COUNTEREXAMPLES RELATED TO THE SATO–TATE CONJECTURE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Daniel Miller

May 2017

COUNTEREXAMPLES RELATED TO THE SATO–TATE CONJECTURE

Daniel Miller, Ph.D.

Cornell University 2017

Let $E_{/\mathbf{Q}}$ be an elliptic curve. The Sato–Tate conjecture, now a theorem, tells us that the angles $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right)$ are equidistributed in $[0,\pi]$ with respect to the measure $\frac{2}{\pi}\sin^2\theta\,\mathrm{d}\theta$ if E is non-CM (resp. $\frac{1}{2\pi}\mathrm{d}\theta + \frac{1}{2}\delta_{\pi/2}$ if E is CM). In the non-CM case, Akiyama and Tanigawa conjecture that the discrepancy

$$D_N = \sup_{x\in[0,\pi]} \left| \frac{1}{\pi(N)} \sum_{p\leqslant N} 1_{[0,x]}(\theta_p) - \int_0^x \frac{2}{\pi}\sin^2\theta\,\mathrm{d}\theta \right|$$

asymptotically decays like $N^{-\frac{1}{2}+\epsilon}$, as is suggested by computational evidence and certain reasonable heuristics on the Kolmogorov–Smirnov statistic. This conjecture implies the Riemann hypothesis for all L-functions associated with E. It is natural to assume that the converse ("generalized Riemann hypothesis implies discrepancy estimate") holds, as is suggested by analogy with Artin L-functions. We construct, for CM abelian varieties, "fake Satake parameters" yielding L-functions which satisfy the generalized Riemann hypothesis, but for which the discrepancy decays like $N^{-\epsilon}$. This provides evidence that in the CM case, the converse to "Akiyama–Tanigawa conjecture implies generalized Riemann hypothesis" does not hold.

We also show that there are Galois representations $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Z}_l)$, ramified at an arbitrarily thin set of primes, whose Satake parameters can be made to converge at any specified rate to any fixed measure μ on $[0,\pi]$ for which $\cos_* \mu$ is absolutely continuous with bounded derivative.

BIOGRAPHICAL SKETCH

Daniel Miller was born in St. Paul, Minnesota. He completed his Bachelor of Science at the University of Nebraska Omaha. In addition to his studies there, he played the piano competitively and attended Cornell's Summer Mathematics Institute. He started his Ph.D. at Cornell planning on a career in academia. Halfway through he had a change of heart, and will be joining Microsoft's Analysis and Experimentation team as a data scientist after graduation. He is happily married to Ivy Lai Miller, and owns a cute but grumpy cat named Socrates.

This thesis is dedicated to my undergraduate adviser, Griff Elder. He is the reason I considered a career in mathematics, and his infectious enthusiasm for number theory has inspired me more than I can say.

ACKNOWLEDGEMENTS

I could not have completed this thesis without help and support from many people. I would like to offer my sincerest thanks to the following people.

My parents Jay and Cindy, for noticing and fostering my mathematical interests early on, and for being unfailingly loving and supportive.

My undergraduate thesis advisor, Griffith Elder. Without his encouragement and inspiration I probably would have never considered a career in math.

Tara Holm, Jason Boynton, and Anthony Weston, for making Cornell's 2011 Summer Mathematics Institute the fantastic introduction to research it was.

My fellow graduate students Sasha Patotski, Balázs Elek, and Sergio Da Silva, for sharing my early love of algebraic geometry, laughing with me at the absurdities of academic life, and listening to my ramblings about number theory.

The mathematics department at Cornell, where many professors were generous with their time and ideas. I appreciate Yuri Berest, John Hubbard, Farbod Shokrieh, Birget Speh, and David Zywina for letting me bounce ideas off them, helping me add rigor to half-baked ideas, and pointing my research in new and interesting directions.

My adviser Ravi Ramakrishna. He kindled my first love for number theory, stayed supportive as my research bounced all over the place, and kept me focused, grounded, and concrete when I needed to be.

Most importantly, my wife Ivy for being there for me through the highs and the lows, when I prematurely thought my thesis was complete, and when I thought my results were completely in shambles. I couldn't have done it without her.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivation from classical analytic number theory

Start with an old problem central to number theory: counting prime numbers. As usual, let $\pi(x)$ be the prime counting function and $\mathrm{Li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}$ be the Eulerian logarithmic integral. The prime number theorem tells us that as $x \to \infty$, $\frac{\pi(x)}{\mathrm{Li}(x)} \to 1$. The standard approach to proving the prime number theorem is by showing that the Riemann ζ-function has non-vanishing meromorphic continuation to $\Re = 1$.

Theorem 1.1.1. *The function $\zeta(s)$ admits a non-vanishing meromorphic continuation to $\Re = 1$ with at most a simple pole at $s = 1$, if and only if* $\lim_{x\to\infty} \frac{\pi(x)}{\mathrm{Li}(x)} = 1$.

Since $\zeta(s)$ does have the desired properties, the prime number theorem is true. It is natural to try to bound the difference $\pi(x) - \mathrm{Li}(x)$. Numerical experiements dating back to Gauss suggest that $|\pi(x) - \mathrm{Li}(x)| \ll x^{-\frac{1}{2}+\epsilon}$, and in fact we have the following result.

Theorem 1.1.2 ([Edw74, Th., p. 90])**.** *The Riemann hypothesis is true if and only if $|\pi(x) - \mathrm{Li}(x)| \ll x^{-\frac{1}{2}+\epsilon}$.*

Neither side of this equivalence is known for certain to be true!

The above discussion generalizes naturally to Artin L-functions. Let $K/\mathbf{Q}$ be a finite Galois extension with group $G = \mathrm{Gal}(K/\mathbf{Q})$. For any rational prime p at which K is unramified, let fr_p be the conjugacy class of the Frobenius at p in G. For any irreducible representation ρ of G, there is a corresponding L-function defined as

$$L(\rho, s) = \prod_p \det\left(1 - \rho(\mathrm{fr}_p)p^{-s}\right)^{-1},$$

where here (and for the remainder of this thesis) we tacitly omit from the product those primes at which ρ is ramified. Given a cutoff x, there is a natural empirical measure $P_x = \frac{1}{\pi(x)} \sum_{p \leqslant x} \delta_{\mathrm{fr}_p}$ on $G^\natural$, the set of conjugacy classes in G. Let μ be the (normalized) Haar measure on $G^\natural$, and let $\mathrm{D}(P_x) = \max_{S \subset G^\natural} |P_x(S) - \mu(S)|$. Then P_x converges weakly to the Haar measure on $G^\natural$ if and only if $\mathrm{D}(P_x) \to 0$. Recall that weak convergence of P_x to μ means $\int f \, \mathrm{d}P_x \to \int f$ for all continuous functions f on $G^\natural$. Since $G^\natural$ is a finite set, all functions on $G^\natural$ are continuous, but later on we will consider weak convergence on more general spaces.

Theorem 1.1.3 ([Ser89, Th. 2 Cor., A.1])**.** *The measures P_x converge weakly to the Haar measure on $G^\natural$ if and only if the function $L(\rho, s)$ admits a non-vanishing analytic continuation to $\Re = 1$ for all nontrivial ρ.*

Both sides of this equivalence are true, and known as the Chebotarev density theorem. Moreover, there is a version of the strong prime number theorem in this context. In this thesis, the Riemann hypothesis for a Dirichlet series $L(s)$ is the statement that $L(s)$ admits a non-vanishing analytic continuation to $\Re > \frac{1}{2}$.

Theorem 1.1.4. *The bound $\mathrm{D}(P_x) \ll x^{-\frac{1}{2}+\epsilon}$ holds if and only if each $L(\rho, s)$, ρ nontrivial, satisfies the Riemann hypothesis.*

The forward implication follows from Theorem 3.2.1, while the reverse implication is a result of Serre [Ser81, Th. 4]. This whole discussion generalizes to a more complicated set of Galois representations—those arising from elliptic curves and more general motives.

1.2 Discrepancy and the Riemann hypothesis for elliptic curves

For background on the Galois representations and L-functions associated to elliptic curves, see [Sil09, III§7, C§17]. Throughout this thesis, what we call the L-function of an elliptic curve (motive, etc.) is the normalized (i.e. analytic instead of algebraic) L-function. Let $E_{/\mathbf{Q}}$ be a non-CM elliptic curve. For any prime l, the l-adic Tate module of E induces a continuous representation $\rho_l \colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$. It is known that the quantities $a_p = \operatorname{tr} \rho_l(\mathrm{fr}_p)$ lie in $\mathbf{Z}$ and satisfy the Hasse bound $|a_p| \leqslant 2\sqrt{p}$. For each unramified prime p, the corresponding Satake parameter for E is $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right) \in [0,\pi]$. These parameters are packaged into an L-function as follows:

$$L(E,s) = \prod_p \frac{1}{(1-e^{i\theta_p}p^{-s})(1-e^{-i\theta_p}p^{-s})} = \prod_p \det\left(1 - \begin{pmatrix} e^{i\theta_p} & \\ & e^{-i\theta_p} \end{pmatrix} p^{-s}\right)^{-1}.$$

More generally we have, for each irreducible representation sym^k of $\mathrm{SU}(2)$, the k-th symmetric power L-function:

$$L(\mathrm{sym}^k E, s) = \prod_p \prod_{j=0}^{k} \frac{1}{1-e^{i(k-2j)\theta_p}p^{-s}} = \prod_p \det\left(1 - \mathrm{sym}^k \begin{pmatrix} e^{i\theta_p} & \\ & e^{-i\theta_p} \end{pmatrix} p^{-s}\right)^{-1}.$$

Numerical experiments suggest that the Satake parameters are equidistributed with respect to the Sato–Tate distribution $\mathrm{ST} = \frac{2}{\pi}\sin^2\theta \, \mathrm{d}\theta$. Indeed, for any cutoff x, let P_x be the empirical measure $\frac{1}{\pi(x)}\sum_{p\leqslant x}\delta_{\theta_p}$. The convergence of P_x to the Sato–Tate measure is closely related to the analytic properties of the $L(\mathrm{sym}^k E, s)$. First, here is the famous Sato–Tate conjecture (now a theorem) in our notation.

Theorem 1.2.1 (Taylor et. al.)**.** *The measures P_x converge weakly to* ST.

Theorem 1.2.2 (Serre)**.** *The Sato–Tate conjecture holds for E if and only if each of the functions $L(\mathrm{sym}^k E, s)$ have analytic continuation to $\Re = 1$.*

The stunning recent proof of the Sato–Tate conjecture [HSBT10] showed that the functions $L(\operatorname{sym}^k E, s)$ have the desired analytic continuation. In fact, they show that for all k, $L(\operatorname{sym}^k E, s)$ has meromorphic continuation to the whole complex plane. Even better, when k is odd, the L-function is potentially automorphic. See Theorem 5.3.4 for a result in this thesis where more can be said about odd symmetric power L-functions than even ones.

The Riemann hypothesis, and its analogue for Artin L-functions, has a natural generalization to elliptic curves. In this context, the discrepancy of the set $\{\theta_p\}_{p\leqslant x}$ is

$$\mathrm{D}_x\left(E, \mathrm{ST}\right) = \sup_{t\in[0,\pi]} \left|P_x[0,t] - \mathrm{ST}[0,t]\right|.$$

The following conjecture is made in [AT99].

Conjecture 1.2.3 (Akiyama–Tanigawa)**.** $\mathrm{D}_x\left(E, \mathrm{ST}\right) \ll x^{-\frac{1}{2}+\epsilon}$.

Akiyama and Tanigawa go on to prove a special case of the following theorem, proved in full generality by Mazur.

Theorem 1.2.4 ([Maz08, §3.4])**.** *If* $\mathrm{D}_x\left(E, \mathrm{ST}\right) \ll x^{-\frac{1}{2}+\epsilon}$*, then all the functions* $L(\operatorname{sym}^k E, s)$ *satisfy the Riemann hypothesis.*

This discussion also makes sense when E has complex multiplication (for simplicity, we consider $E_{/F}$ where F is the field of definition of the complex multiplication). The Sato–Tate measure for such E is the Haar measure on $\mathrm{SO}(2)$, i.e. the uniform measure on $[0,\pi]$. Instead of symmetric power L-functions, there is an L-function for each character of $\mathrm{SO}(2)$. Once again, there is a theorem "Akiyama–Tanigawa conjecture implies Riemann hypothesis." For a precise statement and proof, see Theorem 4.3.2.

It is natural to assume that the converse to the implication "Akiyama–Tanigawa conjecture implies Riemann hypothesis" holds. Zywina first suggested to the author that it might not. In this thesis, we construct a range of counterexamples to the implication "Akiyama–Tanigawa conjecture implies generalized Riemann hypothesis" for the case of CM abelian varieties. Moreover, we generalize the results of [Pan11] to show that there can be no purely Galois-theoretic proof of the Sato–Tate conjecture, for there are Galois representations with arbitrary Sato–Tate distributions! We also show that some of the results of [Sar07] about sums of the form $\sum_{p\leqslant x}\frac{a_p}{\sqrt{p}}$ cannot be generalized to general—in particular, infinitely ramified—Galois representations.

1.3 Notation conventions

If S is a set, 1_S is the characteristic function of S.

Whenever l is mentioned it is a rational prime $\geqslant 7$.

Write $f \ll g$ if $f = O(g)$, i.e. there is a constant $C > 0$ such that $f \leqslant Cg$.

Write $f = \Omega(g)$ (in the convention of Hardy–Littlewood) if $\limsup \frac{f}{g} > 0$.

The symbol $f = \Theta(g)$ means there exist constants $0 < C_1 < C_2$ such that $C_1 g \leqslant f \leqslant C_2 g$. Equivalently, $g \ll f$ and $f \ll g$.

If μ is a measure on $\mathbf{R}$, then write $\mu[a, b]$ for $\mu([a, b])$, and similarly for $[a, b)$, $(a, b]$, etc. Whenever it simplifies the notation, we will write μS for $\mu(S)$.

If μ is a measure on $\mathbf{R}$, then *cumulative distribution function (cdf) of* μ is given by $\mathrm{cdf}_\mu(x) = \mu[-\infty, x]$.

If $z \in \mathbf{C}$, write $\Re z$ for the real part of z.

If $\alpha \in \mathbf{R}$, we write $\Re > \alpha$ for the half-plane of complex numbers with real part $> \alpha$. So a function has analytic continuation to the half-plane $\{z \in \mathbf{C} : \Re z > \alpha\}$ if and only if the function extends to an analytic function on $\Re > \alpha$.

We write $\boldsymbol{x} = (x_1, x_2, \ldots)$ for infinite sequences and $\vec{x} = (x_1, \ldots, x_d)$ for vectors. Sometimes we will have a sequence of vectors, written as $\vec{\boldsymbol{x}} = (\vec{x}_1, \vec{x}_2, \ldots)$.

If $\boldsymbol{x} = (x_1, x_2, \ldots)$ is a sequence, write $P_{\boldsymbol{x},N} = \frac{1}{N} \sum_{n \leqslant N} \delta_{x_n}$ for the corresponding empirical measure. If $\boldsymbol{x} = (x_\alpha)$ is instead indexed by some other discrete subset of $\mathbf{R}^+$ (for example a subset of the primes), write $P_{\boldsymbol{x},N} = \frac{1}{\#\{\text{indices } \leqslant N\}} \sum_{\alpha \leqslant N} \delta_{x_\alpha}$.

Omitted entries in matrices are zero, i.e. $\left(\begin{smallmatrix} a & \\ & b \end{smallmatrix}\right)$ means $\left(\begin{smallmatrix} a & 0 \\ 0 & b \end{smallmatrix}\right)$.

CHAPTER 2

DISCREPANCY

2.1 Equidistribution

Discrepancy (also known as the Kolmogorov–Smirnov statistic) is a way of measuring how closely sample data fits a predicted distribution. It has many applications in computer science and statistics, but here we will focus on only its basic properties, such as how discrepancy changes when sequences are "tweaked" and combined.

First, recall that discrepancy provides a way of sharpening the soft convergence results in [Ser89, A.1]. Let X be a compact topological space, $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ a sequence in X indexed by the rational primes.

Definition 2.1.1. Let μ be a continuous probability measure on X. The sequence $\boldsymbol{x}$ is *equidistributed* with respect to μ if for all $f \in C(X)$, we have

$$\lim_{N\to\infty} \frac{1}{\pi(N)} \sum_{p \leqslant N} f(x_p) = \int f \, \mathrm{d}\mu.$$

In other words, $\boldsymbol{x}$ is μ-equidistributed if the empirical measures $P_{\boldsymbol{x},N} = \frac{1}{\pi(N)} \sum_{p\leqslant N} \delta_{x_p}$ converge to μ in the weak topology. It is easy to see that $\boldsymbol{x}$ is μ-equidistributed if and only if $\left|\sum_{p\leqslant N} f(x_p)\right| = o(\pi(N))$ for all $f \in C(X)$ having $\int f \, \mathrm{d}\mu = 0$. One can restrict to any set of functions which generates a dense subpace of $C(X)^{\mu=0}$.

In the discussion in [Ser89, A.1], X is the space of conjugacy classes in a compact Lie group, and f is allowed to range over the characters of irreducible, nontrivial, unitary representations of the group. Serre's results can be generalized to a much broader class of Dirichlet series, which are of the form

$$L_f(\boldsymbol{x}, s) = \prod_p \frac{1}{1 - f(x_p) p^{-s}}.$$

In fact, in light of the following theorem, we can consider functions f which are only continuous almost everywhere. This allows us to consider step functions like $1_{[0,\pi/2)} - 1_{(\pi/2,\pi]}$ on $[0, \pi]$.

Theorem 2.1.2. *Let X be a compact topological space, μ a Radon probability measure on X, and $f \colon X \to \mathbf{C}$ bounded and continuous μ-almost everywhere. If $\boldsymbol{x}$ is μ-equidistributed, then*

$$\lim_{N\to\infty} \frac{1}{\pi(N)} \sum_{p \leqslant N} f(x_p) = \int f \, \mathrm{d}\mu.$$

Proof. We prove the more general result that if $\{\mu_n\}$ is a sequence of Radon probability measures on X which converges weakly to μ, then $\mu_n(f) \to \mu(f)$ for all f which are bounded and continuous μ-almost everywhere.

