

ON THE DISTRIBUTION OF MATRIX ELEMENTS FOR THE QUANTUM CAT MAP

PÄR KURLBERG AND ZEÉV RUDNICK

ABSTRACT. For many classically chaotic systems it is believed that the quantum wave functions become uniformly distributed, that is the matrix elements of smooth observables tend to the phase space average of the observable. In this paper we study the fluctuations of the matrix elements for the desymmetrized quantum cat map. We present a conjecture for the distribution of the normalized matrix elements, namely that their distribution is that of a certain weighted sum of traces of independent matrices in $SU(2)$. This is in contrast to generic chaotic systems where the distribution is expected to be Gaussian. We compute the second and fourth moment of the normalized matrix elements and obtain agreement with our conjecture.

1. INTRODUCTION

A fundamental feature of quantum wave functions of classically chaotic systems is that the matrix elements of smooth observables tend to the phase space average of the observable, at least in the sense of convergence in the mean [14, 2, 16] or in the mean square [17]. In many systems it is believed that in fact *all* matrix elements converge to the micro-canonical average, however this has only been demonstrated for a couple of arithmetic systems: For “quantum cat maps” [11], and conditional on the Generalized Riemann Hypothesis¹ also for the modular domain [15], in both cases assuming that the systems are desymmetrized by taking into account the action of “Hecke operators”.

As for the approach to the limit, it is expected that the *fluctuations* of the matrix elements about their limit are Gaussian with variance given by classical correlations of the observable [7, 5]. In this note we

Date: February 18, 2003.

This work was supported in part by the EC TMR network “Mathematical aspects of Quantum Chaos” (HPRN-CT-2000-00103). P.K. was also supported in part by the NSF (DMS 0071503), the Royal Swedish Academy of Sciences and the Swedish Research Council. Z.R. was also supported in part by the US-Israel Bi-National Science Foundation.

¹An unconditional proof was recently announced by Elon Lindenstrauss.

study these fluctuations for the quantum cat map. Our finding is that for this system, the picture is very different.

We recall the basic setup [8, 3, 4, 11] (see section 2 for further background and any unexplained notation): The classical mechanical system is the iteration of a linear hyperbolic map $A \in SL(2, \mathbf{Z})$ of the torus $\mathbf{T}^2 = \mathbf{R}^2/\mathbf{Z}^2$ (a “cat map”). The quantum system is given by specifying an integer N , which plays the role of the inverse Planck constant. In what follows, N will be restricted to be a prime. The space of quantum states of the system is $\mathcal{H}_N = L^2(\mathbf{Z}/N\mathbf{Z})$. Let $f \in C^\infty(\mathbf{T}^2)$ be a smooth, real valued observable and $\text{Op}_N(f) : \mathcal{H}_N \rightarrow \mathcal{H}_N$ its quantization. The quantization of the classical map A is a unitary map $U_N(A)$ of \mathcal{H}_N .

In [11] we introduced *Hecke operators*, a group of commuting unitary maps of \mathcal{H}_N , which commute with $U_N(A)$. The space \mathcal{H}_N has an orthonormal basis consisting of joint eigenvectors $\{\psi_j\}_{j=1}^N$ of $U_N(A)$, which we call *Hecke eigenfunctions*. The matrix elements $\langle \text{Op}_N(f)\psi_j, \psi_j \rangle$ converge² to the phase-space average $\int_{\mathbf{T}^2} f(x)dx$ [11]. Our goal is to understand their fluctuations around their limiting value.

Our main result is to present a conjecture for the limiting distribution of the normalized matrix elements

$$F_j^{(N)} := \sqrt{N} \left(\langle \text{Op}_N(f)\psi_j, \psi_j \rangle - \int_{\mathbf{T}^2} f(x)dx \right).$$

For this purpose, define a binary quadratic form associated to A by

$$Q(x, y) = cx^2 + (d - a)xy - by^2, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

For an observable $f \in C^\infty(\mathbf{T}^2)$ and an integer ν , set

$$f^\#(\nu) := \sum_{\substack{n=(n_1, n_2) \in \mathbf{Z}^2 \\ Q(n)=\nu}} (-1)^{n_1 n_2} \widehat{f}(n)$$

where $\widehat{f}(n)$ are the Fourier coefficients of f .

Conjecture 1. *As $N \rightarrow \infty$ through primes, the limiting distribution of the normalized matrix elements $F_j^{(N)}$ is that of the random variable*

$$X_f := \sum_{\nu \neq 0} f^\#(\nu) \text{tr}(U_\nu)$$

where U_ν are independently chosen random matrices in $SU(2)$ endowed with Haar probability measure.

²For arbitrary eigenfunctions, that is ones which are not Hecke eigenfunctions, this need not hold, see [6].

This conjecture predicts a radical departure from the Gaussian fluctuations expected to hold for generic systems [7, 5]. Our first result confirms this conjecture for the variance of these normalized matrix elements.

Theorem 2. *As $N \rightarrow \infty$ through primes, the variance of the normalized matrix elements $F_j^{(N)}$ is given by*

$$(1.1) \quad \frac{1}{N} \sum_{j=1}^N |F_j^{(N)}|^2 \rightarrow \mathbf{E}(X_f^2) = \sum_{\nu \neq 0} |f^\#(\nu)|^2 .$$

For a comparison with the variance expected for the case of *generic* systems, see Section 6.1. A similar departure from this behaviour of the variance was observed recently by Luo and Sarnak [13] for the modular domain. For another analogy with that case, see section 6.2.

We also compute the fourth moment of $F_j^{(N)}$ and find agreement with Conjecture 1:

Theorem 3. *The fourth moment of the normalized matrix elements is given by*

$$\frac{1}{N} \sum_{j=1}^N |F_j^{(N)}|^4 \rightarrow \mathbf{E}(|X_f|^4) = 2 \sum_{\nu \neq 0} |f^\#(\nu)|^4$$

as $N \rightarrow \infty$ through primes.

In the case of *split* primes, that is primes N for which the cat map A is diagonalizable modulo N , the matrix elements are given by one-variable character sums (see Section 6.3) and one may hope to attack Conjecture 1 in that case via a monodromy argument as in [9].

Acknowledgements: We thank Peter Sarnak for discussions on his work with Wenzhi Luo [13].

2. BACKGROUND

The full details on the cat map and its quantization can be found in [11]. For the reader's convenience we briefly recall the setup:

2.1. Classical dynamics. The classical dynamics are given by a hyperbolic linear map $A \in SL(2, \mathbf{Z})$ so that $x = \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbf{T}^2 \mapsto Ax$ is a symplectic map of the torus. Given an observable $f \in C^\infty(\mathbf{T}^2)$, the classical evolution defined by A is $f \mapsto f \circ A$, where $(f \circ A)(x) = f(Ax)$.

2.2. Kinematics: The space of states. As the Hilbert space of states, we take distributions $\psi(q)$ on the line \mathbf{R} which are periodic in both the position and the momentum representation. This restricts h , Planck's constant, to take only inverse integer values. With $h = 1/N$, the space of states, denoted \mathcal{H}_N , is of dimension N and consists of periodic point-masses at the coordinates $q = Q/N$, $Q \in \mathbf{Z}$. We identify \mathcal{H}_N with $L^2(\mathbf{Z}/N\mathbf{Z})$, where the inner product $\langle \cdot, \cdot \rangle$ is given by

$$\langle \phi, \psi \rangle = \frac{1}{N} \sum_{Q \bmod N} \phi(Q) \bar{\psi}(Q).$$

2.3. Observables: The basic observables are given by the operators $T_N(n_1, n_2)$ acting on $\psi \in L^2(\mathbf{Z}/N\mathbf{Z})$ via:

$$(2.1) \quad (T_N(n_1, n_2)\psi)(Q) = e^{\frac{i\pi n_1 n_2}{N}} e\left(\frac{n_2 Q}{N}\right) \psi(Q + n_1).$$

where

$$e(x) = e^{2\pi i x}.$$

Note that

$$(2.2) \quad T_N(n + 2N) = T_N(n)$$

For any smooth classical observable $f \in C^\infty(\mathbf{T}^2)$ with Fourier expansion

$$f(x) = \sum_{n_1, n_2 \in \mathbf{Z}} \widehat{f}(n_1, n_2) e(n_1 p + n_2 q), \quad x = \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbf{T}^2,$$

its quantization, $\text{Op}_N(f)$, is given by

$$\text{Op}_N(f) := \sum_{n_1, n_2 \in \mathbf{Z}} \widehat{f}(n_1, n_2) T_N(n_1, n_2)$$

