NOTES ON THE SATO-TATE CONJECTURE

DIMITRIS KOUKOULOPOULOS

Abstract. Let E be a non-CM elliptic curve over $\mathbb{Q}$. We show that the automorphy of the symmetric powers of the Hasse-Weil L-function attached to E implies the Sato-Tate conjecture.

1. Introduction

Let $E : y^2 = x^3 + ax + b$, where $a, b \in \mathbb{Z}$, be a non-CM elliptic curve of discriminant $\Delta = -(27a^2 + 4b^3)$. For every prime $p \nmid \Delta$, the reduction of E mod p is an elliptic curve over $\mathbb{F}_p$. As Hasse proved, if we set

$$\#E(\mathbb{F}_p) = p + 1 - a_p(E),$$

then

$$|a_p(E)| \le 2\sqrt{p}.$$

In fact, more is known: we can write

$$a_p(E) = \alpha_p + \overline{\alpha}_p, \quad \text{where} \quad |\alpha_p| = \sqrt{p}.$$

The complex numbers α_p and $\overline{\alpha}_p$ are the eigenvalues of the Frobenius automorphism, which can be then identified with the conjugacy of the matrix

$$\begin{pmatrix} e^{i\theta_p} & 0 \\ 0 & e^{-i\theta_p} \end{pmatrix}$$

in the group of 2×2 unitary matrices SU(2), where

$$\alpha_p =: \sqrt{p}e^{i\theta_p}.$$

By possibly switching the roles of α_p and its conjugate, we may assume that $\theta_p \in [0, \pi]$. The group SU(2) is naturally equipped with a Haar measure, which then induces a measure on the conjugacy classes of SU(2). Identifying the latter with the interval $[0, \pi]$ via the map $\theta \to \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$, we find that this measure is given by $\frac{2}{\pi}\sin^2\theta \mathrm{d}\theta$ (which defines a probability measure on θ). The Sato-Tate conjecture (now a theorem proven in the series of papers [1, 4, 8] written (in various combinations) by Clozel, Harris, Shepherd-Barron and Taylor) postulates that the angles θ_p are equidistributed with respect to this measure:

The Sato-Tate conjecture. *Let E and θ_p be as above. For each fixed $[\alpha, \beta] \subset [0, \pi]$, we have*

$$\frac{\#\{p \le x : p \nmid \Delta,\ \alpha \le \theta_p \le \beta\}}{\pi(x)} \sim \frac{2}{\pi}\int_\alpha^\beta \sin^2\theta \mathrm{d}\theta \quad (x \to \infty).$$

Date: September 29, 2017.

2. The method of moments

In order to build a strategy towards proving the Sato-Tate conjecture, we quickly review the method of moments from Probability Theory. Our goal is to prove that a certain sequence of random variables $(X_n)_{n\ge1}$ converges weakly (i.e. in distribution) to a random variable X, which means that

$$\lim_{n\to\infty}\mathbb{P}(\alpha\le X_n\le\beta)=\mathbb{P}(\alpha\le X\le\beta) \tag{2.1}$$

for all α and β where the distribution of X is continuous.

The random variables $X, X_1, X_2, \dots$ need not live in the same probability space. For example, in our case, we will take as our random variables X_n to be functions $p\to\cos\theta_p$, with p living in the probability space $\Omega_n=\{p\le n : p\nmid\Delta\}$ equipped with the uniform counting probability measure. Our target random variable X lives some ambient probability space Ω (which is irrelevant) and has the Sato-Tate distribution, i.e. $\mathbb{P}(\alpha\le X\le\beta)=\frac{2}{\pi}\int_\alpha^\beta\sin^2\theta\mathrm{d}\theta$.

The method of moments allows to show that X_n converges in distribution to X by showing instead that the moments of X_n converge to the moments of X, i.e. that

$$\lim_{n\to\infty}\mathbb{E}[X_n^k]=\mathbb{E}[X^k] \tag{2.2}$$

for each fixed $k\in\mathbb{Z}_{\ge0}$. We sketch the argument: assume we want to prove (2.1). Realizing the left hand side as

$$\mathbb{E}[\mathbf{1}_{[\alpha,\beta]}(X_n)],$$

we majorize and minorize $\mathbf{1}_{[\alpha,\beta]}$ by smooth functions $f^\pm$ for which we assume that

$$\begin{cases} f^+(x)=1 & \text{if } \alpha\le x\le\beta,\\ 0\le f^+(x)\le1 & \text{if } x\in[\alpha-\varepsilon,\alpha]\cup[\beta,\beta+\varepsilon],\\ f^+(x)=0 & \text{otherwise,}\end{cases}$$

and

$$\begin{cases} f^-(x)=1 & \text{if } \alpha+\varepsilon\le x\le\beta-\varepsilon,\\ f^-(x)\in[0,1] & \text{if } x\in[\alpha,\alpha+\varepsilon]\cup[\beta-\varepsilon,\beta],\\ f^-(x)=0 & \text{otherwise,}\end{cases}$$

with $\varepsilon>0$ small enough to be chosen later. Then we have

$$\mathbb{E}[f^-(X_n)]\le\mathbb{E}[\mathbf{1}_{[\alpha,\beta]}(X_n)]\le\mathbb{E}[f^+(X_n)].$$

