=1}^{j} B_i \cdots B_{i-1} \quad (3.4)$$

since the diagonal matrix $\iota(B)c^*$ multiplies the rows of $wx_\iota(y)$ by the corresponding diagonal entries. Clearly $\Delta_j(wx) \geq 1$ since one of the minors is always 1, and clearly

$$|\det(\iota(B)c^*wx_\iota(y))| = |\det(\iota(B)c^*y)| \asymp y B_1^{n-1} \cdots B_{n-2} B_{n-1}. \quad (3.5)$$

From (3.4), and from (3.3) in combination with (3.5), we therefore obtain

$$c_j \leq c_j \Delta_j(wx) \asymp_y c_j \Delta_j(wx_\iota(y)) = \prod_{i=1}^{j} B_i^{-j-i} \quad (3.6)$$

This shows the first statement of the lemma, and the proof of the second is completed by observing that (2.8) and (3.2) imply

$$\frac{1}{c_j} = \prod_{i=1}^{n-1} y(c^*)_i^{s(i,j)}$$

for $1 \leq j \leq n-1$, so that (3.6) implies $\Delta_j(wx) \asymp_y Y \prod_{i=1}^{n-1} (y(c^*)_i B_i)^{s(i,j)}$ as desired. \hfill $\square$

We shall see in a moment that the only Weyl elements contributing to the Kuznetsov formula are of the form

$$w = \begin{pmatrix} \cdot & \cdot & \cdots & \cdot & I_{d_1} \\ \cdot & \cdots & \cdots & \cdot & I_{d_2} \\ \cdot & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdots & \cdots & \cdots & \cdots \\ I_{d_r} & \cdots & \cdots & \cdots & \cdots \\ \end{pmatrix} \quad (3.7)$$

with identity matrices $I_{d_j}$ of dimension $d_j$ (i.e. $d_1 + \cdots + d_r = n$), so without loss of generality we restrict our attention to such matrices. The following technical result computes the Jacobi determinant for a certain change of variables.

Lemma 2. Let $N \in \mathbb{N}^{n-1}$, $w \in W$ of the form (3.7). For $x \in U_w(\mathbb{R})$ define $x' = \iota(N)x_\iota(N)^{-1} \in U_w(\mathbb{R})$. Then

$$\frac{dx'}{dx} = (\iota(N)_N)^n$$

where the left hand side denotes the Jacobi determinant $\det Dx'(x)$. 


Proof. Since \(i(N)(x_{ij})i(N)^{-1} = (x_{ij}N_1 \cdots N_{n-i}(N_1 \cdots N_{n-j})^{-1})_{ij}\) and (2.6), we have to show

\[
(3.8) \quad \prod_{1 \leq j < i \leq n} N_{n-w(j)+1} \cdots N_{n-w(i)} = \prod_{j=1}^{n-1} \left( \frac{N_j}{N_1 \cdots N_{n-w(n-j)}} \right)^{\eta_j}
\]

for an arbitrary \(w\) as in (3.7). We use induction on \(r\) and write \(w' = \left( I_d \begin{array}{c} w \end{array} \right)\), so that \(n + d - w'(j) = n - w(j)\) for all \(1 \leq j \leq n\). We call \(L(w)\) the left hand side of (3.8) and \(R(w)\) the right hand side. We consider first the quotient \(L(w')/L(w)\). The pairs \(1 \leq j < i \leq n\) cancel, and for \(i > n\) only \(j \leq n\) satisfy the summation condition \(w'(i) < w'(j)\). We conclude

\[
(3.9) \quad \frac{L(w')}{L(w)} = \prod_{j=1}^{n} \prod_{i=n+1}^{n+d} N_{n+d-w'(j)+1} \cdots N_{n+d-w'(i)} = \prod_{j=1}^{n-1} N_j^{d_j} \prod_{j=n}^{n+d-1} N_j^{n(n+d-j)}.
\]

On the other hand,

\[
R(w) = \prod_{j=1}^{n-1} N_j^{\eta_j} \prod_{i=1}^{n} (N_1 \cdots N_{n-w(i)})^{\eta_{n-i+1} - \eta_{n-i}} = \prod_{j=1}^{n-1} N_j^{j(n-j)} \times \prod_{i=1}^{n} (N_1 \cdots N_{n-w(i)})^{2i-1 - \frac{n^2+n-1}{2}}
\]

so \(R(w')/R(w)\) equals

\[
= \prod_{j=1}^{n-1} N_j^{d_j} \prod_{j=n}^{n+d-1} N_j^{n(n+d-j) - \frac{n^2+n-1}{2}} \prod_{i=n+1}^{n+d} (N_1 \cdots N_{n-w'(i)})^{2i-1 - \frac{n^2+n-1}{2}}
\]

which is easily seen to equal the right hand side of (3.9). Since trivially \(L(I_d) = R(I_d) = 1\), the induction is complete.

Lemma 3. Let \(B \in \mathbb{R}_{>0}^{n-1}\), \(w = w_* \in W\) as in (1.5). Then

\[
\text{vol}\{x \in U_w(\mathbb{R}) \mid \Delta_j(wx) \leq B_j, 1 \leq j \leq n - 1\} \ll_{\varepsilon} (B_1 \cdots B_{n-1})^{1+\varepsilon}
\]

for any \(\varepsilon > 0\).

Proof. We can assume without loss of generality that \(B_j \geq 1\), otherwise the volume is 0 as seen in the proof of Lemma 1. For \(x \in U_w(\mathbb{R})\) we have

\[
wx = \begin{pmatrix}
1 \\
I_{n-2} \\
\vdots \\
x_{12} & \cdots & x_{n-1,n} \\
x_{1n}
\end{pmatrix},
\]

so that by considering the lower right minors we obtain in particular the inequalities

\[
|x_{1n}| \leq B_1, \quad -x_{1,n} + \sum_{i=n+1-j}^{n-1} x_{1i,n} \leq B_j, \quad j = 2, \ldots, n - 1,
\]
and we also have $|x_{ij}| \leq b := 1 + \max(B_1, \ldots, B_{n-1})$. If $I \subseteq \mathbb{R}$ is any interval of length $|I| \geq 1$, then

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

Thus if $|x_{1n}| \leq B_1$ is fixed, the volume of $(x_{1, n-1}, x_{n-1, n})$ is $O(B_2 \log b)$, and if these are fixed, the volume of $(x_{1, n-2}, x_{n-2, n})$ is $O(B_3 \log b)$, etc. Inductively we obtain the desired bound.

Most likely the statement holds for all $w$, but the proof is particularly simple for $w_*$ which is all we need.