2.4. Dynamics: We let $\Gamma(4, 2N) \subset SL(2, \mathbf{Z})$ be the subgroup of matrices that are congruent to the identity matrix modulo 4 (resp., 2) if N is even (resp., odd). For $A \in \Gamma(4, 2N)$ we can assign unitary operators $U_N(A)$, acting on $L^2(\mathbf{Z}/N\mathbf{Z})$, having the following important properties:

- “Exact Egorov”: For all observables $f \in C^\infty(\mathbf{T}^2)$

$$U_N(A)^{-1} \text{Op}_N(f) U_N(A) = \text{Op}_N(f \circ A).$$

- The quantization depends only on A modulo $2N$: if $A, B \in \Gamma(4, 2N)$ and $A \equiv B \pmod{2N}$ then

$$U_N(A) = U_N(B)$$

- The quantization is multiplicative: if $A, B \in \Gamma(4, 2N)$, then

$$(2.3) \quad U_N(AB) = U_N(A)U_N(B)$$

2.5. Hecke eigenfunctions. Let α, α^{-1} be the eigenvalues of A . Since A is hyperbolic, α is a unit in the real quadratic field $K = \mathbf{Q}(\alpha)$. Define an order \mathfrak{D} of K by letting $\mathfrak{D} = \mathbf{Z}[\alpha]$. (Note that \mathfrak{D} is not necessarily equal to the full ring of integers in K .) Let $v = (v_1, v_2) \in \mathfrak{D}^2$ be a vector such that $vA = \alpha v$. Let $I := \mathbf{Z}[v_1, v_2] \subset \mathfrak{D}$. Then I is an \mathfrak{D} -ideal, and the matrix of α acting on I by multiplication in the basis v_1, v_2 is precisely A . The choice of basis of I gives an identification $I \cong \mathbf{Z}^2$ and the action of \mathfrak{D} on the ideal I by multiplication gives a ring homomorphism

$$\iota : \mathfrak{D} \rightarrow \text{Mat}_2(\mathbf{Z})$$

with the property that the determinant of $\iota(\beta)$, $\beta \in \mathfrak{D}$, is given by $\mathcal{N}(\beta)$, where $\mathcal{N} : \mathbf{Q}(\alpha) \rightarrow \mathbf{Q}$ is the norm map.

Reducing the norm map modulo $2N$ gives a well defined map

$$\mathcal{N}_{2N} : \mathfrak{D}/2N\mathfrak{D} \rightarrow \mathbf{Z}/2N\mathbf{Z},$$

and we let $C(2N)$ be the elements in the kernel of this map that are congruent to 1 modulo $4\mathfrak{D}$ (resp., $2\mathfrak{D}$) if N is even (resp., odd).

Now, reducing ι modulo $2N$ gives a map

$$\iota_{2N} : C(2N) \rightarrow SL_2(\mathbf{Z}/2N\mathbf{Z}).$$

Since $C(2N)$ is commutative, the properties in section 2.4 imply that

$$\{U_N(\iota_{2N}(\beta)) : \beta \in C\}$$

forms a family of commuting operators. Analogously with modular forms, we call these *Hecke operators*, and functions $\psi \in \mathcal{H}_N$ that are simultaneous eigenfunctions of all the Hecke operators are denoted *Hecke eigenfunctions*. Note that a Hecke eigenfunction is an eigenfunction of $U_N(\iota_{2N}(\alpha)) = U_N(A)$.

We note an invariance property of matrix elements, namely that they are invariant under the Hecke operators:

$$\langle \text{Op}_N(f)\psi_j, \psi_j \rangle = \langle \text{Op}_N(f \circ B)\psi_j, \psi_j \rangle, \quad B \in C(2N)$$

This follows from ψ_j being eigenfunctions of the Hecke operators $C(2N)$. In particular, taking $f(x) = e(nx)$ we see that

$$(2.4) \quad \langle T_N(n)\psi_j, \psi_j \rangle = \langle T_N(nB)\psi_j, \psi_j \rangle$$

Moreover, since $-I \in C(2N)$, we have

$$\overline{\langle T_N(n)\psi_j, \psi_j \rangle} = \langle \psi_j, T_N(n)\psi_j \rangle = \langle T_N(-n)\psi_j, \psi_j \rangle = \langle T_N(n)\psi_j, \psi_j \rangle,$$

and this implies that the matrix elements are real.

2.6. The quadratic form associated to A : We define a binary quadratic form associated to $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by

$$Q(x, y) = cx^2 + (d - a)xy - by^2$$

The rationale for it is as follows: Let $\alpha^{\pm 1} = \frac{a+d \pm \sqrt{(a+d)^2 - 4}}{2}$ be the eigenvalues of A and $\mathfrak{D} = \mathbf{Z}[\alpha]$ the order associated to A . Let $v = (v_1, v_2) \in \mathfrak{D}^2$ be an eigenvector for A with eigenvalue α : $vA = \alpha v$. We may take $v = (c, \alpha - a)$. Let $I := \mathbf{Z}[v_1, v_2] = \mathbf{Z}[c, \alpha - a] \subset \mathfrak{D}$. Then I is an \mathfrak{D} -ideal, and the matrix of α acting on I by multiplication in the basis v_1, v_2 is precisely A .

We now consider the quadratic form induced by the norm form on the ideal I . There is some leeway in its definition corresponding to changes of basis and multiplication by integers. One choice is to take

$$\frac{\mathcal{N}(xv_1 + yv_2)}{\mathcal{N}(I)}$$

where $\mathcal{N}(I) = \#\mathfrak{D}/I$. In our case, since $I = \mathbf{Z}[c, \alpha - a]$ and $\mathfrak{D} = \mathbf{Z}[1, \alpha]$ we have $\mathcal{N}(I) = |c|$. A computation shows that the quadratic form is then

$$\frac{1}{|c|} (c^2x^2 + c(d - a)xy - bcy^2) = \text{sign}(c) (cx^2 + (d - a)xy - by^2)$$

Up to sign, this is the quadratic form Q above.

By virtue of the definition of Q as a norm form, we see that A and the Hecke operators are isometries of Q , and since they have unit norm they actually land in the special orthogonal group of Q . That is we find that under the above identifications, $C(2N)$ is identified with $\{B \in SO(Q, \mathbf{Z}/2N\mathbf{Z}) : B \equiv I \pmod{2}\}$.

2.7. A rewriting of the matrix elements. We now show that when ψ is a Hecke eigenfunction, the matrix elements $\langle \text{Op}_N(f)\psi, \psi \rangle$ have a modified Fourier series expansion which incorporates some extra invariance properties.

Lemma 4. *If $m, n \in \mathbf{Z}^2$ are such that $Q(m) = Q(n)$, then for all sufficiently large primes N we have $m \equiv nB \pmod{N}$ for some $B \in SO(Q, \mathbf{Z}/N\mathbf{Z})$.*

Proof. We may clearly assume $Q(m) \neq 0$ because otherwise $m = n = 0$ since Q is anisotropic over the rationals. We take N a sufficiently large odd prime so that Q is non-degenerate over the field $\mathbf{Z}/N\mathbf{Z}$. If $N > |Q(m)|$ then $Q(m) \not\equiv 0 \pmod{N}$ and then the assertion reduces to the fact that if Q is a non-degenerate binary quadratic form over

the finite field $\mathbf{Z}/N\mathbf{Z}$ ($N \neq 2$ prime) then the special orthogonal group $SO(Q, \mathbf{Z}/N\mathbf{Z})$ acts transitively on the hyperbolas $\{Q(n) = \nu\}$, $\nu \neq 0 \pmod N$. This is easy to check since the quadratic form is either equivalent to the split form x_1x_2 or to the norm form of a quadratic extension of $\mathbf{Z}/N\mathbf{Z}$. \square

Lemma 5. *Fix $m, n \in \mathbf{Z}^2$ such that $Q(m) = Q(n)$. If N is a sufficiently large odd prime and ψ a Hecke eigenfunction, then*

$$(-1)^{n_1n_2} \langle T_N(n)\psi, \psi \rangle = (-1)^{m_1m_2} \langle T_N(m)\psi, \psi \rangle$$

Proof. For ease of notation, set

$$\epsilon(n) := (-1)^{n_1n_2}$$

By Lemma 4 it suffices to show that if $m \equiv nB \pmod N$ for some $B \in SO(Q, \mathbf{Z}/N\mathbf{Z})$ then $\epsilon(n) \langle T_N(n)\psi, \psi \rangle = \epsilon(m) \langle T_N(m)\psi, \psi \rangle$.