Since the functions $f^\pm$ are continuous (in fact smooth), they can be approximated at arbitrary precision by polynomials in any fixed compact set. If the random variables X_n do not take abnormally large values too often (and for the purposes of the Sato-Tate conjecture, we are in the particularly convenient situation where the random variables X_n are uniformly bounded), then we can in fact show that there are polynomials $P^\pm$ such that

$$|\mathbb{E}[f^\pm(X_n)]-\mathbb{E}[P^\pm(X_n)]|\le\varepsilon'$$

for each fixed ε', as well as that

$$|\mathbb{E}[f^\pm(X)]-\mathbb{E}[P^\pm(X)]|\le\varepsilon'.$$

We thus conclude that

$$\mathbb{E}[\mathbf{1}_{[\alpha,\beta]}(X_n)]\le\mathbb{E}[P^+(X_n)]+\varepsilon'.$$

Since $P^+(x)$ is a finite linear combination of powers of j, if we know (2.2), then we deduce that

$$\limsup_{n\to\infty} \mathbb{P}(\alpha \le X_n \le \beta) \le \mathbb{E}[P^+(X)] + \varepsilon' \le \mathbb{E}[f^+(X)] + 2\varepsilon'.$$

Since f^+ was chosen so that it approximates well-enough $\mathbf{1}_{[\alpha,\beta]}$, choosing ε small enough in terms of ε' implies that $\mathbb{E}[f^+(X)] \le \varepsilon' + \mathbb{P}(\alpha \le X \le \beta)$, provided that the distribution function of X is continuous at α and β (i.e. there are no Dirac masses lying at these points). We have thus shown that

$$\limsup_{n\to\infty} \mathbb{P}(\alpha \le X_n \le \beta) \le \mathbb{P}(\alpha \le X \le \beta) + 3\varepsilon'.$$

Letting $\varepsilon' \to 0^+$, we conclude that

$$\limsup_{n\to\infty} \mathbb{P}(\alpha \le X_n \le \beta) \le \mathbb{P}(\alpha \le X \le \beta).$$

Similarly, we may prove that

$$\liminf_{n\to\infty} \mathbb{P}(\alpha \le X_n \le \beta) \ge \mathbb{P}(\alpha \le X \le \beta),$$

thus completing the deduction of (2.1) from (2.2).

3. The Sato-Tate measure and Chebyshev polynomials

By the method of moments, establishing the Sato-Tate conjecture is reduced to proving that

$$\frac{1}{\pi(x)} \sum_{p\le x,\ p\nmid\Delta} (\cos\theta_p)^k \sim \frac{2}{\pi}\int_0^\pi (\cos\theta)^k (\sin\theta)^2 \mathrm{d}\theta. \tag{3.1}$$

Actually, it turns out that it is much more convenient to consider a different basis of polynomials over $\mathbb{R}[x]$: instead of taking the canonical basis $1, x, x^2, \dots$, we consider the Chebyshev polynomials of the second kind, defined by the relation

$$U_k(\cos\theta) := \frac{\sin((k+1)\theta)}{\sin\theta}.$$

One can easily check that the right hand side is indeed a polynomial of degree k in $\cos\theta$. [1] Our task then becomes to show that

$$\begin{aligned} \frac{1}{\pi(x)} \sum_{p\le x,\ p\nmid\Delta} U_k(\cos\theta_p) &\sim \frac{2}{\pi}\int_0^\pi U_k(\cos\theta)\sin^2\theta \mathrm{d}\theta \\ &= \frac{2}{\pi}\int_0^\pi \sin((k+1)\theta)\sin\theta \mathrm{d}\theta = \mathbf{1}_{k=0}. \end{aligned} \tag{3.2}$$

[1] Actually, the easiest way to do this is by also introducing the Chebyshev polynomials of the first kind, defined by $T_k(\cos\theta) = \cos(k\theta)$. Using the relation $\sin((k+1)\theta) = \sin\theta\cos(k\theta) + \cos\theta\sin(k\theta)$, we then find that $U_k(x) = T_k(x) + xU_{k-1}(x)$. Similarly, $T_{k+1}(x) = xT_k(x) - (1-x^2)U_{k-1}(x)$, and we may show inductively that U_k and T_k are both polynomials of degree k.