4. Kloosterman sums

Properties of Kloosterman sums for $\text{SL}_n(\mathbb{Z})$ have been obtained and summarized in [Fr]. They generalize in an obvious way to the congruence subgroup $\Gamma_0(q)$. The Bruhat decomposition gives $\text{GL}_n(\mathbb{Q}) = \bigcup_{w \in W} G_w(\mathbb{Q})$ with $G_w := UTwU_w$ as a disjoint union. Let $N, M, c \in \mathbb{Z}^{n-1}$, $w \in W$, $v \in V$. Then provided that

$$\theta_M(c^* w x w^{-1} (c^*)^{-1}) = \theta_N^v(x)$$

for all $x \in w^{-1}U(\mathbb{Q})w \cap U(\mathbb{Q})$ [this set is a “complement” of $U_w$ in $U$], the Kloosterman sum

$$S^w_{q, w}(M, N, c) = \sum_{w^* x w y \in U(\mathbb{Z}) \setminus G_w(\mathbb{Q}) \cap \Gamma_0(q)/U_w(\mathbb{Z})} \theta_M(x) \theta_N^v(y)$$

is well-defined, see [Fr, Proposition 1.3]. If (4.1) is not met, we define $S^w_{q, w}(M, N, c) = 0$. If $v = \text{id}$, we drop it from the notation. By [Fr, p. 175], the Kloosterman sum is non-zero only if $w$ is of the form (3.7). If $\gamma = x_1 c^* w x_2 \in \Gamma_0(q)$ is a matrix occurring in the sum on the right hand side of (4.2), then any minor of $\gamma$, and hence of $c^* w$, obtained by deleting at least the first row and the last column is divisible by $q$. Hence if $w$ is of the form (3.7), then the summation condition in (4.2) can only be met if

$$q \mid c_1, \quad q \mid c_2, \quad \ldots, \quad q \mid c_{n-d_1}.$$  

Observing that

$$\theta_M(x) = \theta(i(M)x t(M)^{-1})$$

and recalling (2.8), we see that (4.1) is equivalent to

$$M_{n-i} \frac{c_{n-i} + 1}{c_{n-i}} = \frac{v_{w(i)+1}}{v_{w(i)}} N_{n-w(i)}$$

for all $1 \leq i \leq n-1$ satisfying $w(i) + 1 = w(i+1)$ with the above convention $c_0 = c_n = 1$ and $v = \text{diag}(v_1, \ldots, v_n)$. If $w$ is of the form (3.7), these are precisely the $i \notin \{d_1, d_1 + d_2, \ldots, d_1 + d_2 + \ldots + d_{r-1}\}$. If $w = \text{id}$, then $x c^* w y = x c^* y$ can only be in $\Gamma_0(q)$ if $c_1 = \ldots = c_{n-1} = 1$, in which the case we conclude from (4.5) that $M_j = \pm N_j$.

Kloosterman sums for $\text{SL}_n(\mathbb{Z})$ enjoy certain multiplicativity properties in the moduli, cf. [Fr, Proposition 2.4]. We state only one particular case. Let $q$ be prime, suppose that
(c_1 \cdots c_{n-1}, q) = 1 \text{ and write } qc = (qc_1, \ldots, qc_{n-1}). \text{ Suppose that } w(1) = n \text{ and } w(n) = 1. \text{ Then }

\begin{align*}
S^w_{q,w}(M, N, qc) &= S^w_{q,w}(M, N', (q, \ldots, q))S^w_{1,w}(M, (\bar{q}N_1, N_2, \ldots, N_{n-2}, \bar{q}N_{n-1}), c) \\
\end{align*}

with

\[ N'_n \equiv N_n - i c_{n-w(i)} c_{n-w(i+1)} + 1 (c_{n-w(i)} + 1) (c_{n-w(i+1)} + 1) (\bmod q). \]

By [DR, Theorem 0.3(i)] we have the trivial bound

\[ |S^w_{q,w}(M, N, c)| \leq |U(Z) \setminus G_w(Q) \cap SL_n(Z)/U_w(Z)| \ll (c_1 \cdots c_{n-1})^{1+\varepsilon}. \]

We now give the proof of Theorem 3 from the introduction, which is the first non-trivial bound for a $GL_n$ Kloosterman sum other than a hyper-Kloosterman sum. For $n = 3$, the statement is essentially contained in [BBM, Lemma 6(c)]. We wish to compute $S^w_{q,w}(M, N, (q, \ldots, q)) = 0$ where $M, N \in Z^{n-1}$ have entries coprime to $q$.

As mentioned before, we can assume that $w$ is of the form (3.7), otherwise the Kloosterman sum vanishes by definition. Next assume that $d_1 > 1$ in (3.7). Applying (4.5) with $i = 1$ we obtain $M_{n-1} = \pm N_{d_1-1} q$, a contradiction. In the same way we exclude the case $d_r > 1$. For $w$ of the form (3.7) with $d_1 = d_r = 1$ and $c^* = \text{diag}(1/q, 1, \ldots, 1, q)$, we recall the definition (4.2) and consider $\gamma = xc^*wy \in G_w(Q) \cap \Gamma_0(q)$ with uniquely determined $x \in U(Z) \setminus U(Q)$, $y \in U_w(Q)/U_w(Z)$. A system of representatives for $U(Z) \setminus U(Q)$ consists of matrices with rational entries in $[0, 1)$ above the diagonal, and similarly we choose a system of representatives of $U_w(Q)/U_w(Z)$ where all relevant entries are restricted to $[0, 1)$. We now determine those representatives $x, y$ that satisfy $xc^*wy \in \Gamma_0(q)$. We have

\[ x = \begin{pmatrix}
1 & * & \cdots & * & x_{1,n} \\
1 & * & \cdots & * & x_{2,n} \\
\vdots \\
1 & * & x_{n-2,n} \\
1 & x_{n-1,n} & 1
\end{pmatrix}, \quad c^*wy = \begin{pmatrix}
I_{d_2} & 1/q \\
I_{d_3} & * \\
\vdots & \vdots \\
I_{d_{r-1}} & * & \cdots & * & y_{w(n-1),n} \\
q & y_{12} & \cdots & * & qy_{1n}
\end{pmatrix}. \]

Since $\gamma \in \Gamma_0(q)$, we must have $y_{12}, \ldots, y_{1,n-1} \in Z$, hence by our choice of representatives $y_{12} = \cdots = y_{1,n-1} = 0$.

Next we consider the $(n-1)$-st row

\[ (q x_{n-1,n}, \cdots, * \cdots *, y_{w(n-1),n} + qy_{1n} x_{n-1,n}) \in Z^n \]

of $\gamma$, where the stars are the same as the stars in the $(n-1)$-st row of $c^*wy$. We conclude that all star-ed entries in the $(n-1)$-st row of $c^*wy$ must be integral, hence 0. We continue with the $(n-2)$-nd row of $\gamma$. By the same argument we first have $qx_{n-2,n} \in Z$ and then also star-ed entry in $(n-2)$-nd row of $x$ is integral (hence 0) as well as all star-ed entries in the $(n-2)$-nd row of $c^*wy$. Continuing in this way, all star-ed entries must vanish. In
other words, \( x \text{ diag}(1/q, 1, \ldots, 1, q) wy \in \Gamma_0(q) \) with \( x \in U(Q), \ y \in U_w(Q) \) and all relevant entries in \([0, 1)\) implies

\[
\begin{pmatrix}
1 & \cdots & x_1/q \\
\vdots & \ddots & \vdots \\
x_{n-1}/q & \cdots & 1
\end{pmatrix}, \quad
y = \begin{pmatrix}
1 & \cdots & y_1/q \\
\vdots & \ddots & \vdots \\
y_{n-1}/q & \cdots & 1
\end{pmatrix}
\]

with \( x_i, y_i \in \{0, \ldots, q-1\} \). For these \( x, y \) we compute

\[
xe^*wy = \begin{pmatrix}
x_1 & \cdots & x_1/q \\
x_2 & \cdots & \frac{(x_1y_1 + 1)/q}{I_2} \\
\vdots & \ddots & \vdots \\
x_{n-1} & \cdots & \frac{(x_{n-1}y_1 + y_w(n-1))/q}{y_n}
\end{pmatrix}.
\]