Let D be the set of points at which f is not continuous. For every $\epsilon > 0$, there exists an open $U_\epsilon \supset D$ with $\mu(U_\epsilon) < \epsilon$, and $f_\epsilon \in C(X)$ such that $|f_\epsilon|_\infty \leqslant |f|_\infty$, $f_\epsilon|_D = 0$, and $f_\epsilon|_{X \smallsetminus U_\epsilon} = f|_{X \smallsetminus U_\epsilon}$. Note that

$$|\mu_n f - \mu f| \leqslant |\mu_n f - \mu_n f_\epsilon| + |\mu_n f_\epsilon - \mu f_\epsilon| + |\mu f_\epsilon - \mu f|. \tag{2.1}$$

Now $|\mu_n f - \mu f| \leqslant \mu_n(U_\epsilon)|f|_\infty$. Since U_ϵ is open, this converges to $\mu(U_\epsilon)|f|_\infty < \epsilon|f|_\infty$. The second term in (2.1) converges to zero because f_ϵ is continuous, and the third term can be bounded as $|\mu f_\epsilon - \mu f| \leqslant \mu(U_\epsilon)|f|_\infty < \epsilon|f|_\infty$. We have shown that $\limsup_{n\to\infty} |\mu_n f - \mu f| \leqslant 2\epsilon|f|_\infty$. Since ϵ was arbitrary, the result follows. □

2.2 Definitions and first results

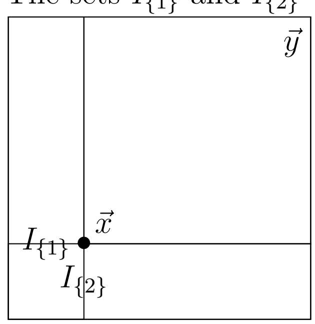
We will define discrepancy for measures on the d-dimensional half-open box $[\vec{0}, \vec{\infty}) = [\vec{0}, \infty)^d \subset \mathbf{R}^d$. For vectors $\vec{x}, \vec{y} \in [\vec{0}, \vec{\infty})$, we say $\vec{x} < \vec{y}$ if $x_i < y_i \, \forall i$, and in that case write $[\vec{x}, \vec{y})$ for the half-open box $[x_1, y_1) \times \cdots \times [x_d, y_d)$.

Definition 2.2.1. Let μ, ν be probability measures on $[\vec{0}, \vec{\infty})$. The *discrepancy between* μ *and* ν is $\mathrm{D}(\mu, \nu) = \sup_{\vec{x} < \vec{y}} |\mu[\vec{x}, \vec{y}) - \nu[\vec{x}, \vec{y})|$, where $\vec{x}$ and $\vec{y}$ range over $[\vec{0}, \vec{\infty})$. The *star discrepancy between* μ *and* ν is $\mathrm{D}^\star(\mu, \nu) = \sup_{\vec{0} \leqslant \vec{y}} \left| \mu[\vec{0}, \vec{y}) - \nu[\vec{0}, \vec{y}) \right|$, where $\vec{y}$ ranges over $[\vec{0}, \vec{\infty})$.

Lemma 2.2.2. $\mathrm{D}^\star(\mu, \nu) \leqslant \mathrm{D}(\mu, \nu) \leqslant 2^d \, \mathrm{D}^\star(\mu, \nu)$.

Proof. The first inequality holds because the supremum defining discrepancy is taken over a larger set than that defining star discrepancy. To prove the second inequality, let $\vec{x} < \vec{y}$ be in $[\vec{0}, \vec{\infty})$. For $S \subset \{1, \ldots, d\}$, let $I_S = \{\vec{t} \in [\vec{0}, \vec{y}) : t_i < x_i \, \forall \, i \in S\}$. Inclusion-exclusion tells us that $\mu[\vec{x}, \vec{y}) = \sum_{S \subset \{1, \ldots, d\}} (-1)^{\# S} \mu(I_S)$, and

Figure 2.1: The sets $I_{\{1\}}$ and $I_{\{2\}}$ when $d = 2$.



similarly for ν. Since each of the I_S are half-open boxes intersecting the origin, we know that $|\mu(I_S) - \nu(I_S)| \leqslant \mathrm{D}^\star(\mu, \nu)$. It follows that

$$|\mu[\vec{x}, \vec{y}) - \nu[\vec{x}, \vec{y})| \leqslant \sum_{S \subset \{1, \ldots, d\}} |\mu(I_S) - \nu(I_S)| \leqslant 2^d \, \mathrm{D}^\star(\mu, \nu).$$

For a discussion and related context, see [KN74, Ch. 2 Ex. 1.2]. $\square$

Since we are only interested in the asymptotics of discrepancy, we will sometimes gloss over the distinction between discrepancy and star discrepancy, using whichever type of discrepancy makes a proof easier to follow.

We are usually interested in comparing empirical measures and their conjectured asymptotic distribution. Let $\vec{\boldsymbol{x}} = (\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots)$ be a sequence in $[\vec{0}, \vec{\infty})$, and μ a probability measure on $[\vec{0}, \vec{\infty})$. For any $N \geqslant 1$, the empirical measure associated to the truncated sequence $\vec{\boldsymbol{x}}_{\leqslant N} = (\vec{x}_n)_{n \leqslant N}$ is $P_{\vec{\boldsymbol{x}},N} = \frac{1}{N} \sum_{n \leqslant N} \delta_{\vec{x}_n}$. Write $\mathrm{D}_N(\vec{\boldsymbol{x}}, \mu) = \mathrm{D}_N\left(P_{\vec{\boldsymbol{x}},N}, \mu\right)$, and similarly for star discrepancy. In this context,

$$\mathrm{D}_N^\star(\vec{\boldsymbol{x}}, \mu) = \sup_{\vec{y} \in [\vec{0}, \vec{\infty})} \left| \frac{\#\{n \leqslant N : \vec{x}_n \in [\vec{0}, \vec{y})\}}{N} - \int_{[\vec{0}, \vec{y})} \mathrm{d}\mu \right| .$$

When the measure μ is clear from the context, we will refer to $\mathrm{D}_N(\vec{\boldsymbol{x}}, \mu)$ (resp. $\mathrm{D}_N^\star(\vec{\boldsymbol{x}}, \mu)$) as the discrepancy (resp. star discrepancy) of the sequence $\boldsymbol{x}$.

If the measure μ is only defined on a Borel subset of $[\vec{0}, \vec{\infty})$, we tacitly extend it by zero to $\mathbf{R}^d$. It is also possible to define discrepancy for sequences lying in compact Lie groups. For example, if the sequence $\vec{\boldsymbol{x}}$ lies in a real torus T, choose an isomorphism $(\mathbf{R}/\mathbf{Z})^d \simeq T$, and using that isomorphism identify the torus with $[0,1)^d \subset [\vec{0}, \vec{\infty})$. This gives a definition of discrepancy for sequences in T. Two different isomorphisms $T \simeq (\mathbf{R}/\mathbf{Z})^d$ will give two different definitions of discrepancy, but asymptotically they will be bounded above and below by constant multiples of each other as long as the measure in question is the normalized Haar measure on the torus. In that case, we write $\mathrm{D}_N(\vec{\boldsymbol{x}})$ in place of $\mathrm{D}_N(\vec{\boldsymbol{x}}, \mu)$.

Now let G be a compact, connected Lie group, and consider a sequence lying in the space $G^\natural$ of conjugacy classes. Choose a maximal torus $T \subset G$, and recall that $G^\natural = T/W$, for W the Weyl group of T. There is a half-open box in $\mathfrak{t} = \mathrm{Lie}(T)$ which maps bijectively to T under the exponential map. Choose a "half-open" polyhedral set Q which is a fundamental domain for the action of W on T. Then $Q \subset \mathfrak{t}$ and, if we choose a basis for $\mathfrak{t}$ mapping to zero in T, we can identify $G^\natural$ with a subset of $[0,1)^d$, $d = \dim T$. This gives a definition of discrepancy for a sequence in $G^\natural$ with respect to the Haar measure. Of course this definition depends on the choice of T, Q, and the basis of $\mathfrak{t}$, but asymptotically these definitions are all the same. The paper [Ros13] has a different definition of discrepancy which only works for semisimple simply-connected groups, but also proves an Erdös–Turán inequality in that context. It is likely that a reasonable application of isotropic discrepancy would render these definitions equivalent, at least for asymptotic purposes, but as the two definitions coincide for SU(2), we do not explore this further.

Sometimes the sequence $\boldsymbol{x}$ will not be indexed by the natural numbers, but by the rational primes, or some other discrete subset of $\mathbf{R}^+$. In that case we will still use the notations $\mathrm{D}_N(\boldsymbol{x}, \mu)$, $\boldsymbol{x}_{\leqslant N} = (\vec{x}_1, \dots, \vec{x}_N)$, $\boldsymbol{x}_{\geqslant N} = (\vec{x}_N, \vec{x}_{N+1}, \dots)$, etc., keeping in mind that the set $\{\vec{x}_\alpha : \alpha \leqslant N\}$ is involved, and that in formulas $\frac{1}{N}$ is replaced by $\#\{\text{indices } \leqslant N\}^{-1}$.

Why half-open boxes? The choice of sets of the form $[\vec{x}, \vec{y})$ in the definition of discrepancy seems rather arbitrary, and it is. Discrepancy can easily be defined as a supremum over all open (or closed) balls—and those definitions generalize to arbitrary metric spaces. There are also more subtle definitions involving suprema over open or closed convex sets (isotropic discrepancy). See [KN74] for a discussion and comparison of these differing definitions. In this thesis, we restrict to half-open boxes because they are computationally tractable, fit well with Diophantine approximation on tori, and the theory is most well-developed for this definition.

2.3 Statistical heuristics

Let Ω be a probability space, and let $\{\theta_p\}$ be a collection of prime-indexed independent, identically distributed, random variables on Ω with continuous joint distribution μ. By this, we mean each $\theta_p \colon \Omega \to \mathbf{R}$ is measurable, and if P is the probability measure on Ω, then $\left(\prod_{p\in S}\theta_p\right)_* P = \prod_{p\in S}\mu$ for all sets S of primes. In the language of statistics, $\{\theta_p\}$ is a sequence of iid random variables with joint distribution μ. For the sake of concreteness, the reader may take $\mu = \frac{2}{\pi}\sin^2\theta\,\mathrm{d}\theta$, supported on $[0,\pi]$. Then the discrepancy (known as the Kolmogorov–Smirnov statistic in this context) is the random variable

$$D_N = \sup_{x\in[0,\pi]} \left| \frac{1}{\pi(N)} \sum_{p\leqslant N} 1_{[0,x]} \circ \theta_p - \int 1_{[0,x]}\,\mathrm{d}\mu \right|.$$

Kolmogorov and Smirnov proved that the inside of the absolute value, a function-valued random variable, converges to zero. The Glivenko–Cantelli theorem says that $D_N \to 0$ almost everywhere, and even better, the normalized discrepancy $\sqrt{\pi(N)}D_N$ approaches a limiting distribution (supremum of the Brownian bridge) which does not depend on μ. The rate of convergence of $\sqrt{\pi(N)}D_N$ to that distribution is quantified by the Dvoretzky–Kiefer–Wolfowitz inequality, which tells us that $\mathrm{P}\left(\sqrt{\pi(N)}D_N > t\right) \leqslant 2e^{-2t^2}$.

Now let $E_{/\mathbf{Q}}$ be a non-CM elliptic curve, and suppose the Satake parameters θ_p are being randomly drawn from the distribution ST. Then the above theorems suggest that as $N\to\infty$, we should have $\mathrm{D}_N(\boldsymbol{\theta},\mathrm{ST}) \sim \pi(N)^{-\frac{1}{2}}$, or, if we are not so lucky, at least $\mathrm{D}_N(\boldsymbol{\theta},\mathrm{ST}) \ll N^{-\frac{1}{2}+\epsilon}$. Ideally, the normalized discrepancy $\sqrt{\pi(N)}\,\mathrm{D}_N(\boldsymbol{\theta},\mathrm{ST})$ would also be equidistributed, but sadly, numerical experiments conducted by the author suggest this is not the case.

2.4 The Koksma–Hlawka inequality

In this section we summarize the results of the paper [Ö], generalizing them as needed for our context. Recall that a function f on $[\vec{0}, \vec{\infty}) \subset \mathbf{R}^d$ is said to be of *bounded variation* (in the measure-theoretic sense) if there is a finite Radon measure (i.e., finite and inner regular) ν such that $f(\vec{x}) - f(\vec{0}) = \nu[\vec{0}, \vec{x}]$. In such a case we write $\operatorname{Var}(f) = |\nu|$. If $f \in C^d(\mathbf{R}^d)$, then $\operatorname{Var}(f) = \int_{[\vec{0},\vec{\infty})} \left| \frac{\mathrm{d}^d f}{\mathrm{d}t_1 \ldots \mathrm{d}t_d} \right|$.

Theorem 2.4.1 (Koksma–Hlawka)**.** *Let μ be a probability measure on $[\vec{0}, \vec{\infty})$, f of bounded variation. For any sequence $\vec{\boldsymbol{x}} = (\vec{x}_1, \vec{x}_2, \ldots)$ in $[\vec{0}, \vec{\infty})$, we have*

$$\left| \frac{1}{N} \sum_{n \leqslant N} f(\vec{x}_n) - \int f \,\mathrm{d}\mu \right| \leqslant \operatorname{Var}(f)\, \mathrm{D}_N(\vec{\boldsymbol{x}}, \mu).$$

Proof. By assumption, there is a finite Radon (that is, inner regular) measure ν such that $f(\vec{y}) - f(\vec{0}) = \nu[\vec{0}, \vec{y}]$. What follows relies on the fact that $1_{[\vec{0},\vec{x}]}(\vec{y}) = 1_{[\vec{y},\vec{\infty})}(\vec{x})$.

$$\begin{aligned}
\frac{1}{N} \sum_{n \leqslant N} f(\vec{x}_n) - \int f \,\mathrm{d}\mu &= \frac{1}{N} \sum_{n \leqslant N} \Big(f(\vec{x}_n) - f(\vec{0}) \Big) - \int \Big(f(\vec{x}) - f(\vec{0}) \Big) \,\mathrm{d}\mu(\vec{x}) \\
&= \frac{1}{N} \sum_{n \leqslant N} \int 1_{[\vec{0},\vec{x}_n]}(\vec{y}) \,\mathrm{d}\nu(\vec{y}) - \int \int 1_{[\vec{0},\vec{x}]}(\vec{y}) \,\mathrm{d}\nu(\vec{y}) \,\mathrm{d}\mu(\vec{x}) \\
&= \int \left(\frac{1}{N} \sum_{n \leqslant N} 1_{[\vec{y},\vec{\infty})}(\vec{x}_n) - \int 1_{[\vec{y},\vec{\infty})} \,\mathrm{d}\mu \right) \mathrm{d}\nu(\vec{y}).
\end{aligned}$$

It follows that

$$\left| \frac{1}{N} \sum_{n \leqslant N} f(\vec{x}_n) - \int f \,\mathrm{d}\mu \right| \leqslant \sup_{\vec{y} \in [\vec{0},\vec{\infty})} \left| \frac{1}{N} \sum_{n \leqslant N} 1_{[\vec{y},\vec{\infty})}(\vec{x}_n) - \int 1_{[\vec{y},\vec{\infty})} \,\mathrm{d}\mu \right| \cdot |\nu|.$$

The supremum is bounded above by $\mathrm{D}_N(\vec{\boldsymbol{x}}, \mu)$, so the proof is complete. □

This theorem is proved in a somewhat restrictive setting, and there are more general versions of the theorem for less restrictive notions of bounded variation.

For example, f a function on $\mathbf{R}^+$ that is bounded variation in the traditional sense (for example, piecewise continuous) and μ a continuous probability measure, the inequality

$$\left|\frac{1}{N}\sum_{n\leqslant N} f(x_n) - \int f\,\mathrm{d}\mu\right| \leqslant \mathrm{Var}(f)\,\mathrm{D}^\star_N(\boldsymbol{x},\mu)$$

still holds [KN74, Ch. 2, Th. 5.1]. In particular, when μ is the Sato–Tate measure and f is piecewise continuous, we can apply this inequality. When $d > 1$, the non-measure-theoretic definition of variation is more general, but much more complicated, than the measure-theoretic one. See [KN74, 2§5] for a discussion of this.

2.5 Comparing and combining sequences

Throughout this section, λ is the Lebesgue measure on $\mathbf{R}$. Recall that for another measure μ on $\mathbf{R}$, the *Radon–Nikodym derivative* of μ with respect to λ (when it exists) is uniquely determined by $\mathrm{cdf}_\mu(x) = \int_{-\infty}^x \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t)\,\mathrm{d}t$. The Radon–Nikodym derivative exists whenever μ is absolutely continuous with respect to λ, i.e. $\mu(S) = 0$ whenever $\lambda(S) = 0$. The following result, a generalization of [KN74, Ch. 2 Th. 4.1], quantifies how much the discrepancy of a sequence changes when the elements of the sequence are changed by a small amount.

Lemma 2.5.1. *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be sequences in $[0,\infty)$. Suppose μ is an absolutely continuous probability measure on $[0,\infty)$ with bounded Radon–Nikodym derivative $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$. Let $\epsilon > 0$ be arbitrary. Then*

$$|\mathrm{D}^\star_N(\boldsymbol{x},\mu) - \mathrm{D}^\star_N(\boldsymbol{y},\mu)| \leqslant \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right\|_\infty \epsilon + \frac{\#\{n\leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N}.$$

Proof. Fix $\epsilon > 0$, and let $t \in [0,\infty)$ be arbitrary. For all $n \leqslant N$ such that $y_n < t$, either $x_n < t + \epsilon$ or $|x_n - y_n| \geqslant \epsilon$. It follows that

$$P_{\boldsymbol{y},N}[0,t) \leqslant P_{\boldsymbol{x},N}[0,t+\epsilon) + \frac{\#\{n \leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N}.$$

Moreover, we have $|P_{\boldsymbol{x},N}[0,t+\epsilon) - \mu[0,t+\epsilon)| \leqslant \mathrm{D}_N^\star(\boldsymbol{x},\mu)$. Putting these together, we get:

$$\begin{aligned} P_{\boldsymbol{y},N}[0,t) - \mu[0,t) &\leqslant P_{\boldsymbol{x},N}[0,t+\epsilon) - \mu[0,t) + \frac{\#\{n \leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N} \\ &\leqslant \mu[t,t+\epsilon) + \mathrm{D}_N^\star(\boldsymbol{x},\mu) + \frac{\#\{n \leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N} \\ &\leqslant \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right\|_\infty \epsilon + \mathrm{D}_N^\star(\boldsymbol{x},\mu) + \frac{\#\{n \leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N} \end{aligned}$$

This tells us that $\mathrm{D}_N^\star(\boldsymbol{y},\mu) \leqslant \left\|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right\|_\infty \epsilon + \mathrm{D}_N^\star(\boldsymbol{x},\mu) + \frac{\#\{n \leqslant N : |x_n - y_n| \geqslant \epsilon\}}{N}$. Reversing the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$, we obtain the desired result. $\square$

This next result shows that if a sequence is transformed by an isometry of $\mathbf{R}$, the discrepancy of the transformed sequence is roughly the same as the discrepancy of the original sequence.

