MODULAR SIGNS, OR YET ANOTHER RECOGNITION PROBLEM FOR MODULAR FORMS

E. KOWALSKI

There are many results in the arithmetic of modular forms which are, more or less, concerned with various ways of characterizing a given (primitive) cusp form f from its siblings, starting from the fact that Fourier coefficients, hence the L-function, determine uniquely a cusp form f relative to a congruence subgroup Γ of $SL(2,\mathbf{Z})$[1], going through stronger forms of the multiplicity one theorem for automorphic representations, and then to various explicit forms of these statements, where only finitely many coefficients are required (say at primes $p \leqslant X$, for some explicit X depending on the parameters defining f), and to "statistic" versions of the latter, where X can be reduced drastically, provided one accepts some possible exceptions. Some of these statements were strongly suggested by the analogy with the problem of the least quadratic-non-residue, which is a problem of great historic importance in analytic number theory.[2]

Recently, Lau and Wu [LW2] have found, for real characters, a precise "threshold" $y(Q)$ for which the upper bound on the number of real characters of conductor $q \leqslant Q$ with value $+1$ for primes $p \leqslant y(Q)$ almost coincides with a lower bound for this number. Then they proved in [LW1] a corresponding upper-bound for recognition problems for modular forms. However, as they mention, it is doubtful that the analogue of the lower bound holds, because there are much less restrictions on the values of Fourier coefficients (or Hecke eigenvalues) than on values of real characters (the latter take value in $\{-1,0,1\}$, whereas, for instance, there are about $4\sqrt{p}$ values for the p-th Fourier coefficient of the modular form of weight 2 attached to an elliptic curve).

One may be tempted, at least out of curiosity, to try to remove this discrepancy by considering only the signs of the Fourier coefficients, and the corresponding recognition problem. In this short note, we consider simply the first basic question:

Question. *Assume f has real coefficients. Is it true that f is determined uniquely by the sequence of* signs *of its Fourier coefficients $\lambda_f(p)$ (where "sign" is interpreted in a relaxed way so that 0 has the same sign as both positive and negative numbers)?*

As we will see, the answer is "Yes", and one can relax the assumption to hold only for most primes, allowing a (small) exceptional set. There is no claim here that this is a particularly deep question, although we use some sophisticated tools; if one wishes to do so, one can see this as the study of a new type of $\{\pm 1\}$ sequence which are likely to be quite random.

Theorem 1. *Let $q_1, q_2 \geqslant 1$ be integers, let $k_1, k_2 \geqslant 2$ be even integers and $f_1 \in S_{k_1}(q_1)^*$, $f_2 \in S_{k_2}(q_2)^*$ be primitive holomorphic cusp forms for the congruence subgroups $\Gamma_0(q_1)$, $\Gamma_0(q_2)$ and weight k_1, k_2 respectively, with trivial nebentypus. Let $\lambda_{f_1}(p)$ and $\lambda_{f_2}(p)$ denote the Hecke eigenvalues of f_1 and f_2 for p prime. Assume neither of f_1 or f_2 is of CM type.*

[1] A fact which is obvious but in fact depends crucially on the existence of cusps, or in other words, of unipotent elements in Γ.

[2] For instance, as the first application, and presumably motivation, for Linnik's large sieve...

If we have

$$\lambda_{f_1}(p) \geqslant 0 \text{ if and only if } \lambda_{f_2}(p) \geqslant 0, \tag{1}$$

for every prime p, except those in a set S of analytic density κ, with $\kappa < 729/2000000 = 5 \cdot 3^6 \cdot 10^{-7}$, then in fact $f_1 = f_2$, and of course $q_1 = q_2$, $k_1 = k_2$.

To clarify the notation, note that the $\lambda_f(p)$ are the analytically normalized Hecke eigenvalues, i.e., the Fourier expansion of f at the cusp ∞ is given by

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

(but the conclusion of the theorem is also valid with $\lambda_f(p)$ replaced by the actual Fourier coefficients since the assumption (1) is not altered by multiplying by $p^{(k_1-1)/2}$ or $p^{(k_2-1)/2}$). The eigenvalues $\lambda_f(p)$ are real numbers, since the nebentypus is trivial.

Let us finally mention that the analytic density we use is the Dirichlet density: a set S of primes has density κ if and only if

$$\sum_{p \in S} \frac{1}{p^\sigma} \sim \kappa \sum_p \frac{1}{p^\sigma} \sim -\kappa \log(\sigma - 1), \qquad \text{as } \sigma \to 1$$

(of course the statement is also valid with natural density, but this turns out to be the most convenient to work with).