4. Fun with Chebyshev polynomials

We make a small digression to discuss some key properties of the Chebyshev polynomials. Firstly, we note that they form an orthonormal family (actually, an orthonormal basis) with respect to the Sato-Tate measure:

$$\frac{2}{\pi}\int_0^\pi U_k(\cos\theta)U_\ell(\cos\theta)\sin^2\theta\mathrm{d}\theta = \frac{2}{\pi}\int_0^\pi \sin((k+1)\theta)\sin((\ell+1)\theta)\mathrm{d}\theta = \mathbf{1}_{k=\ell}.$$

We now obtain a different formula for U_k, noticing the similarity of U_k with the Dirichlet kernel:

$$U_k(\cos\theta) = \frac{e^{i(k+1)\theta}-e^{-i(k+1)\theta}}{e^{i\theta}-e^{-i\theta}} = \frac{e^{-i(k+1)\theta}}{e^{-i\theta}}\cdot\frac{e^{2(k+1)\theta}-1}{e^{2i\theta}-1} = e^{-ik\theta}\sum_{\ell=0}^{k}e^{2\ell i\theta} = \sum_{a+b=k}e^{ai\theta}e^{-bi\theta}. \tag{4.1}$$

Thus we also have that

$$\begin{aligned} U_k(\cos\theta) &= \frac{1}{2}\sum_{a+b=k}e^{ai\theta}e^{-bi\theta} + \frac{1}{2}\sum_{a+b=k}e^{ai\theta}e^{-bi\theta} \\ &= \frac{1}{2}\sum_{a=0}^{k}e^{i(2a-k)\theta} + \frac{1}{2}\sum_{b=0}^{k}e^{i(k-2b)\theta} \\ &= \sum_{a=0}^{k}\cos((2a-k)\theta). \end{aligned} \tag{4.2}$$

Finally, we consider the product U_kU_{k+h}, where $k,h\in\mathbb{Z}_{\ge0}$. We know it is a linear combination of $U_0,U_1,\dots,U_{2k+h}$, say $\sum_{0\le j\le 2k+h}c_jU_j$, and we want to calculate the coefficients of this expansion. By orthogonality and (4.2), we have

$$\begin{aligned} c_j &= \frac{2}{\pi}\int_0^\pi U_k(\cos\theta)U_{k+h}(\cos\theta)U_j(\cos\theta)\sin^2\theta\mathrm{d}\theta \\ &= \frac{2}{\pi}\int_0^\pi \sin((k+1)\theta)\sin((k+h+1)\theta)U_j(\cos\theta)\mathrm{d}\theta \\ &= \frac{1}{\pi}\int_0^\pi (\cos(h\theta)-\cos((2k+h+2)\theta))\sum_{a=0}^{j}\cos((2a-j)\theta)\mathrm{d}\theta. \end{aligned}$$

Since $0\le j\le 2k+h$, we have that $2k+h+2>|2a-j|$ for all $a\in[0,j]$, so that

$$c_j = \frac{1}{\pi}\sum_{a=0}^{j}\int_0^\pi \cos(h\theta)\cos((2a-j)\theta)\mathrm{d}\theta.$$

The only a's for which the above integral is non-zero are those with $2a-j=\pm h$, i.e. $a=(j\pm h)/2$. For these a's to be integers, we must have that $j\equiv h\,(\mathrm{mod}\,2)$. Note though that if $j<h$, then $(j-h)/2<0$ and $(j+h)/2>j$, so that no such a can lie in $[0,j]$. Thus we only have solutions when $j\ge h$ with $j\equiv h\,(\mathrm{mod}\,2)$, in which case we find that $a=(j\pm h)/2\in\mathbb{Z}\cap[0,j]$. When $h=0$, we only find one solution, $a=j$, and $c_j=\frac{1}{\pi}\int_0^\pi\mathrm{d}\theta=1$. When $h\ge1$, we have two values of a contributing, so that $c_j=\frac{2}{\pi}\int_0^\pi\cos^2(h\theta)\mathrm{d}\theta=1$. In any case, $c_j = 1$. We conclude that

$$U_k U_{k+h} = \sum_{\substack{h \le j \le 2k+h \\ j \equiv h \pmod 2}} U_j = \sum_{\ell=0}^{k} U_{h+2\ell}. \tag{4.3}$$

5. Symmetric powers

For each prime $p \nmid \Delta$, we define

$$f_k(p) := U_k(\cos\theta_p).$$

When $k = 1$, we note that

$$f_1(p) = 2\cos\theta_p = e^{i\theta_p} + e^{-i\theta_p} = \frac{a_p(E)}{\sqrt{p}}$$

by the definition of θ_p. Notice that these are the prime values of the coefficients of $L(s, E)$, the Hasse-Weil L-function re-normalized so that its line of symmetry is $\mathrm{Re}(s) = 1/2$. We then extend f_1 to all integers via the formula

$$\sum_{n=1}^{\infty} \frac{f_1(n)}{n^s} = L(s, E) = \prod_{p \nmid \Delta} \left(1 - \frac{e^{i\theta_p}}{p^s}\right)^{-1} \left(1 - \frac{e^{-i\theta_p}}{p^s}\right)^{-1} \prod_{p \mid \Delta} (\cdots).$$

It thus becomes a multiplicative function.