Lemma 2.5.2. *Let σ be an isometry of $\mathbf{R}$, and $\boldsymbol{x}$ a sequence in $[0,\infty)$ such that $\sigma(\boldsymbol{x})$ is also in $[0,\infty)$. Let μ be an absolutely continuous measure on $[0,\infty)$ such that $\sigma_*\mu$ is supported on $[0,\infty)$. Then $|\mathrm{D}_N(\boldsymbol{x},\mu) - \mathrm{D}_N(\sigma_*\boldsymbol{x},\sigma_*\mu)| \leqslant \frac{2}{N}$.*

Proof. Every isometry of $\mathbf{R}$ is a composition of a translation and a reflection. The statement is clear if σ is a translation, as then the two discrepancies are equal. So, suppose $\sigma(t) = a - t$ for some $a > 0$. Since μ is absolutely continuous, $\mu\{t\} = 0$ for all $t \geqslant 0$. In particular, $\mu[s,t) = \mu(s,t]$. By definition, $P_{\boldsymbol{x},N}\{t\} \leqslant \frac{1}{N}$. For any interval $[s,t)$ in $[0,\infty)$, we know that $|P_{\boldsymbol{x},N}[s,t) - P_{\boldsymbol{x},N}(s,t]| \leqslant \frac{2}{N}$, hence

$$|P_{\boldsymbol{x},N}[s,t) - \mu[s,t) - (P_{\sigma_*\boldsymbol{x},N}[a-t,a-s) - \sigma_*\mu[a-t,a-s))| \leqslant \frac{2}{N}.$$

This proves the result. $\square$

A technique we will use throughout this thesis involves comparing the discrepancy of a sequence with the discrepancy of a pushforward sequence, with respect to the pushforward measure.

Lemma 2.5.3. *Let I, J be closed connected intervals and $f\colon I \twoheadrightarrow J$ a continuous monotonic map. If $\boldsymbol{x}$ is a sequence in I and μ is an absolutely continuous probability measure on I, then $|\mathrm{D}_N(\boldsymbol{x},\mu) - \mathrm{D}_N(f_*\boldsymbol{x}, f_*\mu)| \leqslant \frac{4}{N}$.*

Proof. Because f is continuous and monotonic, given $[u,v) \subset I$, there exists $[x,y) \subset J$ such that $f[u,v)$ differs from $[x,y)$ by at most two elements. Compute:

$$|P_{f_*\boldsymbol{x},N}[x,y) - f_*\mu[x,y) - (P_{\boldsymbol{x},N}[u,v) - \mu[u,v))| \leqslant \frac{4}{N},$$

the equality $f_*\mu[u,v) = \mu[x,y)$ following from the continuity of μ. Since $[u,v) \subset I$ was arbitrary, it follows that $\mathrm{D}_N(\boldsymbol{x},\mu) \leqslant \mathrm{D}_N(f_*\boldsymbol{x}, f_*\mu) + \frac{4}{N}$.

Similarly, for any $[x,y) \subset J$, there exists $[u,v) \subset I$ such that $f^{-1}[x,y)$ differs from $[u,v)$ by at most two elements. Compute:

$$|P_{\boldsymbol{x},N}[u,v) - \mu[u,v) - (P_{f_*\boldsymbol{x},N}[x,y) - f_*\mu[x,y))| \leqslant \frac{4}{N}.$$

Here, $[x,y)$ is arbitrary, so $\mathrm{D}_N(f_*\boldsymbol{x}, f_*\mu) \leqslant \mathrm{D}_N(\boldsymbol{x},\mu) + \frac{4}{N}$. The result follows. □

Now we show that the discrepancy behaves as expected when two sequences are interleaved.

Definition 2.5.4. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be sequences in $[0,\vec{\infty}) \subset \mathbf{R}^d$. We write $\boldsymbol{x} \wr \boldsymbol{y}$ for the interleaved sequence $(x_1, y_1, x_2, y_2, x_3, y_3, \dots)$.

Write $P_{\boldsymbol{x}\wr\boldsymbol{y},N} = \frac{1}{2}(P_{\boldsymbol{x},N} + P_{\boldsymbol{y},N})$ for the combined empirical measure of the interleaved sequence $\boldsymbol{x} \wr \boldsymbol{y}$.

Theorem 2.5.5. *Let I and J be disjoint open boxes in $[0, \vec{\infty})$, and let μ, ν be absolutely continuous probability measures on I and J, respectively. Let $\boldsymbol{x}$ be a sequence in I and $\boldsymbol{y}$ be a sequence in J. Then*

$$\max\{\mathrm{D}_N(\boldsymbol{x}, \mu), \mathrm{D}_N(\boldsymbol{y}, \nu)\} \leqslant \mathrm{D}_N(\boldsymbol{x} \wr \boldsymbol{y}, \mu + \nu) \leqslant \mathrm{D}_N(\boldsymbol{x}, \mu) + \mathrm{D}(\boldsymbol{y}, \nu)$$

Proof. Any half-open box in $[0, \vec{\infty})$ can be split by a coordinate hyperplane into two disjoint half-open boxes $[\vec{a}, \vec{b}) \sqcup [\vec{s}, \vec{t})$, each of which intersects at most one of I and J. We may assume that $[\vec{a}, \vec{b}) \cap J = \varnothing$ and $[\vec{s}, \vec{t}) \cap I = \varnothing$. Then

$$\begin{aligned}\left|P_{\boldsymbol{x}\wr\boldsymbol{y},N}([\vec{a}, \vec{b}) \sqcup [\vec{s}, \vec{t})) - (\mu + \nu)([\vec{a}, \vec{b}) \sqcup [\vec{s}, \vec{t}))\right| &\leqslant |P_{\boldsymbol{x},N}[\vec{a}, \vec{b}) - \mu[\vec{a}, \vec{b})| + |P_{\boldsymbol{y},N}[\vec{s}, \vec{t}) - \nu[\vec{s}, \vec{t})| \\ &\leqslant \mathrm{D}_N(\boldsymbol{x}, \mu) + \mathrm{D}_N(\boldsymbol{y}, \nu).\end{aligned}$$

This yields the second inequality in the statement of the theorem. To see the first, assume that the maximum discrepancy is $\mathrm{D}_N(\boldsymbol{x}, \mu)$, and let $[\vec{s}, \vec{t})$ be a half-open box such that $|P_{\boldsymbol{x},N}[\vec{s}, \vec{t}) - \mu[\vec{s}, \vec{t})|$ is within some arbitrary ϵ of $\mathrm{D}_N(\boldsymbol{x}, \mu)$. Use the coordinate hyperplane between I and J to "cut off" the part of $[\vec{s}, \vec{t})$ that does not intersect I. Replacing $[\vec{s}, \vec{t})$ with this smaller box, we may assume it does not intersect J. Under the assumption $[\vec{s}, \vec{t}) \cap J = \varnothing$, we have $\left|P_{\boldsymbol{x}\wr\boldsymbol{y},N}[\vec{s}, \vec{t}) - (\mu + \nu)[\vec{s}, \vec{t})\right| = |P_{\boldsymbol{x},N}[\vec{s}, \vec{t}) - \mu[\vec{s}, \vec{t})|$, which yields the result. □

2.6 Examples

Historically, one of the first interesting examples of an equidistributed sequence is the set of translates of an irrational number modulo one.

Theorem 2.6.1 (Weyl, Sierpiński, Bohl)**.** *Let $a \in \mathbf{R}$ be irrational. Then the sequence $\boldsymbol{x} = (a \mod 1, 2a \mod 1, 3a \mod 1, \dots)$ is equidistributed in $[0, 1)$.*

We will prove this result in Chapter 4. It is known, and we will prove, that sequences of this form have discrepancy which decays roughly like $N^{-\alpha}$, for some $\alpha \in \left(0, \frac{1}{2}\right)$ which controls the "goodness" of rational approximations of x. It is useful to have a sequence whose discrepancy decays faster. The best known rate of decay is achieved by the following example.

Definition 2.6.2. For $n \in \mathbf{N}$, write n in base 2 as $n = \sum a_i 2^i$ and put $v_n = \sum a_i 2^{-(i+1)}$. The *van der Corput sequence* is $\boldsymbol{v} = (v_1, v_2, v_3, \dots)$.

The van der Corput sequence has generalizations to other bases and higher dimensions, but we will not use them. The discrepancy of the van der Corput sequence has extremely fast convergence to zero.

Lemma 2.6.3 ([KN74, Ch. 2 Th. 3.5])**.** $\mathrm{D}_N(\boldsymbol{v}) \leqslant \frac{\log(N+1)}{N \log 2}$.

The van der Corput sequence is uniformly distributed (equidistributed with respect to the Lebesgue measure). We can use the results of the previous section to construct sequences equidistributed with respect to more general measures.

Theorem 2.6.4. *Let μ be an absolutely continuous probability measure on an interval I whose cdf is strictly increasing on I. Then there exists a sequence $\boldsymbol{x} = (x_1, x_2, \dots)$ in I such that $\mathrm{D}_N(\boldsymbol{x}, \mu) \ll \frac{\log(N)}{N}$.*

Proof. Since $(\mathrm{cdf}_\mu)_* \mu$ is the Lebesgue measure, Lemma 2.5.3 tells us that $\left|\mathrm{D}_N(\mathrm{cdf}_\mu^{-1}(\boldsymbol{v}), \mu) - \mathrm{D}_N(\boldsymbol{v})\right| \ll N^{-1}$. Since cdf_μ is strictly increasing, it is an order-preserving bijection between I and $[0, 1]$, which gives us the desired result with $\boldsymbol{x} = \mathrm{cdf}_\mu^{-1}(\boldsymbol{v})$. The implied constant in the result does not depend on μ. $\square$

Now that we can construct sequences with discrepancy decaying rapidly (with respect to a fixed measure μ), we use the sequences with rapid discrepancy decay to construct sequences whose discrepancy decays at any specified rate. The $N^{-\alpha}$ in the following theorem could actually be specified by any decreasing function of N which converges to zero, but doesn't decay faster than N^{-1}.

Theorem 2.6.5. *Let μ be an absolutely continuous probability measure, supported on I, whose cdf is strictly increasing on I. Fix $\alpha \in (0,1)$. Then there exists a sequence $\boldsymbol{x} = (x_1, x_2, \dots)$ such that $\mathrm{D}_N(\boldsymbol{x}, \mu) = \Theta(N^{-\alpha})$.*

The proof is similar in concept to the proof that a conditionally (but not absolutely!) convergent sequence may be rearranged to sum to any desired value. We start with a van der Corput sequence with rapidly decaying discrepancy. Our sequence begins by adding van der Corput elements until the discrepancy is smaller than $N^{-\alpha}$, then repeatedly adds the same element to the end of the sequence, pushing up the discrepancy until it is bigger than $N^{-\alpha}$. There are two main difficulties. First, we need to show that repeatedly adding the same element to the end of a sequence eventually forces the discrepancy to increase, and that when doing this, the discrepancy does not increase or decrease too rapidly.

Proof. Let $I = [a,b]$. If $\boldsymbol{x}_{\leqslant N}$ is a sequence of length N, let $\boldsymbol{x}_{\leqslant N} : a^M$ be the sequence $(x_1, \dots, x_N, a, \dots, a)$ (M copies of a). We begin by showing that the discrepancy of $\boldsymbol{x}_{\leqslant N} : a^M$ is eventually large relative to $N^{-\alpha}$. Recalling that $\mu\{a\} = 0$, we have:

$$\mathrm{D}(\boldsymbol{x}_{\leqslant N} : a^M, \mu) \geqslant \left| \frac{\#\{n \leqslant N+M : x_n = a\}}{N+M} - \mu\{a\} \right| \geqslant \frac{M}{N+M}.$$

So for fixed N, if we add enough a's to the end of $\boldsymbol{x}_{\leqslant N}$, the discrepancy

$\mathrm{D}(\boldsymbol{x}_{\leqslant N}; a^M, \mu)$ will be larger than $(N+M)^{-\alpha}$. On the other hand for $J = [s,t) \subset I$,

$$\begin{aligned}\left|P_{\boldsymbol{x}_{\leqslant N}:a^M}(J) - P_{\boldsymbol{x}_{\leqslant N}}(J)\right| &= \frac{\left|\#\{n \leqslant N : x_n \in J\} + M - \frac{M+N}{N}\#\{n \leqslant N : x_n \in J\}\right|}{M+N} \\ &= \frac{\left|M - \frac{M}{N}\#\{n \leqslant N : x_n \leqslant t\}\right|}{M+N} \\ &\leqslant \frac{M}{M+N},\end{aligned}$$

which implies that $\mathrm{D}\left(\boldsymbol{x}_{\leqslant N} : a^M, \mu\right) \leqslant \mathrm{D}\left(\boldsymbol{x}^N, \mu\right) + \frac{M}{M+N}$. This lets us control how rapidly the discrepancy can increase.

Let $\boldsymbol{v}$ be the μ-equidistributed van der Corput sequence of Theorem 2.6.4, possibly transformed linearly to lie in $[a,b]$. We know that $\mathrm{D}(\boldsymbol{v}^N, \mu) \ll \frac{\log N}{N}$, which converges to zero faster than $N^{-\alpha}$.

We construct the sequence $\boldsymbol{x}$ via the following recipe. Start with $(x_1 = v_1, x_2 = v_2, \dots)$ until, for some N_1, $\mathrm{D}_{N_1}(\boldsymbol{x}, \mu) < N_1^{-\alpha}$. Then set $x_{N_1+1} = a$, $x_{N_1+2} = a$, $\dots$, until $\mathrm{D}_{N_1+M_1}(\boldsymbol{x}, \mu) > (N_1 + M_1)^{-\alpha}$. Then set $x_{N_1+M_1+1} = v_{N_1+1}$, $x_{N_1+M_1+2} = v_{N_1+2}$, $\dots$, until once again $\mathrm{D}_{N_1+M_1+N_2}(\boldsymbol{x}, \mu) < (N_1 + M_1 + N_2)^{-\alpha}$. Repeat indefinitely. We will show first, that the two steps are possible, and that nowhere does $\mathrm{D}_N(\boldsymbol{x}, \mu)$ differ by too much from $N^{-\alpha}$.

Note that $\frac{M+1}{N+M+1} - \frac{M}{N+M} \leqslant N^{-1}$. This tells us that when we are adding a's at the end of $\boldsymbol{x}_{\leqslant N}$, the discrepancy of $\boldsymbol{x}_{\leqslant N} :^M a$ is eventually increasing, and can increase by at most N^{-1} at each step. So if $\mathrm{D}(\boldsymbol{x}_{\leqslant N}, \mu) < N^{-\alpha}$, we can ensure that $\mathrm{D}(\boldsymbol{x}_{\leqslant N} : a^M, \mu)$ is at most N^{-1} greater than $N^{-\alpha}$. Moreover, we know that $\mathrm{D}(\boldsymbol{x}_{\leqslant N} : a, \mu)$ is at most $\frac{2}{N+1}$ away from $\mathrm{D}(\boldsymbol{x}_{\leqslant N}, \mu)$. So when adding van der Corput elements to the end of the sequence, its discrepancy cannot decay any faster than by $\frac{2}{N+1}$ per a added. This yields

$$\left|\mathrm{D}_N(\boldsymbol{x}, \mu) - N^{-\alpha}\right| \ll N^{-1},$$

which is even stronger than we need. $\square$

CHAPTER 3

DIRICHLET SERIES WITH EULER PRODUCT

3.1 Definitions and motivation

We start by considering a very general class of Dirichlet series: those that admit a product formula with degree 1 factors. The motivating example was suggested to the author by Ramakrishna. Let $E_{/\mathbf{Q}}$ be an elliptic curve and let

$$L_{\mathrm{sgn}}(E,s) = \prod_p \frac{1}{1-\mathrm{sgn}(a_p)p^{-s}}.$$

How much can we say about the behavior of $L_{\mathrm{sgn}}(E,s)$? For example, does it admit analytic continuation to $\Re = 1$? Yes, by the results of [Ser89, A.2]. In fact, we will see later that the Akiyama–Tanigawa conjecture implies the existence of a non-vanishing analytic continuation of $L_{\mathrm{sgn}}(E,s)$ to $\Re > \frac{1}{2}$. Can the rank of E be found from $L_{\mathrm{sgn}}(E,s)$? Theoretically yes, by the following result, which the author learned from Harris.

Theorem 3.1.1. *If E_1 and E_2 are non-CM elliptic curves over $\mathbf{Q}$ with $\operatorname{sgn} a_p(E_1) = \operatorname{sgn} a_p(E_2)$ for all p, then E_1 and E_2 are isogenous.*

Proof. Assume by way of contradiction that E_1 and E_2 are non-isogenous, non-CM elliptic curves over $\mathbf{Q}$ with $\operatorname{sgn} a_p(E_1) = \operatorname{sgn} a_p(E_2)$ for all p. By [Har09, 5.4], the pairs $(\theta_p(E_1), \theta_p(E_2))$ are equidistributed with respect to $\mathrm{ST}\times\mathrm{ST} = \frac{4}{\pi^2}\sin^2\theta_1 \sin^2\theta_2 \, \mathrm{d}\theta_1 \, \mathrm{d}\theta_2$ on $[0,\pi]\times[0,\pi]$.

Recall that if $f(\theta) = 1_{[0,\pi/2)}(\theta) - 1_{(\pi/2,\pi]}(\theta)$, then $\mathrm{sgn}(a_p) = f(\theta_p)$. Moreover, $g(\theta_1,\theta_2) = |f(\theta_1) - f(\theta_2)|$ is non-negative and continuous almost everywhere, and

$g(\theta_p(E_1), \theta_p(E_2)) = 0$ if and only if $\operatorname{sgn} a_p(E_1) = \operatorname{sgn} a_p(E_2)$. It is clear that $\int g \, \mathrm{dST} \times \mathrm{ST} > 0$. Harris' equidistribution result tells us that

$$\int g \, \mathrm{dST} \times \mathrm{ST} = \lim_{N \to \infty} \frac{1}{\pi(N)} \sum_{p \leqslant N} g(\theta_p(E_1), \theta_p(E_2)) = 0,$$

which is a contradiction. $\square$

It follows that if $L_{\mathrm{sgn}}(E_1, s) = L_{\mathrm{sgn}}(E_2, s)$ for all s in some right half-plane, then E_1 and E_2 are isogenous, so in particular they have the same rank. Can we recover the rank of E from the behavior of $L_{\mathrm{sgn}}(E, s)$ at $s = \frac{1}{2}$? For $k \geqslant 1$, let r_k be the order of vanishing of $L(\operatorname{sym}^k E, s)$ at $s = \frac{1}{2}$. The heuristics in [Sar07] suggest that if the Akiyama–Tanigawa (and other natural conjectures) hold, then $L_f(E, s)$ has a zero of order $\sum_{k \geqslant 1} \widehat{f}(\operatorname{sym}^k) \left(-2r_k + (-1)^{k+1}\right)$ at $s = \frac{1}{2}$. Here, $\widehat{f}(\operatorname{sym}^k) = \int_{\mathrm{SU}(2)^\natural} f(x) \operatorname{tr} \operatorname{sym}^k(x) \, \mathrm{d}x$ is the sym^k-Fourier coefficient of f. If $f = \operatorname{tr} \operatorname{sym}^1$ (this is called U_1 in Sarnak's letter), then $\widehat{f}(\operatorname{sym}^k)$ vanishes for $k \geqslant 2$, so the order of vanishing of $L_{U_1}(E, s)$ at $s = \frac{1}{2}$ is $1 - 2r_1$, which, if we assume the Birch and Swinnerton-Dyer conjecture, allows us to "read" the rank of E from the behavior of $L_{U_1}(E, s)$ at $s = \frac{1}{2}$. However, for $f = 1_{[0, \pi/2)} - 1_{(\pi/2, \pi]}$, we have

$$\widehat{f}(\operatorname{sym}^k) = \begin{cases} 0 & k \text{ is even} \\ \frac{2}{\pi} (-1)^{\frac{k-1}{2}} \left(\frac{1}{k} + \frac{1}{k+2}\right) & k \text{ is odd} \end{cases},$$

so it's not clear if the rank of E can be directly read from the behavior of $L_{\mathrm{sgn}}(E, s)$ at $s = \frac{1}{2}$.

