Large Sieve Inequalities for GL(n)-forms In the Conductor Aspect

Akshay Venkatesh *

September 18, 2003

Abstract

Duke and Kowalski in [6] derive a large sieve inequality for automorphic forms on GL(n) via the Rankin-Selberg method. We give here a partial complement to this result: using some explicit geometry of fundamental regions, we prove a large sieve inequality yielding sharp results in a region distinct to that in [6]. As an application, we give a generalization to GL(n) of Duke's multiplicity theorem from [5]; we also establish basic estimates on Fourier coefficients of GL(n) forms by computing the ramified factors for GL(n) $\times$ GL(n) Rankin-Selberg integrals.

1 Introduction

The purpose of this is to develop a "large sieve" inequality for automorphic forms on GL(n). It is sharp in a very short range: when one is considering Fourier coefficients much smaller than the conductor. Nevertheless, this suffices for some applications. In the process, we establish some results on Fourier coefficients on GL(n) that should find application in many investigations of an analytic nature.

There are practically no results of an analytic nature on GL(n), for $n \geq 3$, that do not rest on the properties of Rankin-Selberg L-functions; for example, [9], [10], [6] all use as input only the properties of standard and Rankin-Selberg L-functions. In particular, in [6] Duke-Kowalski derive a large sieve inequality based on properties of GL(n) Rankin-Selberg convolutions.

*Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA. email: akshayv@math.mit.edu. Partially supported by NSF grant DMS-0245606.

In this paper we derive a different large sieve inequality for GL(n) using geometric methods coming from some explicit reduction theory. It seems plausible that with a very careful analysis of the Rankin-Selberg method one might be able to replicate some of the present results, but the overlap and ultimate scope of these techniques is not entirely clear.

In short, this paper is essentially a direct generalization of [5] to the GL(n) setting. In [5], a method of Iwaniec is used to bound Fourier coefficients of modular forms, and thereby a large sieve is derived. In effect, we generalize this method of Iwaniec to GL(n); it is sharp only in a very short range, but even this gives interesting results. On GL(2), this type of method is closely related to Rankin-Selberg; on GL(n) the two diverge. In particular, when $n > 2$, we will use Bessel's inequality in a rather wasteful way: we will make a Fourier expansion of an automorphic form with respect to a maximal unipotent subgroup, but we make no usage of the "non-abelian" part of this expansion.

Applications are given to bounding the number of GL(n)-automorphic forms of Galois type; this generalizes the work of Duke. Other applications are possible and we intend to pursue them elsewhere.

We note that on GL(2) the methods of this paper are far inferior to what can be proved directly. One can derive very sharp inequalities directly from the Petersson-Kuznetsov formula. On a general group, such formulas are not available (and, to the extent that they are, the constituent integral transforms and exponential sums are far less understood).

We conclude with a very brief description of the method and the main problems. On GL(2) a large sieve can be established by estimating the L^2-norms of forms on a region like $\{0 \leq x \leq 1, y \geq Y\}$ and optimizing Y, as in [5]. On GL(n), we mimic this proof by finding an appropriate "large" Siegel domain. However, there are many ways to produce such Siegel domains in GL(n), and they are usually unsuitable for our purpose. A fundamental difference (which manifests itself at the level of Eisenstein series) is that: *a* 2*-dimensional lattice can only have a few short primitive vectors, while a* 3*-dimensional lattice can have many.* It is this geometric issue that forms the main obstacle (albeit in a disguised way).

The reader primarily interested in the central details may wish to immediately skip to Proposition 3, of which the main Theorem is essentially a corollary, and its proof in Section 5.

Acknowledgements: The author would like to thank Bill Duke for interesting conversations and encouragement. He has also benefited from conversations with Farrell Brumley and Peter Sarnak.

Notation: Throughout this paper, the implicit constant in $\ll, \gg, \asymp$

should be understood as depending on n. Here $f \asymp g$ should be read as $f \ll g \ll f$.

We shall often make statements such as the following " If $P \ll Q$ and $R \gg S$, then $T \ll U$." Such statements should understood as: "given C, C', there exists C'' so that $P \leq CQ$ and $R \geq C'S \implies R \leq C''S$."

The phrase "is bounded" should be understood as "bounded by a constant possibly depending on n."

2 Results

Throughout this paper $q \geq 1$ is a positive integer.

We will work with automorphic cuspidal representations of conductor q on $\mathrm{GL}(n)$ over $\mathbb{Q}$. Let π be such a representation; let $\lambda_\pi(n)$ be the coefficient of n^{-s} in the L-function $L(s, \pi)$.

We will fix, once and for all, a compact subset $\mathcal{S} \subset \widehat{\mathrm{GL}(n, \mathbb{R})}$ of the unitary dual of $\mathrm{GL}(n, \mathbb{R})/Z(\mathbb{R})^0$, where $Z(\mathbb{R})^0$ is the connected component of the center of $\mathrm{GL}(n, \mathbb{R})$.

This is a technical generalization of allowing a single π_∞, analogous on $\mathrm{GL}(2)$ to restricting the Laplacian eigenvalue. (Since $\mathrm{GL}(n, \mathbb{R})$ has no discrete series if $n > 2$, we prefer not to restrict π_∞ to a single representation. For the application in Section 6, or other arithmetic applications such as in [6], however, the restricted Theorem would suffice.)

The results we prove, being geometric in origin, are naturally phrased in terms of the L^2-normalization. To convert to Hecke eigenvalues, it will be convenient to include the following weight:

$$\omega_\pi = \lim_{s \to 1} \frac{s-1}{L^{(q,\infty)}(s, \pi \times \tilde{\pi})} \tag{1}$$

where $L^{(q,\infty)}$ denotes the Rankin-Selberg L function, omitting the factors at primes dividing q and at ∞.

It is of course believed that $\omega_\pi \asymp_\epsilon q^\epsilon$. At present neither lower nor upper bounds are known in general.

For π as above, χ_π will denote the central character of π, regarded as a Dirichlet character.

Theorem 1. *Let χ be a Dirichlet character mod q. Let $S_1(q)$ (respectively $S_0(q, \chi)$) denote the set of cuspidal automorphic representations of conductor q (respectively conductor q and central character χ) and with $\pi_\infty \in \mathcal{S}$.*

Let $S = S_1(q)$ or $S_0(q,\chi)$. Set $N = q^{1/(n-1)}, \kappa = n$ in the former case and $N = q^{1/(2n-2)}, \kappa = n-1$ in the latter case. Let $\{a_i : 1 \leq i \leq N\}$ be an arbitrary set of complex numbers.

Then:

$$\sum_{\pi \in S} \omega_\pi \left| \sum_{\substack{n=1 \\ (n,q)=1}}^{N} a_n \lambda_\pi(n) \overline{\chi_\pi(n)} \right|^2 \ll_{\epsilon, \mathcal{S}} q^{\kappa+\epsilon} \sum_{n=1}^{N} |a_n|^2 \tag{2}$$

Remarks.

1. Observe that (so long as $\mathcal{S}$ contains a small open set of tempered representations) the cardinality of S should roughly behave as q^κ as $q \to \infty$. Therefore one believes the bound in the above Theorem to be essentially sharp.

   Indeed, "limit multiplicity formula" (see [4]) suggest that $|S_1(q)|$ is essentially proportional to the index $[\mathrm{GL}_n(\mathbb{Z}) : \Gamma_1(q)]$, where $\Gamma_1(q)$ is defined in Section 3.1. This index grows like q^n (to within q^ϵ). These limit multiplicity formulae are not yet established (to the author's knowledge) for GL_n, but one certainly believes the result to be valid, and we have no explicit need of them other than to determine what the "trivial bound" is.

2. Duke and Kowalski, [6], prove that the corresponding equality is valid when $N \gg q^n |S|^2$ and the constant $q^{\kappa+\epsilon}$ is replaced by $N^{1+\epsilon}$. Therefore the present result is complementary. It should be remarked that [6] is more flexible than Theorem 1 in that it allows one to consider an arbitrary set of forms; on the other hand it only achieves savings when N is large.

3. Averaging over $S_1(q)$ corresponds to averaging over all χ of conductor q. It is therefore not surprising one obtains a slightly better result for $S_1(q)$ than for $S_0(q,\chi)$.

4. The method of proof gives an inequality for all N. We have only stated it, however, in the region where it is expected to be sharp.

For simplicity, we prove the Theorem in the case where $\mathcal{S}$ is a subset of the spherical unitary dual of $\mathrm{PGL}_n(\mathbb{R})$ – in particular, χ is an even Dirichlet character; the proof in general is identical.

