AMPLIFICATION ARGUMENTS FOR LARGE SIEVE INEQUALITIES

E. KOWALSKI

Abstract. We give a new proof of the arithmetic large sieve inequality based on an amplification argument, and use a similar method to prove a new sieve inequality for classical holomorphic cusp forms. A sample application of the latter is also given.

1. The classical large sieve

The classical arithmetic large sieve inequality states that, for any real numbers N, $Q \geqslant 1$, any choice of subsets $\Omega_p \subset \mathbf{Z}/p\mathbf{Z}$ for primes $p \leqslant Q$, we have

$$|\{n \leqslant N \mid n \pmod{p} \notin \Omega_p \text{ for } p \leqslant Q\}| \leqslant \frac{\Delta}{H} \tag{1}$$

where

$$H = \sum_{q \leqslant Q}^{\flat} \prod_{p \mid q} \frac{|\Omega_p|}{p - |\Omega_p|},$$

and Δ is any constant for which the "harmonic" large sieve inequality holds: for any complex numbers (a_n), we have

$$\sum_{q \leqslant Q} \sum_{a \pmod{q}}^{*} \Bigl| \sum_{n \leqslant N} a_n e\Bigl(\frac{an}{q}\Bigr) \Bigr|^2 \leqslant \Delta \sum_{n \leqslant N} |a_n|^2, \tag{2}$$

the notation $\sum^{\flat}$ and $\sum^{*}$ denoting, respectively, a sum over squarefree integers, and one over integers coprime with the (implicit) modulus, which is q here.

By work of Montgomery-Vaughan and Selberg, it is known that one can take

$$\Delta = Q^2 - 1 + N$$

(see, e.g., [7, Th. 7.7]).

There are a number of derivations of (1) from (2); for one of the earliest, see [10, Ch. 3]. The most commonly used is probably the argument of Gallagher involving a "submultiplicative" property of some arithmetic function (see, e.g., [8, §2.2] for a very general version).

2000 *Mathematics Subject Classification.* Primary 11N35, 11F11.

Key words and phrases. Large sieve inequality, modular form, amplification.

This work was completed while on sabbatical leave at the Institute for Advanced Study (Princeton, NJ). This material is based upon work supported by the National Science Foundation under agreement No. DMS-0635607. Any opinions, findinds and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We will show in this note how to prove (1) quite straightforwardly from the *dual* version of the harmonic large sieve inequality: Δ is also any constant for which

$$\sum_{n\leqslant N}\Bigl|\sum_{q\leqslant Q}\ \sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}\beta(q,a)e\Bigl(\frac{an}{q}\Bigr)\Bigr|^2\leqslant\Delta\sum_{q\leqslant Q}\ \sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}|\beta(q,a)|^2, \tag{3}$$

holds for arbitrary complex numbers $(\beta(q,a))$. This is of some interest because, quite often,[1] the inequality (2) is proved by duality from (3), and because, in recent generalized versions of the large sieve (see [8]), it often seems that the analogue of (3) is the most natural inequality to prove – or least, the most easily accessible. So, in some sense, one could dispense entirely with (2) for many applications! In particular, note that both known proofs of the optimal version with $\Delta = N-1+Q^2$ proceed by duality.

Note that some ingredients of many previous proofs occur in this new argument. Also, there are other proofs of (1) working directly from the inequality (3) which can be found in the older literature on the large sieve, usually with explicit connections with the Selberg sieve (see the references to papers of Huxley, Kobayashi, Matthews and Motohashi in [11, p. 561]), although none of those that the author has seen seems to give an argument which is exactly identical or as well motivated. Also, traces of this argument appear earlier in some situations involving modular forms, e.g., in [4]. In Section 2, we will use the same method to obtain a new type of sieve inequality for modular forms; in that case, it doesn't seem possible to adapt easily the classical proofs.

Indeed, maybe the most interesting aspect of our proof is that it is very easy to motivate. It flows very nicely from an attempt to improve the earlier inequality

$$|\{n\leqslant N\ |\ n\,(\mathrm{mod}\,p)\notin\Omega_p \text{ for } p\leqslant Q\}|\leqslant\frac{\Delta}{K},\qquad K=\sum_{p\leqslant Q}\frac{|\Omega_p|}{p}, \tag{4}$$

of Rényi, which is most easily proved using (3) instead of (1), as in [8, §2.4].

We will explain this quite leisurely; one could be much more concise and direct (as in Section 2).

Let

$$S=\{n\leqslant N\ |\ n\,(\mathrm{mod}\,p)\notin\Omega_p \text{ for } p\leqslant Q\},$$

be the sifted set; we wish to estimate from above the cardinality of this finite set. From (3), the idea is to find an "amplifier" of those integers remaining in the sifted set, i.e., an expression of the form

$$A(n)=\sum_{q\leqslant Q}\ \sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}\beta(q,a)e\Bigl(\frac{an}{q}\Bigr)$$

which is *large* (in some sense) when $n\in S$. Then an estimate for $|S|$ follows from the usual Chebychev-type manoeuvre.