Definition 3.1.2. Let $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ be a sequence of complex numbers indexed by the primes. The associated Dirichlet series is $L(\boldsymbol{x}, s) = \prod_p (1 - x_p p^{-s})^{-1}$.

If x_p is defined only for a subset of the primes, we tacitly set $x_p = 0$ (so the Euler factor is 1) at all primes for which x_p is not defined.

Lemma 3.1.3. *Let $\boldsymbol{x}$ be a sequence with $|\boldsymbol{x}|_\infty \leqslant 1$. Then $L(\boldsymbol{x}, s)$ defines a holomorphic function on the region $\Re > 1$. On that region, $\log L(\boldsymbol{x}, s) = \sum_{p^r} \frac{x_p^r}{rp^{rs}}$.*

Proof. Expanding the product for $L(\boldsymbol{x}, s)$ formally, we have $L(\boldsymbol{x}, s) = \sum_{n\geqslant 1} \frac{\prod_p x_p^{v_p(n)}}{n^s}$. An easy comparison with the Riemann zeta function tells us that this sum is holomorphic on $\Re > 1$. By [Apo76, Th. 11.7], the product formula holds in the same region. The formula for $\log L(\boldsymbol{x}, s)$ comes from [Apo76, 11.9 Ex. 2]. □

Abel summation is a commonly-used result that will allow us to turn questions on the analytic continuation and non-vanishing of $L(\boldsymbol{x}, s)$ into questions about the asymptotics of $\sum_{p\leqslant N} x_p$.

Lemma 3.1.4 (Abel summation)**.** *Let $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ be a sequence of complex numbers, f a smooth $\mathbf{C}$-valued function on $\mathbf{R}$. Then*

$$\sum_{p\leqslant N} f(p)x_p = f(N)\sum_{p\leqslant N} x_p - \int_2^N f'(t)\sum_{p\leqslant t} x_p \,\mathrm{d}t.$$

Proof. If $p_1, \dots, p_n$ is an enumeration of the primes $\leqslant N$, we have

$$\begin{aligned}\int_2^N f'(t)\sum_{p\leqslant t} x_p \,\mathrm{d}t &= \sum_{p\leqslant N} x_p \int_{p_n}^N f'(t)\,\mathrm{d}t + \sum_{i=1}^{n-1}\sum_{p\leqslant p_i} x_p \int_{p_i}^{p_{i+1}} f'(t)\,\mathrm{d}t \\ &= (f(N) - f(p_n))\sum_{p\leqslant N} x_p + \sum_{i=1}^{n-1}(f(p_{i+1}) - f(p_i))\sum_{p\leqslant p_i} x_p \\ &= f(N)\sum_{p\leqslant N} x_p - \sum_{p\leqslant N} f(p)x_p,\end{aligned}$$

as desired. □

Theorem 3.1.5. *Let $|\boldsymbol{x}|_\infty \leqslant 1$, and assume $|\sum_{p\leqslant N} x_p| \ll N^{\alpha+\epsilon}$ for some $\alpha \in [\frac{1}{2}, 1]$. Then the series for $\log L(\boldsymbol{x}, s)$ converges conditionally to a holomorphic function on $\Re > \alpha$.*

Proof. Formally split the sum for $\log L(\boldsymbol{x}, s)$ into two pieces:

$$\log L(\boldsymbol{x}, s) = \sum_p \frac{x_p}{p^s} + \sum_p \sum_{r \geqslant 2} \frac{x_p^r}{rp^{rs}}.$$

For each p, we have

$$\left| \sum_{r \geqslant 2} \frac{x_p^r}{rp^{rs}} \right| \leqslant \sum_{r \geqslant 2} p^{-r\Re s} = p^{-2\Re s} \frac{1}{1 - p^{-\Re s}}.$$

Elementary analysis gives $1 \leqslant \frac{1}{1-p^{-\Re s}} \leqslant 2 + 2\sqrt{2}$, so the second piece of $\log L(\boldsymbol{x}, s)$ converges absolutely on $\Re > \frac{1}{2}$. We could simply cite [Ten95, II.1 Th. 10] to finish the proof; instead we prove directly that $\sum \frac{x_p}{p^s}$ converges absolutely to a holomorphic function on the region $\Re > \alpha$.

By Lemma 3.1.4 (Abel summation) with $f(t) = t^{-s}$, we have

$$\begin{aligned} \sum_{p \leqslant N} \frac{x_p}{p^s} &= N^{-s} \sum_{p \leqslant N} x_p + s \int_2^N \sum_{p \leqslant t} x_p \frac{\mathrm{d}t}{t^{s+1}} \\ &\ll N^{-\Re s + \alpha + \epsilon} + |s| \int_2^N t^{\alpha+\epsilon} \frac{\mathrm{d}t}{t^{\Re s + 1}}. \end{aligned} \tag{3.1}$$

Since $\alpha - \Re s < 0$, the first term is converges to zero. Since $\Re s + 1 - \alpha > 1$ and ϵ is arbitrary, the integral converges absolutely, and the proof is complete. $\square$

The proof of Theorem 3.1.5 actually gives an absolutely convergent expression for $\log L(\boldsymbol{x}, s)$ on the region $\Re > \alpha$. Since the term $N^{-s} \sum_{p \leqslant N} x_p$ in (3.1) converges to zero, we get

$$\log L(\boldsymbol{x}, s) = s \int_2^\infty t^{-s-1} \left(\sum_{p \leqslant t} x_p \right) \mathrm{d}t + \sum_p \sum_{r \geqslant 2} \frac{x_p^r}{rp^{rs}}.$$

Let X be a topological space, $f \colon X \to \mathbf{C}$ a function with $|f|_\infty \leqslant 1$, and $\boldsymbol{x} = (x_2, x_3, \dots)$ a sequence in X. Write

$$L_f(\boldsymbol{x}, s) = \prod_p \frac{1}{1 - f(x_p) p^{-s}},$$

for the associated Dirichlet series. In the remainder, we will exclusively focus on Dirichlet series of this type.

3.2 Automorphic and motivic L-functions

Suppose G is a compact group, $G^\natural$ the space of conjugacy classes in G. If $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ is a sequence in $G^\natural$ and ρ is a finite-dimensional unitary representation of G, put

$$L(\rho(\boldsymbol{x}), s) = \prod_p \frac{1}{\det(1 - \rho(x_p)p^{-s})}.$$

Clearly $L((\rho_1 \oplus \rho_2)(\boldsymbol{x}), s) = L(\rho_1(\boldsymbol{x}), s)L(\rho_2(\boldsymbol{x}), s)$. Now, suppose G is a compact connected Lie group, let $T \subset G$ be a maximal torus, and recall that $T \twoheadrightarrow G^\natural$ [Bou05, IX.5 Prop. 5]. The representation $\rho|_T$ decomposes as $\bigoplus \chi^{\oplus m_\chi}$, where χ ranges over characters of T and the entire expression is W-invariant. We may regard the x_p as lying in T/W, so we have

$$L(\rho(\boldsymbol{x}), s) = \prod_\chi L(\chi(\boldsymbol{x}), s)^{m_\chi}.$$

If the trivial representation appears in $\rho|_T$, this product formula will include a copy (possibly several) of $\zeta(s)$. Since $\chi(x_p) \in S^1$, the above formula decomposes $L(\rho(\boldsymbol{x}), s)$ into a product of Dirichlet series of the type considered above. For $G = \mathrm{SU}(2)$, the trivial representation occurs in $\mathrm{sym}^k\big|_T$ if and only if k is even, and the resulting $\zeta(s)$ in the product decomposition of $L(\mathrm{sym}^k \boldsymbol{x}, s)$ may explain why some of the results in this thesis only apply to *odd* symmetric powers.

Theorem 3.2.1. *Let $G^\natural$ be the space of conjugacy classes in a compact group, $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ a sequence in $G^\natural$. If ρ is a nontrivial unitary representation of G and $\left|\sum_{p \leqslant N} \operatorname{tr} \rho(x_p)\right| \ll N^{\alpha+\epsilon}$ for $\alpha \in \left[\frac{1}{2}, 1\right]$, then $L(\rho(\boldsymbol{x}), s)$ admits a non-vanishing analytic continuation to $\Re > \alpha$.*

Proof. On $\Re > 1$, we have $\log L(\rho(\boldsymbol{x}), s) = \sum_{r \geqslant 1} \sum_p \frac{\operatorname{tr} \rho(x_p)^r}{rp^{rs}}$. Just as in the proof of Theorem 3.1.5, we can split the sum into two terms, $\sum_p \frac{\operatorname{tr} \rho(x_p)}{p^s}$

and $\sum_{r\geqslant 2}\sum_p \frac{\operatorname{tr}\rho(x_p)^r}{rp^{rs}}$. Analytic continuation and nonvanishing remain the same when we omit finitely many Euler factors, so we may ignore all primes for which $\frac{\dim(\rho)}{p^{-1/2}} \geqslant 1$. Then the second sum converges on $\Re > \frac{1}{2}$, and assuming $\left|\sum_{p\leqslant N} \operatorname{tr}\rho(x_p)\right| \ll N^{\alpha+\epsilon}$, an argument identical to the one used in the proof of Theorem 3.1.5, using Abel summation, shows that $\sum_p \frac{\operatorname{tr}\rho(x_p)}{p^s}$ converges to a holomorphic function on $\Re > \alpha$. This yields the desired non-vanishing analytic continuation. $\square$

3.3 Discrepancy and the Riemann hypothesis

Definition 3.3.1. We say the *Riemann hypothesis* for $L(\boldsymbol{x}, s)$ holds if the function $\log L(\boldsymbol{x}, s)$ admits analytic continuation to $\Re > \frac{1}{2}$.

Under reasonable analytic hypotheses, namely conditional convergence of the Dirichlet series for $\log L(\boldsymbol{x}, s)$ on $\Re > \frac{1}{2}$, the result [Ten95, II.1 Th. 10] gives an estimate $|\sum_{p\leqslant N} x_p| \ll N^{\frac{1}{2}+\epsilon}$.

Theorem 3.3.2. *Let (X, μ) be a probability space in which discrepancy and Koksma–Hlawka make sense (i.e., Theorem 2.4.1 applies), and let $\boldsymbol{x} = (x_2, x_3, x_5, \dots)$ be a sequence in X with $\mathrm{D}_N(\boldsymbol{x}, \mu) \ll N^{-\frac{1}{2}+\epsilon}$. For any function f on X of bounded variation with $\int f \, \mathrm{d}\mu = 0$, $L_f(\boldsymbol{x}, s)$ satisfies the Riemann hypothesis.*

Proof. By the Koksma–Hlawka inequality (Theorem 2.4.1), the bound on discrepancy yields the estimate $\left|\sum_{p\leqslant N} f(x_p)\right| \ll N^{\frac{1}{2}+\epsilon}$. By Theorem 3.1.5, the Riemann hypothesis holds for $L_f(\boldsymbol{x}, s)$. $\square$

The same proof shows that if $\mathrm{D}_N(\boldsymbol{x}, \mu) \ll N^{-\alpha+\epsilon}$, then $\log L_f(\boldsymbol{x}, s)$ conditionally converges to a holomorphic function on $\Re > 1 - \alpha$. This theorem applied to the function $L_{\mathrm{sgn}}(E, s)$ shows that the Akiyama–Tanigawa conjecture implies the Riemann hypothesis for $L_{\mathrm{sgn}}(E, s)$. The author is unaware of any results, conditional or otherwise, that suggest $L_{\mathrm{sgn}}(E, s)$ has analytic continuation past $\Re = \frac{1}{2}$ or has any kind of functional equation. Also, if $\int f \, \mathrm{d}\mu \neq 0$, the function $L_f(\boldsymbol{x}, s)$ will have a singularity at $s = 1$, but the author is not aware of a way to continue the function $L_f(\boldsymbol{x}, s)$ past $\Re = 1$, even if $\mathrm{D}_N(\boldsymbol{x}, \mu)$ decays rapidly.

Let $F = \mathbf{F}_q(t)$ be a function field, $E_{/F}$ a generic elliptic curve. There is, for every prime $\mathfrak{p}$ of F, a Satake parameter $\theta_{\mathfrak{p}} \in [0, \pi]$, defined in the usual way. It is known [Kat88, Ch. 3] that

$$\left| \sum_{\mathrm{N}(\mathfrak{p}) \leqslant x} \operatorname{tr} \operatorname{sym}^k \begin{pmatrix} e^{i\theta_{\mathfrak{p}}} & \\ & e^{-i\theta_{\mathfrak{p}}} \end{pmatrix} \right| \ll k\sqrt{x}. \tag{3.2}$$

Briefly, let G be a compact Lie group, ρ an irreducible unitary representation of G. For $f \in L^1(G)$, the Fourier coefficient of f at ρ is $\widehat{f}(\rho) = \int f(x)\overline{\operatorname{tr} \rho(x)} \, \mathrm{d}x$. If $f \in L^1(G^\natural)$, then $f = \sum_\rho \widehat{f}(\rho) \operatorname{tr} \rho$. When $G = \mathrm{SU}(2)$, the nontrivial irreducible unitary representations are sym^k for $k \geqslant 1$.

Equation (3.2) tells us that for any $f \in C(\mathrm{SU}(2)^\natural)$ with $\sum_{k \geqslant 1} |\widehat{f}(\operatorname{sym}^k)| < \infty$ and $\widehat{f}(\operatorname{sym}^0) = 0$, the strange Dirichlet series $L_f(\boldsymbol{\theta}, s)$ satisfies the Riemann hypothesis.

The best estimate on discrepancy is found in [Nie91], where it is shown that $D_x \ll N^{-\frac{1}{4}}$ by applying a generalization of the Koksma–Hlawka inequality to $\mathrm{SU}(2)^\natural$. Namely, for any odd r, we have

$$\mathrm{D}_x(\boldsymbol{\theta}, \mathrm{ST}) \ll \frac{1}{r} + \sum_{k=1}^{2r-1} \frac{1}{k} \left| \frac{1}{\pi_F(x)} \sum_{\mathrm{N}(\mathfrak{p}) \leqslant x} \operatorname{tr} \operatorname{sym}^k \begin{pmatrix} e^{i\theta_{\mathfrak{p}}} & \\ & e^{-i\theta_{\mathfrak{p}}} \end{pmatrix} \right|.$$

Using the estimate (3.2) on character sums, Niederreiter is able to derive $\mathrm{D}_x \ll x^{-\frac{1}{4}}$. This fits well with the results of [BK15, RT16] in the number field case, both of which derive estimates of the form $\mathrm{D}_N \ll N^{-\frac{1}{4}+\epsilon}$, assuming both the generalized Riemann hypothesis and the functional equation for all symmetric-power L-functions associated to the (non-CM) elliptic curve in question.

CHAPTER 4

IRRATIONALITY EXPONENTS AND CM ABELIAN VARIETIES

4.1 Definitions and first results

We follow the notation of [Lau09]. Fix a dimension $d \geqslant 1$, and let $\vec{x} = (x_1, \ldots, x_d) \in \mathbf{R}^d$ be such that the x_i are linearly independent over $\mathbf{Q}$. If $d = 1$, the *irrationality exponent* of $x \in \mathbf{R}$ is the supremum of the set of $w \in \mathbf{R}^+$ such that there infinitely many rational numbers $\frac{p}{q}$ with $\left|x - \frac{p}{q}\right| \leqslant q^{-w}$. If x is rational, then it has irrationality exponent 1. If x is an algebraic irrational, then Roth's theorem says its irrationality exponent is 2. Liouiville constructed transcendental numbers with arbitrarily large irrationality exponent. Only a measure-zero set of reals, for example the Louiville number $\sum_{r\geqslant 1} 10^{-r!}$, have infinite irrationality exponent. In the results below, we will only consider reals with finite irrationality exponent. When $d \geqslant 1$, there are a d natural measures of irrationality, but we will use only two of them.

For the remainder of this thesis, $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbf{R}^d$.

Definition 4.1.1. Let $\omega_0(\vec{x})$ (resp. $\omega_{d-1}(\vec{x})$) be the supremum of the set of real numbers w for which there exist infinitely many $(n, \vec{m}) \in \mathbf{Z} \times \mathbf{Z}^d$ such that

$$|n\vec{x} - \vec{m}|_\infty \leqslant |(n, \vec{m})|_\infty^{-w} \qquad \text{(resp.}$$
$$|n + \langle \vec{m}, \vec{x} \rangle| \leqslant |(n, \vec{m})|_\infty^{-w}).$$

It is easy to see that both $\omega_0(\vec{x})$ and $\omega_{d-1}(\vec{x})$ are nonnegative. Even better, by [Lau09, Th. 2 Cor], $\omega_0(\vec{x}) \geqslant \frac{1}{d}$ and $\omega_{d-1}(\vec{x}) \geqslant d$. These two quantities are related by Khintchine's transference principle [Lau09, Th. 2], namely

$$\frac{\omega_{d-1}(\vec{x})}{(d-1)\omega_{d-1}(\vec{x})+d} \leqslant \omega_0(\vec{x}) \leqslant \frac{\omega_{d-1}(\vec{x})-d+1}{d}.$$

Moreover, the second of these inequalities is sharp in a very strong sense.

Theorem 4.1.2 ([Jar36]). *Let $w \geqslant 1/d$. Then there exists $\vec{x} \in \mathbf{R}^d$ such that $\omega_0(\vec{x}) = w$ and $\omega_{d-1}(\vec{x}) = dw + d - 1$.*

We can relate the traditional irrationality exponent and the invariant ω_0 in the special case $d = 1$.

Theorem 4.1.3. *If $d = 1$, then $\omega_0(x) = \mu - 1$, where μ is the traditional irrationality exponent of x.*

Proof. Both μ and ω_0 are invariant under translation by $\mathbf{Z}$, so without loss of generality we may assume $x \in [0, 1)$.