Obviously, this matrix is in \( \Gamma_0(q) \) if and only if the \( n-1 \) congruences

\[
x_1y_1 + 1 \equiv 0 \pmod{q}, \quad x_iy_1 + y_w(i) \equiv 0 \pmod{q}, \quad 2 \leq i \leq n-1
\]

are satisfied. This can be solved easily, and we obtain the explicit expression

\[
S_{q,w}^\prime(M, N, (q, \ldots, q)) = \sum_{x_1, \ldots, x_{n-1} \pmod{q}} e\left(\frac{M_1x_{n-1} + N_1\overline{x_1}x_{w^{-1}(n-1)}}{q}\right).
\]

If \((M_1N_1, q) = 1\), the sum vanishes unless \( n - 1 = w^{-1}(n - 1) \). The latter case happens for \( w \) of the form \((3.7)\) with \( d_1 = d_r = 1 \), if and only if \( w = w_* \) and then the Kloosterman sum equals \( q^{n-2} \).

5. AUTOMORPHIC FORMS AND WHITTAKER FUNCTIONS

We denote by \( \{\varpi\} \) an orthonormal basis of right \( O_n(R)^\dagger \)-invariant automorphic forms for the group \( \Gamma_0(q) \), cuspidal or Eisenstein series. The space \( L^2(\Gamma_0(q)\backslash \mathcal{H}) \) is equipped with the standard inner product \( \langle f, g \rangle = \int_{\Gamma_0(q)\backslash \mathcal{H}} f(xy)\overline{g(xy)}dx \). We denote by \( \int Q d\varpi \) a combined sum/integral over the complete spectrum of \( L^2(\Gamma_0(q)\backslash \mathcal{H}) \). The relevant spectral decomposition is a special case of Langlands’ general theory, see e.g. [Ar] for a convenient summary in adelic language. All \( \varpi \) belong to representations of level \( q' \mid q \) (cf. [JPSS, Théorème]) and we assume that \( \{\varpi\} \) contains all cuspidal newvectors of level \( q' \mid q \). The underlying representation is denoted by \( \pi \), so \( \varpi \in V_\pi \). For notational simplicity let us denote the local archimedean Langlands parameter \( \mu_\pi(\infty) \) simply by \( \mu = (\mu_1, \ldots, \mu_n) \); it satisfies

\[
(5.1) \quad \mu_1 + \cdots + \mu_n = 0, \quad \{\mu_1, \ldots, \mu_n\} = \{-\bar{\mu}_1, \ldots, -\bar{\mu}_n\}.
\]

For a (not necessarily cuspidal) automorphic form \( \varpi \) and \( N \in \mathbb{N}^{n-1} \) we define its \( N \)-th Fourier coefficient \( A_{\varpi}(N) \) by

\[
(5.2) \quad \int_{U(Z)\backslash U(R)} \varpi(xy)\theta_N(-x)dx = \frac{A_{\varpi}(N)}{N^n}W_\mu(N \cdot y(y))
\]

where \( y \in \mathcal{T}(R) \) and \( W_\mu : \mathbb{R}_\geq 0^{-1} \to \mathbb{C} \) is the standard (spherical) Whittaker function, cf. e.g. [St, Section 2].
If \( \varpi \) is a cuspidal newform and \((m,q) = 1\), the \((m,1,\ldots,1)\)-th Fourier coefficient is proportional to the \(m\)-th Hecke eigenvalue \(\lambda_\varpi(m)\) (or \(\lambda_\bar{\varpi}(m)\) depending on normalization), and by Rankin-Selberg theory we obtain
\[
|A_\varpi((m,1,\ldots,1))|^2 \propto_{\mu} \left| \frac{\lambda_\varpi(m)}{[\text{SL}_n(\mathbb{Z}) : \Gamma_0(q)]L(1,\pi,\text{Ad})}\right| \gg_{\mu} |\lambda_\varpi(m)|^2 q^{-(n-1) - \varepsilon}
\]
if \( \varpi \) is \(L^2\)-normalized, cf. e.g. [Ve, Proposition 1]. Here we used the upper bound [Li, Theorem 2] for the \(L\)-value (the residue of the Rankin-Selberg \(L\)-function) on the edge of the critical strip.

Comparing coefficients, we obtain the lemma.

**Lemma 4.** For a prime \(p \nmid q\) and \(\nu > n\) we have
\[
\max_{0 \leq j \leq n-1} |\lambda_\varpi(p^{\nu-j})| \geq (2p^{\sigma_\pi(\nu)})^{1-n} p^{\nu\sigma_\pi(p)}
\]

**Proof.** The following argument is taken from [LS, Lemma 3]. We have an identity of power series
\[
\sum_{\nu=0}^{\infty} \lambda_\varpi(p^\nu) x^\nu = \prod_{j=1}^{n} \left(1 - p^{\mu_\pi(p^j)} x\right)^{-1}.
\]
Without loss of generality let \(\mu_\pi(p,1)\) have the largest real part, i.e. \(\Re \mu_\pi(p,1) = \sigma_\pi(p)\). Then
\[
\sum_{\nu=0}^{\infty} p^{\nu\mu_\pi(p,1)} x^\nu = \prod_{j=2}^{n} \left(1 - p^{\mu_\pi(p^j)} x\right) \sum_{\nu=0}^{\infty} \lambda_\varpi(p^\nu) x^\nu.
\]
Comparing coefficients, we obtain the lemma. \(\square\)