To construct the amplifier $A(n)$, we look first at a single prime $p\leqslant Q$. If $n\in S$, we have $n\,(\mathrm{mod}\,p)\notin\Omega_p$. If we expand the characteristic function of Ω_p in terms of additive

[1] But not always – Gallagher's very short proof, found e.g. in [11, Th. 1, p. 549], proceeds directly.

characters,[2] we have then

$$0 = \mathbf{1}_{\Omega_p}(n) = \sum_{a \,(\mathrm{mod}\, p)} \alpha(p,a) e\Big(\frac{an}{p}\Big), \qquad \alpha(p,a) = \frac{1}{p} \sum_{x \in \mathbf{Z}/p\mathbf{Z}} \mathbf{1}_{\Omega_p}(x) e\Big(\frac{ax}{p}\Big),$$

and the point is that the contribution of the constant function (0-th harmonic) is, indeed, relatively "large", because it is

$$\alpha(p,0) = \frac{|\Omega_p|}{p},$$

and exactly reflects the probability of a random element being in Ω_p. Thus for $n \,(\mathrm{mod}\, p) \notin \Omega_p$, we have

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} \beta(p,a) e\Big(\frac{an}{p}\Big) = c_p \tag{5}$$

with

$$c_p = \frac{|\Omega_p|}{p}, \qquad \beta(p,a) = -\alpha(p,a).$$

If we only use the contribution of the primes in (3), and the amplifier

$$A(n) = \sum_{p \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} \beta(p,a) e\Big(\frac{an}{p}\Big),$$

then by (3), we get

$$\sum_{n \in S} |A(n)|^2 \leqslant \sum_{n \leqslant N} |A(n)|^2 \leqslant \Delta \sum_{p \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} |\beta(p,a)|^2.$$

For $n \in S$, the size of the amplifier is

$$|A(n)|^2 = \Big| \sum_{p \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} \beta(p,a) e\Big(\frac{an}{p}\Big) \Big|^2 = \Big| \sum_{p \leqslant Q} c_p \Big|^2 = K^2,$$

by (5), while on the other hand, by applying the Parseval identity in $\mathbf{Z}/p\mathbf{Z}$, we get

$$\begin{aligned} \sum_{p \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} |\beta(p,a)|^2 &= \sum_{p \leqslant Q} \Big(\frac{1}{p} \sum_{x \in \mathbf{Z}/p\mathbf{Z}} |\mathbf{1}_{\Omega_p}(x)|^2 - \alpha(p,0)^2 \Big) \\ &= \sum_{p \leqslant Q} c_p(1-c_p) \leqslant K. \end{aligned}$$

So we obtain

$$K^2 |S| \leqslant \Delta K,$$

i.e., exactly Rényi's inequality (4), by this technique.

To go further, we must exploit all the squarefree integers $q \leqslant Q$ (and not only the primes) to construct the amplifier. This is most easily described using the Chinese Remainder Theorem to write

$$\mathbf{Z}/q\mathbf{Z} \simeq \prod_{p \mid q} \mathbf{Z}/p\mathbf{Z}, \qquad (\mathbf{Z}/q\mathbf{Z})^\times \simeq \prod_{p \mid q} (\mathbf{Z}/p\mathbf{Z})^\times,$$

[2] We use this specific basis to use (3), but any orthonormal basis containing the constant function 1 would do the job, as in [8].

and putting together the amplifiers modulo primes $p \mid q$: if $n \in S$ then $n \pmod p \notin \Omega_p$ for all $p \mid q$, and hence multiplying out (5) over $p \mid q$, we find constants $\beta(q,a) \in \mathbf{C}$, defined for $(a,q) = 1$ (because $\beta(p,a)$ is defined for a coprime with p), such that

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \beta(q,a) e\Big(\frac{an}{q}\Big) = \prod_{p \mid q} c_p.$$

Moreover, because the product decomposition of the Chinese Remainder Theorem is compatible with the Hilbert space structure involved, we have

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} |\beta(q,a)|^2 = \prod_{p \mod q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} |\beta(p,a)|^2 = \prod_{p \mid q} c_p(1-c_p).$$

Arguing as before, we obtain from (3) – using all squarefree moduli $q \leqslant Q$ this time – that

$$|S| \leqslant \Delta \frac{A}{B^2}, \tag{6}$$

with

$$A = \sideset{}{^\flat}\sum_{q \leqslant Q} \prod_{p \mid q} c_p(1-c_p), \qquad B = \sideset{}{^\flat}\sum_{q \leqslant Q} \prod_{p \mid q} c_p.$$

This is not quite (1), but we have some flexibility to choose another amplifier, namely, notice that this expression is not homogeneous if we multiply the coefficients $\beta(q,a)$ by scalars independent of a, and we can use this to find a better inequality. Precisely, let

$$\gamma(q,a) = \Big(\prod_{p \mid q} \xi_p\Big) \beta(q,a),$$

where ξ_p are arbitrary real coefficients.