First we show that $\omega_0(x) \geqslant \mu - 1$. Suppose there exist infinitely many p/q with $\left|x - \frac{p}{q}\right| \leqslant q^{-w}$. Since $x < 1$ we may assume that for infinitely many of the p/q, $p < q$. Then $|qx - p| \leqslant q^{-(w-1)} = \max(p, q)^{-w}$, which tells us that $\omega_0(x) \geqslant \mu - 1$.

Now, we show that $\mu \geqslant \omega_0(x) + 1$. Suppose there exist infinitely many (n, m) with $|nx - m| \leqslant \max(|n|, |m|)^{-w}$. By the reverse triangle inequality, $||nx| - |m|| \leqslant \max(|n|, |m|)^{-w}$, and since $x < 1$, for n sufficiently large this implies $|n| \geqslant |m|$. It follows that for infinitely many (n, m), we have $\left|x - \frac{m}{n}\right| \leqslant n^{-(w+1)}$, which implies $\mu \geqslant \omega_0(x) + 1$. □

Here is a statement of Roth's theorem in the current context.

Theorem 4.1.4 (Roth). *Let $x \in (\overline{\mathbf{Q}} \cap \mathbf{R}) \smallsetminus \mathbf{Q}$. Then $\omega_0(x) = 1$.*

Proof. This follows directly from [Rot55] and Theorem 4.1.3. $\square$

Given $\vec{x} \in \mathbf{R}^d$, write $\mathrm{d}(\vec{x}, \mathbf{Z}^d) = \min_{\vec{m} \in \mathbf{Z}^d} |\vec{x} - \vec{m}|_\infty$. Note that $\mathrm{d}(\vec{x}, \mathbf{Z}^d) = 0$ if and only if $\vec{x} \in \mathbf{Z}^d$. Moreover, $\mathrm{d}(-, \mathbf{Z}^d)$ is well-defined for elements of $(\mathbf{R}/\mathbf{Z})^d$.

Lemma 4.1.5. *Let $\vec{x} \in \mathbf{R}^d$ with $|\vec{x}|_\infty < 1$ and $\omega_0(\vec{x})$ (resp. $\omega_{d-1}(\vec{x})$) finite. Then*

$$\frac{1}{\mathrm{d}(n\vec{x}, \mathbf{Z}^d)} \ll |n|^{\omega_0(\vec{x})+\epsilon} \qquad \text{for } n \in \mathbf{Z} \text{ (resp.}$$
$$\frac{1}{\mathrm{d}\left(\langle \vec{m}, \vec{x}\rangle, \mathbf{Z}\right)} \ll |\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\epsilon} \qquad \text{for } \vec{m} \in \mathbf{Z}^d\text{).}$$

Proof. Let $\epsilon > 0$. Then there are only finitely many $n \in \mathbf{Z}$ (resp. $\vec{m} \in \mathbf{Z}^d$) such that the inequalities in Definition 4.1.1 hold with $w = \omega_0(x) + \epsilon$ (resp. $w = \omega_{d-1}(\vec{x}) + \epsilon$). In other words, there exist constants $C_0, C_{d-1} > 0$, depending on $\vec{x}$ and ϵ, such that

$$|n\vec{x} - \vec{m}|_\infty \geqslant C_0 |(n, \vec{m})|_\infty^{-\omega_0(\vec{x})-\epsilon},$$
$$|n + \langle \vec{m}, \vec{x}\rangle| \geqslant C_{d-1} |(n, \vec{m})|_\infty^{-\omega_{d-1}(\vec{x})-\epsilon}$$

for all $(n, \vec{m}) \neq 0$ in $\mathbf{Z} \times \mathbf{Z}^d$.

Start with the first inequality. Fix n, and let $\vec{m}$ be a lattice point achieving the minimum $|n\vec{x} - \vec{m}|_\infty$; then $\mathrm{d}(n\vec{x}, \mathbf{Z}^d) \geqslant C_0 |(n, \vec{m})|_\infty^{-\omega_0(\vec{x})-\epsilon}$. Since $|n\vec{x} - \vec{m}|_\infty < 1$, the reverse triangle inequality gives $\left| |n| - \frac{|\vec{m}|_\infty}{|\vec{x}|_\infty} \right| \leqslant \frac{1}{|\vec{x}|_\infty}$. So $|n|$ and $|\vec{m}|$ are bounded above and below by scalar multiples of each other, which tells us that $\mathrm{d}(n\vec{x}, \mathbf{Z}^d) \geqslant C_0' |n|^{-\omega_0(\vec{x})-\epsilon}$ for C_0' depending on $\vec{x}$. It follows that $\frac{1}{\mathrm{d}(n\vec{x}, \mathbf{Z}^d)} \ll |n|^{\omega_0(\vec{x})+\epsilon}$, the implied constant depending on x and ϵ.

Now we consider the second inequality. Note that $\mathrm{d}(\langle \vec{m}, \vec{x}\rangle, \mathbf{Z}) = |n + \langle \vec{m}, \vec{x}\rangle|$ for some n with $|n| \leqslant |\vec{m}|_2 \cdot |\vec{x}|_2 + 1$. Thus $|(n, \vec{m})|_\infty \ll |\vec{m}|_2 \ll |\vec{m}|_\infty$ with the implied constants depending on d and x, because any two norms on a finite-dimensional

Banach space are equivalent. This gives us $\mathrm{d}(\langle\vec{m},\vec{x}\rangle,\mathbf{Z}) \geqslant C'_{d-1}|\vec{m}|_\infty^{-\omega_{d-1}(\vec{x})-\epsilon}$, for some constant C'_{d-1}. This implies

$$\frac{1}{d(\langle\vec{m},\vec{x}\rangle,\mathbf{Z})} \ll |\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\epsilon},$$

the implied constant depending on $\vec{x}$ and ϵ. □

4.2 Irrationality exponents and discrepancy

Let $\vec{x} = (x_1,\dots,x_d) \in \mathbf{R}^d$. The sequence $(\vec{x} \mod \mathbf{Z}^d, 2\vec{x} \mod \mathbf{Z}^d,\dots)$ will be equidistributed in a subtorus of $(\mathbf{R}/\mathbf{Z})^d$ whose rank is equal to the dimension of the $\mathbf{Q}$-vector space spanned by $\{x_1,\dots,x_d\}$. We are interested in the case where this sequence is equidistributed in the whole torus $(\mathbf{R}/\mathbf{Z})^d$, so assume $x_1,\dots,x_d$ are linearly independent over $\mathbf{Q}$ (this condition also makes sense for elements of $(\mathbf{R}/\mathbf{Z})^d$). For $\vec{x} \in (\mathbf{R}/\mathbf{Z})^d$, we wish to control the discrepancy of the sequence $(\vec{x}, 2\vec{x}, 3\vec{x},\dots)$ with respect to the Haar measure of $(\mathbf{R}/\mathbf{Z})^d$.

Theorem 4.2.1 (Erdös–Turán–Koksma. [DT97, Th. 1.21])**.** *Let $\boldsymbol{\vec{x}}$ be a sequence in $(\mathbf{R}/\mathbf{Z})^d$ and h an arbitrary integer. Then*

$$\mathrm{D}_N(\boldsymbol{\vec{x}}) \ll \frac{1}{h} + \sum_{0\leqslant|\vec{m}|_\infty\leqslant h} \frac{1}{r(\vec{m})}\left|\frac{1}{N}\sum_{n\leqslant N} e^{2\pi i\langle\vec{m},\vec{x}_n\rangle}\right|,$$

where the first sum ranges over $\vec{m} \in \mathbf{Z}^d$, $r(\vec{m}) = \prod \max\{1,|m_i|\}$, and the implied constant depends only on d.

Lemma 4.2.2. *Let $x \in \mathbf{R}$. Then $\left|\sum_{n\leqslant N} e^{2\pi inx}\right| \leqslant \frac{2}{\mathrm{d}(x,\mathbf{Z})}$.*

Proof. We begin with an easy bound:

$$\left|\sum_{n\leqslant N} e^{2\pi inx}\right| = \frac{|e^{2\pi i(N+1)x} - e^{2\pi ix}|}{|e^{2\pi ix}-1|} \leqslant \frac{2}{|e^{2\pi ix}-1|}.$$

Since $|e^{2\pi ix} - 1| = \sqrt{2 - 2\cos(2\pi x)}$ and $\cos(2\theta) = 1 - 2\sin^2\theta$, we obtain

$$\left|\sum_{n \leqslant N} e^{2\pi inx}\right| \leqslant \frac{1}{|\sin(\pi x)|}.$$

It is easy to check that $|\sin(\pi x)| \geqslant \mathrm{d}(x, \mathbf{Z})$, whence the result. $\square$

Corollary 4.2.3. *Let $\vec{x} \in (\mathbf{R}/\mathbf{Z})^d$ with $(x_1, \ldots, x_d)$ linearly independent over $\mathbf{Q}$. Then for $\vec{\boldsymbol{x}} = (\vec{x}, 2\vec{x}, 3\vec{x}, \ldots)$, we have*

$$\mathrm{D}_N(\vec{\boldsymbol{x}}) \ll \frac{1}{h} + \frac{1}{N} \sum_{0 < |\vec{m}|_\infty \leqslant h} \frac{2}{r(\vec{m})\mathrm{d}(\langle \vec{m}, \vec{x} \rangle, \mathbf{Z})}$$

for any integer h, with the implied constant depending only on d.

Proof. Apply the Erdös–Turán–Koksma inequality (Theorem 4.2.1), and bound the exponential sums using Lemma 4.2.2. $\square$

We combine the above results to estimate an upper bound on the discrepancy of the sequence $\vec{\boldsymbol{x}}$.

Theorem 4.2.4. *Let $\vec{\boldsymbol{x}} = (\vec{x}, 2\vec{x}, 3\vec{x}, \ldots)$ in $(\mathbf{R}/\mathbf{Z})^d$. Then*

$$\mathrm{D}_N(\vec{\boldsymbol{x}}) \ll N^{-\frac{1}{\omega_{d-1}(\vec{x})+1}+\epsilon}.$$

Proof. Fix $\epsilon > 0$ smaller than $\frac{1}{\omega_{d-1}(\vec{x})-1}$, and choose $\delta > 0$ such that $\frac{1}{\omega_{d-1}(\vec{x})+1+\delta} = \frac{1}{\omega_{d-1}(\vec{x})+1} - \epsilon$. By Corollary 4.2.3, we know that

$$\mathrm{D}_N(\vec{\boldsymbol{x}}) \ll \frac{1}{h} + \frac{1}{N} \sum_{0 < |\vec{m}|_\infty \leqslant h} \frac{1}{r(\vec{m})\mathrm{d}(\langle \vec{m}, \vec{x} \rangle, \mathbf{Z})},$$

and by Lemma 4.1.5, we know that $\mathrm{d}(\langle \vec{m}, \vec{x} \rangle, \mathbf{Z})^{-1} \ll |\vec{m}|^{\omega_{d-1}(\vec{x})+\delta}$. It follows that

$$\mathrm{D}_N(\vec{\boldsymbol{x}}) \ll \frac{1}{h} + \frac{1}{N} \sum_{0 < |\vec{m}|_\infty \leqslant h} \frac{|\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\delta}}{r(\vec{m})}.$$

All that remains is to bound the sum. Clearly

$$\sum_{0<|\vec{m}|_\infty\leqslant h} \frac{|\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\delta}}{r(m)} \ll \int_1^h \int_1^h \cdots \int_1^h \frac{\max(|t_1|,\dots,|t_d|)^{\omega_{d-1}(\vec{x})+\delta}}{t_1\dots t_d}\,\mathrm{d}t_1\dots\mathrm{d}t_d.$$

For each permutation σ of $\{1,\dots,d\}$, call I_σ the set of all $(t_1,\dots,t_d)$ in $[1,\infty)^d$ with $t_{\sigma(1)}<\cdots<t_{\sigma(d)}$. Then $[1,\infty)^d=\bigcup_{\sigma\in S_d} I_\sigma$, and each integral over I_σ is easy to bound. For example, the integral over I_1 is

$$\begin{aligned}\int_1^h \int_1^{t_d} \cdots \int_1^{t_2} \frac{t_d^{\omega_{d-1}(\vec{x})+\delta}}{t_1\dots t_d}\,\mathrm{d}t_1\dots\mathrm{d}t_d &\ll \int_1^h t^{\omega_{d-1}(\vec{x})+\delta-1}\,\mathrm{d}t \prod_{j=1}^{d-1}\int_1^h \frac{\mathrm{d}t}{t}\\ &\ll (\log h)^{d-1} h^{\omega_{d-1}(\vec{x})+\delta}.\end{aligned}$$

It follows that $\mathrm{D}_N(\vec{\boldsymbol{x}}) \ll \frac{1}{h}+\frac{1}{N}(\log h)^{d-1}h^{\omega_{d-1}(\vec{x})+\delta}$. Setting $h\approx N^{\frac{1}{1+\omega_{d-1}(\vec{x})+\delta}}$, we see that $D_N(\vec{\boldsymbol{x}})\ll N^{-\frac{1}{\omega_{d-1}(\vec{x})+1+\delta}} = N^{-\frac{1}{\omega_{d-1}(\vec{x})+1}+\epsilon}$.

For a slightly different proof of a similar result, given as a sequence of exercises, see [KN74, Ch. 2, Ex. 3.15, 16, 17]. Also, this estimate is quite coarse, but a better one would only have a smaller leading coefficient, which no doubt would be useful for computational purposes, but does not strengthen any of the results in this thesis. □

Theorem 4.2.5. *Let $\vec{x}\in\mathbf{R}^d$ be such that $x_1,\dots,x_d$ are linearly independent over $\mathbf{Q}$, and let $\vec{\boldsymbol{x}}=(\vec{x},2\vec{x},3\vec{x},\dots)$ in $(\mathbf{R}/\mathbf{Z})^d$. Then $\mathrm{D}_N(\vec{\boldsymbol{x}})=\Omega\left(N^{-\frac{d}{\omega_0(\vec{x})}-\epsilon}\right)$.*

Proof. We follow the proof of [KN74, Ch. 2, Th. 3.3], modifying it as needed for our context. Given $\epsilon>0$, there exists $\delta>0$ such that $\frac{d}{\omega_0(\vec{x})-\delta}=\frac{d}{\omega_0(\vec{x})}+\epsilon$.

By the definition of $\omega_0(\vec{x})$, there exist infinitely many $(q,\vec{m})$ with $q>0$ such that $|q\vec{x}-\vec{m}|_\infty\leqslant|(q,\vec{m})|_\infty^{-\omega_0(\vec{x})+\delta/2}$. Since $|(q,\vec{m})|_\infty\geqslant q$, we derive the seemingly stronger statement that for infinitely many q, there exists $\vec{m}\in\mathbf{Z}^d$ such that $|q\vec{x}-\vec{m}|_\infty\leqslant q^{-\omega_0(\vec{x})+\delta/2}$ or, equivalently, $|\vec{x}-q^{-1}\vec{m}|\leqslant q^{-1-\omega_0(\vec{x})+\delta/2}$. Fix one such q, and let $N = \lfloor q^{\omega_0(\vec{x})-\delta} \rfloor$. For each $n \leqslant N$, we have

$$\left| n\vec{x} - nq^{-1}\vec{m} \right|_\infty \leqslant n \left| \vec{x} - q^{-1}\vec{m} \right|_\infty \leqslant nq^{-1-\omega_0(\vec{x})+\delta/2} \leqslant q^{-1-\delta/2}.$$

Thus, for each $n \leqslant N$, $n\vec{x}$ is within $q^{-1-\delta/2}$ of the grid $\frac{1}{q}\mathbf{Z}^d \subset (\mathbf{R}/\mathbf{Z})^d$. So no element of $\{\vec{x}, \dots, N\vec{x}\}$ lies in the half-open box $I_q = \left[q^{-1-\delta/3}, q^{-1} - q^{-1-\delta/3}\right)^d$. Moreover, I_q has volume $\left(q^{-1} - 2q^{-1-\delta/3}\right)^d$. For q sufficiently large, the volume of I_q is bounded below by $2^{-d}q^{-d}$, so the discrepancy $\mathrm{D}_N(\boldsymbol{\vec{x}})$ is bounded below by $2^{-d}q^{-d}$. Since $q^{\omega_0(\vec{x})-\delta} \leqslant 2N$, the discrepancy $\mathrm{D}_N(\boldsymbol{\vec{x}})$ is bounded below by

$$2^{-d}\left((2N)^{\frac{1}{\omega_0(\vec{x})+\delta}}\right)^{-d} = 2^{-d-\frac{d}{\omega_0(\vec{x})+\delta}} N^{-\frac{d}{\omega_0(\vec{x})+\delta}} = 2^{-d\left(1+\frac{1}{\omega_0(\vec{x})}\right)-\epsilon} N^{-\frac{d}{\omega_0(\vec{x})}-\epsilon}.$$

Since $\mathrm{D}_N(\boldsymbol{\vec{x}})$ can, as $N \to \infty$, be bounded below by a constant multiple of $N^{-\frac{d}{\omega_0(\vec{x})}-\epsilon}$, the proof is complete. □

4.3 Pathological Satake parameters for CM abelian varieties

We apply the results of the previous sections to L-functions associated to CM abelian varieties. For background on the motivic Galois group and Sato–Tate group of an abelian variety, see [ST68, Ser94, Yu15]. Recall that for E a non-CM elliptic curve, the Akiyama–Tanigawa conjecture implies the Riemann hypothesis for all $L(\mathrm{sym}^k E, s)$, $k \geqslant 1$. The appearance of sym^k is dictated by the classification of irreducible representations of $\mathrm{SU}(2)$, the Sato–Tate group of E. If A is a CM abelian variety, there should be an L-function (and Galois representation) for each irreducible representation of the Sato–Tate group of A, which we denote by $\mathrm{ST}(A)$. In the CM case, $\mathrm{ST}(A)$ is a real torus, so things can be described relatively explicitly.

Let $K/\mathbf{Q}$ be a finite Galois extension, $A_{/K}$ a g-dimensional abelian variety with complex multiplication by F, defined over K, that is, $F = \mathrm{End}_K(A)_{\mathbf{Q}}$. Since the action of F commutes with $\rho_l\colon G_{\mathbf{Q}} \to \mathrm{GL}_{2g}(\mathbf{Q}_l)$, the Galois representation coming from the l-adic Tate module of A takes values in $\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}(\mathbf{Q}_l)$, where $R_{F/\mathbf{Q}}\mathbf{G}_{\mathrm{m}}$ is the Weil restriction of scalars of the multiplicative group from F to $\mathbf{Q}$. The functor of points of $\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}$ is $R \mapsto (R \otimes F)^{\times}$. It follows that the Sato–Tate group of A is a subgroup of the maximal compact torus inside $\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}(\mathbf{C})$.