We need an archimedean analogue of this result, which is a bit more technical. Roughly speaking, the growth of \(W_\mu\) near the origin should capture the size of \(\sigma_\pi(\infty)\) in the same way as the growth of \(\lambda_\varpi(p^\nu)\) captures the size of \(\sigma_\pi(p)\), but this is harder to see as the Mellin transform of \(W_\mu\) is not perfectly understood and the location of poles is subtle. We start by summarizing some properties. As in [St, (3.1), (3.2)] we consider the re-normalized Whittaker function
\[
W^*_\mu(y) = \pi^{(n-1)n(n+1)/12} y^{-n/2} W_{2\mu} \left(\sqrt{y_1/\pi}, \ldots, \sqrt{y_{n-1}/\pi}\right).
\]
The corresponding Mellin transform \(\widehat{W}^*_\mu(s) = \int_{\mathbb{R}_{>0}^{n-1}} W^*_\mu(y) y^s \frac{dy_1}{y_1} \cdots \frac{dy_{n-1}}{y_{n-1}}\) is meromorphic in \(\mu\) and \(s \in \mathbb{C}^{n-1} [FG]\). Explicitly, we have [St, (3.7)]
\[
\widehat{W}^*_\mu(s_1) = \Gamma(s_1 + \mu_1) \Gamma(s_1 + \mu_2), \quad n = 2,
\]
\[
\widehat{W}^*_\mu(s_1, s_2) = \frac{\Gamma(s_1 + \mu_1) \Gamma(s_1 + \mu_2) \Gamma(s_1 + \mu_3) \Gamma(s_2 - \mu_1) \Gamma(s_2 - \mu_2) \Gamma(s_2 + \mu_2)}{\Gamma(s_1 + s_2)}, \quad n = 3,
\]
but in general there do not seem to be such simple formulae. For \(\Re s_2, \ldots, \Re s_{n-1}\) sufficiently large and \(\Re s_1 > \sigma_\pi(\infty)\), the function \(\widehat{W}^*_\mu(s)\) is holomorphic by [St, Theorem 3.1]. If in addition \(\mu_1, \ldots, \mu_n\) are pairwise distinct, then \(\widehat{W}^*_\mu(s)\) has simple poles at \(s_1 = -\mu_j, 1 \leq j \leq n\), with residue
\[
\widehat{W}^*_\mu(s_j) \left(\prod_{1 \leq k \leq n, k \neq j} \Gamma(\mu_k - \mu_j)\right)
\]
Lemma 5. Let \( \alpha \geq 0, c_0, c_1, c_2 > 0, \beta \in \mathbb{C} \). Let \( I = [a, b] \subseteq (0, 1) \) be an interval with \( (1 + c_0)a \leq b \leq 2a \) and \( w : I \to \mathbb{C} \) a smooth function satisfying

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

for \( y \in I \). Then there exist constants \( c'_0, c'_1, c'_2 > 0 \) depending only on \( c_0, c_1, c_2, \beta \) (but not on \( a, b \)) and an interval \( I' = [a', b'] \subseteq I \) with \( (b' - a') \geq c'_0 (b - a) \) such that

\[
|w(y)| \geq c'_1 y^{-\alpha}, \quad |w'(y)| \leq c'_2 \|w'\| y^{-1}
\]

for \( y \in I' \).

**Proof.** Let \( \tilde{w}(y) = w(y)y^{-\beta} \), so that

\[
y^{1+\beta} \tilde{w}'(y) = -D_{\beta}w(y), \quad y^{1+\beta} \tilde{w}''(y) = -\partial_y (D_{\beta}w)(y) + \frac{1+\beta}{y} D_{\beta}w(y).
\]

Then (5.6) implies

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

for \( \tilde{c}_2 = 2^{1+|\Re \beta|} c_2 + |1 + \beta| \) and \( \tilde{\alpha} = \alpha + \Re \beta \). Let \( y_0 = \max_{y \in I} |\tilde{w}'(y)| \). Changing \( \tilde{w} \) by a fourth root of unity if necessary, we can assume that

\[
\Re \tilde{w}'(y_0) \geq \frac{1}{\sqrt{2}} \max \left( c_1 y_0^{-\tilde{\alpha}-1}, \|\tilde{w}'\| \right).
\]
The condition $|\tilde{w}''(y)| \leq \tilde{c}_2 \|\tilde{w}'\|^{-1}$ implies that the slightly weaker inequality
\[
\Re \tilde{w}'(y) \geq \frac{1}{2\sqrt{2}} \max \left( c_1 y_0^{-\tilde{a} - 1}, \|\tilde{w}'\| \right) \times \Re \tilde{w}'(y_0)
\]
holds on some non-empty sub-interval $I_0 = [a_0, b_0] \subseteq I$ containing $y_0$, where $(b_0 - a_0) = (b - a)/\sqrt{8\tilde{c}_2}$. Distinguishing the cases $\Re \tilde{w}'(a_0) > -c_3 y_0 \Re \tilde{w}'(y_0)$ and $\Re \tilde{w}'(a_0) \leq -c_3 y_0 \Re \tilde{w}'(y_0)$ for $c_3 = (b_0 - a_0)/\sqrt{8\tilde{c}_2}$, we confirm in both cases
\[
\max(\Re \tilde{w}(a_0), |\Re \tilde{w}'(b_0)|) \geq \min \left( c_3 y_0 \Re \tilde{w}'(y_0), \left( -c_3 y_0 + \frac{b_0 - a_0}{2\sqrt{2}} \right) \Re \tilde{w}'(y_0) \right) = c_3 y_0 \Re \tilde{w}'(y_0).
\]
Thus there exists a non-empty subinterval $I' = [a', b'] \subseteq I_0$ of length $\gg (b_0 - a_0)$ such that $|\tilde{w}(y)| \geq \frac{1}{4\tilde{c}_2} c_3 c_1 y^{-\tilde{a}}$ on $I'$. Changing back to $w$, we obtain (5.7).

We are now prepared for the following analogue of Lemma 4. For a function $E$ on $\mathbb{R}^{n-1}_{\geq 0}$ and $X \in \mathbb{R}^{n-1}_{> 0}$ define
\[
E(X)(y_1, \ldots, y_{n-1}) = E(X_1 y_1, \ldots, X_{n-1} y_{n-1}).
\]

**Lemma 6.** Assume that $\mu$ varies in some compact set $\Omega$, and let $Z \geq 1$. There exist $r \in \mathbb{N}$ and a compact set $S \subseteq \mathbb{R}^{n-1}_{\geq 0}$ both depending only on $\Omega$ (not on $Z$) and a finite collection of (measurable) functions $E_1, \ldots, E_r : \mathbb{R}^{n-1}_{\geq 0} \to \mathbb{R}$ depending on $\Omega$ and $Z$ that are uniformly bounded (independent of $Z$) and supported in $S$ such that
\[
\sum_{j=1}^r |\langle E_j^{(Z,1,\ldots,1)}, W_\mu \rangle|^2 \gg \Omega^Z Z^{2n+2\sigma_\mu(\infty)}
\]
for $\mu \in \Omega$ and $\eta$ as in (2.7).