It seems for $n = 3$ that an application to mean square values of $L(1/2, \pi)$ might be possible. For large n the allowable size of N restricts the applicability. However, it is always sufficient to get nontrivial estimates in the following type of question: counting forms whose Fourier coefficients have some prescribed behaviour. We give an application in this vein in Section 6.

3 Fourier coefficients: adelic and archimeedean

The derivation of the Theorem will be a computation on a real symmetric space; since most results about GL_n are phrased adelically, we briefly cover aspects of the transition. We define the relevant set of cuspidal representations under consideration and make precise the normalization of Fourier coefficients.

3.1 Cuspidal representations

We refer to [2] for foundational details.

Let $\mathbb{A}$ be the ring of adeles of $\mathbb{Q}$ and $\mathbb{A}_f$ the ring of finite adeles; thus $\mathbb{A} = \mathbb{R} \times \mathbb{A}_f$. Let $K_1(q)$ and $K_0(q)$ be the open compact subgroups of $\mathrm{GL}_n(\mathbb{A}_f)$ corresponding to $\Gamma_1(q)$ and $\Gamma_0(q)$ respectively (these consist of the integral matrices belonging to $\mathrm{GL}_n(\mathbb{Z})$ and with bottom row congruent to $(0, 0, \dots, 1)$ and $(0, 0, \dots, 0, ?)$ mod q respectively). Let $K_\infty = O_n(\mathbb{R}) \subset \mathrm{GL}_n(\mathbb{R})$. Let N be the algebraic subgroup of GL_n consisting of unipotent upper triangular matrices. Finally, let $\psi_{\mathbb{A}}$ be an unramified character of $\mathbb{A}/\mathbb{Q}$, i.e. trivial on $\prod_p \mathbb{Z}_p$. Denote by the same symbol ($\psi_{\mathbb{A}}$) the character of $N(\mathbb{Q})\backslash N(\mathbb{A})$ defined by:

$$n \mapsto \psi(n_{12} + n_{23} + \cdots + n_{n-1,n}) \tag{3}$$

We denote by ψ_v the corresponding character of $N(\mathbb{Q}_v)$; in particular ψ_∞ gives a character of $N(\mathbb{Z})\backslash N(\mathbb{R})$.

Let $A \subset \mathrm{GL}(n, \mathbb{R})$ be the subset of diagonal matrices with positive entries, and let $\rho : A \to \mathbb{R}_+$ be the 1/2-sum of positive roots (for the positive system defined by N). Finally, set $A_T^+ = \{\mathrm{diag}(a_i) \in A : a_i a_{i+1}^{-1} \geq \sqrt{3}/2, 1 \leq i \leq n-2;\ \ a_{n-1}a_n^{-1} \geq T^{-1}\}$. Here $\mathrm{diag}(\dots)$ denotes the diagonal matrix with the specified entries along the diagonal.

Let $Z \subset \mathrm{GL}_n$ be the center. We now fix, once and for all, Haar measures. All discrete groups are endowed with counting measure. If $v = p$ is a finite place, fix the Haar measure on $\mathrm{GL}_n(\mathbb{Q}_p), N(\mathbb{Q}_p)$ and $Z(\mathbb{Q}_p)$ so that the measure of the sets $\mathrm{GL}_n(\mathbb{Z}_p), N(\mathbb{Z}_p), Z(\mathbb{Z}_p)$ are all 1. Fix the Haar measure on $N(\mathbb{R})$ so that $\mathrm{vol}(N(\mathbb{Z})\backslash N(\mathbb{R})) = 1$ and on K_∞ so that $\mathrm{vol}(K_\infty) = 1$. Fix a Haar measure on A and $Z(\mathbb{R})$ and give $G = \mathrm{GL}_n(\mathbb{R})$ the measure arising from the Iwasawa decomposition $G = N(\mathbb{R})AK_\infty$. These choices also induce Haar measures on $\mathrm{PGL}_n(\mathbb{Q}_v)$ as well as the adelic points of $N, \mathrm{PGL}_n, \mathrm{GL}_n$.

Let Y be the set of isomorphism classes of *unitary, generic, spherical* representations of $\mathrm{PGL}_n(\mathbb{R})$. Y can be identified with a subset of $\mathbb{C}^{n-1}/S_n$, where S_n is the symmetric group on n letters; topologize it accordingly. For each $\nu \in Y$, let $\pi(\nu)$ be the corresponding $\mathrm{PGL}_n(\mathbb{R})$-representation. We fix once and for all a spherical Whittaker function W_ν associated to $\pi(\nu)$, transforming on the left under $N(\mathbb{R})$ by ψ_∞, and so that the assignment $\nu \mapsto W_\nu$ is continuous. As remarked in the previous section, we assume for simplicity that $\mathcal{S}$ is a subset of the spherical unitary dual of $\mathrm{PGL}_n(\mathbb{R})$; in particular, we will regard $\mathcal{S}$ as a subset of Y.

Let χ be an even Dirichlet character of conductor dividing q; we identify it with a character of $\mathbb{A}^\times/\mathbb{Q}^\times$. Let $^\circ\mathcal{A}_n$ be the space of cuspidal automorphic forms on the adelic quotient $Z(\mathbb{R})\mathrm{GL}_n(\mathbb{Q})\backslash\mathrm{GL}_n(\mathbb{A})$, and $^\circ\mathcal{A}_n^\chi$ the space of those with central character χ. The set of irreducible subrepresentations $\pi \subset {^\circ\mathcal{A}_n}$, i.e. cuspidal representations, that have a K_∞-fixed vector and have conductor q is denoted $\mathcal{CP}(q)$; those π that additionally have central character χ are denoted $\mathcal{CP}(q)^\chi$.

Such π necessarily have $K_1(q)$-fixed vectors, by a theorem in [7], but the converse is not true on account of "oldforms." Each $\pi \in \mathcal{CP}(q)$ may be expressed as a tensor product: $\pi = \otimes_v \pi_v$, where π_v is a representation of $\mathrm{GL}_n(\mathbb{Q}_v)$ and the tensor product is defined with reference to a choice of spherical vector $^\circ\phi_v \in \pi_v$ for almost all v. For $\pi \in \mathcal{CP}(q)$, we denote by $\nu_\pi \in Y$ the parameter of the archimedean representation π_∞.

Let $^\circ\mathcal{A}_n(q)$ be the space of $K_\infty \times K_1(q)$-invariant functions in $^\circ\mathcal{A}_n$; we regard elements of $^\circ\mathcal{A}_n(q)$ as cuspidal automorphic forms on $\mathrm{PGL}_n(\mathbb{R})/\mathrm{PO}_n(\mathbb{R})$ with respect to $\Gamma_1(q)$.

3.2 Newforms for GL_n

Let $\pi \in \mathcal{CP}(q)$, and regard π as a subrepresentation of $^\circ\mathcal{A}_n$. Choose a vector $\varphi_\pi^{\mathrm{new}}$ in the space of π that is fixed by $K_\infty \times K_1(q)$, and so that for $g = (g_\infty, 1) \in \mathrm{GL}_n(\mathbb{A}) = \mathrm{GL}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{A}_f)$ one has:

$$\int_{N(\mathbb{Q})\backslash N(\mathbb{A})} \varphi_\pi^{\mathrm{new}}(ng)\overline{\psi_{\mathbb{A}}(n)}dn = W_{\nu_\pi}(g_\infty) \tag{4}$$

This uniquely specifies a nonzero vector $\varphi_\pi^{\mathrm{new}}$. We refer to it as *the* new vector. It is factorizable; fix a decomposition of $\varphi_\pi^{\mathrm{new}} = \otimes_v \varphi_{\pi,v}^{\mathrm{new}}$ as a tensor product

Each π_v admits a Whittaker model unique up to scalars, transforming under the character ψ_p of $N(\mathbb{Q}_p)$. We specify it as follows: if v is a finite prime, we specify it so that $W_{\varphi_{\pi,v}^{\mathrm{new}}}$ (the Whittaker function associated to the new vector $\varphi_{\pi,v}^{\mathrm{new}} \in \pi_v$) takes the value 1 at the identity. With this normalization, it agrees with the *essential vector* in [7]. On the other hand, at $v = \infty$, we require that $W_{\varphi_{\pi,\infty}^{\mathrm{new}}} = W_{\nu_\pi}$.