Then we have the new amplification property

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \gamma(q,a) e\Big(\frac{an}{q}\Big) = \prod_{p \mid q} \xi_p c_p$$

with altered "cost" given by

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} |\gamma(q,a)|^2 = \prod_{p \mid q} \xi_p^2 c_p(1-c_p),$$

so that, arguing as before, we get

$$|S| \leqslant \Delta \frac{A_1}{B_1^2}$$

with

$$A_1 = \sideset{}{^\flat}\sum_{q \leqslant Q} \prod_{p \mid q} \xi_p^2 c_p(1-c_p), \qquad B_1 = \sideset{}{^\flat}\sum_{q \leqslant Q} \prod_{p \mid q} \xi_p c_p.$$

By homogeneity, the problem is now to minimize a quadratic form (namely A_1) under a linear constraint given by B_1. This is classical, and is done by Cauchy's inequality: writing

$$c_q = \prod_{p \mid q} c_p, \qquad \tilde{c}_q = \prod_{p \mid q} (1-c_p), \qquad \xi_q = \prod_{p \mid q} \xi_p$$

for ease of notation, we have

$$B_1^2 = \Bigl(\sum_{q\leqslant Q}{}^{\flat}\, \xi_q c_q\Bigr)^2 \leqslant \Bigl(\sum_{q\leqslant Q}{}^{\flat}\, \xi_q^2 c_q \tilde{c}_q\Bigr)\Bigl(\sum_{q\leqslant Q}{}^{\flat}\, \frac{c_q}{\tilde{c}_q}\Bigr) = A_1 H,$$

with equality if and only if ξ_p is proportional to

$$\xi_p = \frac{1}{1-c_p} = \frac{p}{p-|\Omega_p|},$$

in which case

$$c_p\xi_p = \xi_p^2 c_p(1-c_p) = \frac{|\Omega_p|}{p-|\Omega_p|},$$

and we get $A_1 = B_1 = H$, hence $|S| \leqslant \Delta H^{-1}$, which is (1).

Remarks. (1) The last optimization step is reminiscent of the Selberg sieve (see, e.g., [7, p. 161, 162]). Indeed, it is well known that the Selberg sieve is related to the large sieve, and particularly with the dual inequality (3), as explained in [5, p. 125]. Note however that the coefficients we optimize for, being of an "amplificatory" nature, are different from the coefficents λ_d typically sought for in Selberg's sieve, which are akin to the Möbius function and of a "mollificatory" nature.

(2) The argument does not use any particular feature of the classical sieve, and thus extends immediately to provide a proof of the general large sieve inequality of [8, Prop. 2.3] which is directly based on the dual inequality [8, Lemma 2.8]; readers interested in the formalism of [8] are encouraged to check this.

Example. What are the amplifiers above in some simple situations? In the case – maybe the most important – where we try to count primes, we then take $\Omega_p = \{0\}$ to detect integers free of small primes by sieving, and (5) becomes

$$\sum_{a\,(\mathrm{mod}\,p)}{}^{*} \Bigl(-\frac{1}{p}\Bigr)\cdot e\Bigl(\frac{an}{p}\Bigr) = \frac{1}{p},$$

if $p \nmid n$. Then, for q squarefree, the associated detector is the identity

$$\sum_{a\,(\mathrm{mod}\,q)}{}^{*} \frac{\mu(q)}{q} e\Bigl(\frac{an}{q}\Bigr) = \frac{1}{q},$$

if $(n, q) = 1$, or in other words, it amounts to the well-known formula

$$\sum_{a\,(\mathrm{mod}\,q)}{}^{*} e\Bigl(\frac{an}{q}\Bigr) = \mu(q)$$

for the values of a Ramanujan sum with coprime arguments. Note that in this case, the optimization process above replaced $c_p = \frac{1}{p}$ with

$$\xi_p c_p = \frac{1}{p-1},$$

which is not a very big change – and indeed, for small sieves, the bound (6) is not far from (1), and remains of the right order of magnitude.

On the other hand, for an example in a large sieve situation, we can take Ω_p to be the set of squares in $\mathbf{Z}/p\mathbf{Z}$. The characteristic function (for odd p) is

$$\mathbf{1}_{\Omega_p}(x) = \sum_{a \,(\mathrm{mod}\, p)} \tau(p,a) e\Big(\frac{ax}{p}\Big)$$

with coefficients given – essentially – by Gauss sums

$$\tau(p,a) = \frac{1}{p}\Big(1 + \frac{1}{2} \sideset{}{^*}\sum_{x \,(\mathrm{mod}\, p)} e\Big(\frac{ax^2}{p}\Big)\Big).$$

Then c_p tends to $1/2$ as $p \to +\infty$, while $\xi_p c_p$ tends to 1. This difference leads to a discrepancy in the order of magnitude of the final estimate: using standard results on bounds for sums of multiplicative functions, (6) and taking $Q = \sqrt{N}$, we get

$$|S| \ll \sqrt{N}(\log N)^{1/4},$$

instead of $|S| \ll \sqrt{N}$ that follows from (1).