By the Chinese Remainder Theorem,

$$SO(Q, \mathbf{Z}/2N\mathbf{Z}) \simeq SO(Q, \mathbf{Z}/N\mathbf{Z}) \times SO(Q, \mathbf{Z}/2\mathbf{Z})$$

(recall N is odd) and so

$$C(2N) \simeq \{B \in SO(Q, \mathbf{Z}/2N\mathbf{Z}) : B \equiv I \pmod 2\} \simeq SO(Q, \mathbf{Z}/N\mathbf{Z}) \times \{I\}$$

Thus if $m \equiv nB \pmod N$ for $N \in SO(Q, \mathbf{Z}/N\mathbf{Z})$ then there is a unique $\tilde{B} \in C(2N)$ so that $m \equiv n\tilde{B} \pmod N$.

We note that $\epsilon(n)T_N(n)$ has period N , rather than merely $2N$ for $T_N(n)$ (see (2.2)). Then since $m = n\tilde{B} \pmod N$,

$$\epsilon(m)T_N(m) = \epsilon(n\tilde{B})T_N(n\tilde{B}) = \epsilon(n)T_N(n\tilde{B})$$

(recall that $\tilde{B} \in C(2N)$ preserves parity: $n\tilde{B} \equiv n \pmod 2$, so $\epsilon(n\tilde{B}) = \epsilon(n)$).

Thus for ψ a Hecke eigenfunction,

$$\epsilon(m) \langle T_N(m)\psi, \psi \rangle = \epsilon(n) \langle T_N(n\tilde{B})\psi, \psi \rangle = \epsilon(n) \langle T_N(n)\psi, \psi \rangle$$

the last equality by (2.4). \square

Define for $\nu \in \mathbf{Z}$

$$f^\#(\nu) := \sum_{n \in \mathbf{Z}^2: Q(n)=\nu} (-1)^{n_1n_2} \hat{f}(n)$$

and

$$(2.5) \quad V_\nu(\psi) := \sqrt{N} (-1)^{n_1n_2} \langle T_N(n)\psi, \psi \rangle$$

where $n \in \mathbf{Z}^2$ is a vector with $Q(n) = \nu$ (if it exists) and set $V_\nu(\psi) = 0$ otherwise. By Lemma 5 this is well-defined, that is independent of the choice of n . Then we have

Proposition 6. *If ψ is a Hecke eigenfunction, f a trigonometric polynomial, and $N \geq N_0(f)$, then*

$$\sqrt{N} \langle \text{Op}_N(f)\psi, \psi \rangle = \sum_{\nu \in \mathbf{Z}} f^\#(\nu) V_\nu(\psi)$$

To simplify the arguments, in what follows we will restrict ourself to dealing with observables that are trigonometric polynomials.

3. ERGODIC AVERAGING

We relate mixed moments of matrix coefficients to traces of certain averages of the observables: Let

$$(3.1) \quad D(n) = \frac{1}{|C(2N)|} \sum_{B \in C(2N)} T_N(nB)$$

The following shows that $D(n)$ is essentially diagonal when expressed in the Hecke eigenbasis.

Lemma 7. *Let \tilde{D} be the matrix obtained when expressing $D(n)$ in terms of the Hecke eigenbasis $\{\psi_i\}_{i=1}^N$. If N is inert in K , then \tilde{D} is diagonal. If N splits in K , then \tilde{D} has the form*

$$\tilde{D} = \begin{pmatrix} D_{11} & D_{12} & 0 & 0 & \dots & 0 \\ D_{21} & D_{22} & 0 & 0 & \dots & 0 \\ 0 & 0 & D_{33} & 0 & \dots & 0 \\ 0 & 0 & 0 & D_{44} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & D_{NN} \end{pmatrix}$$

where ψ_1, ψ_2 correspond to the quadratic character of $C(2N)$. Moreover, in the split case, we have

$$|D_{ij}| \ll N^{-1/2}$$

for $1 \leq i, j \leq 2$.

Proof. If N is inert, then the Weil representation is multiplicity free when restricted to $C(2N)$ (see Lemma 4 in [10].) If N is split, then $C(2N)$ is isomorphic to \mathbb{F}_N^\times (i.e., the invertible elements of \mathbb{F}_N , where \mathbb{F}_N is the finite field with N elements), and the trivial character occurs with multiplicity one, the quadratic character occurs with multiplicity two, and all other characters occur with multiplicity one (see [12], section 4.1.)

As for the bound on in the split case, assume first that $f(x, y) = e(\frac{n_1x+n_2y}{N})$ for some $n_1, n_2 \in \mathbf{Z}$, such that $n = (n_1, n_2)$ is not an eigenvector of A modulo N . We may give an explicit construction of the

Hecke eigenfunctions as follows (see [12], section 4 for more details): there exists $M \in SL_2(\mathbf{Z}/2N\mathbf{Z})$ such that the eigenfunctions ψ_1, ψ_2 can be written as

$$\psi_1 = \sqrt{N} \cdot U_N(M)\delta_0$$

and

$$\psi_2 = \sqrt{\frac{N}{N-1}} \cdot U_N(M)(1 - \delta_0)$$

where $\delta_0(x) = 1$ if $x \equiv 0 \pmod{N}$, and $\delta_0(x) = 0$ otherwise. Now,

$$D_{ij} = \langle T_N((n_1, n_2))\psi_i, \psi_j \rangle$$

and if we let $\phi_1 = \sqrt{N}\delta_0$ and $\phi_2 = \sqrt{\frac{N}{N-1}}(1 - \delta_0)$, exact Egorov gives

$$\langle T_N((n_1, n_2))\psi_i, \psi_j \rangle = \langle T_N((n'_1, n'_2))\phi_i, \phi_j \rangle$$

where $(n'_1, n'_2) \equiv (n_1, n_2)M \pmod{N}$. Since n is assumed not to be an eigenvector of A , we have $n'_1 \not\equiv 0 \pmod{N}$ and $n'_2 \not\equiv 0 \pmod{N}$. Hence

$$\begin{aligned} D_{11} &= \langle T_N((n'_1, n'_2))\phi_1, \phi_1 \rangle = \frac{N}{N} \sum_{x=1}^N (T_N((n'_1, n'_2))\delta_0)(x)\delta_0(x) \\ &= e\left(\frac{n'_1 n'_2}{2N}\right)\delta_0(0 + n'_1) = 0 \end{aligned}$$

since $n'_1 \not\equiv 0 \pmod{N}$. Similarly,

$$\begin{aligned} D_{22} &= \langle T_N((n'_1, n'_2))\phi_2, \phi_2 \rangle \\ &= \frac{N}{N-1} \frac{1}{N} e\left(\frac{n'_1 n'_2}{2N}\right) \sum_{x=1}^N e\left(\frac{n'_2 x}{N}\right) (1 - \delta_0)(x + n'_1) (1 - \delta_0)(x) \\ &= \frac{1}{N-1} e\left(\frac{n'_1 n'_2}{2N}\right) \sum_{\substack{1 \leq x \leq N-1 \\ x \neq -n'_1}} e\left(\frac{n'_2 x}{N}\right) \end{aligned}$$

which is $O(1/N)$ since $n'_2 \not\equiv 0 \pmod{N}$. Finally,

$$\begin{aligned} D_{21} &= \langle T_N((n'_1, n'_2))\phi_2, \phi_1 \rangle = \\ &= \frac{N}{\sqrt{N-1}} \frac{1}{N} \sum_{x=1}^N (T_N((n'_1, n'_2))(1 - \delta_0))(x)\delta_0(x) = \\ &= \frac{1}{\sqrt{N-1}} e\left(\frac{n'_1 n'_2}{2N}\right) e\left(\frac{n'_2 \cdot 0}{N}\right) (1 - \delta_0)(0 + n'_1) = O\left(\frac{1}{\sqrt{N-1}}\right), \end{aligned}$$

and a similar argument shows that $D_{21} = O(N^{-1/2})$. \square

Remark: In the split case, it is still true that $D_{ij} \ll N^{-1/2}$ for all i, j , but this requires the Riemann hypothesis for curves, whereas the above is elementary.