**Proof.** For $Z \ll 1$ this is [BBM, Lemma 1]. For convenience we repeat the short argument in a slightly modified fashion that we will need later. There exists $Z_0 > 0$ (depending only on $\Omega$) such that for each $\mu \in \Omega$ we can choose an open set $S_\mu \subseteq \mathbb{R}^{n-1}_{> 0}$ such that $|\Re W_\mu(y)| > 2Z_0$ for all $y \in S_\mu$ or $|\Im W_\mu(y)| > 2Z_0$ for all $y \in S_\mu$. Next choose open neighbourhoods $U_\mu$ about $\mu$ such that $|\Re W_\mu(y)| > Z_0$ for all $y \in S_\mu$ and all $\mu^* \in U_\mu$ or $|\Im W_\mu(y)| > Z_0$ for all $y \in S_\mu$ and all $\mu^* \in U_\mu$. By compactness we pick a finite collection of such neighbourhoods $U_{\mu_1}, \ldots, U_{\mu_i}$ covering $\Omega$, and define the corresponding $E_j$ to be real-valued functions with support on $S_{\mu_j}$ and non-vanishing on the interior $S_{\mu_j}$. Now suppose that $Z$ is sufficiently large (in terms of $\Omega$). We try to mimic the proof of Lemma 4. Assume (without loss of generality by (5.1) and Weyl group symmetry) that $\Re(-\mu_1) = \sigma_\mu(\infty)$, and with the notation as above let
\[
\tilde{w}_\mu(s) := \frac{\hat{W}_\mu^*(s)}{s + \mu_1} = \hat{W}_\mu^*(s) \prod_{j=2}^n (s + \mu_j).
\]
Taking inverse Mellin transforms, we obtain
\[
\tilde{w}_\mu(y) = D_{\mu_2} \cdots D_{\mu_n} W_\mu^*(y)
\]
where the differential operators are applied to the first variable $y_1$. On the other hand, by Mellin inversion and (5.5) we have the asymptotic expansion
\[
y_1 \partial_{y_1} \tilde{w}_\mu(y) = \mu_1 y_1^{\Re_{01} + 1/2} W_\mu^*(y, \ldots, y_{n-1}) + O_{y_2, \ldots, y_{n-1}, \mu} \left( y_1^{\Re_{01} + 1/2} \right)
\]
for \( y_1 \to 0 \) and \( j \in \{0,1\} \) where

\[
W^{*\ast}_\mu(y_2, \ldots, y_{n-1}) = W^{*\ast}_\mu(y_2, \ldots, y_{n-1}) \prod_{j=2}^{n-1} y_j^{\mu_1} \prod_{k=2}^{n} (1 + \mu_k - \mu_1).
\]

Whenever \(|W^{*\ast}_\mu(y_2, \ldots, y_{n-1})| \geq Z^{-1/2} \), say, and \( y_2, \ldots, y_n \gg 1 \) (with implied constants depending only on \( \Omega \)), we can apply repeatedly Lemma 5 to \( W_\mu(y) \) with \( \beta = \mu_2, \ldots, \mu_n \) to obtain two constants \( 1/2 < \gamma_1 < \gamma_2 < 1 \) with

\[
|W^{*}_\mu(y)| \gg y_1^{-\sigma_\pi(\infty)}|W^{*\ast}_\mu(y_2, \ldots, y_{n-1})|
\]

for \( y_1 \in \left[\gamma_1/Z^2, \gamma_2/Z^2\right] \). By the same argument as in the beginning of the proof, we can now choose a finite collection of functions \( E^{*\ast}_j : \mathbb{R}^{n-2} \to \mathbb{C} \) depending on \( \Omega \) (but not on \( Z \), provided that \( Z \) is sufficiently large in terms of \( \Omega \)) such that \( \sum_j |(E^{*\ast}_j, W^{*\ast}_\mu)|^2 \gg 1 \) for \( \mu \in \Omega \), the inner product being restricted to the last \( n-2 \) coordinates. Next define

\[
E^{*}_j(y_1, \ldots, y_n) = \delta_{\gamma_1 \leq y_1 \leq \gamma_2} E^{*\ast}_j(y_2, \ldots, y_{n-1}),
\]

so that

\[
\sum_j \left| \int_{\mathbb{R}^{n-1}_+} E^{*}_j(Z^2 y_1, y_2, \ldots, y_{n-1}) W^{*}_\mu(y) \frac{dy_1}{y_1} \cdots \frac{dy_{n-1}}{y_{n-1}} \right|^2 \gg Z^{4\sigma_\pi(\infty)}.
\]

Finally changing variables \( y_j \leftarrow y_j^{1/2}/\pi \) as in (5.4), we obtain

\[
Z^{4\sigma_\pi(\infty)} \ll \sum_j \left| \int_{\mathbb{R}^{n-1}_+} y^{-\eta} E^{*}_j(Z^2 \pi^2 y_1^2, \pi^2 y_2^2, \ldots, \pi^2 y_{n-1}^2) W^{*}_\mu(y) \frac{dy_1}{y_1} \cdots \frac{dy_{n-1}}{y_{n-1}} \right|^2 = Z^{-2\eta} \sum_j |(E^{(Z,1,\ldots,1)}_j, W^{*\ast}_\mu)|^2
\]

upon defining \( E_j(y_1, \ldots, y_n) = y^\eta E^{*}_j(\pi^2 y_1^2, \pi^2 y_2^2, \ldots, \pi^2 y_{n-1}^2) \). Re-normalizing \( \mu \) and \( \sigma_\pi(\infty) \) by division by 2, we obtain the lemma. \( \square \)

6. Poincaré series and the Kuznetsov formula

Let \( E \) be a fixed compactly supported (measurable) function on \( \mathbb{R}^{n-1}_+ \), \( X \in \mathbb{R}^{n-1}_+ \) a “parameter” and define the right \( O_n(\mathbb{R})Z^+ \) invariant function \( F^{(x)} : \text{GL}(n) \to \mathbb{C} \) by

\[
F^{(x)}(xy_\kappa) = \theta(x) E^{(x)}(y(y))
\]

for \( x \in U(\mathbb{R}), y \in \tilde{T}(\mathbb{R}), \kappa \in O_n(\mathbb{R}), \alpha \in Z^+ \) and \( \theta = \theta(1,\ldots,1) \) as in (2.3), \( E^{(x)} \) as in (5.8). For \( N \in \mathbb{N}^{n-1} \) we consider the Poincaré series

\[
P^{(x)}_N(xy) = \sum_{\gamma \in U(Z) \backslash U_\alpha(q)} F^{(x)}(\iota(N)\gamma xy).
\]

Note that \( F^{(x)}(\iota(N)xy) = \theta_N(x) E(X \cdot N \cdot y(y)), \) cf. (4.4). Let \( N, M \in \mathbb{N}^{n-1} \). By [Fr, Theorem A] (with \( \rho = \text{triv}, \nu_1 = \ldots = \nu_{n-1} = 0 \)) we have

\[
\int_{U(Z) \backslash U(\mathbb{R})} P^{(x)}_M(xy) \theta_N(-x) dx = \sum_{w \in W} \sum_{v \in V} \sum_{c \in \mathbb{N}^{n-1}} S^w_{q,w}(M, N, c) \int_{U_v(\mathbb{R})} F^{(x)}(c^w w xy) \theta_N(-x) dx.
\]
For fixed $y$ and fixed compact support of $E$, it follows from the two bounds in Lemma 1 with $M$ in place of $B$ that the $c$-sum runs over a finite set (depending on $M$, $y$ and the support of $F$), and the $U_w(T)\langle \xi \rangle$-integral runs over a compact domain (again depending on $M$, $y$ and the support of $F$). In particular the right hand side is absolutely convergent (and the assumption $R_N > 2/n$ in [Fr, Theorem A] can be dropped; Friedberg works more generally with bounded $E$ rather than compactly supported $E$). Without loss of generality we can assume that $w$ is of the form (3.7).