Before going on to the proof, here is how one can see that the answer should be what we claim, at least when the exceptional set is empty: in that case, the assumption (1) translates to

$$\lambda_{f_1}(p)\lambda_{f_2}(p) \geqslant 0$$

for all primes p. But it is well-known (from Rankin-Selberg theory) that if $f_1 \neq f_2$, we have

$$\sum_p \frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma} = O(1) \tag{2}$$

as $\sigma \to 1$. Thus we only need to find a lower bound for the left-hand side (which is a sum of non-negative terms) that grows as X grows to obtain a contradiction. Since Rankin-Selberg theory also gives

$$\sum_p \frac{\lambda_{f_1}(p)^2}{p^\sigma} \sim -\log(\sigma - 1), \qquad \text{as } \sigma \to 1, \tag{3}$$

the only difficulty is that one might fear that the coefficients of f_1 and f_2 are such that whenever $\lambda_{f_1}(p)$ is not small, the value of $\lambda_{f_2}(p)$ *is* very small. In other words, what we must show is that the smaller order of magnitude of (2) compared with (3) is not due to the small size of the summands, but to sign compensations.

If we assume that the Fourier coefficients obey the Sato-Tate conjecture (for analytic density suffices), we can immediately see that this can not happen. Indeed, in that case the sets

$$\{p \mid \lambda_{f_i}(p) \in [a, b]\}$$

of primes have analytic density

$$\frac{1}{\pi} \int_a^b \sqrt{1 - \frac{t^2}{4}} dt,$$

for any fixed $a < b$ in the interval $[-2, 2]$. In particular, for a small enough, the sets

$$S_i = \{p \mid |\lambda_{f_i}(p)| > a\}$$

have density $> 1/2$, say, $= 3/4$, and then the intersection $S_1 \cap S_2$ itself has density $\geqslant 1/2$, with

$$\sum_p \frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma} \geqslant \sum_{p \in S_1 \cap S_2} \frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma} \geqslant a^2(-\log(\sigma - 1) + O(1))$$

for $\sigma > 1$. This does indeed contradict (2).

So we need to work around the fact that we don't know that the Sato-Tate law holds, except in the cases recently treated by Taylor (see Mazur's survey [M]).

For this, the idea is similar to an earlier technique used by Serre (see [Sh]) to prove consequences of the Sato-Tate Conjecture using the first few symmetric power L-functions only.

The main lemma is the following result, which is (modestly) of independent interest.

Lemma 2. *Let q be an integer, let $k \geqslant 2$ be an even integer and $f \in S_k(q)^*$ be a primitive holomorphic cusp form for the congruence subgroup $\Gamma_0(q)$, with weight k and trivial nebentypus. Let $\lambda_f(p)$ for p prime denote the Hecke eigenvalues of f. Assume that f is not of CM type. Then there exists a constant $\alpha > 0$ and $\delta > 1/2$ such that*

$$\sum_{|\lambda_f(p)| > \alpha} \frac{1}{p^\sigma} \geqslant (\delta + o(1)) \sum_p \frac{1}{p^\sigma}$$

for $\sigma > 1$. In fact, one can take

$$\alpha = 0.27, \qquad \delta = 1/2 + 1/100.$$

The lemma does not hold for CM forms, because we then have $\lambda_f(p) = 0$ for a set of primes of density $1/2$. However, one could prove a version of Theorem 1 when f_1 (or f_2) is of CM-type using the (better known) distribution of the eigenvalues in those cases.

Proof. We recall first that for any prime $p \nmid q$, denoting by α_p and β_p the two roots of the quadratic polynomial

$$X^2 - \lambda_f(p)X + 1,$$

we can write

$$\lambda_f(p^n) = \alpha_p^n + \alpha_p^{n-1}\beta_p + \cdots + \alpha_p\beta_p^{n-1} + \beta_p^n = X_n(\lambda_f(p))$$

for some polynomial $X_n \in \mathbf{R}[X]$, independent of p (a slightly modified Chebychev polynomial of the second kind). The first few such polynomials with even indices are given by

$$X_0 = 1, \quad X_2 = X^2 - 1, \quad X_4 = X^4 - 3X^2 + 1, \quad X_6 = X^6 - 5X^4 + 6X^2 - 1, \tag{4}$$

(it will be clear later on why we do not consider odd indices n).