2. Sieving for modular forms

To illustrate the possible usefulness of the proof given in the first section, we use the same technique to prove a new type of large sieve inequality for classical (holomorphic) modular forms. The originality consists in using known inequalities for Fourier coefficients (due to Deshouillers-Iwaniec) as a tool to obtain a sieve where the cusp forms are the objects of interest, i.e., to bound from above the number of cusp forms of a certain type satisfying certain local conditions.

Let $k \geqslant 2$ be a fixed even integer. For any integer $q \geqslant 1$, let $S_k(q)^*$ be the finite set of primitive holomorphic modular forms of level q and weight k, with trivial nebentypus (more general settings can be studied, but we restrict to this one for simplicity). We denote by

$$f(z) = \sum_{n \geqslant 1} n^{(k-1)/2} \lambda_f(n) e(nz)$$

the Fourier expansion of a form $f \in S_k(q)^*$ at the cusp at infinity.

We consider on this finite set the "measure" $\mu = \mu_q$ defined by

$$\mu_q(\{f\}) = \frac{(k-1)!}{(4\pi)^{k-1}\langle f, f\rangle},$$

where $\langle \cdot, \cdot \rangle$ is the Petersson inner product. This is the familiar "harmonic weight", and we denote

$$\boldsymbol{E}_q(\alpha) = \sum_{f \in S_k(q)^*} \mu_q(\{f\}) \alpha(f), \quad \boldsymbol{P}_q(\mathcal{P} \text{ is true}) = \sum_{\substack{f \in S_k(q)^* \\ \mathcal{P}(f) \text{ is true}}} \mu_q(\{f\}) \tag{7}$$

the corresponding averaging operator and "probability", for an arbitrary property $\mathcal{P}(f)$ referring to the modular forms $f \in S_k(q)^*$. (Note that it is only asymptotically that this is a probability measure, as $q \to +\infty$).

Imitating the notation in [8, Ch. 1], we now denote by

$$\rho_p \,:\, \begin{cases} S_k(q)^* \to \mathbf{R} \\ f \mapsto \lambda_f(p), \end{cases}$$

the p-th Fourier coefficient maps, which we see as giving "global-to-local" data, similar to reduction maps modulo primes for integers. If $d \geqslant 1$ is a squarefree integer coprime with q, we denote

$$\rho_d \,:\, \begin{cases} S_k(q)^* \to \mathbf{R}^{\omega(d)} \\ f \mapsto (\rho_p(f))_{p|d} = (\lambda_f(p))_{p|d}, \end{cases}$$

which we emphasize is a *tuple* of Fourier coefficients, that should not be mistaken with the single number $\lambda_f(d)$.

The basic relation with sieve is the following idea: provided Q is small enough, the $(\rho_p(f))_{p\leqslant Q}$ become equidistributed as $q \to +\infty$ for the product Sato-Tate measure

$$\nu_Q = \bigotimes_{p\leqslant Q} \mu_{ST},$$

where

$$\mu_{ST} = \frac{1}{\pi}\mathbf{1}_{[-2,2]}(t)\sqrt{1-\frac{t^2}{4}}\; dt,$$

and this is similar to the equidistribution of arithmetic sequences like the integers or the primes modulo squarefree d, and the independence due to the Chinese Remainder Theorem.

The quantitative meaning of this principle is easy to describe if Q is bounded (independently of q), but requires some care when it grows with q. For our purpose, we express it as given by uniform bounds for Weyl-type sums associated with a suitable orthonormal basis of $L^2(\nu_Q)$. The latter is easy to construct. Indeed, recall first the standard fact that the Chebychev polynomials X_m, $m \geqslant 0$, defined by

$$X_m(2\cos\theta) = \frac{\sin((m+1)\theta)}{\sin\theta}, \qquad \theta \in [0,\pi], \tag{8}$$

form an orthonormal basis of $L^2(\mu_{ST})$. Then standard arguments show that for $Q \geqslant 2$ and ν_Q the measure above on $[-2,2]^{\pi(Q)}$, the functions

$$\Lambda_d(x) = \prod_p X_{m_p}(x_p), \qquad \text{for all } x = (x_p)_p \in [-2,2]^{\pi(Q)},$$

defined for any Q-friable integer[3] $d \geqslant 1$, factored as

$$d = \prod_{p\leqslant Q} p^{m_p},$$

form an orthonormal basis of $L^2(\nu_Q)$. (In particular we have $\Lambda_1 = 1$, the constant function 1.)

[3] I.e., integer only divisible by primes $\leqslant Q$.