Recall, following [Ser94], that the motivic Galois group of A should be a subgroup $G_A \subset \mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}$ such that for all primes l, the image $\rho_l(G_{\mathbf{Q}})$ lies inside $G_A(\mathbf{Q}_l)$, and is open in $G_A(\mathbf{Q}_l)$. For general abelian varieties, the existence of the motivic Galois group is a matter of conjecture, but for CM abelian varieties, it can be described directly. Let $\mathfrak{a} = \mathrm{Lie}(A)$, and $\det_{\mathfrak{a}}\colon \mathrm{R}_{K/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}} \to \mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}$ be the map induced by the determinant of the action of K on $\mathfrak{a}$ (viewed as an F-vector space). Then $G_A = \mathrm{im}(\det_{\mathfrak{a}})$ [Yu15], and $\mathrm{ST}(A)$ is a maximal compact subgroup of $G_A^1(\mathbf{C}) = G_A^{\mathrm{N}_{F/\mathbf{Q}}=1}(\mathbf{C})$. So $\mathrm{ST}(A) \simeq (\mathbf{R}/\mathbf{Z})^d$ for some $1 \leqslant d \leqslant g$, and every unitary character of $\mathrm{ST}(A)$ is induced by an algebraic character of G_A^1. Any character of a subtorus extends to the whole torus, so any character of G_A^1 is the restriction of a character of $\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_{\mathrm{m}}$.

Let $\mathfrak{p}$ be a prime of K at which A has good reduction. Then $F = \mathrm{End}(A)_{\mathbf{Q}} \hookrightarrow \mathrm{End}(A_{/\mathbf{F}_{\mathfrak{p}}})_{\mathbf{Q}}$, and the Frobenius element $\mathrm{fr}_{\mathfrak{p}} \in \mathrm{End}(A_{/\mathbf{F}_{\mathfrak{p}}})_{\mathbf{Q}}$ comes from an element $\pi_{\mathfrak{p}} \in F$. In other words, $\rho_l(\mathrm{fr}_{\mathfrak{p}}) = \pi_{\mathfrak{p}}$. The element $\pi_{\mathfrak{p}} \in F$ is $\mathfrak{p}$-Weil of weight 1, i.e. $|\sigma(\pi_{\mathfrak{p}})| = \mathrm{N}(\mathfrak{p})^{1/2}$ for all embeddings $\sigma\colon F \hookrightarrow \mathbf{C}$. The normalized element $\theta_{\mathfrak{p}} = \frac{\pi_{\mathfrak{p}}}{\mathrm{N}(\mathfrak{p})^{1/2}}$ lies in $\mathrm{ST}(A)$, and we call this the Satake parameter for A at $\mathfrak{p}$. For the Satake parameters to be equidistributed in $\mathrm{ST}(A)$, it is necessary and sufficient for the L-function $L(r \circ \rho_l, s)$ to have non-vanishing analytic continuation to $\Re = 1$ for each $r \in \mathrm{X}^*(\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_\mathrm{m})$ which has nontrivial restriction to $\mathrm{ST}(A)$. By the Wiener–Ikehara Tauberian theorem, this is equivalent to an estimate $\left|\sum_{\mathrm{N}(fp)\leqslant x} r(\theta_\mathfrak{p})\right| = o(\pi_K(x))$.

Theorem 4.3.1 (Shimura–Taniyama, Weil, Hecke)**.** *The elements $\theta_\mathfrak{p} \in \mathrm{ST}(A)$ are equidistributed with respect to the Haar measure.*

Proof. By [ST68, Th. 10, 11], for every $r \in \mathrm{X}^*(\mathrm{R}_{F/\mathbf{Q}}\,\mathbf{G}_\mathrm{m})$ induced by $\sigma\colon F \hookrightarrow \mathbf{C}$, there exists a Hecke character ω_r of K such that $L(r \circ \rho_l, s) = L(s, \omega_r)$. For $r = \sum m_\sigma \sigma$, we have $L(r \circ \rho_l, s) = \prod L(\sigma \circ \rho_l, s)^{m_\sigma}$, so the general result follows. Moreover ω_r is nontrivial if and only if $r|_{\mathrm{ST}(A)}$ is. Since L-functions of Hecke characters have the desired analytic continuation and nonvanishing, the result follows. $\square$

Recall that $L(r \circ \rho_l, s) = \prod \left(1 - r(\theta_\mathfrak{p})\,\mathrm{N}(\mathfrak{p})^{-s}\right)^{-1}$ (this is the normalized L-function, not the algebraic L-function). As in Chapter 2, the choice of an isomorphism $(\mathbf{R}/\mathbf{Z})^d \simeq \mathrm{ST}(A)$ yields a definition of discrepancy for sequences in $\mathrm{ST}(A)$. We call the "Akiyama–Tanigawa conjecture for A" the estimate $\mathrm{D}_N(\boldsymbol{\theta}) \ll N^{-\frac{1}{2}+\epsilon}$, where $\boldsymbol{\theta} = (\theta_\mathfrak{p})_\mathfrak{p}$ is the sequence of Satake parameters of A.

Theorem 4.3.2. *The Akiyama–Tanigawa conjecture for A implies the Riemann hypothesis for all $L(r \circ \rho_l, s)$ with $r|_{\mathrm{ST}(A)}$ nontrivial.*

Proof. The Akiyama–Tanigawa estimate implies, via the Koksma–Hlawka inequality, an estimate $\left|\sum_{\mathrm{N}(\mathfrak{p})\leqslant N} r(\theta_\mathfrak{p})\right| \ll N^{-\frac{1}{2}+\epsilon}$. By Theorem 3.2.1, the function $L(r \circ \rho_l, s)$ satisfies the Riemann hypothesis. $\square$

It is natural to ask: does the Riemann hypothesis for all $L(r \circ \rho_l, s)$ imply the Akiyama–Tanigawa conjecture for A? We proceed to construct L-functions coming from "fake Satake parameters" which provide evidence to the contrary for nonmotivic (non-automorphic, in fact) Satake parameters.

Give $(\mathbf{R}/\mathbf{Z})^d$ the Haar measure normalized to have total mass one. Recall that for any $f \in L^1((\mathbf{R}/\mathbf{Z})^d)$, the Fourier coefficients of f are, for $\vec{m} \in \mathbf{Z}^d$:

$$\widehat{f}(\vec{m}) = \int_{(\mathbf{R}/\mathbf{Z})^d} e^{2\pi i \langle \vec{m}, \vec{x} \rangle} \, \mathrm{d}\vec{x},$$

where $\langle \vec{m}, \vec{x} \rangle = m_1 x_1 + \cdots + m_d x_d$ is the usual inner product. Typically, if f is a function on $(\mathbf{R}/\mathbf{Z})^d$, sums of the form $\sum_{n \leqslant N} f(n\vec{x})$ will be $o(N)$. When f is a character of the torus, there is a much stronger bound.

Theorem 4.3.3. *Fix $\vec{x} \in (\mathbf{R}/\mathbf{Z})^d$ with $\omega_{d-1}(\vec{x})$ finite. Then*

$$\left| \sum_{n \leqslant N} e^{2\pi i \langle \vec{m}, n\vec{x} \rangle} \right| \ll |\vec{m}|_\infty^{\omega_{d-1}(\vec{x}) + \epsilon}$$

as $\vec{m}$ ranges over $\mathbf{Z}^d \smallsetminus 0$.

Proof. From Lemma 4.2.2 we know that $\left| \sum_{n \leqslant N} e^{2\pi i \langle \vec{m}, n\vec{x} \rangle} \right| \ll \mathrm{d}(\langle \vec{m}, \vec{x} \rangle, \mathbf{Z})^{-1}$, and from Lemma 4.1.5, we know that $\mathrm{d}(\langle \vec{m}, \vec{x} \rangle, \mathbf{Z})^{-1} \ll |\vec{m}|_\infty^{\omega_{d-1}(x) + \epsilon}$. The result follows. $\square$

By writing any function as a Fourier series, we can apply this result to sums of the form $\sum_{n \leqslant N} f(n\vec{x})$.

Theorem 4.3.4. *Let $\vec{x} \in \mathbf{R}^d$ with $\omega_{d-1}(\vec{x})$ finite. Fix $f \in C^\infty((\mathbf{R}/\mathbf{Z})^d)$ with $\widehat{f}(\vec{0}) = 0$. Then $\left| \sum_{n \leqslant N} f(n\vec{x}) \right| \ll 1$.*

Proof. Write f as a Fourier series: $f(\vec{x}) = \sum_{\vec{m} \in \mathbf{Z}^d} \widehat{f}(\vec{m}) e^{2\pi i \langle \vec{m}, \vec{x} \rangle}$. Since $\widehat{f}(\vec{0}) = 0$, we can compute:

$$\begin{aligned}\left|\sum_{n\leqslant N} f(n\vec{x})\right| &= \left|\sum_{n\leqslant N}\sum_{\vec{m}\in\mathbf{Z}^d\smallsetminus\vec{0}} \widehat{f}(\vec{m})e^{2\pi in\langle\vec{m},\vec{x}\rangle}\right| \\ &\leqslant \sum_{\vec{m}\in\mathbf{Z}^d\smallsetminus\vec{0}} \left|\widehat{f}(\vec{m})\right|\cdot\left|\sum_{n\leqslant N} e^{2\pi in\langle\vec{m},\vec{x}\rangle}\right| \\ &\ll \sum_{\vec{m}\in\mathbf{Z}^d\smallsetminus\vec{0}} \left|\widehat{f}(\vec{m})\right|\cdot|\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\epsilon}.\end{aligned}$$

The sum converges since the Fourier coefficients $\widehat{f}(\vec{m})$ converge to zero faster than the reciprocal of any polynomial. $\square$

The function f does not need to be smooth—so long as its Fourier coefficients decay sufficiently rapidly as to force $\sum|\widehat{f}(\vec{m})|\cdot|\vec{m}|_\infty^{\omega_{d-1}(\vec{x})+\epsilon}$ to converge, the proof works.

Enumerate the primes of K with increasing norms as $\mathfrak{p}_1,\mathfrak{p}_2,\mathfrak{p}_3,\dots$. Let $\vec{y}\in\mathbf{R}^d$ with $y_1,\dots,y_d$ linearly independent over $\mathbf{Q}$. The associated sequence of "fake Satake parameters" is $\vec{\boldsymbol{x}} = (\vec{y},2\vec{y},3\vec{y},4\vec{y},\dots)$, where we put $\vec{x}_{\mathfrak{p}_n} = n\vec{y} \mod \mathbf{Z}^d$. For any fixed $w\geqslant\frac{1}{d}$, by Theorem 4.1.2, we can find $\vec{y}$ with $\omega_0(\vec{y}) = w$ and $\omega_{d-1}(\vec{y}) = dw+d-1$.

Theorem 4.3.5. *The sequence $\vec{\boldsymbol{x}}$ is equidistributed in $(\mathbf{R}/\mathbf{Z})^d$, with discrepancy decaying as $\mathrm{D}_N(\vec{\boldsymbol{x}})\ll N^{-\frac{1}{dw+d}+\epsilon}$, and for which $\mathrm{D}_N(\vec{\boldsymbol{x}}) = \Omega\left(N^{-\frac{d}{w}-\epsilon}\right)$. However, for any $f\in C^\infty((\mathbf{R}/\mathbf{Z})^d)$ with $\widehat{f}(0)=0$, the Dirichlet series $L_f(\boldsymbol{x},s)$ satisfies the Riemann hypothesis.*

Proof. The upper bound on discrepancy is Theorem 4.2.4, and the lower bound is Theorem 4.2.5. For the functions f in question, Theorem 4.3.4 gives an estimate (stronger than) $\left|\sum_{\mathrm{N}(\mathfrak{p})\leqslant N} f(\vec{x}_{\mathfrak{p}})\right|\ll N^{\frac{1}{2}}$, and Theorem 3.2.1 tells us this estimate implies the Riemann hypothesis. $\square$

This shows that even if each $L(r_*\rho_l, s)$ satisfies the Riemann hypothesis, we may not conclude that the Akiyama–Tanigawa conjecture holds for A. Note also that Theorem 3.2.1 does *not* tell us that $L_f(\vec{\boldsymbol{x}}, s)$ has analytic continuation to $\Re > 0$, or that there are no zeros in $\Re > 0$. For, the term $\sum_{\mathfrak{p}} \sum_{r \geqslant 2} \frac{f(\vec{x}_{\mathfrak{p}})^r}{r\,\mathrm{N}(\mathfrak{p})^{rs}}$ will not converge past $\Re > \frac{1}{2}$.

CHAPTER 5

PATHOLOGICAL GALOIS REPRESENTATIONS

5.1 Notation and supporting results

In this section we loosely summarize and adapt the results of [KLR05, Pan11]. Throughout, if F is a field and M a G_F-module, we write $\mathrm{H}^\bullet(F, M)$ in place of $\mathrm{H}^\bullet(G_F, M)$. All Galois representations will take values in $\mathrm{GL}_2(\mathbf{Z}/l^n)$ or $\mathrm{GL}_2(\mathbf{Z}_l)$ for l a (fixed) rational prime, and all deformations will have fixed determinant. So we consider the cohomology of $\mathrm{Ad}^0 \bar{\rho}$, the induced representation on trace-zero matrices by conjugation.

If S is a set of rational primes, $\mathbf{Q}_S$ denotes the largest extension of $\mathbf{Q}$ unramified outside S. So $\mathrm{H}^i(\mathbf{Q}_S, -)$ is what is usually written as $\mathrm{H}^1(G_{\mathbf{Q},S}, -)$. If M is a $G_{\mathbf{Q}}$-module and S a finite set of primes, denote the corresponding Tate–Shafarevich group by

$$Ш^i_S(M) = \ker\left(\mathrm{H}^i(\mathbf{Q}_S, M) \to \prod_{p\in S} \mathrm{H}^i(\mathbf{Q}_p, M)\right).$$

If l is a rational prime and S a finite set of primes containing l, then for any $\mathbf{F}_l[G_{\mathbf{Q}_S}]$-module M, write $M^\vee = \hom_{\mathbf{F}_l}(M, \mathbf{F}_l)$ with the obvious $G_{\mathbf{Q}_S}$-action, and write $M^* = M^\vee(1)$ for the Cartier dual of $M^\vee$. By [NSW08, Th. 8.6.7], there is an isomorphism $Ш^1_S(M^*) \simeq Ш^2_S(M)^\vee$. As a result, if $Ш^1_S(M)$ and $Ш^2_S(M)$ are trivial, and $S \subset T$, then $Ш^1_T(M)$ and $Ш^2_T(M)$ are also trivial.

Definition 5.1.1. A *good residual representation* is an odd, absolutely irreducible, weight-2 representation $\bar{\rho}\colon G_{\mathbf{Q}_S} \to \mathrm{GL}_2(\mathbf{F}_l)$, where $l \geqslant 7$ is a rational prime.

Recall that $\bar{\rho}$ is weight-2 if $\det\bar{\rho}$ is the mod-l cyclotomic character. Similarly, $\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$ is weight-2 if $\det\rho$ is the l-adic cyclotomic character. Roughly, "good residual representations" have enough properties that we can prove meaningful theorems about their lifts without assuming the modularity results of Khare–Wintenberger.

Theorem 5.1.2. *Let $\bar{\rho}\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_l)$ be a good residual representation. Then there exists a weight-2 lift of $\bar{\rho}$ to $\mathbf{Z}_l$, ramified at the same set of primes as $\bar{\rho}$.*

Proof. This is [Ram02, Th. 1], taking into account that the paper in question allows for arbitrary fixed determinants. □

Definition 5.1.3. Let $\bar{\rho}\colon G_{\mathbf{Q}_S} \to \mathrm{GL}_2(\mathbf{F}_l)$ be a good residual representation. A prime $p \not\equiv \pm 1 \pmod{l}$ is *nice* if $\mathrm{Ad}^0 \bar{\rho} \simeq \mathbf{F}_l \oplus \mathbf{F}_l(1) \oplus \mathbf{F}_l(-1)$, i.e. if the eigenvalues of $\bar{\rho}(\mathrm{fr}_p)$ have ratio p.

Taylor allows $p \equiv -1 \pmod{l}$, but the results of [Pan11] require $p \not\equiv -1 \pmod{l}$. The following theorem gives a complete description of the versal deformation ring for $\bar{\rho}|_{G_{\mathbf{Q}_p}}$ when p is nice.

Theorem 5.1.4 ([Ram99])**.** *Let $\bar{\rho}$ be a good residual representation and p a nice prime. Then any deformation of $\bar{\rho}|_{G_{\mathbf{Q}_p}}$ is induced by $G_{\mathbf{Q}_p} \to \mathrm{GL}_2(\mathbf{Z}_l[\![a,b]\!]/\langle ab\rangle)$, sending*

$$\mathrm{fr}_p \mapsto \begin{pmatrix} p(1+a) & \\ & (1+a)^{-1} \end{pmatrix} \qquad \tau_p \mapsto \begin{pmatrix} 1 & b \\ & 1 \end{pmatrix},$$

where $\tau_p \in G_{\mathbf{Q}_p}$ is a generator for tame inertia.

We close this section by introducing some new terminology and notation to condense the lifting process used in [KLR05].

Fix a good residual representation $\bar{\rho}$. We will consider weight-2 deformations of $\bar{\rho}$ to $\mathbf{Z}/l^n$ and $\mathbf{Z}_l$. Call such a deformation a "lift of $\bar{\rho}$ to $\mathbf{Z}/l^n$ (resp. $\mathbf{Z}_l$)." We will often restrict the local behavior of such lifts, i.e. the restrictions of a lift to $G_{\mathbf{Q}_p}$ for p in some set of primes. The necessary constraints are captured in the following definition.

Definition 5.1.5. Let $\bar{\rho}$ be a good residual representation, $h\colon \mathbf{R}^+ \to \mathbf{R}_{\geqslant 1}$ an increasing function. An *h-bounded lifting datum* is a tuple $(\rho_n, R, U, \{\rho_p\}_{p\in R\cup U})$, where

1. $\rho_n\colon G_{\mathbf{Q}_R} \to \mathrm{GL}_2(\mathbf{Z}/l^n)$ is a lift of $\bar{\rho}$.
2. R and U are finite sets of primes, R containing l and all primes at which ρ_n ramifies.
3. $\pi_R(x) \leqslant h(x)$ for all x.
4. $Ш^1_R(\mathrm{Ad}^0\,\bar{\rho})$ and $Ш^2_R(\mathrm{Ad}^0\,\bar{\rho})$ are trivial.
5. For all $p \in R\cup U$, $\rho_p \equiv \rho_n|_{G_{\mathbf{Q}_p}} \pmod{l^n}$.
6. For all $p \in R$, ρ_p is ramified.
7. ρ_n admits a lift to $\mathbf{Z}/l^{n+1}$.

If $(\rho_n, R, U, \{\rho_p\})$ is an h-bounded lifting datum, we call another h-bounded lifting datum $(\rho_{n+1}, R', U', \{\rho_p\})$ a *lift of* $(\rho_n, R, U, \{\rho_p\})$ if $U \subset U'$, $R \subset R'$, and for all $p \in R \cup U$, the two possible ρ_p agree.