Now we claim that there exists a polynomial

$$P = a_0 + a_2X_2 + a_4X_4 + a_6X_6 \in \mathbf{R}[X]$$

with the following properties: (1) $a_0 > 1/2$; (2) for some $\alpha > 0$, and $x \in [-2, 2]$, we have

$$P(x) \leqslant \chi_A(x), \quad \text{where} \quad A = \{x \in [-2, 2] \mid |x| > \alpha\}. \tag{5}$$

Assuming this, we conclude as follows: by (2), we have

$$\sum_{|\lambda_f(p)| > \alpha} \frac{1}{p^\sigma} \geqslant \sum_{p \nmid q} \frac{P(\lambda_f(p))}{p^\sigma} = a_0 \sum_{p \nmid q} \frac{1}{p^\sigma} + \sum_{i=1}^{3} \sum_{p \nmid q} \frac{X_{2i}(\lambda_f(p))}{p^\sigma}.$$

By the holomorphy and non-vanishing at $s = 1$ of the (partial) second, fourth and sixth symmetric power L-functions (see [KS, Th. 3.3.7, Prop. 4.3] for the last two), since $X_n(\lambda_f(p))$ is exactly the p-th coefficient of the n-symmetric power for $p \nmid q$, we know that

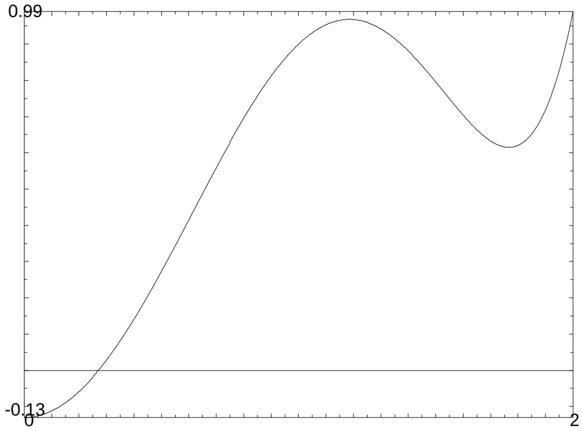
$$\sum_{p \nmid q} \frac{X_{2i}(\lambda_f(p))}{p^\sigma} = O(1)$$

for $\sigma \geqslant 1$ and $i = 1, 2, 3$. Hence the result follows with $\delta = a_0 > 1/2$.

Now to check the claim, and verify the values of α and δ, we just exhibit a suitable polynomial, namely

$$P = \frac{1}{2} + \frac{1}{100} + \frac{1}{4}X_2 - \frac{1}{4}X_4 + \frac{14}{100}X_6 = \frac{7}{50}X^6 - \frac{19}{20}X^4 + \frac{46}{25}X^2 - \frac{13}{100},$$

which is even and the graph of which on $[0, 2]$ is as follows:



The origin of this particular example is explained in a remark below; the value of α is an approximation to the real root

$$\alpha_0 = 0.27093173464429513973194337 92\ldots$$

of P in $[0, 2]$; the maximum value of P on $[0, 2]$ is attained at $x = 2$ and is equal to 99/100. $\square$

Before using this lemma to conclude the proof of Theorem 1, let's explain how the polynomial comes about, and in particular why we have used the sixth symmetric power. Indeed, one can not argue similarly using only the second and fourth symmetric power. In other words, there is no polynomial

$$P = a_0 + a_2X_2 + a_4X_4 \in \mathbf{R}[X]$$

such that $P \leqslant \chi_I$ for some interval $I = [\alpha, 2] \subset]0, 2]$ and with $a_0 > 1/2$. Indeed, note that if such a polynomial exists, it must satisfy (1) $P(0) \leqslant 0$; (2) $P \leqslant 1$ on $[0, 2]$; (3) $a_0 > 1/2$.

But then, expressing P in the basis of powers of X, we find

$$P = a_0 - a_2 + a_4 + (a_2 - 3a_4)X^2 + a_4X^4$$

and in particular, we check that

$$P(0) + P(\sqrt{2}) = (a_0 - a_2 + a_4) + (a_0 - a_2 + a_4 + 2a_2 - 6a_4 + 4a_4) = 2a_0$$

so condition (3) leads to $P(0)+P(\sqrt{2})>1$, and if $P(0)\leqslant 0$, this means that $P(\sqrt{2})>1$, showing that (1) and (2) are then incompatible.