We have also the following fact which gives the link between this orthonormal basis and our local data $(\rho_p)_p$: for any integer $m \geqslant 1$ coprime with q and divisible only by primes $p \leqslant Q$, and any $f \in S_k(q)^*$, we have

$$\Lambda_m(\rho_d(f)) = \lambda_f(m), \qquad \text{where} \qquad d = \prod_{p|m} p. \tag{9}$$

This is simply a reformulation of the Hecke multiplicativity relations between Fourier coefficients of primitive forms.

Remark. Our situation is similar to that of classical sieve problems, where (in the framework of [8]) we have a set X (with a finite measure μ) and surjective maps $X \xrightarrow{\rho_\ell} Y_\ell$ with *finite* target sets Y_ℓ, each equipped with a probability density ν_ℓ, so that the equidistribution can be measured by the size of the remainders $r_\ell(y)$ defined by

$$\mu(\rho_\ell^{-1}(y)) = \mu(X)\nu_\ell(y) + r_\ell(y)$$

and the independence by using finite sets $m = \{\ell_1, \ldots, \ell_k\}$, and

$$Y_m = \prod_{\ell\in m} Y_\ell, \qquad \rho_m = \prod_{\ell\in m} \rho_\ell \,:\, X \to Y_m, \qquad \nu_m(y_1, \ldots, y_k) = \nu_{\ell_1}(y_1)\cdots\nu_{\ell_k}(y_k),$$

and looking at

$$\mu(\rho_m^{-1}(y)) = \mu(X)\nu_m(y) + r_m(y).$$

Here the compact set $[-2, 2]$ requires the use of infinitely many functions to describe an orthonormal basis. Another (less striking) difference is that our local information lies in the same set $[-2, 2]$ for *all* primes, whereas classical sieves typically involve reduction modulo primes, which lie in different sets.

We now state the analogue, in this language, of the dual large sieve inequality (3).

Proposition 1. *With notation as above, for all $Q \geqslant 1$, all integers $N \geqslant 1$, all complex numbers $\alpha(m)$ defined for m in the set $\Psi_q(N, Q)$ of Q-friable integers $\leqslant N$ coprime with q, we have*

$$\boldsymbol{E}_q\Bigl(\Bigl|\sum_{m\in\Psi_q(N,Q)} \alpha(m)\Lambda_m(\rho_d(f))\Bigr|^2\Bigr) \ll (1 + Nq^{-1}) \sum_m |\alpha(m)|^2, \tag{10}$$

where the implied constant depends only on k and d on the left-hand side is the radical $\prod_{p|m} p$.

Proof of Proposition 1. This is in fact simply a consequence of one of the well-known large sieve inequalities for Fourier coefficients of cusp forms (as developped by Iwaniec and by Deshouillers–Iwaniec, see [3]). The point is that because of (9), the left-hand side of (10) can be rewritten

$$\begin{aligned} S &= \sum_{f\in S_k(q)^*} \mu_q(\{f\}) \Bigl|\sum_{m\in\Psi_q(N,Q)} \alpha(m)\lambda_f(m)\Bigr|^2 \\ &= \frac{(k-1)!}{(4\pi)^{k-1}} \sum_{f\in S_k(q)^*} \Bigl|\sum_{m\in\Psi_q(N,Q)} \alpha(m)\frac{\lambda_f(m)}{\|f\|}\Bigr|^2 \end{aligned}$$

We can now enlarge this by positivity; remarking that

$$\{\frac{f}{\|f\|}\ \mid\ f \in S_k(q)^*\}$$

can be seen as a subset of an orthonormal basis of the space $S_k(q)$ of cusp forms of weight k and level q, and selecting any such basis $\mathcal{B}_{k,q} \supset S_k(q)^*$, we have therefore

$$S \leqslant \frac{(k-1)!}{(4\pi)^{k-1}} \sum_{\varphi\in\mathcal{B}_{k,q}} \Bigl|\sum_{m\leqslant N} \alpha(m)\lambda_\varphi(m)\Bigr|^2,$$

where we put $\alpha(m) = 0$ if $m \notin \Psi_q(N, Q)$, and where the $\lambda_\varphi(m)$ are the Fourier coefficients, so that

$$\varphi(z) = \sum_{m\geqslant 1} m^{(k-1)/2}\lambda_\varphi(m)e(nz),$$

(as earlier for Hecke forms). Now by the large sieve inequality in [7, Theorem 7.26], taking into account the slightly different normalization,[4] we have

$$\frac{(k-1)!}{(4\pi)^{k-1}} \sum_{\varphi\in\mathcal{B}_{k,q}} \Bigl|\sum_{1\leqslant m\leqslant N} \alpha(m)\lambda_\varphi(m)\Bigr|^2 \ll (1 + Nq^{-1}) \sum_m |\alpha(m)|^2 \tag{11}$$

with an absolute implied constant, and this leads to (10). □

Remark 2. In terms of equidistribution (which are hidden in this proof), the basic statement for an individual prime p is that

$$\lim_{q\to+\infty} \boldsymbol{E}_q(X_m(\rho_p(f))) = 0,$$

for all $m \geqslant 1$. Such results are quite well-known and follow in this case from the Petersson formula. There is an implicit version already present in Bruggeman's work (see [1, §4], where it is shown that, on average, "most" Maass forms with Laplace eigenvalue $\leqslant T$, satisfy the Ramanujan-Petersson conjecture), and the first explicit result goes back to Sarnak [13], still in the case of Maass forms.[5] Serre [14] and Conrey, Duke and Farmer [2] gave similar statements for holomorphic forms, and Royer [12] described quantitative versions in that case.