Now let $\varpi$ be a not necessarily cuspidal automorphic form occurring in the spectrum of $L^2(\Gamma_0(q)\setminus H)$. By unfolding, (5.2) and a change of variables $y \leftrightarrow i(N)\mu$, we have

$$\langle \varpi, P_{M}^{(X)} \rangle = \int_{T(\mathbb{R})} \int_{U(\mathbb{Z})\setminus U(\mathbb{R})} \varpi(xy)\theta_N(-x)E(X)(N \cdot y(y))dx \ d^*y = N^\eta A_{\varpi}(N) \langle W_\mu, E^{(X)} \rangle$$

where as before $\mu = \mu_(\infty)$. By Parseval we obtain

$$\langle P_{M}^{(X)}, P_{N}^{(X)} \rangle = N^\eta M^{\eta} \int_{(q)} A_\varpi(M)A_\varpi(N) \langle W_\mu, E^{(X)} \rangle^2 d\varpi.$$ 

On the other hand, by unfolding and (6.2) we can express $\langle P_{M}^{(X)}, P_{N}^{(X)} \rangle$ as

$$\int_{T(\mathbb{R})} \int_{U(\mathbb{Z})\setminus U(\mathbb{R})} P_{M}^{(X)}(xy)\theta_N(-x)E(X)(N \cdot y(y))dx \ d^*y$$

$$= \sum_{w \in W} \sum_{\ell \in V} \sum_{c \in \mathbb{N}^{n-1}} S_{q,w}^{\ell}(M, N, c) \int_{T(\mathbb{R})} \int_{U_w(\mathbb{R})} F^{X}(\ell(M)c^*wxy)\theta_N(-x)E(X \cdot N \cdot y(y))dx \ d^*y.$$ 

Let

$$(6.3) \quad A = i(X \cdot M)c^*w(X \cdot N)^{-1}w^{-1} = i(X \cdot M \cdot w(X \cdot N))c^* \in T(\mathbb{R}),$$

so that $y(A)^\eta c_1 \cdots c_{n-1} = (X \cdot M \cdot w(X \cdot N))^{\eta}$ by (2.9). We change variables $y \leftrightarrow i(X \cdot N)y$, $x \leftrightarrow i(X \cdot N)x_i(X \cdot N)^{-1}$. By Lemma 2 we obtain

$$\sum_{w \in W} \sum_{\ell \in V} \sum_{c \in \mathbb{N}^{n-1}} S_{q,w}^{\ell}(M, N, c) \frac{(X \cdot M)^{\eta}(X \cdot N)^{\eta}}{c_1 \cdots c_{n-1} y(A)^{\eta}}$$

$$\times \int_{T(\mathbb{R})} \int_{U_w(\mathbb{R})} F^{X}(i(X)^{-1}A wxy)\theta_N^\ell(-x)E(y(y))dx \ d^*y.$$ 

We conclude the following Kuznetsov-type formula.

**Lemma 7.** Let $M, N \in \mathbb{N}^{n-1}$, $X \in \mathbb{R}^{n-1}$, $E$ a compactly supported function on $\mathbb{R}^{n-1}$ and define $F^{X}$ as in (6.1). Then

$$(6.4) \quad \int_{(q)} A_\varpi(M)A_\varpi(N) \langle W_\mu, E^{(X)} \rangle^2 d\varpi$$

$$= \sum_{w \in W} \sum_{\ell \in V} \sum_{c \in \mathbb{N}^{n-1}} S_{q,w}^{\ell}(M, N, c) \frac{X^{2\eta}}{c_1 \cdots c_{n-1} y(A)^{\eta}} \int_{T(\mathbb{R})} \int_{U_w(\mathbb{R})} F^{X}(i(X)^{-1}A wxy)\theta_N^\ell(-x)E(y(y))dx \ d^*y$$

with $A$ as in (6.3).

As mentioned before, the Kloosterman sum $S_{q,w}^{\ell}(M, N, c)$ vanishes unless $w$ is of the form (3.7), in which case we have the additional conditions (4.3), as well as (4.5) for $i \not\in \{d_1, \ldots, d_1 + \ldots + d_{n-1}\}$. The $c$-sum is restricted by Lemma 1 and the support of $E$. 
7. Proofs of Theorems 1, 2, 4

We start with the proof of Theorem 2. We specialize Lemma 7 to

\[ M = N = (m, 1, \ldots, 1), \quad X = (Z, 1, \ldots, 1) \]

with \((m, q) = 1\). We need to bound the spectral side from below and the Kloosterman side from above. By (5.3) and positivity we have

\[
(7.1) \quad \sum_{\pi \in \mathcal{F}_i(q)} |\lambda_{\pi}(m)|^2 Z^{2\eta + 2\sigma_e(\infty)} \ll I q^{n-1+\varepsilon} \int (q) |A_{\pi}(M)|^2 Z^{2\eta + 2\sigma_e(\infty)} \delta_{\lambda = 1} \, d\varpi.
\]

By Lemma 6 there is a finite set of compactly supported functions \(E_j\) such that

\[
Z^{2\eta + 2\sigma_e(\infty)} \delta_{\lambda = 1} \ll \sum_j |\langle W_{\mu_{\varpi}}, E_j^{(X)} \rangle|^2.
\]

Thus in order to bound the left hand side of (7.1) it suffices to consider the right hand side of (6.4) for a fixed \(E^{(X)} = E_j^{(X)}\), and we are left with bounding

\[
q^{n-1+\varepsilon} \sum_{w \in W} \sum_{v \in V} \sum_{\varepsilon \in \mathbb{N}^{n-1}} \frac{S_{q,w}^n(M, N, c)}{c_1 \cdots c_{n-1}} \frac{X^{2\eta}}{y(A)^n} 
\]

\[
\times \int_{\tilde{T}(\mathbb{R})} \int_{U_w(\mathbb{R})} F^{(X)}(\mu(X)^{-1} A w x y) \theta^\varepsilon(-x) E(y) \, dx \, dy.
\]

For \(w = \mathrm{id}\) we have \(c_1 = \ldots = c_{n-1} = 1\) and hence \(A = I_n\), and the contribution is \(O(q^{n-1+\varepsilon} Z^{2\eta})\).