Looking at this argument, however, reveals quickly that it "almost" works: precisely, this suggests looking at the value $P(\sqrt{2})$ for a general polynomial as above, and then one notices quickly that the polynomial

$$P_0=\frac{1}{2}+\frac{1}{4}X_2-\frac{1}{4}X_4=X^2-\frac{X^4}{4}=X^2(1-X/2)(1+X/2)$$

is "borderline": we have $0\leqslant P_0\leqslant 1$ on $[0,2]$ and $a_0=1/2$; in fact

$$P_0(0)=0,\quad a_0=1/2,\quad \max_{x\in[0,2]}P_0(x)=P_0(\sqrt{2})=1.$$

So we looked for, and found, our polynomial by simply "deforming" slightly this example, increasing slightly a_0 to have it $>1/2$, and compensating with a small multiple of X_6. But we did not really try to optimize this deformation argument.

We can now use this lemma to conclude the proof of Theorem 1. The assumption (1) implies that

$$\lambda_{f_1}(p)\lambda_{f_2}(p)\geqslant 0,$$

for all primes $p\notin S$. In addition, we have

$$|\lambda_{f_1}(p)\lambda_{f_2}(p)|\leqslant 4,$$

for all primes p by the Deligne bound. Hence we find that (with $\alpha>0$ and $\delta>1/2$ as in Lemma 2), we have

$$\begin{aligned}\sum_p\frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma}&=\sum_{p\notin S}\frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma}+\sum_{p\in S}\frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma}\\
&\geqslant\alpha^2\sum_{\substack{|\lambda_{f_1}(p)|>\alpha,\ |\lambda_{f_2}(p)|>\alpha\\ p\notin S}}\frac{1}{p^\sigma}-4\kappa|\log(\sigma-1)|+O(1)\\
&\geqslant\alpha^2\Big\{\sum_p\frac{1}{p^\sigma}-\sum_{p\in S}\frac{1}{p^\sigma}-\sum_{|\lambda_{f_1}(p)|\leqslant\alpha}\frac{1}{p^\sigma}\\
&\qquad\qquad-\sum_{|\lambda_{f_1}(p)|\leqslant\alpha}\frac{1}{p^\sigma}\Big\}-4\kappa|\log(\sigma-1)|+O(1)\\
&\geqslant\alpha^2\Big\{(1-2(1-\delta))\sum_p\frac{1}{p^\sigma}+O(1)\Big\}-4\kappa|\log(\sigma-1)|+O(1)\\
&=(\alpha^2(2\delta-1)-4\kappa)|\log(\sigma-1)|+O(1)\end{aligned}$$

for any $\sigma>1$.

Since $2\delta>1$, we find that the left-hand side goes to $+\infty$ as $\sigma\rightarrow 1$ under the condition

$$\kappa<\frac{\alpha^2(2\delta-1)}{4}=\frac{729}{2000000}$$

(with the values of Lemma 2). However, as already mentioned, the theory of Rankin-Selberg L-functions shows that if $f_1\neq f_2$, we have

$$\sum_p\frac{\lambda_{f_1}(p)\lambda_{f_2}(p)}{p^\sigma}=O(1)$$

since there is no pole (or zero) of $L(f_1\times f_2,s)$ at $s=1$. So we must indeed have $f_1=f_2$.

Remark 3. As previously noted, Lemma 2 is somewhat "dual" to well-known investigations of consequences of holomorphy of the first symmetric power L-functions towards the Sato-Tate conjecture, explained in particular in Serre's letter to Shahidi (Appendix to [Sh]), and refined most recently by Kim and Shahidi [KS, §4]. The difference is that, in those works, one is interested in finding $c \in [-2, 2]$, as large as possible, such that $\lambda_f(p) > c$ for infinitely many p and $\lambda_f(p) < -c$ for infinitely many p. In the lemma, the value of c (i.e., α) is not important, but the density of the set of primes has to be large.