Now let $\phi \in \pi$ be an arbitrary *factorizable* vector, i.e. $\phi = \otimes \phi_v$ with $\phi_v \in \pi_v$. Let $W_{\phi,\mathbb{A}}$ be the "$N(\mathbb{A})$-Fourier-coefficient" associated to ϕ, i.e.

$$W_{\phi,\mathbb{A}}(g) = \int_{N(\mathbb{Q})\backslash N(\mathbb{A})} \phi(ng)\overline{\psi_{\mathbb{A}}(n)}dn \tag{5}$$

Our choices now guarantee that given $g = (g_v) \in \mathrm{GL}_n(\mathbb{A})$

$$W_{\phi,\mathbb{A}}(g) = \prod_v W_{\phi_v}(g_v) \tag{6}$$

Here $W_{\phi_v}(g_v)$ is the Whittaker function associated to $\phi_v \in \pi_v$, evaluated at g_v. To verify (6) note that both sides are Whittaker models for π, and they agree, by (4) and (5), in the case where $\phi = \varphi^{\mathrm{new}}$ and $g = (g_\infty, 1) \in \mathrm{GL}_n(\mathbb{A})$.

Let m be a positive integer coprime to q. Let $\delta_m = \mathrm{diag}(m, m, \ldots, m, 1)$, considered as an element of $\mathrm{GL}_n(\mathbb{Q})$ (and thus of $\mathrm{GL}_n(\mathbb{A})$ and $\mathrm{GL}_n(\mathbb{Q}_v)$ for each v; it will be clear from context in which group it is considered as lying.) Let $\lambda_\pi(m)$ be the mth coefficient in the L-series of π; then it is a consequence of Shintani's formula [11] that:

$$\prod_{p \text{ prime}} W_{\varphi_{\pi,p}^{\mathrm{new}}}(\delta_m) = m^{-(n-1)/2}\chi(m)\overline{\lambda_\pi(m)} \tag{7}$$

Observe in deriving this that $W_{\varphi_{\pi,p}^{\mathrm{new}}}(\delta_m) = 1$ if $p|q$, on account of the $K_1(q)$-invariance of $\varphi_\pi^{\mathrm{new}}$.

3.3 Fourier coefficients

We will now think more in terms of the *real* symmetric space $\Gamma_1(q)\backslash\mathrm{PGL}_n(\mathbb{R})$ rather than the adelic one, and will generally use f rather than ϕ or φ to suggest a function on the real symmetric space.

Suppose $\pi \in \mathcal{CP}(q)$; set f_π^{new} to be the L^2-normalized form:

$$f_\pi^{\mathrm{new}}(g) = \frac{\varphi_\pi^{\mathrm{new}}(g)}{\sqrt{\int_{\mathrm{PGL}_n(\mathbb{Q})\backslash\mathrm{PGL}_n(\mathbb{A})} |\varphi_\pi^{\mathrm{new}}(g)|^2 dg}} \tag{8}$$

which is a function on $\mathrm{GL}_n(\mathbb{A})$, but we shall think of it in terms of its restriction to $\mathrm{GL}_n(\mathbb{R})$, i.e. as a "classical" automorphic form.

Now let $f \in {}^\circ\mathcal{A}_n(q)$ be arbitrary. Set, for $g_\infty \in GL(n,\mathbb{R})$ and m a positive integer:

$$W_f(m, g_\infty) = \int_{N(\mathbb{Z})\backslash N(\mathbb{R})} f(ng_\infty)\overline{\psi_\infty(\delta_m n \delta_m^{-1})}dn \tag{9}$$

In particular, if f_π^{new} is the new vector for $\pi \in \mathcal{CP}(q)$ with infinity type ν_π, $W_{f_\pi^{\mathrm{new}}}(m, g_\infty)$ is, by the uniqueness of real Whittaker models, a multiple of $W_{\nu_\pi}(\delta_m g)$. Accordingly:

Definition 1. *Let $\pi \in \mathcal{CP}(q)$ with infinity type $\nu_\pi \in \mathcal{S}$. Let $(m,q) = 1$. We define the mth Fourier coefficient $a_\pi(m) \in \mathbb{C}$ via the formula*

$$W_{f_\pi^{\mathrm{new}}}(m, g_\infty) = a_\pi(m) m^{-(n-1)/2} W_{\nu_\pi}(\delta_m g_\infty) \tag{10}$$

The next result is of independent importance and requires an understanding of factors in the Rankin-Selberg integral for $\mathrm{GL}(n)$ at ramified places. The proof is postponed until the final section.

Proposition 1. *Let $\pi \in \mathcal{CP}(q)$ with infinity type $\nu_\pi \in \mathcal{S}$. Then*

$$a_\pi(m) = \omega_\pi^{1/2}\overline{\lambda_\pi(m)}\chi_\pi(m)c_\pi$$

where $q^{-\epsilon} \ll_{\epsilon,\mathcal{S}} c_\pi \ll_{\epsilon,\mathcal{S}} q^\epsilon$, and ω_π is as in (1).

Proof. In Section 7.

We also need to deal with the fact that we are allowing a range of infinity types π_∞, i.e. "Laplacian eigenvalues"; the point is that, roughly speaking, although the Whittaker functions W_ν vary with $\nu \in Y$, they do not vary too much.

Proposition 2. *For each $\nu_0 \in Y$, there exists an open set $U \subset Y$ containing ν_0 and a function $c(\nu) : U \to \mathbb{C}$ with $1/2 < |c(\nu)| < 1$ so that the following holds (for* any *q):*

Let $\Pi = \{\pi \in \mathcal{CP}(q) : \nu_\pi \in U\}$ and m a positive integer prime to q. Let $(b_\pi)_{\pi\in\Pi}$ be complex numbers, and set $f = \sum_{\pi\in U} b_\pi f_\pi^{\mathrm{new}}$. Thus $f \in {}^\circ\mathcal{A}_n(q)$. Define $W_f(m,g)$ as in (9). Then

$$\int_{Z(\mathbb{R})\backslash A_m^+} |W_f(m,a)|^2 a^{-2\rho} da \gg_{\nu_0} |\sum_{\pi\in\Pi} c(\nu_\pi) b_\pi a_\pi(m)|^2 \tag{11}$$

(Here the implicit constant depends on ν_0 and the choice of U and $c(\nu)$, but not on q and m.)

The function $c(\nu)$ appears rather peculiar, but it will vanish in the final statement of the large sieve.

Proof. The left-hand side of (11), in view of (10), is

$$m^{-(n-1)} \int_{Z(\mathbb{R})\backslash A_m^+} |\sum_{\pi\in\Pi} b_\pi a_\pi(m) W_{\nu_\pi}(\delta_m a)|^2 a^{-2\rho} da = \tag{12}$$

$$\int_{Z(\mathbb{R})\backslash A_1^+} |\sum_{\pi\in\Pi} b_\pi a_\pi(m) W_{\nu_\pi}(a)|^2 da \tag{13}$$

where we make the substitution $a \leftarrow \delta_m^{-1} a$.

Denote by X any compact subset of $Z(\mathbb{R})\backslash A_1^+$ with nonempty interior. Pointwise on X, we have $W_{\nu'} \to W_{\nu_0}$ as $\nu' \to \nu_0$. Set

$$c(\nu) = \frac{\int_X W_\nu(a)\overline{W_{\nu_0}(a)}da}{\int_X |W_{\nu_0}(a)|^2 da}$$

It is well-defined as W_{ν_0} does not vanish identically on any open set, by real-analyticity. We choose U so small that $|c(\nu)| > 1/2$. Now:

$$\int_{Z(\mathbb{R})\backslash A_1^+} |\sum_{\pi\in\Pi} b_\pi a_\pi(m) W_{\nu_\pi}(a)|^2 da$$

$$\geq \int_X |\sum_{\pi\in\Pi} b_\pi a_\pi(m) W_{\nu_\pi}(a)|^2 da \geq \int_X |\sum_{\pi\in\Pi} c(\nu_\pi) b_\pi a_\pi(m)|^2 |W_{\nu_0}(a)|^2 da \gg_{\nu_0, X}$$

$$|\sum_{\pi\in\Pi} c(\nu_\pi) b_\pi a_\pi(m)|^2 \tag{14}$$

where the second inequality utilises Cauchy-Schwarz in the Hilbert space $L^2(X)$, and the implicit constant in the third inequality depends on ν_0, X; the important point is that it is independent of q. □

4 Proof of Main Theorem

In this section, we prove the main theorem, contingent on some results from reduction theory that we prove in the following section. We follow the notation of the previous section.

Let ω_N be the compact subset of $N(\mathbb{R})$ consisting of $\{n_{ij} \in N : -1/2 \leq n_{ij} < 1/2\}$; thus ω_N is a fundamental domain for $N(\mathbb{Z})$ acting on $N(\mathbb{R})$. Define $\mathfrak{S}(T) \subset \mathrm{GL}_n(\mathbb{R})$ as

$$\mathfrak{S}(T) = \omega_N A_T^+ K_\infty$$

where A_T^+ is as defined in Section 3.1.