More generally, (4.1) implies that, when $p \nmid \Delta$, then $f_k(p)$ is equal to the prime values of the coefficients of $L(s, \mathrm{sym}^k E)$, the k-th symmetric power of $L(s, E)$ (again, our normalization here means that the line of symmetry stays fixed at $\mathrm{Re}(s) = 1/2$). We then extend f_k to all integers via the formula

$$\sum_{n=1}^{\infty} \frac{f_k(n)}{n^s} = L(s, \mathrm{sym}^k E) = \prod_{p \nmid \Delta} \prod_{a+b=k} \left(1 - \frac{e^{(a-b)i\theta_p}}{p^s}\right)^{-1} \prod_{p \mid \Delta} (\cdots).$$

It thus becomes a multiplicative function, i.e.

$$\sum_{n=1}^{\infty} \frac{f_k(n)}{n^s} = \prod_p \left(1 + \frac{f_k(p)}{p^s} + \frac{f_k(p^2)}{p^{2s}} + \cdots\right).$$

We will often write

$$L(s, f_k) \quad \text{instead of} \quad L(s, \mathrm{sym}^k E).$$

We also let Λ_{f_k} be the coefficients of the logarithmic derivative, i.e.

$$\sum_{n=1}^{\infty} \frac{\Lambda_{f_k}(n)}{n^s} = -\frac{L'}{L}(s, \mathrm{sym}^k E).$$

In particular, we have

$$\Lambda_{f_k}(p) = f_k(p) \log p.$$

We have not specified the Euler factors at primes $p \mid \Delta$ because from the point of view of proving (3.2) they are completely irrelevant. Indeed, we have

$$\sum_{n \le x} \Lambda_{f_k}(n) = \sum_{p \le x,\ p \nmid \Delta} f_k(p) \log p + O_{E,k}(\sqrt{x}),$$

so that (3.2) for $k \geq 1$ becomes

$$\sum_{n \leq x} \Lambda_{f_k}(n) = o_{x \to \infty}(x). \tag{5.1}$$

6. How to prove a prime number theorem

We now arrive to the importance of automorphy of $L(s, \mathrm{sym}^k E)$. The reason why we built the Dirichlet series $L(s, \mathrm{sym}^k E)$ out of the $f_k(p)$'s was in the hope that we could say something about it. To understand this better, let us consider a more down to earth example, the case of Dirichlet characters. To prove the prime number theorem in arithmetic progressions, it suffices to study $\sum_{p \leq x} \chi(p)$. To do so, we consider the L-function

$$L(s, \chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

The reason why this is such a nice object to consider is that understanding $\chi(n)$ on average over integers n is very easy because χ is a periodic function. Automorphy is a generalization of periodicity to more complicated objects. Knowing that $L(s, \mathrm{sym}^k E)$ is automorphic would immediately imply all sorts of very nice properties (such as the existence of a functional equation) for which we refer the reader to [5] and [3]. The one property that is relevant for our purposes is that the automorphy of $L(s, \mathrm{sym}^k E)$ implies that

$$\sum_{n \leq x} f_k(n) \ll x^{1-\delta} \quad (x \geq 2), \tag{6.1}$$

where we allow the implied constant to depend on k and E. To prove it, we simply write

$$\sum_{n \leq x} f_k(n) = \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=2} L(s, \mathrm{sym}^k E) \frac{x^s}{s} \mathrm{d}$$

and use the analyticity of $L(s, \mathrm{sym}^k E)$ to shift the contour to a line $\mathrm{Re}(s) = 1 - \delta - o(1)$. Ignoring issues of convergence, and bounding the integrant trivially on this new line of integration yields (6.1).

It is relation (6.1) that is the crucial one and that we are going to use (together with an analogous property of the Rankin-Selberg convolutions $f_k \otimes f_k$ and $f_k \otimes f_2$, as we will see later). We already saw how it follows by knowing the analyticity of $L(s, \mathrm{sym}^k E)$ in a region $\mathrm{Re}(s) > 1 - \delta - o(1)$. Conversely, if we knew (6.1), then we could easily analytically continue $L(s, f_k)$ to the half-plane $\mathrm{Re}(s) > 1 - \delta$. Indeed, partial summation implies that

$$L(s, f_k) = \int_{1^-}^{\infty} \frac{1}{x^s} \mathrm{d} \sum_{n \leq x} f_k(n) = s \int_1^{\infty} \sum_{n \leq x} f_k(n) \frac{\mathrm{d}x}{x^{s+1}}.$$

Under the assumption of (6.1), the right hand side is an absolutely convergent integral for $\mathrm{Re}(s) > 1 - \delta$, so that it defines an analytic function in this region. This establishes the claimed analytic continuation.