We can now derive the analogues of the arithmetic inequality (1) and of Rényi's inequality (4). The basic "sieve" questions we look at is to bound from above the cardinality (or rather, μ_q-measure) of sets of the type

$$S = \{f \in S_k(q)^*\ \mid\ \lambda_f(p) = \rho_p(f) \notin \Omega_p \text{ for } p \leqslant Q,\ p \nmid q\},$$

for $\Omega_p \subset [-2, 2]$. Because the expansion of the characteristic function of Ω_p in terms of Chebychev polynomials involves infinitely many terms, we restrict to a simple type of condition sets Ω_p of the following type:

$$\Omega_p = \{x \in [-2, 2]\ \mid\ Y_p(x) \leqslant \beta_{p,0} - \delta_p\}, \tag{12}$$

where

$$Y_p = \beta_{p,0} + \beta_{p,1}X_1 + \cdots + \beta_{p,s}X_s$$

[4] The case $k = 2$ requires adding a factor $\log N$.

[5] This is the only result we know that discusses the issue of independence of the coefficients at various primes.

is a real-valued polynomial and $\delta_p > 0$ (the degree s is assumed to be the same for all p). Note that $\beta_{p,0}$ is the μ_{ST}-average of Y_p, so our sets S are those where the Fourier coefficients for $p \leqslant Q$ are "away" from the putative average value according to the Sato-Tate measure.

Denote also by

$$\sigma_p^2 = \sum_{1 \leqslant i \leqslant s} \beta_{p,i}^2 = \int_{-2}^{2} Y_p^2 d\mu_{ST} - \Bigl(\int_{-2}^{2} Y_p d\mu_{ST}\Bigr)^2,$$

the variance of Y_p.

Then the analogue of (4) is

$$\boldsymbol{E}_q\Bigl(\Bigl(\sum_{p \leqslant Q} (Y_p(\lambda_f(p)) - \beta_{p,0})\Bigr)^2\Bigr) \ll (1 + Q^s q^{-1}) \sum_{p \leqslant Q} \sigma_p^2, \tag{13}$$

where the implied constant depends only on k, and that of (1) is

$$\boldsymbol{P}_q\Bigl(Y_p(\lambda_f(p)) \leqslant \beta_{p,0} - \delta_p \text{ for all } p \leqslant Q\Bigr) \ll (1 + N^s q^{-1}) H^{-1} \tag{14}$$

where $\delta_p > 0$, $N \geqslant 1$ is arbitrary and

$$H = \sum_{m \in \Psi_q(N,Q)} \prod_{p \mid m} \frac{\delta_p^2}{\sigma_p^2},$$

the implied constant depending again only on k.

To prove (13), we apply (10) with $N = Q^s$ and $\alpha(m) = 0$ unless $m = p^j$ with $1 \leqslant j \leqslant s$ and $p \leqslant Q$, $p \nmid q$, in which case

$$\alpha(p^j) = \beta_{p,j}.$$

By definition of $Y_p(x)$ and of Λ_d, we get

$$\sum_{m \in \Psi_q(N,Q)} \alpha(m) \Lambda_m(\rho_d(f)) = \sum_{p \leqslant Q} (Y_p(\rho_p(f)) - \beta_{p,0}),$$

showing that (13) is indeed a special case of (10).

To prove (14), we use the "amplification" method of the previous section. The basic observation is that if, for some prime p, we have

$$Y_p(\lambda_f(p)) \leqslant \beta_{0,p} - \delta_p, \tag{15}$$

then it follows that

$$\sum_{1 \leqslant i \leqslant s} (-\beta_{p,i}) X_i(\lambda_f(p)) \geqslant \delta > 0.$$