Proposition 3. *Suppose $T \ll q^{1/(n-1)}$. Then the map $\mathfrak{S}(T) \to \Gamma_1(q)\backslash \mathrm{GL}_n(\mathbb{R})$ has fibers of size $\ll 1$ (see remarks on notation, end of first section). If $T \ll q^{1/(2n-2)}$, the map $\mathfrak{S}(T) \to \Gamma_0(q)\backslash \mathrm{GL}_n(\mathbb{R})$ has fibers of size $\ll 1$.*

Proof. In next section. □

Proof. (of Theorem 1). We now turn to the proof of the Theorem in the case $S = S_1(q)$, the other case being similar. Fix $\nu_0 \in \mathcal{S}$, let U be a neighbourhood of ν_0 as in Proposition 2, and set $\Pi = \{\pi \in \mathcal{CP}(q) : \nu_\pi \in U\}$.

For some complex numbers $\mathbf{b} = (b_\pi)_{\pi\in\Pi}$, set $f = \sum_{\pi\in\Pi} b_\pi f_\pi^{\mathrm{new}} \in {}^\circ\mathcal{A}_n(q)$. Let $||\cdot||_2$ denote the L^2-norm on ${}^\circ\mathcal{A}_n(q)$ obtained by integrating over $\Gamma_1(q)\backslash \mathrm{PGL}_n(\mathbb{R})$. Then $||f||_2^2 = c[\mathrm{GL}_n(\mathbb{Z}) : \Gamma_1(q)] \sum_\pi |b_\pi|^2$, for an appropriate constant c that depends only on the choice of measures. This follows from (8); the extra factor $[\mathrm{GL}_n(\mathbb{Z}) : \Gamma_1(q)]$ arises in comparing measures on $\Gamma_1(q)\backslash \mathrm{PGL}_n(\mathbb{R})$ and $\mathrm{PGL}_n(\mathbb{Q})\backslash \mathrm{PGL}_n(\mathbb{A})$. Note that $q^{n-\epsilon} \ll_\epsilon [\mathrm{GL}_n(\mathbb{Z}) : \Gamma_1(q)] \ll q^n$.

It is then evident from Proposition 3 that if $T \ll q^{1/(n-1)}$, we have

$$\int_{\mathfrak{S}(T)} |f(g)|^2 dg \ll ||f||_2^2 \ll q^n \sum_{\pi\in\Pi} |b_\pi|^2 \tag{15}$$

Now (in the notation of the previous section) ψ_∞ defines a character of $N(\mathbb{Z})\backslash N(\mathbb{R})$. Let $\delta_m \in \mathrm{GL}_n(\mathbb{R})$ be as in the previous section. Then $n \mapsto \psi_\infty(\delta_m n \delta_m^{-1})$ define characters of $N(\mathbb{Z})\backslash N(\mathbb{R})$; these characters are orthogonal for distinct m. As in the previous section, we set

$$W_f(m, g) = \int_{N(\mathbb{Z})\backslash N(\mathbb{R})} \overline{\psi_\infty(\delta_m n \delta_m^{-1})} f(ng) dn$$

Applying Bessel's inequality, we obtain $\int_{N(\mathbb{Z})\backslash N(\mathbb{R})} |f(ng)|^2 dn \geq \sum_{m=1}^\infty |W_f(m, g)|^2$. Integrating over A_T^+ and using the Iwasawa decomposition, we obtain:

$$\int_{Z(\mathbb{R})\backslash\mathfrak{S}(T)} |f(g)|^2 dg \geq \sum_{m=1}^\infty \int_{Z(\mathbb{R})\backslash A_T^+} a^{-2\rho} |W_f(m, a)|^2 da \tag{16}$$

Combining (15) and (16), noting that $A_m^+ \subset A_T^+$ if $m \leq T$, and applying Proposition 2, we obtain:

$$\sum_{\substack{m\leq T\\(m,q)=1}} |\sum_{\pi\in\Pi} c(\nu_\pi) b_\pi a_\pi(m)|^2 \ll_{\nu_0} q^{n+\epsilon} \sum_{\pi\in\Pi} |b_\pi|^2$$

Dualizing, we obtain that for any sequence $\mathbf{d} = (d_m)_{1 \leq m \leq T}$, we have:

$$\sum_{\pi \in \Pi} | \sum_{\substack{m \leq T \\ (m,q)=1}} d_m c(\nu_\pi) a_\pi(m)|^2 \ll_{\nu_0} q^{n+\epsilon} \sum_{m \leq T} |d_m|^2 \tag{17}$$

Proposition 1 relates $a_\pi(m)$ to the coefficients of the L-series; and to remove the restriction $\pi \in \Pi$, we cover the original $\mathcal{S}$ with a finite number of Us as in Proposition 2. This latter step is independent of q. Using these, one deduces the Theorem from (17) (in the case $S = S_1(q)$). □

We observe that the use of Bessel's inequality is extremely wasteful for $n > 2$, as we "capture" only a small part – the abelian part – of the spectrum of $N(\mathbb{Z}) \backslash N(\mathbb{R})$. The Rankin-Selberg method is less wasteful; however, there are obstacles in using it in this context which the author does not know how to overcome.

5 Some lattice reduction theory

In this section, we prove Proposition 3, by translating it to a statement about lattices and using some reduction theory. A *lattice* for our purpose is a free $\mathbb{Z}$-module L of finite rank endowed with a positive definite quadratic form (i.e. $L \otimes \mathbb{R}$ is given the structure of a Euclidean space). We denote by $||\lambda||$ the length of a vector $\lambda \in L$.

Let L be a lattice and L' a subgroup of L, not necessarily of full rank. Then (by a slight abuse of notation) we denote by L/L' the lattice that is the projection of L onto $(L' \otimes \mathbb{R})^\perp$, the perpendicular being taken inside $(L \otimes \mathbb{R})$. If $x \in L$, we will denote by $||x||_{L/L'}$ the norm of the coset $x + L'$ in L/L'. It is possible for $||x||_{L/L'} = 0$ for $x \notin L'$ if L' is not "saturated," i.e. if $\mathbb{Q}.L' \cap L \neq L'$. Given $\lambda \in L$, we say λ is primitive if $\mathbb{Q}.\lambda \cap L = \mathbb{Z}\lambda$.

Generally, given a subset S of an abelian group, we set $\langle S \rangle$ to be the subgroup generated by S.

It is convenient to fix bases. Let $V_n = \mathbb{R}^n, V_{n,\mathbb{Z}} = \mathbb{Z}^n$. Let $e_1, \ldots, e_n$ be the standard basis vectors. Let Q be the "standard" quadratic form on V_n, so that $Q(e_i, e_j) = \delta_{ij}$. By means of Q, $V_{n,\mathbb{Z}}$ and all the discrete subgroups of V that we consider will become lattices. Let $G = GL(V_n) \equiv \mathrm{GL}_n(\mathbb{R}), \Gamma = GL(V_{n,\mathbb{Z}}) \equiv \mathrm{GL}_n(\mathbb{Z}), K = O(Q)$. We regard G as acting on V_n on the right, thus identifying G with the space of bases for V_n, via $g \mapsto (e_1 g, e_2 g, \ldots, e_n g)$. Then $\Gamma \backslash G/K$ is identified with the space of lattices of rank n. Let $N(\mathbb{R}), A$ be as before upper triangular and diagonal matrices, respectively, in the identification of G with $\mathrm{GL}_n(\mathbb{R})$.

One may also identify the Iwasawa decomposition with the Gram-Schmidt "orthogonalization" process. Indeed, given g with rows $x_1, \ldots, x_n$, let $g = ntk$ (with $n \in N(\mathbb{R}), t \in A, k \in K$); let $(y_i)_{1 \leq i \leq n}$ be the row vectors of k. Let $n_{j,l}$ and t_j denote the (j, l) entry of n (respectively the (j, j) entry of t). Then one has

- $||y_i|| = 1$, for $1 \leq i \leq n$, and $y_i \perp y_j$ if $i \neq j$.
- For each i, $y_i \in \langle x_i, x_{i+1}, \ldots, x_n \rangle$ but $y_i \perp \langle x_{i+1}, \ldots, x_n \rangle$.
- $x_i = t_i y_i + n_{i,i+1} t_{i+1} y_{i+1} + n_{i,i+2} t_{i+2} y_{i+2} + \ldots t_n n_{i,n} y_n$ for $1 \leq i \leq n$.

We use the following notion of reduced basis; it should be noted that it differs from the Minkowski notion of "reduced basis."