Assuming that we know the analytic continuation of $L(s, \mathrm{sym}^k f)$ to the region $\mathrm{Re}(s) > 1 - \delta$, the logarithmic derivative is a meromorphic function in $\mathrm{Re}(s) > 1 - \delta$. We then use Perron's formula to write the partial sums of Λ_{f_k} in terms of it:

$$\sum_{n \leq x} \Lambda_{f_k}(n) = \frac{1}{2\pi i} \int_{\mathrm{Re}(s)=2} \left(-\frac{L'}{L}\right) (s, \mathrm{sym}^k E) \frac{x^s}{s} \mathrm{d}s.$$

The important thing to notice here is that $|x^s| = x^{\mathrm{Re}(s)}$ for all s. So $|x^s| = x^2$ on the line of integration $\mathrm{Re}(s) = 2$. However, we are after a bound on $\sum_{n\leq x} \Lambda_{f_k}(n)$ that is $o(x)$ as $x \to \infty$. So we shift the line of integration to $\mathrm{Re}(s) = 1 - \varepsilon$ for some $\varepsilon \in (0, \delta)$ (ignoring all annoying convergence issues when doing so). By Cauchy's residue theorem, when doing so, we pick up poles in the region enclosed by the lines $\mathrm{Re}(s) = 2$ and $\mathrm{Re}(s) = 1 - \varepsilon$ (which is actually a compact region in the Riemann sphere). We thus arrive to the formula

$$\sum_{n\leq x} \Lambda_{f_k}(n) = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=1-\varepsilon}\left(-\frac{L'}{L}\right)(s, \mathrm{sym}^k E)\frac{x^s}{s}\mathrm{d}s + \text{residues}.$$

The new integral is an error term, because $|x^s| = x^{1-\varepsilon} = o(x)$ here. If we could show the residues don't contribute a lot, we would be done. Since $L(s, \mathrm{sym}^k E)$ is analytic for $\mathrm{Re}(s) > 1 - \delta$, the only potential residues of the integrant $(-L'/L)(s, \mathrm{sym}^k E)x^s/s$ come from potential zeroes of $L(s, \mathrm{sym}^k E)$. If we could somehow prove there were no zeroes with $\mathrm{Re}(s) = 1$, we would be done. In the next section, we will show how to prove that $L(1 + it, \mathrm{sym}^k E) \neq 0$ for all $k \geq 1$ and for all $t \in \mathbb{R}$.

We conclude this section by stating without proof a more precise result that shows that, as a matter of fact, we only need a much weaker version of (6.1), together with non-vanishing on the 1-line, to deduce the Sato-Tate conjecture (for us, $g = f_k$ and $D = k + 1$):

Theorem 1. *Let g be a multiplicative function with Dirichlet series $L(s, g)$. Let Λ_g be the coefficients of $-L'(s, g)/L(s, g)$, and assume that there is some D such that $|\Lambda_g| \leq D\Lambda$ (in particular, $|g(p)| \leq D$). Assume further that*

$$\sum_{n\leq x} g(n) \ll \frac{x}{(\log x)^{D+1+\varepsilon}} \quad (x \geq 2). \tag{6.2}$$

If we list the zeroes of $L(s, g)$ on the 1-line (with multiplicity), say $1 + i\gamma_1, \ldots, 1 + i\gamma_m$, then

$$\sum_{p\leq x}(g(p) + p^{i\gamma_1} + \cdots + p^{i\gamma_m}) = o_{x\to\infty}(x/\log x).$$

In particular, if $L(1 + it, g) \neq 0$ for all t, then

$$\sum_{p\leq x} g(p) = o_{x\to\infty}(x/\log x).$$

7. de la Vallée-Poussin's argument

Now, we are going to show how automorphy can be used to prove that $L(1+it, \mathrm{sym}^k E) \neq 0$ for all t. (A suitable adaptation of the argument can be made to work under weaker hypotheses of the form (6.2).)

We begin by outlining the argument for the Riemann zeta function, that goes back to de la Vallée-Poussin (though we will present a more modern version of it using the theory of *pretentious multiplicative functions*). If $\zeta(1 + it) = 0$, then there is $c \in \mathbb{C}$ such that $\zeta(\sigma + it) \sim c \cdot (\sigma - 1) \sim c/\zeta(\sigma) \ll 1/\zeta(\sigma)$ as $\sigma \to 1^+$. Since

$$\zeta(\sigma)\zeta(\sigma + it) = \prod_p \left(1 + \frac{1 + p^{-it}}{p^\sigma} + O\left(\frac{1}{p^2}\right)\right),$$

taking logarithms and then real parts yields

$$\sum_p \frac{1+\mathrm{Re}(p^{-it})}{p^\sigma} \le O(1).$$

On the other hand, the LHS is ≥ 0, so that

$$\sum_p \frac{1+\mathrm{Re}(p^{-it})}{p^\sigma} = O(1).$$

Dividing by $\sum_p 1/p^\sigma$, we infer that

$$\frac{\sum_p \mathrm{Re}(p^{-it})/p^\sigma}{\sum_p 1/p^\sigma} \sim -1 \quad (\sigma \to 1^+).$$

We express this conceptually by writing that

$$\mathbb{E}[\mathrm{Re}(p^{-it})] = -1, \tag{7.1}$$

where it is understood that given a sequence a_p indexed by primes p we define its average value by

$$\mathbb{E}[a_p] = \lim_{\sigma\to 1^+} \frac{\sum_p a_p/p^\sigma}{\sum_p 1/p^\sigma},$$

provided that this limit exists.