Now let ξ_p, for $p \leqslant Q$, be arbitrary auxiliary positive real numbers, and let

$$\xi_d = \prod_{p \mid d} \xi_p$$

for $d \mid P(Q)$, the product of all primes $p \leqslant Q$. If (15) holds for all $p \leqslant Q$ coprime with q, then we find by multiplying out that, for any integer $m \in \Psi_q(N,Q)$, i.e., such that

$$d \leqslant N, \quad d \mid P(Q), \quad (d,q) = 1, \quad d = p_1 \cdots p_k, \text{ (say)}, \tag{16}$$

and for such (ξ_p), we have

$$\xi_d \sum_{1\leqslant j_1,\ldots,j_k\leqslant s}\cdots\sum (-1)^k \beta_{p_1,j_1}\cdots\beta_{p_k,j_k} X_{j_1}(\lambda_f(p_1^{j_1}))\cdots X_{j_k}(\lambda_f(p_k^{j_k})) \geqslant \xi_d \delta^{\omega(d)},$$

which translates to

$$\xi_d \sum_{m\in S_d} \alpha(m)\Lambda_m((\lambda_f(p))_{p\leqslant Q}) \geqslant \xi_d \prod_{p|d} \delta_p,$$

where m runs over the set S_d of integers of the type

$$m = \prod_{p|d} p^{v_p(n)}, \quad \text{with} \quad 1 \leqslant v_p(n) \leqslant s, \quad \prod_{p|m} p \leqslant N,$$

so $m \leqslant N^s$, and

$$\alpha(m) = \prod_{p|d} (-\beta_{p,v_p(m)}).$$

Thus, summing over d subject to (16), squaring, then averaging over f and applying (10), we find that the probability

$$\mathcal{P} = \boldsymbol{P}_q\Big(Y_p(\lambda_f(p)) \leqslant \beta_{p,0} - \delta_p \text{ for all } p \leqslant Q\Big)$$

satisfies

$$\mathcal{P} \ll (1 + N^s q^{-1})\frac{A_1}{B_1^2},$$

where

$$A_1 = \sum_d \xi_d^2 \sum_{m\in S_d} |\alpha(m)|^2 = \sum_d \xi_d^2 \prod_{p|d} \sigma_p^2 \quad B_1 = \sum_d \xi_d \prod_{p|d} \delta_p$$

Cauchy's inequality shows that $B_1^2 \leqslant HA_1$, with equality if

$$\xi_p = \frac{\delta_p}{\sigma_p^2}, \text{ for all } p \leqslant Q,$$

and the inequality above, with this choice, leads to

$$\mathcal{P} \ll (1 + N^s q^{-1})H^{-1},$$

as desired.

Remark. If one tries to adapt, for instance, the standard proof in [8], one encounters problems because the latter would (naively at least) involve the problematic expansion of a Dirac measure at a fixed $x \in [-2, 2]$ in terms of Chebychev polynomials.

Here is an easy application of (14), for illustration (stronger results for that particular problem follow from the inequality of Lau and Wu [9], as will be explained with other related results in a forthcoming joint work): it is well-known that for $f \in S_k(q)^*$, the sequence of real numbers $(\lambda_f(p))_p$ changes sign infinitely often, and there has been some recent interest (see, e.g., the paper [6] of Iwaniec, Kohnen and Sengupta) in giving quantitative bounds on the first sign change. We try instead to show that this first sign-change is quite small on average over f (compare with [4]): fix $A > 0$, and let

$$S_{q,A} = \{f \in S_k(q)^* \mid \lambda_f(p) \leqslant 0 \text{ for all } p \leqslant (\log q)^A\}$$

(any other combination of signs is permissible). This is a "sifted set", and we claim that

$$|S_{q,A}| \ll q^{1/2+1/(2A)+\varepsilon}$$

for any $\varepsilon > 0$, where the implied constant depends only on k and ε. Since $S_k(q)^*$ is of size about q (for fixed k), this is a non-trivial bound for all $A > 1$. Moreover, to prove this bound, it suffices to show

$$\boldsymbol{P}_q(S_{q,A}) \ll q^{-1/2+1/(2A)+\varepsilon}$$

since we have the well-known upper bound $\mu_q(\{f\}) \gg q^{-1-\varepsilon}$ for any $\varepsilon > 0$ (see, e.g., [7, p. 138]).

The sets $\Omega_p =]0,2]$ used in $S_{q,A}$ are not exactly in the form (12), so we use some smoothing: we claim there exists a real polynomial Y of degree $s = 2$ such that

$$Y(x) \leqslant \operatorname{sgn}(x), \quad \text{for all } x \in [-2,2], \quad \text{and} \quad \beta_0 = \int_{-2}^{2} Y d\mu_{ST} > -1. \tag{17}$$