Definition 2. *A basis $(l_1, l_2, \ldots, l_n)$ for a lattice $L \subset V_n$ is* reduced *if the corresponding matrix $g = \begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix}$ belongs to $\mathfrak{S}(\sqrt{3}/2)$.*

If X is any lattice of rank n, let ι be any isometry $X \otimes_{\mathbb{Z}} \mathbb{R} \to V_n$; then a basis $(x_1, \ldots, x_n)$ is reduced if $(\iota(x_1), \iota(x_2), \ldots, \iota(x_n))$ is reduced as above.

One verifies easily that the above definition is independent of ι. Similarly one defines "reduced" for lattices of rank $< n$.

Reduction theory shows every lattice has a reduced basis, and the number of reduced bases is finite (Siegel's property). The lengths $||x_n||, ||x_{n-1}||, \ldots$ of this basis differ from the successive minima (in the sense of Minkowski) by amounts that are bounded only in terms of n. Further, if $(x_i)_{1 \leq i \leq n}$ is a reduced basis for $L \subset V_n$, and one decomposes $g = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ in the Iwasawa decomposition as $g = ntk$, as above, then one has $t_j \leq ||x_j|| \ll t_j$. One sees that if $(x_1, \ldots, x_n)$ is a reduced basis of X, then $(x_2, \ldots, x_n)$ is a reduced basis for $\langle x_2, \ldots, x_n \rangle$ and $(\overline{x_1}, \ldots, \overline{x_{n-1}})$ is a reduced basis for $X/\langle x_n \rangle$; here $\overline{x_j}$ denotes the image of x_j in $X/\langle x_n \rangle$.

Lemma 1. *Let (x_i) be a reduced basis for X, and set $a_i = ||x_i||$. Let $\lambda \in X$ be expressed as $\lambda = \sum_{i=1}^n \lambda^{(i)} x_i$. Then $|\lambda^{(i)}| \ll a_i^{-1} ||\lambda||$.*

Proof. This follows inductively. First, project onto $X/\langle x_2, x_3, \ldots, x_n \rangle$; let $\overline{\lambda}$ and $\overline{x_1}$ be the images of λ and x_1. Then $||\overline{x_1}||_{X/\langle x_2, x_3, \ldots, x_n \rangle} \gg a_1$, whereas $||\overline{\lambda}||_{X/\langle x_2, x_3, \ldots, x_n \rangle} \leq ||\lambda||$. It follows that $|\lambda^{(1)}| \ll a_1^{-1} ||\lambda||$.

Now replace λ by $\lambda' = \lambda - \lambda^{(1)}x_1$. It lies in the lattice $\langle x_2, \ldots, x_n \rangle$, and we may repeat the same argument to show that $|\lambda^{(2)}| \ll a_2^{-1}||\lambda'|| \ll a_2^{-1}||\lambda||$. Continuing in this way demonstrates the Lemma. □

Let $B_T = \{x \in V : ||x|| \leq T\}$, the ball of radius T. For any lattice X, let $\mathrm{vol}(X)$ be the covolume of X (with the volume on $X \otimes \mathbb{R}$ associated to the quadratic form). If $L' \subset L$ is saturated, one has $\mathrm{vol}(L')\mathrm{vol}(L/L') = \mathrm{vol}(L)$. If X has reduced basis $(x_i)_{1 \leq i \leq n}$, we have $\mathrm{vol}(X) \asymp \prod_{i=1}^n ||x_i||$.

Lemma 2. *Let the lattice X have reduced basis (x_i) with lengths $||x_i|| = a_i$. Then*

$$\#(X \cap B_T) \ll \prod_{i=1}^{n} \frac{T + a_i}{a_i} \ll \frac{T^n + a_1^n}{\mathrm{vol}(X)}$$

Proof. This follows immediately from the preceding remarks and the previous Lemma. □

Lemma 3. *Let X be as in Lemma 2. Let $x \in X$, $q \in \mathbb{Z}$, and let $X' = x + qX$. Then*

$$\#(X' \cap B_T) \ll \frac{1}{\mathrm{vol}(X)} \prod_{i=1}^{n} (T/q + a_i)$$

Proof. Given $x_1, x_2 \in X' \cap B_T$, we see that $x_1 - x_2 \in (qX) \cap B_{2T}$. We are reduced to Lemma 2, with X replaced by qX, T replaced by $2T$. □

For any lattice Y, let $\min(Y)$ be the minimal length of a nonzero vector from Y. Remark that if $y_1, \ldots, y_r$ is a reduced basis for Y, then $\min(Y) \asymp ||y_r||$, see comments after Definition 2.

Lemma 4. *Let X be a lattice with reduced basis $(x_1, x_2, \ldots, x_n)$. Let $x \in X, q \in \mathbb{Z}$ and let $X' = x + qX$; thus X' is a coset of qX in X.*

Let $T \ll q^{1/(n-1)}$. Then the number of primitive $\lambda \in X$ which satisfy conditions (1), (2), (3) below is $\ll 1$. Here:

1. $\lambda \notin \langle x_2, x_3, \ldots, x_n \rangle$.
2. $\lambda \in X'$.
3. $\min(X/\langle \lambda \rangle) \gg T^{-1}||\lambda||$.

(Recall remarks on notation at end of first section.)

Proof. Let N be the number of such λ. Suppose λ satisfies (1), (2), (3). We may write $\lambda = \sum_{i=1}^n \lambda^{(i)} x_i$ with $\lambda^{(1)} \neq 0$. Minkowski's result (on the first minimum of a lattice) shows that

$$\min(X/\langle \lambda \rangle) \ll (\mathrm{vol}(X)||\lambda||^{-1})^{1/(n-1)}$$

thus, if λ satisfies (3), we must have:

$$\mathrm{vol}(X)^{1/(n-1)}||\lambda||^{-1/(n-1)} \gg T^{-1}||\lambda||$$

This implies that

$$||\lambda|| \ll T^{(n-1)/n}\mathrm{vol}(X)^{1/n} \tag{18}$$

Lemma 3 now implies that:

$$N \ll \frac{1}{\mathrm{vol}(X)} \prod_{i=1}^{n} (T^{(n-1)/n}\mathrm{vol}(X)^{1/n}q^{-1} + a_i) \tag{19}$$

On the other hand, we may (for the purpose of bounding it) assume that $N \neq 0$. In particular, any element λ satisfying condition (1) must satisfy $||\lambda|| \gg a_1$; combining with (18):

$$T^{(n-1)/n}\mathrm{vol}(X)^{1/n} \gg ||\lambda|| \gg a_1$$

In particular, we may assume that $a_1 \ll T^{(n-1)/n}\mathrm{vol}(X)^{1/n}$. However, since $a_1 \gg a_2 \gg \cdots \gg a_n$ and $\prod_{i=1}^{n} a_i \asymp \mathrm{vol}(X)$, this implies that

$$a_n \gg \mathrm{vol}(X)a_1^{-(n-1)} \gg \mathrm{vol}(X)^{1/n}T^{-(n-1)^2/n} \tag{20}$$

Thus if $q > T^{n-1}$, we see that $T^{(n-1)/n}\mathrm{vol}(X)^{1/n}q^{-1} \ll a_n$. The bound (19) then shows that $N \ll 1$ as required. □

We also require a variant:

Lemma 5. *Let X be a lattice with reduced basis $(x_1, x_2, \ldots, x_n)$. Let $x \in X$ and let $X' = \langle x \rangle + qX$. Let $T \ll q^{\frac{1}{2n-2}}$.*

Then the number of primitive $\lambda \in X$ which satisfy conditions (1), (2), (3) above is $\ll 1$.

Proof. For $y \in X'$, we may write $y = \sum_{i=1}^{n} y^{(i)}x_i$. We first claim that, for any c:

$$\#\{y \in X' : y \text{ primitive in } X, |y^{(i)}| \leq c\sqrt{q}\} \ll 1 \tag{21}$$

Here the implicit constant may depend on c.