Relation (7.1) says that $p^{-it} \sim -1$ on average. But then we should have that $p^{2it} \sim 1$, and performing a similar argument as above should imply that $\zeta(1+it) = \infty$. More precisely, we claim that $\mathbb{E}[\mathrm{Re}(p^{2it})] = 1$, i.e. that

$$\sum_p \frac{1-\mathrm{Re}(p^{-2it})}{p^\sigma} = o\left(\sum_p \frac{1}{p^\sigma}\right).$$

Since $\log\zeta(\sigma+2it) = \sum_p 1/p^{\sigma+2it} + O(1)$ and $\log\zeta(\sigma) = \sum_p 1/p^\sigma + O(1) = -\log(\sigma-1) + O(1)$, we thus find hat $|\zeta(\sigma+2it)| \ge 1/(\sigma-1)^{1-o(1)}$ as $\sigma \to 1^+$, whence $\zeta(1+2it) = \infty$. But this can only happen when $t = 0$, since the only pole of ζ is at $s = 1$. But we started with the assumption that $\zeta(1+it) = 0$, so $t \ne 0$. We have thus arrived at a contradiction.

The above argument completes the proof of the non-vanishing of ζ on the line $\mathrm{Re}(s) = 1$, minus a small detail: how do we actually prove that $\mathbb{E}[\mathrm{Re}(p^{-2it})] = 1$ when $\mathbb{E}[\mathrm{Re}(p^{-it})] = -1$? Note that

$$\begin{aligned}\left(\sum_p \frac{1+\mathrm{Re}(p^{-it})}{p^\sigma}\right)^{1/2} &= \frac{1}{\sqrt{2}}\left(\sum_p \frac{|1+p^{-it}|^2}{p^\sigma}\right)^{1/2} \\ &= \frac{1}{2\sqrt{2}}\left(\sum_p \frac{|1+p^{-it}|^2}{p^\sigma}\right)^{1/2} + \frac{1}{2\sqrt{2}}\left(\sum_p \frac{|1+p^{it}|^2}{p^\sigma}\right)^{1/2} \\ &\ge \frac{1}{2\sqrt{2}}\left(\sum_p \frac{|p^{-it}-p^{-it}|^2}{p^\sigma}\right)^{1/2} = \frac{1}{2}\left(\sum_p \frac{1-\mathrm{Re}(p^{-2it})}{p^\sigma}\right)^{1/2},\end{aligned}$$

where we used Minkowski's inequality. This completes the proof that $\zeta(1+it) \ne 0$ for all t.

8. Non-vanishing of $L(s, \operatorname{sym}^k E)$ on the line $\operatorname{Re}(s) = 1$

We imitate the argument for ζ: having a zero of $L(s, \operatorname{sym}^k E)$ at $s = 1 + it$ means that $L(\sigma + it, \operatorname{sym}^k E) \sim c \cdot (\sigma - 1) \sim c/\zeta(\sigma)$ as $\sigma \to 1^+$, so that $|L(\sigma + it, \operatorname{sym}^k E)\zeta(\sigma)| \ll 1$. Taking logarithms, we find that

$$\sum_p \frac{1 + \operatorname{Re}(f_k(p)p^{-it})}{p^\sigma} \leq O(1). \tag{8.1}$$

In fact, if the order vanishing of $L(s, \operatorname{sym}^k E)$ at $s = 1 + it$ is $m \geq 1$, then we can prove that

$$\sum_p \frac{m + \operatorname{Re}(f_k(p)p^{-it})}{p^\sigma} = O(1), \tag{8.2}$$

whence

$$\mathbb{E}[\operatorname{Re}(f_k(p)p^{-it})] = -m. \tag{8.3}$$

We want to use the trick of de la Vallée-Poussin and say that since $\operatorname{Re}(f_k(p)p^{-it})$ is $-m$ on average, then it must be the case that $\operatorname{Re}(f_k(p)^2 p^{-2it}) \geq 1$, but this not as straightforward now because $|f_k(p)|$ could potentially be large (as large as $k + 1$).

We consider the Rankin-Selberg L-function $L(s, f_k \otimes f_k)$. It is the Dirichlet series of a multiplicative function whose coefficients at primes $p \nmid \Delta$ are given by $f_k(p)^2$. Since f_k is real-valued, we know from the theory of Rankin-Selberg convolutions that $L(s, f_k \otimes f_k)$ is meromorphic with its only singularity being a simply pole at $s = 1$. In particular, $L(\sigma, f_k \otimes f_k) \sim c/(\sigma - 1) \sim c\zeta(\sigma)$ as $\sigma \to 1^+$, whence

$$\sum_p \frac{f_k(p)^2 - 1}{p^\sigma} = O(1) \tag{8.4}$$

as $\sigma \to 1^+$. We claim that this implies that $m = 1$.