Theorem 5.1.6. *Let $\bar{\rho}$ be a good residual representation, $h\colon \mathbf{R}^+ \to \mathbf{R}_{\geqslant 1}$ increasing to infinity. If $(\rho_n, R, U, \{\rho_p\})$ is an h-bounded lifting datum, $U' \supset U$ is a finite set of primes disjoint from R, and $\{\rho_p\}_{p\in U'}$ extends $\{\rho_p\}_{p\in U}$, then there exists an h-bounded lift $(\rho_{n+1}, R', U', \{\rho_p\})$ of $(\rho_n, R, U, \{\rho_p\})$.*

Proof. By [KLR05, Lem. 8], there exists a finite set N of nice primes such that the map

$$\mathrm{H}^1(\mathbf{Q}_{R\cup N}, \mathrm{Ad}^0\,\bar{\rho}) \to \prod_{p\in R} \mathrm{H}^1(\mathbf{Q}_p, \mathrm{Ad}^0\,\bar{\rho}) \times \prod_{p\in U'} \mathrm{H}^1_{\mathrm{nr}}(\mathbf{Q}_p, \mathrm{Ad}^0\,\bar{\rho}) \tag{5.1}$$

is an isomorphism. In fact, $\#N = \dim \mathrm{H}^1(\mathbf{Q}_{R\cup N}, \mathrm{Ad}^0\,\bar{\rho}^*)$, and the primes in N are chosen, one at a time, from Chebotarev sets. Since $\pi_R(x)$ is eventually constant and $h(x)$ increases to infinity, $h(x) \geqslant \pi_R(x)+1$ for all $x \geqslant C_1$ for some C_1. Choose the first prime p in N to be $\geqslant C_1$; then $\pi_{R\cup\{p\}}(x) \leqslant h(x)$ for all x. Repeat this process for allor all the other primes in N. We can ensure that the bound $\pi_{R\cup N}(x) \leqslant h(x)$ continues to hold. We also choose the primes in N to be larger than any prime in U'.

By our hypothesis, ρ_n admits a lift to $\mathbf{Z}/l^{n+1}$; call one such lift ρ^*. For each $p \in R\cup U'$, $\mathrm{H}^1(\mathbf{Q}_p, \mathrm{Ad}^0\,\bar{\rho})$ acts transitively on lifts of $\rho_n|_{G_{\mathbf{Q}_p}}$ to $\mathbf{Z}/l^{n+1}$. In particular, there are cohomology classes $f_p \in \mathrm{H}^1(\mathbf{Q}_p, \mathrm{Ad}^0\,\bar{\rho})$ such that $f_p \cdot \rho^* \equiv \rho_p \pmod{l^{n+1}}$ for all $p \in R\cup U'$. Moreover, for all $p \in U'$, the class f_p is unramified. Since the map (5.1) is an isomorphism, there exists $f \in \mathrm{H}^1(\mathbf{Q}_{R\cup N}, \mathrm{Ad}^0\,\bar{\rho})$ such that $f\cdot\rho^*|_{G_{\mathbf{Q}_p}} \equiv \rho_p \pmod{l^{n+1}}$ for all $p \in R\cup U'$.

Clearly $f\cdot\rho^*|_{G_{\mathbf{Q}_p}}$ admits a lift to $\mathbf{Z}_l$ for all $p\in R\cup U'$, but it does not necessarily admit such a lift for $p \in N$. By repeated applications of [Pan11, Prop. 3.10], there exists a set $N' \supset N$, with $\#N' \leqslant 2\#N$, of nice primes and $g \in \mathrm{H}^1(\mathbf{Q}_{R\cup N'}, \mathrm{Ad}^0\,\bar{\rho})$ such that $(g+f)\cdot\rho^*$ still agrees with ρ_p for $p \in R\cup U'$, and $(g+f)\cdot\rho^*$ is nice for all $p \in N'$. As above, the primes in N' are chosen one at a time from Chebotarev sets, so we can continue to ensure the bound $\pi_{R\cup N'}(x) \leqslant h(x)$ and also that all primes in N' are larger than those in U'. Let $\rho_{n+1} = (g+f)\cdot\rho^*$. Let $R' = R\cup\{p\in N' : \rho_{n+1} \text{ is ramified at } p\}$. For each $p \in R'\smallsetminus R$, choose a lift ρ_p of $\rho_{n+1}|_{G_{\mathbf{Q}_p}}$ to $\mathbf{Z}_l$.

Since $\rho_{n+1}|_{G_{\mathbf{Q}_p}}$ admits a lift to $\mathbf{Z}/l^{n+2}$ (in fact, it admits a lift to $\mathbf{Z}_l$) for each p, and $Ш^1_{R'}(\mathrm{Ad}^0\bar{\rho})$, $Ш^2_{R'}(\mathrm{Ad}^0\bar{\rho})$ are trivial, the deformation ρ_{n+1} admits a lift to $\mathbf{Z}/l^{n+2}$. Thus $(\rho_{n+1}, R', U', \{\rho_p\})$ is the desired lift of $(\rho_n, R, U, \{\rho_p\})$. □

5.2 Galois representations with specified Satake parameters

Fix a good residual representation $\bar{\rho}$, and consider weight-2 deformations of $\bar{\rho}$. The final deformation, $\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$, will be constructed as the inverse limit of a compatible collection of lifts $\rho_n\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}/l^n)$. At any given stage, we will be concerned with making sure that there exists a lift to the next stage, and that there is a lift with the necessary properties. Fix a sequence $\boldsymbol{x} = (x_1, x_2, \dots)$ in $[-1, 1]$. The set of unramified primes of ρ is not determined at the beginning, but at each stage there will be a large finite set U of primes which we know will remain unramified. Reindexing $\boldsymbol{x}$ by these unramified primes, we will construct ρ so that for all unramified primes p, $\mathrm{tr}\,\rho(\mathrm{fr}_p) \in \mathbf{Z}$, satisfies the Hasse bound, and has $\mathrm{tr}\,\rho(\mathrm{fr}_p) \approx x_p$. Moreover, we can ensure that the set of ramified primes has density zero in a very strong sense (controlled by a parameter function h) and that our trace of Frobenii are very close to specified values.

Given any deformation ρ, write $\pi_{\mathrm{ram}(\rho)}(x)$ for the function which counts ρ_n-ramified primes $\leqslant x$. Since we will have $\pi_{\mathrm{ram}(\rho)}(x) \ll h(x)$ and bounds of this form are only helpful if $h(x) = o(\pi(x))$, we will usually assume $h(x) \ll x^{\epsilon}$, e.g. $h(x) = \log x$ or something which grows even slower (for example, the inverse of the Ackerman function).

Theorem 5.2.1. *Let l, $\bar{\rho}$, $\boldsymbol{x}$ be as above. Fix a function $h\colon \mathbf{R}^+ \to \mathbf{R}_{\geqslant 1}$ which* *increases to infinity. Then there exists a weight-*2 *deformation* ρ *of* $\bar{\rho}$*, such that:*

1. $\pi_{\mathrm{ram}(\rho)}(x) \ll h(x)$.
2. *For each unramified prime* p, $a_p = \operatorname{tr}\rho(\mathrm{fr}_p) \in \mathbf{Z}$ *and satisfies the Hasse bound.*
3. *For each unramified prime* p, $\left|\frac{a_p}{2\sqrt{p}} - x_p\right| \leqslant \frac{lh(p)}{2\sqrt{p}}$.

Proof. Begin with $\rho_1 = \bar{\rho}$. By [KLR05, Lem. 6], there exists a finite set R, containing the set of primes at which $\bar{\rho}$ ramifies, such that $\text{Ш}^1_R(\operatorname{Ad}^0\bar{\rho})$ and $\text{Ш}^2_R(\operatorname{Ad}^0\bar{\rho})$ are trivial. Let R_1 be the union of R and all primes p with $\frac{l}{2\sqrt{p}} > 2$. Since $\frac{l}{2\sqrt{p}} \to 0$ as $p \to \infty$, the set R_1 is finite. For all $p \notin R_1$ and any $a \in \mathbf{F}_l$, there exists $a_p \in \mathbf{Z}$ satisfying the Hasse bound with $a_p \equiv a \pmod{l}$. In fact, given any $x_p \in [-1,1]$, there exists $a_p \in \mathbf{Z}$ satisfying the Hasse bound such that $\left|\frac{a_p}{2\sqrt{p}} - x_p\right| \leqslant \frac{l}{2\sqrt{p}}$. Choose, for all primes $p \in R_1$, a ramified lift ρ_p of $\rho_1|_{G_{\mathbf{Q}_p}}$. Let U_1 be the set of primes p not in R_1 such that $\frac{l^2}{2\sqrt{p}} > \min\left(2, \frac{lh(p)}{2\sqrt{p}}\right)$; this is finite because $\frac{l^2}{2\sqrt{p}} \to 0$ and also eventually $h(p) \geqslant l$. If U_1 is empty, then the next few sentences of the proof are superfluous, but the theorem still holds. For each $p \in U_1$, there exists $a_p \in \mathbf{Z}$, satisfying the Hasse bound, such that

$$\left|\frac{a_p}{2\sqrt{p}} - x_p\right| \leqslant \frac{l}{2\sqrt{p}} \leqslant \frac{lh(p)}{2\sqrt{p}},$$

and moreover $a_p \equiv \operatorname{tr}\bar{\rho}(\mathrm{fr}_p) \pmod{l}$. For each $p \in U_1$, let ρ_p be an unramified lift of $\bar{\rho}|_{G_{\mathbf{Q}_p}}$ with $\operatorname{tr}\rho_p$ being the desired a_p. It may not be that $\pi_{R_1}(x) \leqslant h(x)$ for all x. Let $C = \max\{\pi_{R_1}(x)\}$; this is finite because R_1 is and $\pi_{R_1}(x)$ is constant past the largest prime in R_1. Then for $h^* = Ch$, we have $\pi_{R_1}(x) \leqslant h^*(x)$ for all x.

We have constructed our first h^*-bounded lifting datum $(\rho_1, R_1, U_1, \{\rho_p\})$. We proceed to construct $\rho = \varprojlim \rho_n$ inductively, by constructing a new h^*-bounded lifting datum for each n. We ensure that U_n contains all primes for which $\frac{l^{n+1}}{2\sqrt{p}} > \min\left(2, \frac{lh(p)}{2\sqrt{p}}\right)$, so there are always integral a_p satisfying the Hasse bound which satisfy any mod-l^{n+1} constraint, and that can always choose these a_p so as to preserve statement 2 in the theorem.

The base case is already complete, so suppose we are given $(\rho_{n-1}, R_{n-1}, U_{n-1}, \{\rho_p\})$. We may assume that U_{n-1} contains all primes for which $\frac{l^n}{2\sqrt{p}} > \min\left(2, \frac{lh(p)}{2\sqrt{p}}\right)$. Let U_n be the set of all primes not in R_{n-1} such that $\frac{l^{n+1}}{2\sqrt{p}} > \min\left(2, \frac{lh(p)}{2\sqrt{p}}\right)$. For each $p \in U_n \smallsetminus U_{n-1}$, there is an integer a_p, satisfying the Hasse bound, such that $a_p \equiv \rho_n(\mathrm{fr}_p) \pmod{l^n}$, and moreover $\left|\frac{a_p}{2\sqrt{p}} - x_p\right| \leqslant \frac{lh(p)}{2\sqrt{p}}$. For such p, let ρ_p be an unramified lift of $\rho_n|_{G_{\mathbf{Q}_p}}$ such that $\operatorname{tr}\rho_n(\mathrm{fr}_p)$ is the desired a_p. By Theorem 5.1.6, there exists an h^*-bounded lifting datum $(\rho_n, R_n, U_n, \{\rho_p\})$ extending and lifting $(\rho_{n-1}, R_{n-1}, U_{n-1}, \{\rho_p\})$. This completes the inductive step. $\square$

The implied constant in the bound $\pi_{\mathrm{ram}(\rho)}(x) \ll h(x)$ depends on $\bar{\rho}$ (and hence l) but not on h. We will apply this theorem to construct Galois representations with specified Sato–Tate distributions in the next section, but for now here is a small consequence, which addresses the results in [Sar07]. Sarnak remarks that for $E_{/\mathbf{Q}}$ a non-CM elliptic curve with rank r, the partial sums $\frac{\log x}{\sqrt{x}} \sum_{p \leqslant x} \frac{a_p}{\sqrt{p}}$ approach a limiting distribution with mean $1 - 2r$.

Corollary 5.2.2. *Let $L \in [-\infty, \infty]$ and $\epsilon > 0$ be given. Then there exists a weight 2 Galois representation $\rho\colon G \to \mathrm{GL}_2(\mathbf{Z}_l)$, such that each $a_p = \operatorname{tr}\rho(\mathrm{fr}_p) \in \mathbf{Z}$ satisfies the Hasse bound,*

$$L = \lim_{N \to \infty} \frac{\log N}{\sqrt{N}} \sum_p \frac{a_p}{\sqrt{p}}$$

and $\pi_{\mathrm{ram}(\rho)}(x) \ll \log(x)$.

Proof. Begin with a sequence (x_p) in $[-1, 1]$ such that $\lim_{N\to\infty} \frac{\log N}{\sqrt{N}} \sum_{p \leqslant N} x_p = L$.

If $L = \pm\infty$, we can choose $x_p = \pm 1$. By Theorem 5.2.1, there exists $\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$ with $\pi_{\mathrm{ram}(\rho)}(x) \ll \log(x)$, and such that for each unramified p, $a_p = \operatorname{tr}\rho(\mathrm{fr}_p) \in \mathbf{Z}$, satisfies the Hasse bound, has $\left|\frac{a_p}{2\sqrt{p}} - x_p\right| < \frac{l\log p}{\sqrt{p}}$, and by inspecting the proof, we can even ensure that $\sum\left(\frac{a_p}{2\sqrt{p}} - x_p\right)$ converges conditionally (make sure the sign of $\frac{a_p}{2\sqrt{p}} - b_p$ alternates). Note that

$$\begin{aligned}\left|\frac{\log N}{\sqrt{N}}\sum_{p\leqslant N}\frac{a_p}{2\sqrt{p}} - \frac{\log N}{\sqrt{N}}\sum_{p\leqslant N} x_p\right| &\leqslant \frac{\log N}{\sqrt{N}}\left|\sum_{\substack{p\leqslant N\\ p \text{ ramified}}}\left(\frac{a_p}{2\sqrt{p}} - x_p\right) + \sum_{\substack{p\leqslant N\\ p \text{ unramified}}}\left(\frac{a_p}{2\sqrt{p}} - x_p\right)\right| \\ &\ll \frac{\log N}{\sqrt{N}}\left(\pi_{\mathrm{ram}(\rho)}(N) + \left|\sum_{p\leqslant N}\left(\frac{a_p}{2\sqrt{p}} - x_p\right)\right|\right),\end{aligned}$$

which converges to zero. When $L \neq \pm\infty$, this shows that the limit in question exists and is L. When $L = \pm\infty$, this shows that the the sums in question diverge to L. $\square$

5.3 Galois representations with specified Sato–Tate distributions

For $k \geqslant 1$, let

$$U_k(\theta) = \operatorname{tr}\operatorname{sym}^k\begin{pmatrix} e^{i\theta} & \\ & e^{-i\theta}\end{pmatrix} = \frac{\sin((k+1)\theta)}{\sin\theta}.$$

Then $U_k(\cos^{-1} t)$ is the k-th Chebyshev polynomial of the 2nd kind. Moreover, $\{1\} \cup \{U_k\}$ forms an orthonormal basis for $L^2([0,\pi], \mathrm{ST}) = L^2(\mathrm{SU}(2)^\natural)$.

This section has two parts. First, for any reasonable measure μ on $[0,\pi]$ invariant under the same "flip" automorphism as the Sato–Tate measure, there is a sequence (a_p) of integers satisfying the Hasse bound $|a_p| \leqslant 2\sqrt{p}$, such that for $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right)$, the discrepancy $\mathrm{D}_N(\boldsymbol{\theta}, \mu)$ behaves like $\pi(N)^{-\alpha}$ for predetermined

$\alpha \in \left(0, \frac{1}{2}\right)$, while for any odd k, the strange Dirichlet series $L_{U_k}(\boldsymbol{\theta}, s)$, which we will write as $L(\mathrm{sym}^k \boldsymbol{\theta}, s)$, satisfies the Riemann hypothesis. In the second part of this section, we associate Galois representations to these fake Satake parameters.

Definition 5.3.1. Let $\mu = f(t)\,\mathrm{d}t$ be an absolutely continuous measure with continuous cdf on $[0, \pi]$. If $f(t) \ll \sin(t)$, then μ is a *Sato–Tate compatible measure.*

The key facts about Sato–Tate compatible measures are that $\cos_* \mu$ satisfies the hypotheses of Theorem 2.6.5, so there are "$N^{-\alpha}$-decaying van der Corput sequences" for $\cos_* \mu$, and also that since $\cos\colon [0, \pi] \to [-1, 1]$ is an strictly decreasing, we know that for any sequence $\boldsymbol{x}$ on $[-1, 1]$, $\mathrm{D}_N(\boldsymbol{x}, \cos_* \mu) \approx \mathrm{D}_N(\cos^{-1} \boldsymbol{x}, \mu)$. Finally, the Radon–Nikodym derivative μ (and also $\cos_* \mu$) is bounded , so Lemma 2.5.1 applies. Recall that for deceasing functions φ_1, φ_2, we write $\varphi_1(N) = \Theta(\varphi_2(N))$ if there exists constants $0 < C_1 < C_2$ such that $C_1\varphi_2(N) \leqslant \varphi_1(N) \leqslant C_2\varphi_2(N)$.

Theorem 5.3.2. *Let μ be a Sato–Tate compatible measure, and fix $\alpha \in \left(0, \frac{1}{2}\right)$. Then there exists a sequence of integers a_p satisfying the Hasse bound, such that if we set $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right)$, then $\mathrm{D}_N(\boldsymbol{\theta}, \mu) = \Theta(\pi(N)^{-\alpha})$.*

Proof. Apply Theorem 2.6.5 to find a sequence $\boldsymbol{x}$ such that $\mathrm{D}_N(\boldsymbol{x}, \cos_* \mu) = \Theta(\pi(N)^{-\alpha})$. For each prime p, there exists an integer a_p such that $|a_p| \leqslant 2\sqrt{p}$ and $\left|\frac{a_p}{2\sqrt{p}} - x_p\right| \leqslant p^{-1/2}$. Let $y_p = \frac{a_p}{2\sqrt{p}}$, and apply Lemma 2.5.1 with $\epsilon = N^{-1/2}$. We obtain

$$|\mathrm{D}_N(\boldsymbol{x}, \cos_* \mu) - \mathrm{D}_N(\boldsymbol{y}, \cos_* \mu)| \ll N^{-1/2} + \frac{\pi(N^{1/2})}{\pi(N)},$$

which tells us that $\mathrm{D}_N(\boldsymbol{y}, \cos_* \mu) = \Theta(\pi(N)^{-\alpha})$. Now let $\boldsymbol{\theta} = \cos^{-1}(\boldsymbol{y})$. Apply Lemma 2.5.3 to $\boldsymbol{\theta} = \cos^{-1}(\boldsymbol{y})$, and we see that $\mathrm{D}_N(\boldsymbol{\theta}, \mu) = \Theta(\pi(N)^{-\alpha})$. $\square$

We can improve this example by controlling the behavior of the sums $\sum_{p \leqslant N} U_k(\theta_p)$ for odd k. Let σ be the involution of $[0, \pi]$ given by $\sigma(\theta) = \pi - \theta$. Note that $\sigma_* \mathrm{ST} = \mathrm{ST}$. Moreover, note that for any odd k, $U_k \circ \sigma = -U_k$, so $\int U_k \, \mathrm{dST} = 0$. Of course, $\int U_k \, \mathrm{dST} = 0$ for the reason that U_k is the trace of a non-trivial unitary representation, but we will directly use the "oddness" of U_k in what follows.