We endow X with a new quadratic form by declaring the x_i to be an orthonormal basis; we denote by $||\cdot||'$ the corresponding length function. Now let $x'_1, x'_2, \ldots, x'_n$ be a reduced basis of X' with respect to $||\cdot||'$ and set $a'_j = ||x'_j||'$. Let $\mathcal{R} \subset X = \{\sum_{i=1}^{n} \lambda^{(i)}x_i : |\lambda^{(i)}| < q/2\}$. Then $\mathcal{R}$ injects into X/qX, whereas the image of X' in X/qX has size $\leq q$; thus $|\mathcal{R} \cap X'| \leq q$. On the other hand, if $C = C(n)$ is chosen sufficiently small, $\mathcal{R}$ contains all elements of X, and therefore of X', with $||\cdot||'$-norm less than $C(n)q$. Thus $|\mathcal{R}\cap X'| \gg \frac{q^2}{a'_{n-1}a'_n}$ by inspection, so $a'_{n-1}a'_n \gg q$. In particular, since $a'_{n-1} \gg a'_n$, one notes $a'_{n-1} \gg \sqrt{q}$. Now suppose $y \in X'$ is in the set defined by the left hand side of (21); note that $||y||' \ll \sqrt{q}$. What we have just shown implies that, except for at most $\ll 1$ cases, any such y is a multiple of x'_n. However, only two multiples of x'_n are primitive. This implies (21).

We now modify the previous Proof. As before, any λ satisfying (1),(2),(3) also satisfies (18), and we may assume (as in (20)) that $a_n \gg \mathrm{vol}(X)^{1/n}T^{-(n-1)^2/n}$. Combine this with (21); we see that, with at most $\ll 1$ exceptions, any primitive $y \in X'$ has (usual) norm $||y|| \gg \sqrt{q}\cdot \mathrm{vol}(X)^{1/n}T^{-(n-1)^2/n}$. By assumption $q > T^{2(n-1)}$, thus $||y|| \gg T^{(n-1)/n}\mathrm{vol}(X)^{1/n}$ with at most $\ll 1$ exceptions. In view of (18) we are done. □

Proposition 4. *Let X be any lattice, and let $x \in X$. Set either $X' = x + q\mathbb{Z}$ or $X' = \langle x\rangle + q\mathbb{Z}$. In the former case, assume T is so that $T \ll q^{1/(n-1)}$; in the latter case, assume $T \ll q^{1/(2n-2)}$.*

Then the number of primitive $\lambda \in X'$ so that $\min(X/\langle\lambda\rangle) \gg T^{-1}||\lambda||$ is $\ll 1$.

Proof. Let $(x_1, \ldots, x_n)$ be a reduced basis for X. There is a minimal k, with $1 \le k \le n$, for which:

$$\lambda \in \langle x_k, \ldots, x_n\rangle \text{ but } \lambda \notin \langle x_{k+1}, \ldots, x_n\rangle$$

Set $Y = \langle x_k, \ldots, x_n\rangle$; note then that $\min(Y/\langle\lambda\rangle) \ge \min(X/\langle\lambda\rangle) \gg T^{-1}||\lambda||$. Note that $(x_k, \ldots, x_n)$ is a reduced basis for Y, where Y is endowed with the quadratic form induced from X. Note that the intersection $X' \cap Y$ is either empty, or of the form either $y + qY$ or $\langle y\rangle + qY$.

Now $\lambda \in Y$ satisfies the conditions of the previous Lemmas, with X replaced by Y, X' replaced by $X' \cap Y$.

This shows that – given k – the number of possibilities for λ is $\ll 1$; since $k \in \{1, 2, \ldots, n\}$ we see that the total number of possibilites for λ is still $\ll 1$. □

We may now complete the proof of Proposition 3.

Proof. (Of Proposition 3).

Fix $g \in \mathrm{GL}(n, \mathbb{R})$; set $v_i = e_i g$ to be the ith row of g, and let $X \subset \mathbb{R}^n$ be the lattice spanned by $\langle v_i\rangle_{i=1}^n$. Under the identification of $\mathrm{GL}(n, \mathbb{R})$ with bases for V, the translates γg, for $\gamma \in \Gamma_1(q)$, correspond to $\mathbb{Z}$-bases $(x_1, \ldots, x_n)$ for the lattice X so that $x_n \in v_n + qX$. Similarly the translates

γg, for $\gamma \in \Gamma_0(q)$, correspond to bases $(x_1, x_2, \ldots, x_n)$ so that $x_n \in \langle v_n \rangle + q.X$.

Then, using the relationship of the Iwasawa decomposition and the Gram-Schmidt process, we see that:

$$\gamma g \in \mathfrak{S}(T)$$

if and only if the following properties hold:

1. $||x_{n-1}||_{X/\langle x_n \rangle} \geq T^{-1}||x_n||$,
2. The images of $x_1, x_2, \ldots, x_{n-1}$ form a reduced basis for $X/\langle x_n \rangle$.
3. The projection of each x_i onto x_n, for $1 \leq i \leq n-1$, has length $\leq \frac{1}{2}||x_n||$.

Now, once x_n is chosen, it follows from (2) and the Siegel property that the images $\overline{x_1}, \overline{x_2}, \ldots, \overline{x_{n-1}}$ of $x_1, \ldots, x_{n-1}$ in $X/\langle x_n \rangle$ are specified up to a bounded number of possibilities (i.e., depending only on n). (3) now shows that there are a bounded number of possibilities for $x_1, \ldots, x_{n-1}$.

It now suffices to show that there is a bounded number of possibilities for x_n. In view of (2) one has $\min(X/\langle x_n \rangle) \asymp ||\overline{x_{n-1}}||_{X/\langle x_n \rangle}$; (1) then gives that $\min(X/\langle x_n \rangle) \gg T^{-1}||x_n||$. Now apply the the previous Proposition. □

6 Application

Theorem 1 may be applied to bound the number of forms with conductor q whose Fourier coefficients have prescribed behaviour. As an instance of this, we will use it to give a certain generalization of Duke's theorem from [5].

We have chosen to assume GRH to obtain a rather general result (although well short of the best result one would like). It is possible to give unconditional results, but the author does not know how to obtain them in the same generality or uniformity.

We say that a cuspidal automorphic representation π on $\mathrm{GL}_n(\mathbb{A})$ is associated to a Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$ if they match at all places under the local Langlands correspondence. Given a finite subgroup $G \leq \mathrm{GL}(n, \mathbb{C})$, let $N^G(n, q, \chi)$ be the number of such π on GL_n with conductor q and central character χ, such that the associated Galois representation ρ has image conjugate to G. For example, for GL(2), the possible G are partitioned into types: dihedral, tetrahedral, octahedal, icosahedral, and different bounds were given in [5] in each case.

Proposition 5. *Assume GRH. Let $e(G)$ be the exponent of the abelianization of G, i.e. $e(G) = \inf\{e : x^e \in [G,G]\ \forall x \in G\}$.*

Then there exists $\delta = \delta(n)$ so that $N^G(n,q,\chi) \ll_G q^{n-1-\frac{\delta}{e(G)}}$.

Neither the exponent nor the implicit constant is independent of G. The more serious of these – the exponent dependency – is not too bad, especially since in many interesting cases (e.g. icosahedral with $n = 2$) the group G is close to being simple and has very small abelianization.

Proof. (Sketch) Let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$ be a Galois representation, and let G be its image. By Jordan's theorem, there exists a normal abelian subgroup A in G of index bounded by $f(n)$, a function only of n. A then fixes a line in $\mathbb{C}^n$. By averaging over G/A we find an element $p \in \mathrm{Sym}^{[G:A]}\mathbb{C}^n$ so that $gp = \chi(g)p$ for some character χ of G. In particular, $\mathrm{Sym}^r \rho$ contains a 1-dimensional representation for some $r \leq f(n)$. Call it $\theta : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathbb{C}^\times$. θ factors through G; in view of the definitions, the character $\theta^{e(G)}$ is trivial. In particular, $\mathrm{Sym}^{re(G)}\rho$ contains a trivial subrepresentation.

Consider now the Artin L-function $L(s, \mathrm{Sym}^{re(G)}\rho)$. It has a pole (possibly a multiple pole) at $s = 1$ and its conductor is bounded by a power of q. The assumption of *GRH* shows that $\sum_{p \leq P} \lambda_\pi(p^{re(G)}) \gg_G P/\log(P)$, so long as $P \geq q^\alpha$ for some positive α. (One can deduce this even from [8]: if L is the field extension of $\mathbb{Q}$ defined by the kernel of ρ, one checks that the discriminant of L is bounded by a power of q depending only on G and n). Now one may proceed as in [5], making crucial use of the large-sieve inequality of Theorem 1. Note that the assumption $\omega_\pi \gg q^\epsilon$ is automatic for forms associated to Galois representations, the argument being identical in general to that in [5]. (Indeed, this is true for any π such that all local constituents are tempered.) □

Remarks.

1. This may be regarded as a generalization of Duke's bound, [5], for $n = 2$. In general q^{n-1} may be regarded as the "trivial bound," as is suggested by limit multiplicity formulas, c.f. discussion after Theorem 1; thus for any *particular* G, the above result improves upon this, showing the scarceness of such forms.