Lemma 8. *Let $\{\psi_i\}_{i=1}^N$ be a Hecke basis of \mathcal{H}_N , and let $k, l, m, n \in \mathbf{Z}^2$. Then*

$$\sum_{i=1}^N \langle T_N(m)\psi_i, \psi_i \rangle \overline{\langle T_N(n)\psi_i, \psi_i \rangle} = \text{tr}(D(m)D^*(n)) + O(N^{-1})$$

Moreover,

$$\begin{aligned} \sum_{i=1}^N \langle T_N(k)\psi_i, \psi_i \rangle \overline{\langle T_N(l)\psi_i, \psi_i \rangle} \langle T_N(m)\psi_i, \psi_i \rangle \overline{\langle T_N(n)\psi_i, \psi_i \rangle} \\ = \text{tr}(D(k)D^*(l)D(m)D^*(n)) + O(N^{-2}) \end{aligned}$$

Proof. By definition

$$\sum_{i=1}^N \langle T_N(m)\psi_i, \psi_i \rangle \overline{\langle T_N(n)\psi_i, \psi_i \rangle} = \sum_{i=1}^N D(m)_{ii} \overline{D(n)_{ii}}$$

On the other hand, by lemma 7,

$$\text{tr}(D(m)D(n)^*) = D_{12}(m)\overline{D_{21}(n)} + D_{21}(m)\overline{D_{12}(n)} + \sum_{i=1}^N D_{ii}(m)\overline{D_{ii}(n)}$$

where $D_{12}(m), D_{21}(m), D_{12}(n)$ and $D_{21}(n)$ are all $O(N^{-1/2})$. Thus

$$\sum_{i=1}^N \langle T_N(m)\psi_i, \psi_i \rangle \overline{\langle T_N(n)\psi_i, \psi_i \rangle} = \text{tr}(D(m)D(n)^*) + O(N^{-1})$$

The proof of the second assertion is similar. \square

4. PROOF OF THEOREM 2

In order to prove Theorem 2 it suffices, by Proposition 6, to show that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N V_\nu(\psi_j) \overline{V_\mu(\psi_j)} = \mathbf{E}(\text{tr}(U_\nu) \text{tr}(U_\mu)) = \begin{cases} 1 & \text{if } \mu = \nu, \\ 0 & \text{if } \mu \neq \nu, \end{cases}$$

where $U_\mu, U_\nu \in SU_2$ are random matrices in SU_2 that are independent if $\nu \neq \mu$.

To proceed we will need to evaluate the trace of $T_N(nB_1)T_N(mB_2)^*$.

Lemma 9. *If N is odd and $B_1, B_2 \in C(2N)$ then*

$$\mathrm{tr}(T_N(nB_1)T_N(mB_2)^*) = \begin{cases} (-1)^{m_1m_2+n_1n_2}N & \text{if } nB_1 \equiv mB_2 \pmod{N}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Recall from [11, section 2.3] that

$$(4.1) \quad T_N(mB_2)^* = T_N(-mB_2),$$

$$(4.2) \quad T_N(nB_1)T_N(-mB_2) = e(\omega(nB_1, -mB_2)/2N)T_N(nB_1 - mB_2)$$

where $\omega(x, y) = x_1y_2 - x_2y_1$, and that

$$(4.3) \quad \mathrm{tr}(T_N(x)) = \begin{cases} 0 & \text{if } x \not\equiv (0, 0) \pmod{N}, \\ e(\frac{-x_1x_2}{2})N & \text{if } x \equiv (0, 0) \pmod{N}. \end{cases}$$

(Note that $e(\frac{-x_1x_2}{2N}) = e(\frac{-x_1x_2}{2})$ if $x \equiv (0, 0) \pmod{N}$.) Since $B_1 \equiv B_2 \equiv I \pmod{2}$ and $nB_1 \equiv mB_2 \pmod{N}$, we find that

$$e\left(\frac{\omega(nB_1, -mB_2)}{2N}\right) = e\left(\frac{\omega(n, -m)}{2}\right) = e\left(\frac{n_2m_1 - n_1m_2}{2}\right)$$

and

$$\mathrm{tr}(T_N(nB_1 - mB_2)) = e\left(\frac{-(n_1 - m_1)(n_2 - m_2)}{2}\right)N.$$

Thus

$$\begin{aligned} \mathrm{tr}(T_N(nB_1)T_N(mB_2)^*) &= e\left(\frac{n_2m_1 - n_1m_2 - (n_1 - m_1)(n_2 - m_2)}{2}\right)N = \\ &= e\left(\frac{m_1m_2 - n_1n_2}{2}\right)N = (-1)^{m_1m_2 - n_1n_2}N = (-1)^{m_1m_2 + n_1n_2}N \end{aligned}$$

□

Proposition 10. *Let $\{\psi_i\}_{i=1}^N$ be a Hecke basis of \mathcal{H}_N . If $N \geq N_0(\mu, \nu)$ is prime and $\mu, \nu \not\equiv 0 \pmod{N}$, then*

$$\frac{1}{N} \sum_{j=1}^N V_\nu(\psi_j) \overline{V_\mu(\psi_j)} = \begin{cases} 1 + O(N^{-1}) & \text{if } \mu = \nu, \\ O(N^{-1}) & \text{otherwise.} \end{cases}$$

Proof. Choose $m, n \in \mathbf{Z}^2$ such that $Q(m) = \mu$ and $Q(n) = \nu$. By (2.5) and Lemma 8 we find that

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N V_\nu(\psi_j) \overline{V_\mu(\psi_j)} &= (-1)^{m_1m_2 + n_1n_2} \sum_{j=1}^N \langle T_N(n)\psi_j, \psi_j \rangle \overline{\langle T_N(m)\psi_j, \psi_j \rangle} \\ &= (-1)^{m_1m_2 + n_1n_2} \mathrm{tr}(D(n)D(m)^*) + O(N^{-1}) \end{aligned}$$

Now,

$$D(n)D(m)^* = \frac{1}{|C(2N)|^2} \sum_{B_1, B_2 \in C(2N)} T_N(nB_1)T_N(mB_2)^*$$

Taking the trace of both sides and applying Lemma 9, we get

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N V_\nu(\psi_j) \overline{V_\mu(\psi_j)} &= \\ &= \frac{(-1)^{m_1 m_2 + n_1 n_2}}{|C(2N)|^2} \sum_{\substack{B_1, B_2 \in C(2N) \\ nB_1 \equiv mB_2 \pmod{N}}} (-1)^{m_1 m_2 + n_1 n_2} N + O(N^{-1}) \\ &= \frac{N}{|C(2N)|} \cdot |\{B \in C(2N) : n \equiv mB \pmod{N}\}| + O(N^{-1}) \end{aligned}$$

which, since $|C(2N)| = N \pm 1$, equals

$$= \begin{cases} 1 + O(N^{-1}) & \text{if there exists } B \in C(2N) \text{ such that } n \equiv mB \pmod{N}, \\ O(N^{-1}) & \text{otherwise.} \end{cases}$$

Finally, for N large enough (i.e., $N \geq N_0(\mu, \nu)$), Lemma 4 gives that $n \equiv mB \pmod{N}$ for some $B \in C(2N)$ is equivalent to $\mu = \nu$. \square

5. PROOF OF THEOREM 3

5.1. Reduction. In order to prove Theorem 3 it suffices to show that

$$\begin{aligned} (5.1) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N V_\kappa(\psi_j) \overline{V_\lambda(\psi_j)} V_\mu(\psi_j) \overline{V_\nu(\psi_j)} &= \\ &= \mathbf{E}(\operatorname{tr}(U_\kappa) \operatorname{tr}(U_\lambda) \operatorname{tr}(U_\mu) \operatorname{tr}(U_\nu)) \end{aligned}$$

where $U_\kappa, U_\lambda, U_\mu$ and U_ν are random matrices in SU_2 .

Let $S \subset \mathbf{Z}^4$ be the set of four-tuples $(\kappa, \lambda, \mu, \nu)$ such that $\kappa = \lambda, \mu = \nu$, or $\kappa = \mu, \lambda = \nu$, or $\kappa = \nu, \lambda = \mu$, but **not** $\kappa = \lambda = \mu = \nu$.

Proposition 11. *Let $\{\psi_i\}_{i=1}^N$ be a Hecke basis of \mathcal{H}_N and let $\kappa, \lambda, \mu, \nu \in \mathbf{Z}$. If N is a sufficiently large prime, then*

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N V_\kappa(\psi_j) \overline{V_\lambda(\psi_j)} V_\mu(\psi_j) \overline{V_\nu(\psi_j)} &= \\ &= \begin{cases} 2 + O(N^{-1}) & \text{if } \kappa = \lambda = \mu = \nu, \\ 1 + O(N^{-1}) & \text{if } (\kappa, \lambda, \mu, \nu) \in S, \\ O(N^{-1/2}) & \text{otherwise.} \end{cases} \end{aligned}$$

Given Proposition 11 it is straightforward to deduce (5.1), we need only to note that

$$\mathbf{E}((\operatorname{tr} U)^4) = 2, \quad \mathbf{E}((\operatorname{tr} U)^2) = 1, \quad \mathbf{E}(\operatorname{tr} U) = 0.$$

Since the proof of Proposition 11 will occupy the remainder of this section, we give a brief outline of the proof for the convenience of the reader:

- (1) Express the left hand side of (5.1) as the trace of averaged observables.
- (2) Rewrite the trace as an exponential sum.
- (3) Show that the exponential sum is quite small unless pairwise equality of $\kappa, \lambda, \mu, \nu$ occurs, in which case the exponential sum is given by the number of solutions (modulo N) of a certain equation.
- (4) Determine the number of solutions.