Here are some natural issues that it may be of interest to look at now:
– What is the optimal density κ one can obtain in Theorem 1? If one assumes that f_1 and f_2 obey the pair-Sato-Tate conjecture:

$$\{p \mid \lambda_{f_1}(p) \in [a_1, b_1] \text{ and } \lambda_{f_2}(p) \in [a_2, b_2]\}$$

has density $\mu_{ST}([a_1, b_1])\mu_{ST}([a_2, b_2])$ (where μ_{ST} is the Sato-Tate distribution; in other words, the Fourier coefficients are independently Sato-Tate distributed), one may easily get the result for any $\kappa < 1/2$ (corresponding to the probability for $\mu_{ST} \otimes \mu_{ST}$ of having the same sign). But this can only hold if f_1 and f_2 are not related by quadratic twists, of course (and in that case, for elliptic curves, Harris is currently making progress on the problem, according to [M, Footnote 12]). If, on the other hand, $f_2 = f_1 \otimes \chi$ for a real character χ, the coefficients are of the same sign for a set of primes of density exactly 1/2.
– The limit of the argument we used is fairly easy to determine: if we have all even symmetric power L-functions at our disposal, we can use polynomials P approximating arbitrarily closely to $\chi_A(x)$ (see (5)) for any $\alpha \in]0, 2]$. The value $\delta = a_0$ is then the probability under the Sato-Tate distribution of $A = \{x \mid |x| \geqslant \alpha\}$, namely

$$a_0 = \frac{2}{\pi} \arccos\Big(\frac{\alpha}{2}\Big),$$

and we are led to maximize over $[0, 2]$ the quantity

$$\kappa = \frac{\alpha^2}{4}\Big(\frac{4}{\pi} \arccos\Big(\frac{\alpha}{2}\Big) - 1\Big),$$

and we find numerically that the best value is around $\alpha \simeq 0.971$, allowing to take $\kappa < 0.083571\ldots$ (i.e., knowing only the individual Sato-Tate conjecture, we can allow about 8% of the primes to be in the exceptional set).
– What happens with Maass forms? The problem there is that the exceptional set might correspond to primes where the Fourier coefficients are large, since we do not know the Ramanujan-Petersson bound in this case. This means a direct adaptation must assume much stronger conditions on the set S; for instance, using

$$|\lambda(p)| \leqslant 2p^{7/64}$$

(the Kim-Sarnak bound) for eigenfunctions of Hecke operators for a Maass form, Theorem 1 holds if

$$|\{p \in S \mid p \leqslant X\}| \ll X^{25/32-\delta_1}$$

for $X \geqslant 2$ and some $\delta_1 > 0$ (where the implied constant may depend on f_1, f_2 and δ_1).
– What is the size, as a function of the weight and conductor, of the smallest prime for which the sign of $\lambda_{f_1}(p)$ and $\lambda_{f_2}(p)$ are different? If f_2 is taken (in effect) to be an Eisenstein series with positive Fourier coefficients (e.g., $d(p)$, the divisor function, for a non-holomorphic Eisenstein series, or $\theta(z)^4$, where θ is the theta function of weight 1/2, for $\Gamma_0(4)$), then the question is to find the first negative Hecke eigenvalue for f_1, and there have been some works done recently on this issue (see for instance the work [IKS]

of Iwaniec, Kohnen and Sengupta, where this smallest sign-change is found to be $\ll (k_1 q_1^2)^{29/60}$).
– Can one prove "statistical" estimates for the number of forms of given weight and conductor $\leqslant Q$ for which the sequence of signs for $p \leqslant y(Q)$ is fixed, for some functions $y(Q)$ which grow slowly (e.g., $y(Q) = Q^{\varepsilon}$ for $\varepsilon > 0$ arbitrarily small, or $y(Q) = (\log Q)^A$ for some A)?
– Can one then prove *lower bounds* for this type of number? If one allows non-squarefree conductor, we can get one by fixing an f_1 and considering twists $f_1 \otimes \chi$ by all real characters such that $\chi(p) = 1$ for $p \leqslant y(Q)$. Thus the Lau-Wu lower bound of [LW2] applies here also. But if we restrict, say, to squarefree conductors, constructing a lower bound seems to be a genuinely $GL(2)$-type question.

References

[IKS] H. Iwaniec, W. Kohnen and J. Sengupta: *The first negative Hecke eigenvalue*, International J. Number Theory 3 (2007), 355–363.

[M] B. Mazur: *Finding meaning in error terms*, Bull. A.M.S 45 (2008), 185–228.

[KS] H. Kim and F. Shahidi: *Cuspidality of symmetric powers with applications*, Duke Math. J. 112 (2002), 177–197.

[LW1] 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`.

[LW2] Y.K. Lau and J. Wu: *On the least quadratic non-residue*, Intl. J. Number Theory (to appear).

[Sh] 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`