Let us now consider the remaining \(w\) of the form (3.7). First we bound the moduli \(c_j\). To this end we apply Lemma 1 with \(B = X \cdot M \cdot \mu(X) N\), so that by (2.6) we obtain \(B_1 = B_{n-d_1} = mZ, B_{n-d_1+1} = 1/(mZ)\) if \(d_1 > 1\) and \(B_j = 1\) for all other indices. This gives

\[
(7.2) \quad c_j \ll (mZ)^{s(1,j)+s(n-d_1,j)-s(n-d_1+1,j)} = \begin{cases} mZ, & j \leq n - d_1, \\ 1, & j > n - d_1. \end{cases}
\]

We assume that \(mZ \ll q^2\) with a sufficiently small implied constant, so that \(c_j < q^2\) for all \(j\). We may also assume that \(q\) is sufficiently large, otherwise there is nothing to prove. Now suppose that \(d_1 > 1\) (but \(d_1 < n\) since \(w \neq \mathrm{id}\)). Then by (4.5) with \(i = d_1 - 1\) we have \(c_{n-d_1+2} \cdots c_{n-d_1} = \pm c_{n-d_1+1}^2 m\). Using (4.3) and comparing the \(q\)-adic valuation on both sides, we conclude from (4.3) that both \(c_{n-d_1+1}\) and \(c_{n-d_1+2}\) are divisible by \(q\), which contradicts (7.2) for \(q\) sufficiently large. Hence \(d_1 = 1\), and we see from (4.3) that all \(c_j\) are divisible by \(q\). We write \(c_j = q c'_j\). By (4.6) we obtain

\[
(7.3) \quad S_{q,w}^n(M, M, c) = S_{q,w}(\ast, \ast, (q, \ldots, q), S_{q,w}(\ast, (qm, 1, \ldots, 1, \tilde{q}), c')
\]

where \(*\) is coprime to \(q\). By Theorem 3 the first factor on the right hand side vanishes unless \(w = w_*\), in other words, only the trivial Weyl element and \(w_*\) survive.

(As an aside: if we only wanted to prove Sarnak’s density original density hypothesis with an exponent \(n - 1 - 2\sigma + \varepsilon\) in Theorem 1, then upon choosing \(mZ \ll q\) with a sufficiently small constant, all Weyl elements except the identity would vanish and no further analysis would be necessary. That in the stronger set-up \(mZ \ll q^2\) only \(w_*\) needs to be considered is an artefact of \(q\) being prime.)
Our next aim is to estimate
\[
\left| \int_{\mathbb{R}} \int_{U_{w}(\mathbb{R})} F(X)(\iota(X)^{-1} Aw_{xy})\theta^n(-x)\overline{E(y(x))}dx\,dy \right| \\
\leq \int_{\mathbb{R}} \int_{U_{w}(\mathbb{R})} |E(y(Aw_{xy}))E(y(x))|dx\,dy.
\]
By Lemma 1 and then Lemma 3 the right hand side is bounded by
\[
\ll_E \text{vol}\{ x \in U_{w} \mid \Delta_j(w_{x}x) \ll_E \prod_{i=1}^{n-1} y(A)^{s(i,j)}, 1 \leq j \leq n - 1 \}
\]
(7.4)
\[
\ll_E \prod_{i=1}^{n-1} \prod_{j=1}^{n-1} y(A)^{s(i,j)(1+\varepsilon)} = y(A)^{n(1+\varepsilon)}
\]
since \( \sum_{i} s(i,j) = \eta_{j} \) by (3.2) and (2.7).

Summarizing the previous estimations (and changing the value of \( \varepsilon \)) and applying Theorem 3 to the first factor in (7.3), we obtain
\[
\sum_{\pi \in \mathcal{F}_1(q)} |\lambda_{\pi}(m)|^2 Z^{2\sigma_{\pi}(\infty)}
\ll_I Z^{2n} q^{n-1+\varepsilon} \left( 1 + q^{-2} \sum_{v \in V} c_{1}^{\prime} \cdots c_{n-1}^{\prime} \ll mZ/q \right)
\]
For the Weyl element \( w_{x} \) the consistency relations (4.5) impose serious restrictions on the moduli \( c_{1}^{\prime}, \ldots, c_{n-1}^{\prime} \). We apply (4.5) with \( i = 2, \ldots, n - 2 \) getting \( (c_{i}^{\prime})^2 = c_{i-1}^{\prime} c_{i+1}^{\prime} \) for \( i = 2, \ldots, n - 2 \). If \( n \geq 4 \), then \( c_{3}^{\prime} \) fixes \( c_{1}^{\prime} \) and \( c_{3}^{\prime} \) up to a divisor function, and inductively also \( c_{4}^{\prime}, \ldots, c_{n-1}^{\prime} \). Using the trivial bound (4.7), we finally obtain (again changing the value of \( \varepsilon \))
\[
\sum_{\pi \in \mathcal{F}_1(q)} |\lambda_{\pi}(m)|^2 Z^{2\sigma_{\pi}(\infty)} \ll_I q^{n-1+\varepsilon} \left( 1 + \frac{q^{-2}}{q^{n-1}} \sum_{c_{3}^{\prime} \ll mZ/q} 1 \right) \ll q^{n-1+\varepsilon}
\]
provided \( mZ \ll q^{2} \). In the case \( n = 3 \) we quote from [BBM, (4.2) with \( N = 1 \)] the average Weil-type bound
\[
\sum_{c_{1}, c_{2} \leq X} |S_{w_{x}}^{\circ}((m, 1), (\bar{q}m, \bar{q}), c^{'})| \ll X^{3}(Xm)^{\varepsilon}
\]
to obtain again
\[
\sum_{\pi \in \mathcal{F}_1(q)} |\lambda_{\pi}(m)|^2 Z^{2\sigma_{\pi}(\infty)} \ll_I q^{2+\varepsilon} \left( 1 + \frac{1}{q} \cdot \frac{mZ}{q} \right) \ll q^{2+\varepsilon}
\]
for \( mZ \ll q^{2} \). This completes the proof. \( \Box \)
The proof of Theorem 4 is a simple variation. Again by positivity and (5.3) we have
\[ \sum_{\pi \in F_1(q)} \left| \sum_{m \leq x} \alpha(m) \lambda_\pi(m) \right|^2 \lesssim q^{n-1+\varepsilon} \int_{(q)} \left| \sum_{m \leq x} \alpha(m) A_\pi(M) \right|^2 \delta_\lambda \in I \, d\varpi \]
\[ = q^{n-1+\varepsilon} \sum_{m_1, m_2 \leq x \atop (m_1, m_2, q) = 1} \alpha(m_1) \alpha(m_2) \int_{(q)} A_\pi(M_1) A_\pi(M_2) \delta_\lambda \in I \, d\varpi \]
where \( M = (m, 1, \ldots, 1), M_1 = (m_1, 1, \ldots, 1), M_2 = (m_2, 1, \ldots, 1). \) We detect the condition \( \delta_\lambda \in I \) by a finite collection of test functions \( E_j \) with \( Z = 1 \) as in the previous proof and apply Lemma 7. For \( w = \) id the analogue of (7.2) is
\[ c_j \ll m_2^{s(1,j)} m_1^{s(n-d_1,j)-s(n-d_1+1,j)} \lesssim x \]
which contradicts (4.3) for \( x \ll q \) (with a sufficiently small implied constant) since \( d_1 \neq n \).
So only the trivial Weyl element survives, and we obtain the desired bound. \( \square \)

Corollary 5 follows easily Theorem 4 by observing that an approximate functional equation has length \( q^{1/2} \) (see [IK, Section 5]): for all but \( O(1) \) cuspidal representations \( \pi \in F_1(q) \) (and \( \varepsilon < 1/2 \)) we have
\[ |L(1/2 + it, \pi)|^2 \ll_{I,t,n,\varepsilon} q^{\varepsilon} \sum_{2^j = M \leq q^{1/2+\varepsilon}} \frac{1}{M} \left| \sum_{M \leq m \leq 2M} \lambda_\pi(m) \right|^2 \]
and the desired bound follows directly from Theorem 4. Note that the shape of the ramified coefficients (i.e. \( q \mid m \)) is irrelevant and the condition \( (m, q) = 1 \) in Theorem 4 is void.