5.2. Ergodic averaging. Choose $k, l, m, n \in \mathbf{Z}^2$ such that $Q(k) = \kappa, Q(l) = \lambda, Q(m) = \mu, \text{ and } Q(n) = \nu$. Then

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N V_\kappa(\psi_j) \overline{V_\lambda(\psi_j)} V_\mu(\psi_j) \overline{V_\nu(\psi_j)} &= \\ &= (-1)^{k_1 k_2 + l_1 l_2 + m_1 m_2 + n_1 n_2} N \cdot \\ &\quad \cdot \sum_{j=1}^N \langle T_N(k) \psi_j, \psi_j \rangle \overline{\langle T_N(l) \psi_j, \psi_j \rangle} \langle T_N(m) \psi_j, \psi_j \rangle \overline{\langle T_N(n) \psi_j, \psi_j \rangle} \end{aligned}$$

which, by Lemma 8, equals

$$(-1)^{k_1 k_2 + l_1 l_2 + m_1 m_2 + n_1 n_2} N \operatorname{tr} (D(k) D(l)^* D(m) D(n)^*) + O(N^{-1})$$

Now,

$$\begin{aligned} & D(k)D(l)^*D(m)D(n)^* = \\ &= \frac{1}{|C(2N)|^4} \sum_{B_1, B_2, B_3, B_4 \in C(2N)} T_N(kB_1)T_N(lB_2)^*T_N(mB_3)T_N(nB_4)^* \end{aligned}$$

and in order to evaluate the trace we will need the following four variable analogue of Lemma 9:

Lemma 12. *If N is odd, $B_1, B_2, B_3, B_4 \in C(2N)$ and $kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$, then*

$$\begin{aligned} (5.2) \quad & \text{tr}(T_N(kB_1)T_N(lB_2)^*T_N(mB_3)T_N(nB_4)^*) = \\ &= (-1)^{k_1k_2+l_1l_2+m_1m_2+n_1n_2} e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right) N \end{aligned}$$

where $2t \equiv 1 \pmod{N}$.

On the other hand, if $kB_1 - lB_2 + mB_3 - nB_4 \not\equiv 0 \pmod{N}$, then

$$\text{tr}(T_N(kB_1)T_N(lB_2)^*T_N(mB_3)T_N(nB_4)^*) = 0$$

Proof. By (4.1) and (4.2) we have

$$\begin{aligned} (5.3) \quad & T_N(kB_1)T_N(lB_2)^*T_N(mB_3)T_N(nB_4)^* = \\ &= T_N(kB_1)T_N(-lB_2)T_N(mB_3)T_N(-nB_4) = \\ &= e\left(\frac{\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4)}{2N}\right) T_N(kB_1 - lB_2)T_N(mB_3 - nB_4) = \\ &= e\left(\frac{\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4) + \omega(kB_1 - lB_2, mB_3, -nB_4)}{2N}\right) \\ &\quad \cdot T_N(kB_1 - lB_2 + mB_3 - nB_4) \end{aligned}$$

By (4.3), $\text{tr}(T_N(kB_1 - lB_2 + mB_3 - nB_4)) = 0$ unless $kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$, hence the second assertion follows.

As for the first assertion, assume that $kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$. Then $\omega(kB_1 - lB_2, mB_3 - nB_4) \equiv 0 \pmod{N}$, and since $B_1 \equiv B_2 \equiv B_3 \equiv B_4 \equiv I \pmod{2}$, we have

$$e\left(\frac{\omega(kB_1 - lB_2, mB_3 - nB_4)}{2N}\right) = e\left(\frac{\omega(k - l, m - n)}{2}\right).$$

This, together with (4.3) gives

$$(5.4) \quad \begin{aligned} & \text{tr} (T_N(kB_1)T_N(-lB_2)T_N(mB_3)T_N(-nB_4)) = \\ & = (-1)^{(k_1-l_1+m_1-n_1)(k_2-l_2+m_2-n_2)} \cdot e \left(\frac{\omega(k-l, m-n)}{2} + \frac{\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4)}{2N} \right) N \end{aligned}$$

Since $\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4) \equiv \omega(k, -l) + \omega(m, -n) \pmod{2}$, the Chinese Remainder Theorem gives

$$(5.5) \quad \begin{aligned} & e \left(\frac{\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4)}{2N} \right) = \\ & = e \left(\frac{\omega(k, -l) + \omega(m, -n)}{2} \right) \cdot e \left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N} \right) \end{aligned}$$

where $2t \equiv 1 \pmod{N}$. The result now follows since

$$e \left(\frac{\omega(k-l, m-n)}{2} \right) = (-1)^{(k_1-l_1)(m_2-n_2)-(k_2-l_2)(m_1-n_1)}$$

and

$$e \left(\frac{\omega(k, -l) + \omega(m, -n)}{2} \right) = (-1)^{k_1l_2-k_2l_1+m_1n_2-m_2n_1}$$

and thus the sign of (5.4) is given by

$$(5.6) \quad \begin{aligned} & (-1)^{(k_1-l_1+m_1-n_1)(k_2-l_2+m_2-n_2)} e \left(\frac{\omega(k-l, m-n)}{2} + \frac{\omega(k, -l) + \omega(m, -n)}{2} \right) = \\ & = (-1)^{(k_1-l_1+m_1-n_1)(k_2-l_2+m_2-n_2)+(k_1-l_1)(m_2-n_2)-(k_2-l_2)(m_1-n_1)+k_1l_2-k_2l_1+m_1n_2-m_2n_1} = \\ & = (-1)^{k_1k_2+l_1l_2+m_1m_2+n_1n_2} \end{aligned}$$

□

Thus, using Lemma 12 we obtain

$$\begin{aligned}
(5.7) \quad & \frac{1}{N} \sum_{j=1}^N V_\kappa(\psi_j) \overline{V_\lambda(\psi_j)} V_\mu(\psi_j) \overline{V_\nu(\psi_j)} = \\
& = (-1)^{k_1 k_2 + l_1 l_2 + m_1 m_2 + n_1 n_2} \frac{N}{|C(2N)|^4} \cdot \\
& \quad \cdot \sum_{B_1, B_2, B_3, B_4 \in C(2N)} \operatorname{tr} (T_N(kB_1) T_N(lB_2)^* T_N(mB_3) T_N(nB_4)^*) = \\
& = \frac{N^2}{|C(2N)|^4} \cdot \\
& \quad \cdot \sum_{\substack{B_1, B_2, B_3, B_4 \in C(N) \\ kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}}} e \left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N} \right)
\end{aligned}$$

(Note that $e \left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N} \right)$ only depends on B_1, B_2, B_3, B_4 modulo N , and since $|C(N)| = |C(2N)|$ we may sum over $B_i \in C(N)$ instead of $B_i \in C(2N)$.)

5.3. Exponential sums over curves. In order to show that there is quite a bit of cancellation in (5.7) when pairwise equality of norms do not hold, we will need some results on exponential sums over curves. Let X be a projective curve of degree d_1 defined over the finite field \mathbb{F}_p , embedded in n -dimensional projective space \mathbb{P}^n over \mathbb{F}_p . Further, let $R(X_1, \dots, X_{n+1})$ be a homogeneous rational function in \mathbb{P}^n , defined over \mathbb{F}_p , and let d_2 be the degree of its numerator. Define

$$S_m(R, X) = \sum_{x \in X(\mathbb{F}_{p^m})} e \left(\frac{\sigma(R(x))}{p} \right)$$

where σ is the trace from \mathbb{F}_{p^m} to \mathbb{F}_p , and the accent in the summation means that the poles of $R(x)$ are excluded. Bombieri has proved that the following bound on $|S_m(R, X)|$ holds.

Theorem 13 ([1], Theorem 6). *If $d_1 d_2 < p$ and R is not constant on any component Γ of X then*

$$|S_m(R, X)| \leq (d_1^2 + 2d_1 d_2 - 3d_1) p^{m/2} + d_1^2$$

In order to apply Bombieri's Theorem we need to show that the components of a certain algebraic set are at most one dimensional, and in order to do this we show that the number of points defined over \mathbb{F}_N is $O(N)$. (Such a bound can not hold for all N if there are components of dimension two or higher.)

Lemma 14. *Let $a, b \in \mathbb{F}_N[\alpha]$. If $a \neq 0$ and the equation*

$$\gamma_1 = a\gamma_2 + b, \quad \gamma_1, \gamma_2 \in C(N)$$

is satisfied for more than two values of γ_2 , then $b = 0$ and $\mathcal{N}(a) = 1$.