Indeed, by the Cauchy-Schwarz inequality, we have

$$\frac{\sum_p |f_k(p)|/p^\sigma}{\sum_p 1/p^\sigma} \leq \left(\frac{\sum_p f_k(p)^2/p^\sigma}{\sum_p 1/p^\sigma}\right)^{1/2} \leq \left(1 + O\left(\frac{1}{\sum_p 1/p^\sigma}\right)\right)^{1/2} \leq 1 + O\left(\frac{1}{\sum_p 1/p^\sigma}\right),$$

whence

$$\sum_p \frac{|f_k(p)|}{p^\sigma} \leq \sum_p \frac{1}{p^\sigma} + O(1).$$

Since $|f_k(p)| \geq -\operatorname{Re}(f_k(p)p^{-it})$, we also find that

$$\sum_p \frac{|f_k(p)|}{p^\sigma} \geq -\sum_p \frac{\operatorname{Re}(f_k(p)p^{-it})}{p^\sigma} = \sum_p \frac{m}{p^\sigma} + O(1)$$

by (8.2). In particular, $m = 1$ (by comparing the above estimates when $\sigma \to 1^+$) and

$$\sum_p \frac{|f_k(p)| - 1}{p^\sigma} = O(1). \tag{8.5}$$

Feeding this back into (8.1), we find that

$$\sum_p \frac{|f_k(p)| + \operatorname{Re}(f_k(p)p^{-it})}{p^\sigma} \leq O(1),$$

the advantage of the above formula being that its summands are non-negative integers. If $z_p = f_k(p)p^{-it}/|f_k(p)|$, then

$$|f_k(p)| + \mathrm{Re}(f_k(p)p^{-it}) = |f_k(p)|(1+\mathrm{Re}(z_p)),$$

so that repeating the argument from the end of the previous section with Minkowksi's inequality yields that

$$\frac{1}{4}\sum_p \frac{|f_k(p)|\cdot(1-\mathrm{Re}(z_p^2))}{p^\sigma} \le \sum_p \frac{|f_k(p)|\cdot(1+\mathrm{Re}(z_p))}{p^\sigma} \le O(1).$$

(Or, simply note that $1+\mathrm{Re}(z) \ge (1-\mathrm{Re}(z^2))/4$ when $|z|=1$.) Since $|f_k(p)| \le k+1$, we deduce that

$$\sum_p \frac{f_k(p)^2 - \mathrm{Re}(f_k(p)^2p^{-2it})}{p^\sigma} = \sum_p \frac{|f_k(p)|^2\cdot(1-\mathrm{Re}(z_p^2))}{p^\sigma} \le O(1).$$

Together with (8.4), this implies that

$$\sum_p \frac{\mathrm{Re}(f_k(p)^2p^{-2it})}{p^\sigma} = \sum_p \frac{1}{p^\sigma} + O(1).$$

But this would then mean that $L(s, f_k \otimes f_k)$ has a pole at $s = 1+2it$. As in the case of ζ, this can only happen when $t=0$. But unlike the case of ζ, we do not automatically know that $L(1, \mathrm{sym}^k E) \ne 0$, because $L(s, \mathrm{sym}^k E)$ does not have a pole at $s=1$. Disproving this remaining potential zero (an extreme case of a Siegel zero) is trickier and requires a different argument that we give in the next section.

Remark 1. There is another way of carrying out the above argument, which requires a trick: we have

$$\begin{aligned}0 &\le \sum_p \frac{(1+f_k(p)p^{it}+f_k(p)p^{-it})^2}{p^\sigma}\\ &= \sum_p \frac{1+f_k(p)^2p^{2it}+f_k(p)^2p^{-2it}+2f_k(p)p^{it}+2f_k(p)p^{-it}+2f_k(p)^2}{p^\sigma}\\ &= \sum_p \frac{1+2\mathrm{Re}(f_k(p)^2p^{2it})+4\mathrm{Re}(f_k(p)p^{it})+2f_k(p)^2}{p^\sigma}\end{aligned}$$

However, when $t=0$, we know that $\mathrm{Re}(f_k(p)^2p^{2it}) \le 0$ on average by the analyticity of $L(s, f_k\otimes f_k)$ at $s = 1+2it$, whereas $f_k(p)^2 = 1$ on average and we have seen that $\mathrm{Re}(f_k(p)p^{it}) = -1$ on average. Thus we find that the RHS is $\le 1-4+2 = -1$ on average, which is impossible.