Finally we derive Theorem 1 from Theorem 2. Let us first assume that \( v = p \neq q \) is a fixed prime. We choose \( \nu_0 \) maximal so that \( p^{\nu_0} \ll q^2 \) with an implied constant that is admissible for Theorem 2. We conclude from Theorem 2 with \( Z = 1, m = p^\nu \) and Lemma 4 that
\[ N_p(\sigma, F_1(q)) \ll \sum_{\pi \in F_1(q)} \frac{p^{2\nu_0(s_0(p))}}{p^{2\nu_0}} \ll \frac{1}{q^{4\sigma}} \sum_{\nu_0 - n < \nu < \nu_0} \sum_{\pi \in F_1(q)} |\lambda_\pi(p^\nu)|^2 \ll q^{n-1-4\sigma+\varepsilon}. \]
For \( v = \infty \), Theorem 1 follows directly from Theorem 2 with \( m = 1, Z \ll q^2 \) (again with a sufficiently small implied constant). \( \square \)

References

[Ar] J. Arthur, *Eisenstein series and the trace formula*, in: Automorphic forms, representations and L-functions, Corvallis/Oregon 1977, Proc. Symp. Pure Math. 33 (1979), 253-274

[AB] E. Assing, V. Blomer, *The density conjecture for principal congruence subgroups*, arXiv:2204.08868

[Bi] V. Blomer, *Applications of the Kuznetsov formula on GL(3)*, Invent. math. 194 (2013), 673-729

[BB] V. Blomer, F. Brumley, *The role of the Ramanujan conjecture in analytic number theory*, Bull. Amer. Math. Soc. 50 (2013), 267-320

[BBM] V. Blomer, J. Buttcane, P. Maga, *Applications of the Kuznetsov formula on GL(3): the level aspect*, Math. Ann. 369 (2017), 723-759

[BBR] V. Blomer, J. Buttcane, N. Raufi, *A Sato-Tate law for GL(3)*, Comm. Math. Helv. 89 (2014), 895-919

[BM1] R. W. Bruggeman, R. J. Miatello, *Sum formula for SL_2 over a number field and Selberg type estimate for exceptional eigenvalues*, GAFA 8 (1998), 627-655

[BM2] R. W. Bruggeman, R. J. Miatello, *Density results for automorphic forms on Hilbert modular groups*, GAFA 13 (2003), 681-719
20 VALENTIN BLOMER

[DR] R. Dabrowski, M. Reeder, *Kloosterman sets in reductive groups*, J. Number Theory 73 (1998), 228-255

[FM] T. Finis, J. Matz, *On the asymptotics of Hecke operators for reductive groups*, arxiv:1905.09078

[FP] C.M. da Fonseca, J. Petronilho, *Explicit inverses of some tridiagonal matrices*, Linear Algebra Appl. 325 (2001) 7-21

[Fr] S. Friedberg, *Poincaré series for GL(n): Fourier expansion, Kloosterman sums, and algebraic-geometric estimates*. Math. Z. 196 (1987) 165-188.

[FG] S. Friedberg, D. Goldfeld, *Mellin transforms of Whittaker functions*, Bull. Soc. Math. France 121 (1993), 91-107

[Go] D. Goldfeld, *Automorphic forms and L-functions for the group GL(n,R)*, Cambridge studies in advanced mathematics 99 (2006)

[HK] J. Huntley, Y. Katznelson, *Density theorems for congruence groups in real rank 1*, Duke Math. J. 71 (1993), 463-473

[Hu] M. Huxley, *Exceptional eigenvalues and congruence subgroups*, in: The Selberg trace formula and related topics, Contemp. Math. 53 (1986), 341-349

[Iw] H. Iwaniec, *Small eigenvalues of Laplacian for Γ0(N)*, Acta Arith. 56 (1990), 65-82

[IK] H. Iwaniec, E. Kowalski, *Analytic number theory*, AMS Colloq. Publ. 53, Providence, RI, 2004.

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

[JS] H. Jacquet, J. Shalika, *On Euler products and the classification of automorphic representations. I*, Amer. J. Math. 103 (1981), 499-558

[LS] N. W. Lenstra, J. O. Shallit, *Continued fractions and linear recurrences*, Math. Comp. 61 (1993), 351-354

[Li] X. Li, *Upper bounds on L-functions at the edge of the critical strip*, IMRN 2010, 727-755

[LRS] W. Luo, Z. Rudnick, P. Sarnak, *On the generalized Ramanujan conjecture for GL(n)*, Proc. Sympos. Pure Math. 66, 301-310

[Ma] S. H. Man, *A Density Theorem for Sp(4)*, to appear in J. Lond Math. Soc.

[MT] J. Matz, N. Templier, *Sato-Tate equidistribution for families of Hecke-Maass forms on SL(n, R)/SO(n)*, to appear in Algebra & Number Theory

[PS] O. Parzanchevsky, P. Sarnak, *Super-Golden-Gates for PU(2)*, Adv. Math. 327 (2018), 869-901

[Sa1] P. Sarnak, *Statistical properties of eigenvalues of the Hecke operators*, in: Analytic number theory and Diophantine problems, Progr. Math. 70 (1987), 321-331.

[Sa2] P. Sarnak, *Diophantine Problems and Linear Groups*, Proceedings of the ICM Kyoto (1990), 459-471

[Sa3] P. Sarnak, *Definition of families of L-functions*, https://publications.ias.edu/sarnak/paper/507

[Sa4] P. Sarnak, *Letter on optimal strong approximation*, https://publications.ias.edu/sarnak/paper/2637

[Sa5] P. Sarnak, *Letter on Solvay-Kitaev Theorem and Golden Gates*, https://publications.ias.edu/sarnak/paper/2637

[SST] P. Sarnak, S. W. Shin, N. Templier, *Families of L-functions and their symmetry*, in: Families of automorphic forms and the trace formula, 531-578, Simons Symp., Springer 2016

[SX] P. Sarnak, X. Xue, *Bounds for multiplicities of automorphic representations*, Duke Math. J. 64 (1991), 207-227

[St] E. Stade, *Mellin transforms of GL(n,R) Whittaker functions*, Amer. J. Math. 123 (2001), 121-161

[Ve] A. Venkatesh, *Large sieve inequalities for GL(n)-forms in the conductor aspect*, Adv. in Math. 200 (2006), 336-356

Mathematisches Institut, Endenicher Allee 60, 53115 Bonn, Germany

Email address: blomer@math.uni-bonn.de