Proof. Taking norms, we obtain $1 = \mathcal{N}(a) + \mathcal{N}(b) + \text{tr}(\bar{a}b\gamma_2)$ and hence $\text{tr}(\bar{a}b\gamma_2)$ is constant. If $\bar{a}b \neq 0$, this means that the coordinates (x, y) of γ_2 , when regarding γ_2 as an element of \mathbb{F}_N^2 , lies on some line. On the other hand, $\mathcal{N}(\gamma_2) = 1$ corresponds to γ_2 satisfying some quadratic equation, hence the intersection can be at most two points. (In fact, we may identify $C(N)$ with the solutions to $x^2 - Dy^2 = 1$ for $x, y \in \mathbb{F}_N$, and some fixed $D \in \mathbb{F}_N$.) □

Lemma 15. *Fix $k, l, m, n \in \mathbf{Z}^2$ and let X be the set of solutions to*

$$k - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}, \quad B_2, B_3, B_4 \in C(N)$$

If $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \pmod{N}$, then $|X| \leq 3(N+1)$ for N sufficiently large.

Proof. We use the identification of the action of $C(N)$ on \mathbb{F}_N^2 with the action of $C(N)$ on $\mathbb{F}_N[\alpha]$. The equation

$$k - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$$

is then equivalent to

$$\kappa - \lambda\beta_2 + \mu\beta_3 - \nu\beta_4 = 0$$

where $\beta_i \in C(N)$ and $\kappa, \lambda, \mu, \nu \in \mathbb{F}_N[\alpha]$. We may rewrite this as

$$\kappa - \lambda\beta_2 = \nu\beta_4 - \mu\beta_3 = \beta_4(\nu - \mu\beta_3/\beta_4)$$

and letting $\beta' = \beta_3/\beta_4$, we obtain

$$\kappa - \lambda\beta_2 = \beta_4(\nu - \mu\beta')$$

If $\nu - \mu\beta' = 0$ then $\kappa - \lambda\beta_2 = 0$, and since $Q(l), Q(m) \not\equiv 0 \pmod{N}$ implies that λ, μ are nonzero³, we find that β_2 and β' are uniquely determined, whereas β_4 can be chosen arbitrarily. Thus there are at most $|C(N)|$ solutions for which $\nu - \mu\beta' = 0$.

Let us now bound the number of solutions when $\nu - \mu\beta' \neq 0$: after writing

$$\kappa - \lambda\beta_2 = \beta_4(\nu - \mu\beta')$$

as

$$\frac{\kappa}{\nu - \mu\beta'} + \frac{-\lambda}{\nu - \mu\beta'}\beta_2 = \beta_4,$$

³Recall that Q , up to a scalar multiple, is given by the norm.

Lemma 14 gives that there can be at most two possible values of β_2, β_4 for each β' , and hence there are at most $2|C(N)|$ solutions for which $\nu - \mu\beta' \neq 0$.

Thus, in total, X can have at most $|C(N)| + 2|C(N)| \leq 3(N + 1)$ solutions. \square

5.4. Counting solutions. We now determine the components of X on which $e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right)$ is constant.

Lemma 16. *Assume that $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \pmod{N}$, and let $\text{Sol}(k, l, m, n)$ be the number of solutions to the equations*

$$(5.8) \quad kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$$

$$(5.9) \quad \omega(kB_1, -lB_2) + \omega(mB_3, -nB_4) \equiv -C \pmod{N}$$

where $B_i \in C(N)$. If $C \equiv 0 \pmod{N}$ and N is sufficiently large, then

$$(5.10) \quad \text{Sol}(k, l, m, n) = \begin{cases} 2|C(N)|^2 & \text{if } Q(k) = Q(l) = Q(m) = Q(n), \\ |C(N)|^2 + O(|C(N)|) & \text{if } (Q(k), Q(l), Q(m), Q(n)) \in S, \\ O(|C(N)|) & \text{otherwise.} \end{cases}$$

On the other hand, if $C \not\equiv 0 \pmod{N}$ then

$$\text{Sol}(k, l, m, n) = O(|C(N)|).$$

Proof. For simplicity⁴, we will assume that N is inert. It will be convenient to use the language of algebraic number theory; we identify $(\mathbf{Z}/N\mathbf{Z})^2$ with the finite field $\mathbb{F}_{N^2} = \mathbb{F}_N(\sqrt{D})$ by letting $m = (x, y)$ correspond to $\mu = x + y\sqrt{D}$. First we note that if $n = (z, w)$ corresponds to ν then

$$\begin{aligned} \omega(m, n) &= xw - zy = \text{Im}((x - y\sqrt{D})(z + w\sqrt{D})) = \\ &= \overline{\text{Im}((x + y\sqrt{D})(z + w\sqrt{D}))} \end{aligned}$$

where $\text{Im}(a + b\sqrt{D}) = b$, and hence $\omega(m, n) = \text{Im}(\overline{\mu\nu})$.

Thus, with (k, l, m, n) corresponding to $(\nu_1, \nu_2, \nu_3, \nu_4)$, the values of $Q(k), Q(l), Q(m), Q(n)$ modulo N are (up to a scalar multiple) given by $\mathcal{N}(\nu_1), \mathcal{N}(\nu_2), \mathcal{N}(\nu_3), \mathcal{N}(\nu_4)$. Putting $\mu_i = \nu_i\beta_i$ for $\beta_i \in C(N)$, we find that $\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4) = -C$ can be written as

$$\text{Im}(\overline{\mu_1}\mu_2 + \overline{\mu_3}\mu_4) = C.$$

⁴The split case is similar except for possibility of zero divisors, but these do not occur when k, l, m, n are fixed and N is large enough.

Now, $kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}$ is equivalent to $\mu_1 - \mu_2 = \mu_4 - \mu_3$. Taking norms, we obtain

$$\mathcal{N}(\mu_1) + \mathcal{N}(\mu_2) - \text{tr}(\overline{\mu_1}\mu_2) = \mathcal{N}(\mu_4) + \mathcal{N}(\mu_3) - \text{tr}(\overline{\mu_4}\mu_3)$$

and hence

$$\text{tr}(\overline{\mu_4}\mu_3) = \text{tr}(\overline{\mu_1}\mu_2) + N_4 + N_3 - N_1 - N_2$$

if we let $N_i = \mathcal{N}(\nu_i)$. Since $\text{tr}(\mu) = 2\text{Re}(\mu) = 2\text{Re}(\overline{\mu})$, we find that

$$2\text{Re}(\overline{\mu_3}\mu_4) = 2\text{Re}(\mu_1\overline{\mu_2}) + N_4 + N_3 - N_1 - N_2$$

On the other hand, $\text{Im}(\overline{\mu_1}\mu_2 + \overline{\mu_3}\mu_4) = C$ implies that

$$\text{Im}(\overline{\mu_3}\mu_4) = -\text{Im}(\overline{\mu_1}\mu_2) + C = \text{Im}(\mu_1\overline{\mu_2}) + C$$

and thus

$$\overline{\mu_3}\mu_4 = \mu_1\overline{\mu_2} + K$$

where $K = (N_4 + N_3 - N_1 - N_2)/2 + C\sqrt{D}$. Hence we can rewrite (5.8) and (5.9) as

$$\begin{cases} \overline{\mu_3}\mu_4 = \mu_1\overline{\mu_2} + K \\ \mu_1 + \mu_3 = \mu_2 + \mu_4 \\ \mu_i = \nu_i\beta_i, \beta_i \in C(N) \text{ for } i = 1, 2, 3, 4. \end{cases}$$

Case 1 ($K \neq 0$). Since $\mu_i = \nu_i\beta_i$ with $\beta_i \in C(N)$, we can rewrite

$$\overline{\mu_3}\mu_4 = \mu_1\overline{\mu_2} + K$$

as

$$\overline{\nu_3}\nu_4\beta_4/\beta_3 = \nu_1\overline{\nu_2}\beta_1/\beta_2 + K,$$

and hence

$$\beta_4/\beta_3 = \frac{1}{\overline{\nu_3}\nu_4}(\nu_1\overline{\nu_2}\beta_1/\beta_2 + K).$$

Applying lemma 14 with $\gamma_1 = \beta_4/\beta_3$ and $\gamma_2 = \beta_1/\beta_2$ gives that β_1/β_2 , and hence $\mu_1\overline{\mu_2}$, must take one of two values, say C_1 or C_2 . But $\mu_1\overline{\mu_2} = C_1$ implies that $\mu_1 = \mu_2\frac{C_1}{N_2}$ and hence $\mu_4 = \mu_3\frac{C_1+K}{N_3}$. We thus obtain

$$\mu_2\left(1 - \frac{C_1}{N_2}\right) = \mu_1 - \mu_2 = \mu_4 - \mu_3 = \mu_3\left(1 - \frac{C_1+K}{N_3}\right)$$

Now, if $\mu_1 \neq \mu_2$ then both $1 - \frac{C_1}{N_2}$ and $1 - \frac{C_1+K}{N_3}$ are nonzero. Thus μ_2 is determined by μ_3 , which in turn gives that μ_1 as well as μ_4 is determined by μ_3 . Hence, there can be at most $C(N)$ solutions for which $\mu_1 \neq \mu_2$. (The case $\mu_1\overline{\mu_2} = C_2$ is handled in the same way.)