9. And what about Siegel zeroes?

Our goal is to prove that $L(1, \mathrm{sym}^k E) \ne 0$ and complete the proof of the Sato-Tate conjecture. When k is even, this is actually pretty easy: using the formula (4.3) with $h=0$, we have that $U_k^2 = U_0 + U_2 + \cdots + U_{2k}$, whence $f_k(p)^2 = 1 + f_2(p) + \cdots + f_{2k}(p)$ for $p \nmid \Delta$. In fact, more is known (though we don't really need it): we have that

$$L(s, f_k \otimes f_k) = \zeta(s) L(1, \operatorname{sym}^2 f) L(1, \operatorname{sym}^4 f) \cdots L(1, \operatorname{sym}^{2k} f).$$

As we saw above, the left hand side has a simple pole at $s = 1$. So does $\zeta(s)$. So we cannot have that $L(1, \operatorname{sym}^{2k} f) = 0$, because this would remove the singularity at $s = 1$ from the right hand side.

It remains to treat the case of odd symmetric powers. In this case, we apply (4.3) with $h = 2k-1$ to find that $U_2 U_{2k+1} = U_{2k-1} + U_{2k+1}$, whence $f_2(p) f_{2k+1}(p) = f_{2k-1}(p) + f_{2k+1}(p)$. In terms of L-functions, this (essentially) means that

$$L(s, f_2 \otimes f_{2k+1}) = L(s, \operatorname{sym}^{2k-1} E) L(s, \operatorname{sym}^{2k+1} E).$$

In particular, if $L(1, \operatorname{sym}^{2k+1} E) = 0$, then we also have that $L(1, f_2 \otimes f_{2k+1}) = 0$. Now, as we discussed in the previous section, these two facts would mean that

$$\sum_p \frac{1 + f_{2k+1}(p)}{p^\sigma} = O(1) \quad \text{and} \quad \sum_p \frac{1 + f_2(p) f_{2k+1}(p)}{p^\sigma} = O(1).$$

for all $\sigma > 1$. We are almost done, because we infer from the above relations that $\mathbb{E}[f_2(p) f_{2k+1}(p)] = \mathbb{E}[f_{2k+1}(p)] = -1$, so we should have that $\mathbb{E}[f_2(p)] = 1$, which would then induce a pole at $s = 1$ to the symmetric square $L(s, \operatorname{sym}^2 E)$ (and this is absurd). To complete the proof rigorously, we note that:

$$\begin{aligned} 0 \le \sum_p \frac{(f_2(p) + f_{2k+1}(p))^2}{p^\sigma} &= \sum_p \frac{f_2(p)^2 + 2 f_2(p) f_{2k+1}(p) + f_{2k+1}(p)^2}{p^\sigma} \\ &= \sum_p \frac{1 - 2 + 1}{p^\sigma} + O(1) = O(1), \end{aligned}$$

where we also used (8.4). In particular, Cauchy-Schwarz implies that

$$\sum_p \frac{|f_2(p) + f_{2k+1}(p)|}{p^\sigma} \ll \left(\sum_p \frac{1}{p^\sigma} \right)^{1/2}.$$

Thus

$$\sum_p \frac{f_2(p)}{p^\sigma} = O(1) + \sum_p \frac{1}{p^\sigma} + \sum_p \frac{f_2(p) + f_{2k+1}(p)}{p^\sigma} = \sum_p \frac{1}{p^\sigma} + \left(\sum_p \frac{1}{p^\sigma} \right)^{1/2},$$

a contradiction to the analyticity of $L(s, \operatorname{sym}^2 E)$ at $s = 1$.

References

[1] L. Clozel, M. Harris, Michael, and R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations.* With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras. Publ. Math. Inst. Hautes Études Sci. No. 108 (2008), 1–181.

[2] H. Davenport, *Multiplicative number theory.* Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, 74. Springer-Verlag, New York, 2000.

[3] D. Goldfeld, *Automorphic forms and L-functions for the group $GL(n, \mathbb{R})$.* With an appendix by Kevin A. Broughan. Cambridge Studies in Advanced Mathematics, 99. Cambridge University Press, Cambridge, 2006.

[4] M. Harris, N. Shepherd-Barron, and R. Taylor, *A family of Calabi-Yau varieties and potential automorphy.* Ann. of Math. (2) 171 (2010), no. 2, 779–813.

[5] H. Iwaniec and E. Kowalski, *Analytic number theory.* American Mathematical Society Colloquium Publications, 53, American Mathematical Society, Providence, RI, 2004.

[6] D. Koukoulopoulos, *On multiplicative functions which are small on average.* Geom. Funct. Anal., 23 (2013), no. 5, 1569–1630.

[7] D. Koukoulopoulos and K. Soundararajan, *The structure of multiplicative functions with small partial sums.* Preprint (2016).

[8] R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations. II.* Publ. Math. Inst. Hautes Études Sci. No. 108 (2008), 183–239.

Département de mathématiques et de statistique, Université de Montréal, CP 6128 succ. Centre-Ville, Montréal, QC H3C 3J7, Canada

E-mail address: koukoulo@dms.umontreal.ca