Theorem 5.3.3. *Let μ be a σ-invariant Sato–Tate compatible measure. Fix $\alpha \in \left(0, \frac{1}{2}\right)$. Then there is a sequence of integers a_p, satisfying the Hasse bound, such that for $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right)$, we have*

1. $\mathrm{D}_N(\boldsymbol{\theta}, \mu) = \Theta(\pi(N)^{-\alpha})$.
2. *For all odd k,* $\left|\sum_{k \leqslant N} U_k(\theta_p)\right| \ll \pi(N)^{1/2}$.

Proof. The basic ideas is as follows. Enumerate the primes

$$p_1 = 2, q_1 = 3, p_2 = 5, q_2 = 7, p_3 = 11, q_3 = 13, \ldots.$$

Consider the measure $\mu|_{[0,\pi/2)}$. An argument nearly identical to the proof of Theorem 5.3.2 shows that we can choose a_{p_i} satisfying the Hasse bound so that

$$\mathrm{D}_N\left(\{\theta_{p_i}\}, \mu|_{[0,\pi/2)}\right) = \Theta(N^{-\alpha}).$$

We can also choose the a_{q_i} such that $\frac{a_{q_i}}{2\sqrt{q_i}} \in [\pi/2, \pi]$ and $\left|\frac{a_{p_i}}{2\sqrt{p_i}} + \frac{a_{q_i}}{2\sqrt{q_i}}\right| \ll \frac{1}{\sqrt{p_i}}$. Let $\boldsymbol{x}$ be the sequence of the $\frac{a_{p_i}}{2\sqrt{p_i}}$ and $\boldsymbol{y}$ the corresponding sequence with the q_i-s. Then Lemma 2.5.2 tells us that the discrepancy of $\boldsymbol{y}$ decays at the same rate (within $O(N^{-1})$) as $-\boldsymbol{y}$, and then Lemma 2.5.1 with $\epsilon \approx N^{-1/2}$ tells us that the discrepancy of $-\boldsymbol{y}$ decays at the same rate (within $O(N^{-1/2})$) as the discrepancy of $\boldsymbol{x}$. Thus the discrepancies of both $\boldsymbol{x}$ and $\boldsymbol{y}$ decay as $\Theta(N^{-\alpha})$. Finally, Theorem 2.5.5 tell us that $\mathrm{D}_N(\boldsymbol{x} \wr \boldsymbol{y}, \mu) = \Theta(N^{-\alpha})$.

The function $U_k(\cos^{-1} t)$ is an odd polynomial in t, so for $t_1, t_2 \in [-1,1]$,

$$|U_k(\cos^{-1} t_1) + U_k(\cos^{-1} t_2)| = |U_k(\cos^{-1} t_1) - U_k(\cos^{-1}(-t_1))| \ll |t_1 - (-t_2)|.$$

It follows that since $\left|\frac{a_{p_i}}{2\sqrt{p_i}} - \left(-\frac{a_{q_i}}{2\sqrt{q_i}}\right)\right| \ll p_i^{-1/2}$, then $|U_k(\theta_{p_i}) + U_k(\theta_{q_i})| \ll p_i^{-1/2}$. We can then bound

$$\left|\sum_{i\leqslant N} (U_k(\theta_{p_i}) + U_k(\theta_{q_i}))\right| \ll \sum_{p\leqslant N} p^{-1/2} \ll \pi(N)^{1/2}.$$

□

Note that this proof actually shows that for any $f \in C([0,\pi])$ such that $f\circ\cos^{-1}$ is Lipschitz, and $f(\pi - \theta) = -f(\theta)$, the estimate $\left|\sum_{p\leqslant N} f(\theta_p)\right| \ll \pi(N)^{1/2}$ holds.

Theorem 5.3.4. *Let μ be a Sato–Tate compatible σ-invariant measure on $[0,\pi]$. Fix $\alpha \in \left(0, \frac{1}{2}\right)$ and a good residual representation $\bar\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{F}_l)$. Then there exists a weight-2 lift $\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$ of $\bar\rho$ such that*

1. *$\pi_{\mathrm{ram}(\rho)}(x) \ll \log(x)$.*
2. *For each unramified prime p, $a_p = \operatorname{tr}\rho(\mathrm{fr}_p) \in \mathbf{Z}$ and satisfies the Hasse bound.*
3. *If, for unramified p we set $\theta_p = \cos^{-1}\left(\frac{a_p}{2\sqrt{p}}\right)$, then $\mathrm{D}_N(\boldsymbol{\theta},\mu) = \Theta(\pi(N)^{-\alpha})$.*
4. *For each odd k, the function $L(\mathrm{sym}^k \rho, s)$ satisfies the Riemann hypothesis.*

Proof. Let $\boldsymbol{x}$ be an $N^{-\alpha}$-decay van der Corput sequence for $\cos_* \mu|_{[0,\pi/2)}$, so that $\boldsymbol{x}$ is contained in $(0,1]$. Let $\boldsymbol{y} = -\boldsymbol{x}$ (contained in $[-1,0)$), and put $\boldsymbol{z} = \boldsymbol{x} \wr \boldsymbol{y}$, reindexed by the prime numbers. We have $\mathrm{D}_N(\boldsymbol{z}, \cos_*\mu) = \Theta(\pi(N)^{-\alpha})$ just as in the proof of Theorem 5.3.3. Set $h(x) = \log(x)$. By Theorem 5.2.1, there is a $\rho\colon G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$ lifting $\bar\rho$ such that $\pi_{\mathrm{ram}(\rho)}(x) \ll \log x$, the $\operatorname{tr}\rho(\mathrm{fr}_p)$ are integral, satisfy the Hasse bound, and $\left|\frac{a_p}{2\sqrt{p}} - z_p\right| \leqslant \frac{l \log p}{2\sqrt{p}}$. This implies, just as in the proof of Theorem 5.2.1, that the discrepancy of the sequence $\left\{\frac{a_p}{2\sqrt{p}}\right\}$ decays as $\Theta(\pi(N)^{-\alpha})$ and by Lemma 2.5.3, the discrepancies of $\left\{\frac{a_p}{2\sqrt{p}}\right\}$ and $\{\theta_p\}$ decay asymptotically at the same rate.

We've proved statements 1–3 in the theorem, so all that remains is to prove the Riemann hypothesis for odd symmetric powers. The proof of Theorem 5.3.3 gives us an estimate $\left|\sum_{p \leqslant N} U_k(\theta_p)\right| \ll N^{\frac{1}{2}+\epsilon}$, and this combined with Theorem 3.2.1 yields the result. $\square$

This entire discussion also works with absolutely continuous measure μ, supported on a proper subinterval of $[0, \pi]$, so long as cdf_μ is strictly increasing on that interval. Let I be an arbitrarily small subinterval of $[0, \pi]$ (for example, $I = \left[\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon\right]$), let $B_I(t)$ be a bump function for I, normalized to have total mass one. Then Theorem 5.3.4 gives Galois representations with empirical Sato–Tate distribution converging at any specified rate to $\mu_I = B_I(t)\,\mathrm{d}t$. This is a strictly stronger result than [Pan11, Th. 5.2]. Moreover, the proof of Theorem 5.3.3 shows that in fact for any $f \in C([0, \pi])$ with $f \circ \cos^{-1} \in C^1([-1, 1])$ and $f(\pi - \theta) = -f(\theta)$, the Dirichlet series $L_f(\rho, s) = \prod \left(1 - f(\theta_p) p^{-s}\right)^{-1}$ satisfies the Riemann hypothesis.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE DIRECTIONS

6.1 Fake modular forms

The Galois representations of Theorem 5.3.4 have "fake modular forms" associated to them. Namely, there is a representation of $\mathrm{GL}_2(\mathbf{A})$ with the specified Satake parameters at each prime (for now, set $\theta_p = 0$ at ramified primes). It is natural to ask if these "fake modular forms" have any interesting properties. For example, we know that all their odd symmetric powers satisfy the Riemann hypothesis. The author is unaware of any further results (say about analytic continuation or functional equation) concerning these fake modular forms.

6.2 Dense free subgroups of compact semisimple groups

Let G be a compact semisimple Lie group, for example $\mathrm{SU}(2)$. By [BG03], G contains a dense free subgroup $\Gamma = \langle \gamma_1, \gamma_2 \rangle$. We will now follow the argument of [AK63] to hint at how Γ may yield equidistributed sequences with "bad" discrepancy and small character sums.

Given an integer N, let B_N be the "closed ball of size N" in Γ, that is the set of products $\gamma_{\sigma(1)} \dots \gamma_{\sigma(n)}$, where $n \leqslant N$ and $\sigma \colon \{1, \dots, n\} \to \{1, 2\}$ is a function. We will write $\sigma \colon [n] \to [2]$ in this case. Given an irreducible unitary representation $\rho \in \widehat{G}$, we wish to control the behavior of $\sum_{\gamma \in B_N} \operatorname{tr} \rho(\gamma)$, ideally to show an estimate of the form

$$\left| \sum_{\gamma \in B_N} \operatorname{tr} \rho(\gamma) \right| \ll (\# B_N)^{\frac{1}{2}+\epsilon} .$$

In fact, $\#B_N = \sum_{n=0}^{N} 2^n = 2^{N+1} - 1$. We can encode these sums in terms of convolutions of a measure as follows. Let μ be the measure $\delta_{\gamma_1^{-1}} + \delta_{\gamma_2^{-1}}$ on G. If ρ is any unitary representation (not necessarily irreducible or even finite-dimensional) then μ acts on ρ via $\rho(\mu) \int \rho \, \mathrm{d}\mu$. So, if $\rho = L^2(G)$ via the left regular representation, then $(\mu \cdot f)(x) = f(\gamma_1 x) + f(\gamma_2 x)$, while if $\rho \in \widehat{G}$ and $v \in \rho$, then $\mu \cdot v = \rho(\gamma_1)v + \rho(\gamma_2)v$. Note that

$$\mu^{*n} = \sum_{\sigma \colon [n] \to [2]} \delta_{\gamma_{\sigma(1)} \cdots \gamma_{\sigma(n)}}.$$

This tells us that $\sum_{\gamma \in B_N} f(\gamma) = \sum_{n \leqslant N} \mu^{*n}(f)$. So we really only need to study how μ and its powers act on the functions $\operatorname{tr} \rho$, $\rho \in \widehat{G}$.

First note that $\operatorname{tr} \rho$ generates a subrepresentation of $L^2(G)$ which is isomorphic to ρ. On that representation, we claim that μ is invertible, hence $\sum_{n=0}^{N} \mu^{*n} = (\mu^{*(N+1)} - 1)(\mu - 1)^{-1}$.It follows that $\| \sum_{n=0}^{N} \mu^{*n} \| \leqslant \frac{\|\mu\|^{N+1}}{\|\mu - 1\|}$,

Note that $\|\mu\|^{N+1} \leqslant 2^{(N+1)\alpha}$ if and only if $\|\mu\| \leqslant 2^{\alpha}$. In other words, to get the Riemann hypothesis for L-functions coming from Γ, we need $\|\mu\| \leqslant \sqrt{2}$. If $v \in \rho$ has norm 1, then

$$\begin{aligned}
\|\rho(\mu)v\|^2 &= \langle \rho(\gamma_1^{-1})v + \rho(\gamma_2^{-1})v, \rho(\gamma_1^{-1})v + \rho(\gamma_2^{-1})v \rangle \\
&= 2\|v\|^2 + 2\Re\langle \rho(\gamma_2 \gamma_1^{-1})v, v \rangle.
\end{aligned}$$

So, we want $\Re\langle \rho(\gamma_2\gamma_1^{-1})v, v\rangle \leqslant 0$ for all irreducible ρ. Sadly, even for $\mathrm{SU}(2)$, this is not possible.

Write $\gamma = \gamma_2 \gamma_1^{-1}$, then the identity $\langle \rho(\gamma)\rho(\delta)v, \rho(\delta)v \rangle = \langle \rho(\delta^{-1}\gamma\delta)v, v\rangle$ tells us that we can restrict our search to γ of the form $\left(\begin{smallmatrix} a & \\ & \bar{a} \end{smallmatrix}\right)$ with $|a| = 1$. Now

$$\langle \left(\begin{smallmatrix} a & \\ & \bar{a} \end{smallmatrix}\right) \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right), \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) \rangle = \Re(a),$$

which appears to be promising. But a similar computation with sym^2 shows that one can always get $\langle \mathrm{sym}^2 \gamma v, v \rangle = 1$, so the above approach fails.

There may be alternative ways of bounding the sums $\sum \mu^{*n}(\mathrm{tr}\,\rho)$, but we do not investigate them here.

BIBLIOGRAPHY

[AK63] V. I. Arnol'd and A. L. Krylov. Uniform distribution of points on a sphere and certain ergodic properties of solutions of linear ordinary differential equations in a complex domain. *Dokl. Akad. Nauk SSSR*, 148:9–12, 1963.

[Apo76] Tom M. Apostol. *Introduction to analytic number theory*. Springer-Verlag, New York-Heidelberg, 1976. Undergraduate Texts in Mathematics.

[AT99] Shigeki Akiyama and Yoshio Tanigawa. Calculation of values of L-functions associated to elliptic curves. *Math. Comp.*, 68(227):1201–1231, 1999.

[BG03] E. Breuillard and T. Gelander. On dense free subgroups of Lie groups. *J. Algebra*, 261(2):448–467, 2003.

[BK15] Alina Bucar and Kiran Kedlaya. An application of the effective Sato–Tate conjecture, 2015.

[Bou05] Nicolas Bourbaki. *Lie groups and Lie algebras. Chapters 7–9*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 2005. Translated from the 1975 and 1982 French originals by Andrew Pressley.

[DT97] Michael Drmota and Robert F. Tichy. *Sequences, discrepancies and applications*, volume 1651 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1997.

[Edw74] Harold M. Edwards. *Riemann's zeta function*, volume 58 of *Pure and Applied Mathematics*. Academic Press, 1974.

[Har09] Michael Harris. Potential automorphy of odd-dimensional symmetric powers of elliptic curves and applications. In *Algebra, arithmetic, and geometry: in honor of Yu. I. Manin. Vol. II*, volume 270 of *Progr. Math.*, pages 1–21. Birkhäuser Boston, 2009.

[HSBT10] Michael Harris, Nick Shepherd-Barron, and Richard Taylor. A family of Calabi-Yau varieties and potential automorphy. *Ann. of Math. (2)*, 171(2):779–813, 2010.

[Jar36] Vojtěch Jarník. Über eien Satz von A. Khintchine. Zweite Mitteilung. *Acta Arith.*, 2(1):1–22, 1936.

[Kat88] Nicholas M. Katz. *Gauss sums, Kloosterman sums, and monodromy groups*, volume 116 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1988.

[KLR05] Chandrashekhar Khare, Michael Larsen, and Ravi Ramakrishna. Constructing semisimple p-adic Galois representations with prescribed properties. *Amer. J. Math.*, 127(4):709–734, 2005.

[KN74] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Wiley-Interscience, 1974. Pure and Applied Mathematics.

[Lau09] Michel Laurent. On transfer inequalities in Diophantine approximation. In *Analytic number theory*, pages 306–314. Cambridge Univ. Press, Cambridge, 2009.

[Maz08] Barry Mazur. Finding meaning in error terms. *Bull. Amer. Math. Soc. (N.S.)*, 45(2):185–228, 2008.

[Nie91] Harald Niederreiter. The distribution of values of Kloosterman sums. *Arch. Math. (Basel)*, 56(3):270–277, 1991.

[NSW08] Jürgen Neukirch, Alexander Schmidt, and Kay Wingberg. *Cohomology of number fields*, volume 323 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, second edition, 2008.

[Ö] G. Ökten. Error reduction techniques in quasi-monte carlo integration. *Math. Comput. Modelling*, 30(7-8):61–69.

[Pan11] Aftab Pande. Deformations of Galois representations and the theorems of Sato–Tate and Lang–Trotter. *Int. J. Number Theory*, 7(8):2065–2079, 2011.

[Ram99] Ravi Ramakrishna. Lifting Galois representations. *Invent. Math.*, 138(3):537–562, 1999.

[Ram02] Ravi Ramakrishna. Deforming Galois representations and the conjectures of Serre and Fontaine–Mazur. *Ann. of Math. (2)*, 156(1):115–154, 2002.

[Ros13] Zev Rosengarten. An Erdös–Turán inequality for compact simply-connected semisimple lie groups, 2013.

[Rot55] K. F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2:1–20, 1955.

[RT16] Jeremy Rouse and Jesse Thorner. The explicit Sato–Tate conjecture and densities pertaining to Lehmer-type questions, 2016.

[Sar07] Peter Sarnak. Letter to: Barry Mazur on "Chebyshev's bias" for $\tau(p)$, 2007.

[Ser81] Jean-Pierre Serre. Quelques applications du théorème de densité de Chebotarev. *Inst. Hautes Études Sci. Publ. Math.*, (54):323–401, 1981.

[Ser89] Jean-Pierre Serre. *Abelian l-adic representations and elliptic curves.* Advanced Book Classics. Addison-Wesley, second edition, 1989.

[Ser94] Jean-Pierre Serre. Propriétés conjecturales des groupes de Galois motiviques et des représentations l-adiques. In *Motives (Seattle, WA, 1991)*, volume 55 of *Proc. Sympos. Pure Math.*, pages 377–400. Amer. Math. Soc., Providence, RI, 1994.

[Sil09] Joseph H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Graduate Texts in Mathematics*. Springer, second edition, 2009.

[ST68] Jean-Pierre Serre and John Tate. Good reduction of abelian varieties. *Ann. of Math. (2)*, 88:492–517, 1968.

[Ten95] Gérald Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1995.

[Yu15] Chia-Fu Yu. A note on the Mumford–Tate conjecture for CM abelian varieties. *Taiwanese J. Math.*, 19(4):1073–1084, 2015.