Assuming such a polynomial is given, we observe that

$$\lambda_f(p) \geqslant 0 \Rightarrow Y(\lambda_f(p)) \leqslant -1 = \beta_0 - \delta,$$

for some fixed $\delta > 0$. Therefore, by (14) with $N = q^{1/s}$, we get for all Q that

$$\boldsymbol{P}_q(\lambda_f(p) \leqslant 0 \text{ for } p \leqslant Q) \ll H^{-1}$$

where

$$H = \sum_{m \in \Psi_q(q^{1/s}, Q)} \gamma^{\omega(m)}, \qquad \text{with} \qquad \gamma = \frac{(\beta_0+1)^2}{\beta_1^2 + \beta_2^2},$$

and the implied constant depends only on k. By assumption, we have $\gamma > 0$, and an easy lower bound for H follows in the range of interest simply from bounding $\gamma^{\omega(m)} \gg m^{-\varepsilon}$ and using known results on the cardinality of $\Psi_q(y, (\log q)^A)$: we have

$$\sum_{m \in \Psi_q(q^{1/s}, (\log q)^A)} \gamma^{\omega(m)} \gg q^{s^{-1}(1-A^{-1})-\varepsilon},$$

for any $\varepsilon > 0$, the implied constant depending only on A, the choice of Y and ε. This clearly gives the result, and it only remains to exhibit the polynomial Y. One can check easily that

$$Y(x) = -\frac{3}{4}X_0(x) + \frac{1}{2}X_1(x) + \frac{1}{4}X_2(x) = -1 + \frac{x}{2} + \frac{x^2}{4}$$

does the job (see its graph); the numerical values of β_0, δ and γ are given by

$$\beta_0 = -\frac{3}{4}, \qquad \delta = \frac{1}{4}, \qquad \beta_1^2 + \beta_2^2 = \frac{5}{16}, \qquad \gamma = \frac{4}{5}.$$

Remark 3. See the letter of Serre in the Appendix of [15] for previous examples showing how to use limited information towards the Sato-Tate conjecture to prove distribution results for Hecke eigenvalues (of a fixed modular form).

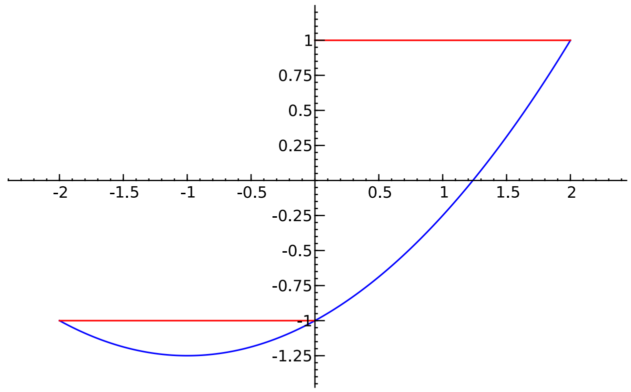

FIGURE 1

REFERENCES

[1] R. Bruggeman: *Fourier coefficients of cusp forms*, Invent. math. 45 (1978), 1–18.
[2] B. Conrey, W. Duke and D. Farmer: *The distribution of the eigenvalues of Hecke operators*, Acta Arith. 78 (1997), 405–409.
[3] J-M. Deshouillers and H. Iwaniec: *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. math. 70 (1982), 220–288.
[4] W. Duke and E. Kowalski: *A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations*, with an Appendix by D. Ramakrishnan, Invent math. 139 (2000), 1–39.
[5] H. Halberstam and H.E. Richert: *Sieve methods*, London Math. Soc. Monograph, Academic Press (London), 1974
[6] H. Iwaniec, W. Kohnen and J. Sengupta: *The first negative Hecke eigenvalue*, International J. Number Theory 3 (2007), 355–363.
[7] H. Iwaniec and E. Kowalski: *Analytic number theory*, AMS Colloquium Publ. 53, 2004.
[8] E. Kowalski: *The large sieve and its applications: arithmetic geometry, random walks and discrete groups*, Cambridge Tracts in Math. 175, 2008.
[9] Y.K. Lau and J. Wu: *A large sieve inequality of Elliott-Montgomery-Vaughan type for automorphic forms and two applications*, International Mathematics Research Notices, Vol. 2008, `doi:10.1093/imrn/rmn162`.
[10] H-L. Montgomery: *Topics in multiplicative number theory*, Lecture Notes Math. 227, Springer-Verlag 1971.
[11] H-L. Montgomery: *The analytic principle of the large sieve*, Bull. A.M.S. 84 (1978), 547–567.
[12] E. Royer: *Facteurs* $\mathbf{Q}$*-simples de* $J_0(N)$ *de grande dimension et de grand rang*, Bull. Soc. Math. France 128 (2000), 219–248.
[13] P. Sarnak: *Statistical properties of eigenvalues of the Hecke operators*, in "Analytic Number Theory and Diophantine Problems" (Stillwater, OK, 1984), Progr. Math. 70, Birkhäuser, 1987, 321–331.
[14] J-P. Serre: *Répartition asymptotique des valeurs propres de l'opérateur de Hecke* T_p, J. American Math. Soc. 10 (1997), 75–102.
[15] F. Shahidi: *Symmetric power L-functions for* $GL(2)$, in: "Elliptic curves and related topics", edited by E. Kishilevsky and M. Ram Murty, CRM Proc. and Lecture Notes 4, 1994, 159–182.

ETH ZÜRICH – D-MATH, RÄMISTRASSE 101, 8092 ZÜRICH, SWITZERLAND
E-mail address: `kowalski@math.ethz.ch`