   This result is, however, considerably less satisfactory than Duke's result and later generalizations. Since there are an infinite number of possible groups G, we have no control of the set of *all* automorphic forms of Galois type at once. (For instance, this means in the setting of Duke's paper that we would obtain good bounds on the number of icosahedral, octahedral and tetrahedral forms; but we have no uniform way of treating all dihedral forms at once).

Underlying this failure to deal with the infinite families of groups G is a fundamental issue: the large sieve we have given, like the large sieve of [6], only controls the coefficients of the standard L-series – but does not control very well the coefficients of (for example) the exterior square L-series. This phenomenon is not important for $n = 2$ but becomes significant for $n > 2$. It seems likely that with this stronger type of large sieve one could get better uniformity.

2. In particular, if one assumes the Strong Artin Conjecture, the theorem implies that *the number of Galois representations of degree n with image conjugate to G, fixed central character, and conductor q* is $\ll q^{n-1-\delta}$ for some $\delta = \delta(G) > 0$.

   This is probably very far from sharp for large n, although it is unclear to the author what the truth of the situation is. It seems conceivable that the number of such Galois representations is $\ll q^C$ where C is independent of n. Indeed this may even be true without specifying the group G. Even more ambitiously (and vaguely) one might hope the answer is $\ll q^{\epsilon}$ if G does not contain "large abelian subgroups,"e.g. dihedral case when $n = 2$. This seems very difficult to prove.

3. Owing to the fact that Theorem 1 does not restrict to a single π_∞, one can also obtain bounds for the number of modular forms associated to Weil group representations. We also refer the reader also to [1] where certain related *finiteness* results are proven.

7 The L^2 norm of the new vector

This is a beautiful exercise in harmonic analysis on $\mathrm{GL}(n, \mathbb{C})$: to compute the L^2 norm of the *essential vector* of Jacquet,Piatetski-Shapiro and Shalika. This is required for Proposition 1.

Let p be a finite prime. We shall now work over $\mathbb{Q}_p$. Let N_n be the upper triangular unipotent matrices in $G_n = \mathrm{GL}_n(\mathbb{Q}_p)$. G_{n-1} is the $\mathrm{GL}_{n-1}(\mathbb{Q}_p)$ embedded in the upper left hand corner of $\mathrm{GL}_n(\mathbb{Q}_p)$. Set $N_{n-1} = N_n \cap G_{n-1}$.

Let π be a unitary generic representation of G_n of (local) conductor p^f. Let W_π be the Whittaker function corresponding to a new vector, normalized as in [7], Theorem 4.1. Let $A \in \mathrm{GL}_{n-1}(\mathbb{C})$ be semisimple, and let σ_A be the spherical representation of G_{n-1} with "Hecke matrix" A; we assume that

A is chosen so that this representation is generic. Let W_A be the spherical Whittaker function associated to σ_A, transforming under the character ψ_p of N_{n-1} (see (3)) and normalized so that $W_A(1) = 1$. Let $\mathbf{f} = (f_1, \ldots, f_{n-1}) \in \mathbb{Z}^{n-1}$; we say $\mathbf{f} \succ 0$ if $f_j \geq f_{j+1}$ for $1 \leq j \leq n-2$. For such $\mathbf{f}$, set $|\mathbf{f}| = \sum_i f_i$. We normalize measures to assign mass 1 to the maximal compact $K_{n-1} \subset G_{n-1}$ consisting of integral matrices, and to also assign mass 1 to $K_{n-1} \cap N_{n-1}$. Note that W_π is K_{n-1}-invariant, and in fact $W_\pi(1) = 1$, as we shall see.

Let $p^{\mathbf{f}} = \mathrm{diag}(p^{f_1}, p^{f_2}, \ldots, p^{f_{n-1}}) \in G_{n-1}$. Let $\rho : \mathrm{diag}(x_1, \ldots, x_{n-1}) \mapsto |x_1^{\frac{n-2}{2}} x_2^{\frac{n-4}{2}} \ldots x_{n-1}^{\frac{2-n}{2}}|_p$ be the half-sum of positive roots for the diagonal torus of G_{n-1}, with respect to the positive system corresponding to N_{n-1}.

Let $\chi_{\mathbf{f}}$ be the character of $\mathrm{GL}(n-1, \mathbb{C})$ associated to the finite dimensional (algebraic) representation with the highest weight $\mathbf{f}$ (e.g. the representation with highest weight $(1, 0, \ldots, 0)$ is the standard representation, and $(0, 0, \ldots, 1)$ is the dual of the standard representation). Shintani's formula gives:

$$W_A(p^{\mathbf{f}}) = \rho(p^{\mathbf{f}}) \chi_{\mathbf{f}}(A)$$

One knows from [7] that $\int_{N_{n-1} \backslash G_{n-1}} W_\pi(g) W_A(g) |\det(g)|^{s-1/2} dg = L(s, \pi \times \sigma_A)$. (Note that taking the $s \to \infty$ limit gives $W_\pi(1) = 1$.) In concrete terms, this expresses the following equality, where both sides are convergent for $\Re(s) \gg 1$:

$$\sum_{\mathbf{f} \succ 0} W_\pi(p^{\mathbf{f}}) \rho(p^{\mathbf{f}})^{-1} p^{-|\mathbf{f}|(s-1/2)} \chi_{\mathbf{f}}(A) = L(s, \pi \times \sigma_A) \tag{22}$$

Let Ξ_s be the formal virtual character of $\mathrm{GL}(n-1, \mathbb{C})$ which contains $\chi_{\mathbf{f}}$ with multiplicity $W_\pi(p^{\mathbf{f}}) \rho(p^{\mathbf{f}})^{-1} p^{-|\mathbf{f}|(s-1/2)}$. Then

$$\int_{N_{n-1} \backslash G_{n-1}} |W_\pi(g)|^2 |\det(g)|^{(s-1/2)} dg = \sum_{\mathbf{f} \succ 0} |W_\pi(p^{\mathbf{f}})|^2 \rho(p^{\mathbf{f}})^{-2} p^{-|\mathbf{f}|(s-1/2)}$$
$$= \langle \Xi_s, \Xi_{1/2} \rangle \tag{23}$$

where the inner product should be regarded formally. We proceed formally, but the computations that follow can be justified when $\Re(s) \gg 1$ by replacing, in the arguments that follow, $\Xi_{1/2}$ and Ξ_s by Ξ_α and Ξ_β so that $\alpha + \overline{\beta} = s + 1/2$, and choosing $\Re(\alpha) \gg 1, \Re(\beta) \gg 1$.

In any case, formally $\langle \Xi_s, \Xi_{1/2} \rangle$ may be computed by integrating $\Xi_s \overline{\Xi_{1/2}}$, considered as a function of $A \in \mathrm{GL}(n-1, \mathbb{C})$, over the space of unitary matrices $U(n-1)$. On the other hand (22) explicitly evaluates Ξ_s; applying it:

$$\int_{N_{n-1}\backslash G_{n-1}} |W_\pi(g)|^2 |\det(g)|^{(s-1/2)} dg = \int_{U(n-1)} L(s, \pi \times \sigma_A) \overline{L(1/2, \pi \times \sigma_A)} dA$$

the integral being taken over $U(n-1)$, where the Haar measure has total mass 1. Let A_π be so that $L(s,\pi) = \det(1 - A_\pi p^{-s})$; thus A_π is a matrix, possibly belonging to $\mathrm{GL}(m, \mathbb{C})$ for $m < n$. Then one obtains:

$$\int_{N_{n-1}\backslash G_{n-1}} |W_\pi(g)|^2 |\det(g)|^{s-1/2} dg$$
$$= \int_{U(n-1)} \det(1 - A_\pi \otimes A p^{-s}) \overline{\det(1 - A_\pi \otimes A p^{-1/2})} dA = \det(1 - A_\pi \otimes \overline{A_\pi} p^{-s-1/2}) \tag{24}$$

Here the last line is a direct computation in invariant theory which is valid only so long as π is ramified, i.e $f \geq 1$. It amounts to decomposing $\mathrm{Sym}^r(\mathbb{C}^p \otimes \mathbb{C}^q)$ under the $\mathrm{GL}_p(\mathbb{C}) \times \mathrm{GL}_q(\mathbb{C})$ action. In the unramified case the result must be modified and in any case can be computed directly from the unramified evaluation of Rankin-Selberg integrals.

As remarked above this computation may be justified when $\Re(s) \gg 1$. Finally, both sides of (24) define positive Dirichlet series, and the equality (24) is valid up to the first (common) pole of both sides (it is even valid everywhere if one interprets both sides in an appropriate sense using meromorphic continuation). In particular, it is valid for $\Re(s) \geq 1/2$.