On the other hand, for $\mu_1 = \mu_2$ we have the family of solutions

$$(5.11) \quad \mu_1 = \mu_2, \quad \mu_4 = \mu_3$$

(note that this implies that $C = \text{Im}(\overline{\mu_1}\mu_2 + \overline{\mu_3}\mu_4) = 0$.)

Case 2 ($K = 0$). Since $K = 0$ and $\mu_1 = \mu_2 + \mu_4 - \mu_3$ we have

$$\overline{\mu_3}\mu_4 = \mu_1\overline{\mu_2} + K = (\mu_2 + \mu_4 - \mu_3)\overline{\mu_2}$$

and hence

$$\mu_4(\overline{\mu_3} - \overline{\mu_2}) = (\mu_2 - \mu_3)\overline{\mu_2}$$

If $\mu_2 - \mu_3 = 0$, we must have $\mu_1 = \mu_4$, and we obtain the family of solutions

$$(5.12) \quad \mu_2 = \mu_3, \quad \mu_1 = \mu_4$$

On the other hand, if $\mu_2 - \mu_3 \neq 0$, we can express μ_4 in terms of μ_2 and μ_3 :

$$\mu_4 = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\overline{\mu_2} = \frac{N_2 - \overline{\mu_2}\mu_3}{N_3 - \overline{\mu_2}\mu_3}\mu_3,$$

which in turn gives that

$$(5.13) \quad \begin{aligned} \mu_1 &= \mu_2 + \mu_4 - \mu_3 = \mu_2 + \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\overline{\mu_2} - \mu_3 \\ &= \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}(\overline{\mu_3} - \overline{\mu_2}) + \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\overline{\mu_2} = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\mu_3 = \frac{\mu_2\overline{\mu_3} - N_3}{\mu_2\overline{\mu_3} - N_2}\mu_2 \end{aligned}$$

Summary. If $K \neq 0$ there can be at most $2|C(N)|$ “spurious” solutions for which $\mu_1 \neq \mu_2$; other than that, we must have

$$\mu_1 = \mu_2, \quad \mu_3 = \mu_4.$$

On the other hand, if $K = 0$, then either

$$\mu_2 = \mu_3, \quad \mu_1 = \mu_4.$$

or

$$\mu_4 = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\overline{\mu_2} = \frac{N_2 - \overline{\mu_2}\mu_3}{N_3 - \overline{\mu_2}\mu_3}\mu_3, \quad \mu_1 = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2}\mu_3 = \frac{\mu_2\overline{\mu_3} - N_3}{\mu_2\overline{\mu_3} - N_2}\mu_2$$

We note that the first case can only happen if $N_1 = N_2$ and $N_3 = N_4$, the second only if $N_2 = N_3$ and $N_1 = N_4$, and the third only if $N_2 = N_4$ and $N_1 = N_3$. Moreover, in all three cases, $C = \text{Im}(K) = \text{Im}(\overline{\mu_1}\mu_2 + \overline{\mu_3}\mu_4) = 0$. We also note that if $N_2 = N_3$, then the third case simplifies to $\mu_1 = \mu_2$ and $\mu_3 = \mu_4$. We thus obtain the following:

If $C \neq 0$ then $K \neq 0$ and there can be at most $O(N)$ “spurious solutions”.

If $C = 0$ and $N_1 = N_2 = N_3 = N_4$ then $K = 0$ and the solutions are given by the two families

$$\mu_2 = \mu_3, \quad \mu_1 = \mu_4$$

and

$$\mu_4 = \frac{N_2 - \overline{\mu_2}\mu_3}{N_3 - \overline{\mu_2}\mu_3}\mu_3 = \mu_3, \quad \mu_1 = \frac{\mu_2\overline{\mu_3} - N_3}{\mu_2\overline{\mu_3} - N_2}\mu_2 = \mu_2$$

If $C = 0$ and $N_1 = N_4 \neq N_2 = N_3$ then $K = 0$ and there is a family of solutions given by

$$\mu_2 = \mu_3, \quad \mu_1 = \mu_4.$$

Similarly, if $C = 0$ and $N_1 = N_3 \neq N_2 = N_4$ then $K = 0$ and there is a family of solutions given by

$$\mu_4 = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2} \mu_2, \quad \mu_1 = \frac{\mu_2 - \mu_3}{\mu_3 - \mu_2} \mu_3$$

If $C = 0$ and $N_1 = N_2 \neq N_3 = N_4$ then $K \neq 0$, in which case we have a family of solutions given by

$$\mu_1 = \mu_2, \quad \mu_3 = \mu_4$$

as well as $O(N)$ “spurious” solutions.

Finally, if $C = 0$ and pairwise equality of norms *do not* hold, then we must have $K \neq 0$ (if $K = 0$ then $\overline{\mu_3} \mu_4 = \mu_1 \overline{\mu_2} + K$ implies that $N_3 N_4 = N_1 N_2$, which together with $N_1 + N_2 = N_3 + N_4$ gives that either $N_1 = N_3, N_2 = N_4$ or $N_1 = N_4, N_2 = N_3$) and in this case there can be at most $O(N)$ “spurious” solutions.

Finally, Lemma 4 gives (for k, l, m, n fixed and N large enough) that pairwise equality of norms modulo N implies pairwise equality of $Q(k), Q(l), Q(m), Q(n)$. \square

5.5. Conclusion. We may now evaluate the exponential sum in (5.7)

Proposition 17. *If $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \pmod{N}$ then, for N sufficiently large, we have*

$$(5.14) \quad \sum_{\substack{B_1, B_2, B_3, B_4 \in C(N) \\ kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}}} e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right) \\ = \begin{cases} 2|C(N)|^2 + O(|C(N)|) & \text{if } Q(k) = Q(l) = Q(m) = Q(n), \\ |C(N)|^2 + O(|C(N)|) & \text{if } (Q(k), Q(l), Q(m), Q(n)) \in S, \\ O(|C(N)|^{3/2}) & \text{otherwise.} \end{cases}$$

Proof. Since both $\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4)$ and $kB_1 - lB_2 + mB_3 - nB_4$ are invariant under the substitution

$$(B_1, B_2, B_3, B_4) \rightarrow (B'B_1, B'B_2, B'B_3, B'B_4)$$

for $B' \in C(N)$, we may rewrite (5.14) as

$$(5.15) \quad |C(N)| \cdot \sum_{\substack{B_2, B_3, B_4 \in C(N) \\ k-lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}}} e\left(\frac{t(\omega(k, -lB_2) + \omega(mB_3, -nB_4))}{N}\right).$$

Let X be the set of solutions to

$$k - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}, \quad B_2, B_3, B_4 \in C(N).$$

By Lemma 15, the dimension of any irreducible component of X is at most 1. The contribution from the zero dimensional components of X is at most $O(|C(N)|)$. As for the one dimensional components, Lemma 16 gives that $\omega(k, -lB_2) + \omega(mB_3, -nB_4)$ cannot be constant on any component unless pairwise equality of norms holds. Thus, if pairwise equality of norms does *not* hold, Bombieri's Theorem gives

$$\begin{aligned} & \sum_{\substack{B_2, B_3, B_4 \in C(N) \\ k - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}}} e\left(\frac{t(\omega(k, -lB_2) + \omega(mB_3, -nB_4))}{N}\right) \\ &= O(N^{1/2}) = O(|C(N)|^{1/2}) \end{aligned}$$

On the other hand, if $\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4)$ equals some constant C modulo N on some one dimensional component, then lemma 16 gives the following: $C \equiv 0 \pmod{N}$, and (5.15) equals $\text{Sol}(k, l, m, n)$, which in turn equals $|C(N)|^2$ or $2|C(N)|^2$ depending on whether $Q(k) \equiv Q(l) \equiv Q(m) \equiv Q(n) \pmod{N}$ or not. \square

Proposition 11 now follows from Proposition 17 on recalling that (see (5.7))

$$\begin{aligned} & \frac{1}{N} \sum_{j=1}^N V_\kappa(\psi_j) \overline{V_\lambda(\psi_j)} V_\mu(\psi_j) \overline{V_\nu(\psi_j)} = \\ &= \frac{N^2}{|C(2N)|^4} \cdot \\ & \cdot \sum_{\substack{B_1, B_2, B_3, B_4 \in C(N) \\ kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \pmod{N}}} e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right) \end{aligned}$$

and that $|C(N)| = |C(2N)| = N \pm 1$.

6. DISCUSSION

6.1. Comparison with generic systems. It is interesting to compare our result for the variance with the predicted answer for generic systems (see [7, 5]), which is

$$(6.1) \quad \sum_{t=-\infty}^{\infty} \int_{\mathbf{T}^2} f_0(x) \overline{f_0(A^t x)} dx$$

where $f_0 = f - \int_{\mathbf{T}^2} f(y)dy$. Using the Fourier expansion this equals

$$\sum_{t=-\infty}^{\infty} \sum_{0 \neq n \in \mathbf{Z}^2} \widehat{f}(n) \overline{\widehat{f}(nAt)}$$

By collecting together frequencies n lying in the same A -orbit, this can be written as

$$\sum_{m \in (\mathbf{Z}^2 - 0)/\langle A \rangle} \left| \sum_{n \in m\langle A \rangle} \widehat{f}(n) \right|^2$$

where $\langle A \rangle$ denotes the group generated by A . We can further massage this expression into a form closer to our formula (1.1) by noticing that the expression $\epsilon(n) := (-1)^{n_1 n_2}$ is an invariant of the A -orbit: $\epsilon(n) = \epsilon(nA)$, because we assume that $A \equiv I \pmod{2}$. Thus we can rewrite the generic variance (6.1) as

$$(6.2) \quad \sum_{m \in (\mathbf{Z}^2 - 0)/\langle A \rangle} \left| \sum_{n \in m\langle A \rangle} (-1)^{n_1 n_2} \widehat{f}(n) \right|^2.$$

The comparison with our answer (1.1), namely

$$\sum_{\nu \neq 0} \left| \sum_{Q(n)=\nu} (-1)^{n_1 n_2} \widehat{f}(n) \right|^2$$

is now clear: Both expressions would coincide if each hyperbola $\{n \in \mathbf{Z}^2 : Q(n) = \nu\}$ consisted of a single A -orbit. It is true that each hyperbola consists of a finite number of A -orbits for $\nu \neq 0$, but that number varies with ν .

6.2. A differential operator. We discuss yet another analogy with the modular domain, pointed out to us by Peter Sarnak: We define a differential operator L on $C^\infty(\mathbf{T}^2)$ by

$$L = -\frac{1}{4\pi^2} Q\left(\frac{\partial}{\partial p}, \frac{\partial}{\partial q}\right)$$

so that $\widehat{Lf}(n) = Q(n)\widehat{f}(n)$.

Given observables f, g , we define a bilinear form $B(f, g)$ by

$$B(f, g) = \sum_{\nu \neq 0} f^\#(\nu) g^\#(\nu)$$

so that (cf. Conjecture 1)

$$B(f, g) = \mathbf{E}(X_f X_g)$$

and by Theorem 2, $B(f, f)$ is the variance of the normalized matrix elements.

It is easy to check that L is self adjoint with respect to B , i.e.,

$$B(Lf, g) = B(f, Lg) .$$

Note that L is also self-adjoint with respect to the bilinear form derived from the expected variance for generic systems (6.1), (6.2). This feature was first observed for the modular domain, where the role of L is played by the Casimir operator [13].

6.3. Connection with character sums. We now explain the connection of Conjecture 1 with the theory of exponential sums in the case of *split* primes, that is primes N for which the cat map A is diagonalizable modulo N . As we show below, in this case the matrix elements are given by one-variable character sums and one may hope to attack Conjecture 1 in that case via a monodromy argument as in [9].

Suppose N is an odd prime for which A is diagonalizable modulo N , that is there is a matrix $M \in SL_2(\mathbf{Z}/2N\mathbf{Z})$ so that $A = MDM^{-1} \pmod{2N}$. In [12] we explained that in that case the normalized Hecke eigenfunctions are given in terms of the Dirichlet characters modulo N as $\psi_\chi := \sqrt{\frac{N}{N-1}} U_N(M) \chi$, and in addition if we denote by δ_0 the Dirac mass at the origin then $\psi_0 = \sqrt{N} U_N(M) \delta_0$ is an additional Hecke eigenfunction. We can write the matrix elements $\langle T_N(n) \psi_\chi, \psi_\chi \rangle$ as characters sums: By Egorov we have

$$\langle T_N(n) \psi_\chi, \psi_\chi \rangle = \frac{N}{N-1} \langle T_N(nM) \chi, \chi \rangle$$

and putting $m = (m_1, m_2) = nM$ this is given by

$$\langle T_N(n) \psi_\chi, \psi_\chi \rangle = e^{\pi i m_1 m_2 / N} \frac{1}{N-1} \sum_{Q \pmod{N}} e\left(\frac{m_2 Q}{N}\right) \chi(Q + m_1) \overline{\chi(Q)}$$

As for the eigenfunction ψ_0 corresponding to the Dirac mass δ_0 , the matrix coefficient $\langle T_N(n) \psi_0, \psi_0 \rangle$ will vanish for N sufficiently large, in fact for all N such that the vector n is not an eigenvector for $A \pmod{N}$. Indeed,

$$\langle T_N(n) \psi_0, \psi_0 \rangle = e^{\pi i m_1 m_2 / N} \sum_{Q \pmod{N}} e\left(\frac{m_2 Q}{N}\right) \delta_0(Q + m_1) \overline{\delta_0(Q)}$$

and for this not to vanish we need $m_1 = 0$, which happens precisely if $m = (0, m_2) = nA$ is an eigenvector of the diagonal matrix D , or equivalently if n is an eigenvector of $A = MDM^{-1}$.

REFERENCES

- [1] E. Bombieri. On exponential sums in finite fields. *Amer. J. Math.*, 88:71–105, 1966.
- [2] Y. Colin de Verdière. Ergodicité et fonctions propres du laplacien. *Comm. Math. Phys.*, 102(3):497–502, 1985.
- [3] M. Degli Esposti. Quantization of the orientation preserving automorphisms of the torus. *Ann. Inst. H. Poincaré Phys. Théor.*, 58(3):323–341, 1993.
- [4] M. Degli Esposti, S. Graffi, and S. Isola. Classical limit of the quantized hyperbolic toral automorphisms. *Comm. Math. Phys.*, 167(3):471–507, 1995.
- [5] B. Eckhardt, S. Fishman, J. Keating, O. Agam, J. Main, and K. Müller. Approach to ergodicity in quantum wave functions. *Phys. Rev. E*, 52(6):5893–5903, 1995.
- [6] F. Faure, S. Nonnenmacher, and S. De Bievre. Scarred eigenstates for quantum cat maps of minimal periods. <http://www.arxiv.org/abs/nlin.CD/0207060>
- [7] M. Feingold and A. Peres. Distribution of matrix elements of chaotic systems. *Phys. Rev. A (3)*, 34(1):591–595, 1986.
- [8] J. H. Hannay and M. V. Berry. Quantization of linear maps on a torus-Fresnel diffraction by a periodic grating. *Phys. D*, 1(3):267–290, 1980.
- [9] N. M. Katz. Sato-Tate equidistribution of Kurlberg-Rudnick sums. *Internat. Math. Res. Notices*, (14):711–728, 2001.
- [10] P. Kurlberg. A local Riemann hypothesis. II. *Math. Z.*, 233(1):21–37, 2000.
- [11] P. Kurlberg and Z. Rudnick. Hecke theory and equidistribution for the quantization of linear maps of the torus. *Duke Math. J.*, 103(1):47–77, 2000.
- [12] P. Kurlberg and Z. Rudnick. Value distribution for eigenfunctions of desymmetrized quantum maps. *Int. Math. Res. Not.*, 2001(18):985–1002, 2001.
- [13] W. Z. Luo and P. Sarnak. *In preparation*.
- [14] A. I. Schnirelman. Ergodic properties of eigenfunctions. *Uspehi Mat. Nauk*, 29(6(180)):181–182, 1974.
- [15] T. Watson. *Princeton Ph.D. thesis*, 2002.
- [16] S. Zelditch. Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.*, 55(4):919–941, 1987.
- [17] S. Zelditch. Quantum ergodicity of C^* dynamical systems. *Comm. Math. Phys.*, 177(2):507–528, 1996.

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY,
SE-412 96 GOTHENBURG, SWEDEN

URL: www.math.chalmers.se/~kurlberg

E-mail address: kurlberg@math.chalmers.se

RAYMOND AND BEVERLY SACKLER SCHOOL OF MATHEMATICAL SCIENCES,
TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL

E-mail address: rudnick@post.tau.ac.il