We are now ready to present the proof of Proposition 1.

Proof. (of Proposition 1.) We follow the notation established in Section 3; in particular $\varphi_\pi^{\mathrm{new}}$ is the new vector for $\pi \in \mathcal{CP}(q)$, and f_π^{new} defined as (8). $\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle$ will denote $\int_{\mathrm{PGL}_n(\mathbb{Q})\backslash \mathrm{PGL}_n(\mathbb{A})} |\varphi^{\mathrm{new}}(g)|^2 dg$. Let $g_\infty \in \mathrm{GL}_n(\mathbb{R})$. For $g = (g_\infty, 1) \in \mathrm{GL}_n(\mathbb{A}) = \mathrm{GL}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{A}_f)$, we have:

$$\begin{aligned} W_{f_\pi^{\mathrm{new}}}(m, g_\infty) &= \int_{N(\mathbb{Z})\backslash N(\mathbb{R})} \overline{\psi_\infty(\delta_m n \delta_m^{-1})} f_\pi^{\mathrm{new}}(n g_\infty) dn \\ &= \frac{1}{\sqrt{\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle}} \int_{N(\mathbb{Q})\backslash N(\mathbb{A})} \overline{\psi_{\mathbb{A}}(\delta_m n \delta_m^{-1})} \varphi_\pi^{\mathrm{new}}(ng) dn \\ &= \frac{1}{\sqrt{\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle}} \int_{N(\mathbb{Q})\backslash N(\mathbb{A})} \overline{\psi_{\mathbb{A}}(n)} \varphi_\pi^{\mathrm{new}}(n \delta_m g) dn = \frac{W_{\varphi_\pi^{\mathrm{new}}, \mathbb{A}}(\delta_m g)}{\sqrt{\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle}} \end{aligned}$$

Here we have used (9), (8), the substitution $n \leftarrow \delta_m^{-1} n \delta_m$, and (5) respectively. (6) and (7) now give:

$$W_{f_\pi^{\mathrm{new}}}(m, g_\infty) = \frac{1}{\sqrt{\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle}} W_{\nu_\pi}(\delta_m g_\infty) m^{-(n-1)/2} \chi(m) \overline{\lambda_\pi(m)} \tag{25}$$

In particular, Definition 1 gives:

$$a_\pi(m) = \frac{\chi(m)\overline{\lambda_\pi(m)}}{\sqrt{\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle}} \tag{26}$$

To analyze $\langle \varphi^{\mathrm{new}}, \varphi^{\mathrm{new}} \rangle$, we utilise the Rankin-Selberg integral due to Jacquet, Piatetski-Shapiro, and Shalika. We refer to Cogdell's survey [3] to fix notation. Let Φ be a Schwarz function on $\mathbb{A}^n$, factorizing as $\Phi = \prod_v \Phi_v$. Fix matters so that for v not dividing ∞ or q, Φ_v is the characteristic function of $\mathbb{Z}_v^n$. Choose Φ_∞ as any positive Schwarz function on $\mathbb{R}^n$, and for $p|q$, set Φ_p to be the characteristic function in $\mathbb{Z}_p^n$ of the inverse image of $(0, \dots, 0, 1) \in (\mathbb{Z}/p^{f_p})^n$, where p^{f_p} is the power of p that occurs in q. We fix the standard Haar measures on $\mathbb{Q}_v$ and $\mathbb{A}$.

For $g \in \mathrm{GL}_n(\mathbb{Q}_v)$, let $b(g) \in (\mathbb{Q}_v)^n$ denote the bottom row of g. Then the work of Jacquet, Piatetski and Shalika gives:

$$\begin{aligned} &\langle \varphi_\pi^{\mathrm{new}}, \varphi_\pi^{\mathrm{new}} \rangle \cdot \int_{\mathbb{A}^n} \Phi(x) dx = \\ &\quad \lim_{s \to 1} \frac{1}{s-1} \prod_v \int_{N(\mathbb{Q}_v) \backslash \mathrm{GL}_n(\mathbb{Q}_v)} |W_{\varphi_{\pi,v}^{\mathrm{new}}}(g_v)|^2 \Phi(b(g_v)) |\det(g)|^s dg_v \\ &= \lim_{s \to 1} \frac{L^{(q,\infty)}(s, \pi \times \tilde{\pi})}{s-1} \prod_{v=\infty \text{ or } v|q} \int_{N(\mathbb{Q}_v) \backslash \mathrm{GL}_n(\mathbb{Q}_v)} |W_{\varphi_{\pi,v}^{\mathrm{new}}}(g_v)|^2 \Phi(b(g_v)) |\det(g)| dg_v \end{aligned} \tag{27}$$

Here $L^{(q,\infty)}$ is the partial L-function, omitting factors at $v|q$ and ∞. Set $I_v = \int_{N(\mathbb{Q}_v) \backslash \mathrm{GL}_n(\mathbb{Q}_v)} |W_{\varphi_{\pi,v}^{\mathrm{new}}}(g_v)|^2 \Phi(b(g_v)) |\det(g)| dg_v$. If $v = \infty$, it is clear (since π_∞ belongs to the compact set $\mathcal{S}$) that $I_\infty \asymp_{\mathcal{S}} 1$. On the other hand, for $p|q$ we may evaluate I_p using (24); one obtains after some routine computation

$$I_p = p^{-f_p n}(1 - p^{-n})^{-1} L(1, A_{\pi_p} \times \overline{A_{\pi_p}})$$

Here A_{π_p} is so that the local factor for $L(s, \pi)$ at p is $\det(1 - A_{\pi_p} p^{-s})$, and $L(1, A_{\pi_p} \times \overline{A_{\pi_p}}) \equiv \det(1 - A_{\pi_p} \otimes \overline{A_{\pi_p}} p^{-1})^{-1}$. The Luo-Rudnick-Sarnak bounds, [9], guarantee that $p^{-\epsilon} \ll L(1, A_{\pi_p} \times \overline{A_{\pi_p}}) \ll p^{\epsilon}$. (Note these bounds are valid even at ramified places, as may be deduced by a formal argument involving base change, due to Clozel.) In particular, one sees that: $q^{-n-\epsilon} \ll_\epsilon \prod_{v|\infty,q} I_v \ll_\epsilon q^{-n+\epsilon}$. On the other hand $\int_{\mathbb{A}^n} \Phi(x)dx = q^{-n} \int_{\mathbb{R}^n} \Phi(x_\infty)dx_\infty$, and we obtain:

$$q^{-\epsilon} \ll_{\epsilon,\mathcal{S}} \omega_\pi \langle \varphi^{\text{new}}, \varphi^{\text{new}} \rangle \ll_{\epsilon,\mathcal{S}} q^{\epsilon}$$

Combining this with (26) completes the proof of Proposition 1. □

References

[1] Greg Anderson, Don Blasius, Robert Coleman, and George Zettler. On representations of the Weil group with bounded conductor. *Forum Math.*, 6(5):537–545, 1994.

[2] A. Borel and H. Jacquet. Automorphic forms and automorphic representations. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1*, Proc. Sympos. Pure Math., XXXIII, pages 189–207. Amer. Math. Soc., Providence, R.I., 1979. With a supplement "On the notion of an automorphic representation" by R. P. Langlands.

[3] James Cogdell. L-functions and converse theorems for $GL(n)$. *To appear.*

[4] David L. DeGeorge and Nolan R. Wallach. Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$. II. The tempered spectrum. *Ann. of Math. (2)*, 109(3):477–495, 1979.

[5] W. Duke. The dimension of the space of cusp forms of weight one. *Internat. Math. Res. Notices*, (2):99–109 (electronic), 1995.

[6] W. Duke and E. Kowalski. A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations. *Invent. Math.*, 139(1):1–39, 2000. With an appendix by Dinakar Ramakrishnan.

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

[8] J. C. Lagarias and A. M. Odlyzko. Effective versions of the Chebotarev density theorem. In *Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975)*, pages 409–464. Academic Press, London, 1977.

[9] Wenzhi Luo, Zeév Rudnick, and Peter Sarnak. On the generalized Ramanujan conjecture for GL(n). In *Automorphic forms, automorphic representations, and arithmetic (Fort Worth, TX, 1996)*, volume 66 of *Proc. Sympos. Pure Math.*, pages 301–310. Amer. Math. Soc., Providence, RI, 1999.

[10] Giuseppe Molteni. Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product. *Duke Math. J.*, 111(1):133–158, 2002.

[11] Takuro Shintani. On an explicit formula for class-1 "Whittaker functions" on GL_n over P-adic fields. *Proc. Japan Acad.*, 52(4):180–182, 1976.