Annals of Mathematics, **161** (2005), 489–507

On the distribution of matrix elements for the quantum cat map

By 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 [15], [2], [17] or in the mean square [18]. 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" [10], and conditional on the Generalized Riemann Hypothesis[1] also for the modular domain [16], 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

*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.

[1]An unconditional proof was recently announced by Elon Lindenstrauss.

correlations of the observable [7], [5]. In this note we 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], [10] (see §2 for further background and any unexplained notation): The classical mechanical system is the iteration of a linear hyperbolic map $A \in \mathrm{SL}(2,\mathbb{Z})$ of the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ (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 $\mathcal{H}_N = L^2(\mathbb{Z}/N\mathbb{Z})$. Let $f \in C^\infty(\mathbb{T}^2)$ be a smooth, real valued observable and $\mathrm{Op}_N(f) : \mathcal{H}_N \to \mathcal{H}_N$ its quantization. The quantization of the classical map A is a unitary map $U_N(A)$ of $\mathcal{H}_N$.

In [10] 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 $\langle \mathrm{Op}_N(f)\psi_j, \psi_j\rangle$ converge[2] to the phase-space average $\int_{\mathbb{T}^2} f(x)dx$ [10]. 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

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

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

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

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

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

where $\widehat{f}(n)$ are the Fourier coefficients of f. (Note that $f^\#$ can be identically zero for nonzero f, e.g., if $f = g - g\circ A$.)

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

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

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

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

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* $N \to \infty$ *through primes*, *the variance of the normalized matrix elements* $F_j^{(N)}$ *is given by*

$$\frac{1}{N}\sum_{j=1}^{N}|F_j^{(N)}|^2 \to \mathbb{E}(X_f^2) = \sum_{\nu\neq 0}|f^{\#}(\nu)|^2 \,. \tag{1.1}$$

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 [12] 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*

$$\frac{1}{N}\sum_{j=1}^{N}|F_j^{(N)}|^4 \to \mathbb{E}(|X_f|^4)$$

as $N \to \infty$ *through primes.*

Acknowledgements. We thank Peter Sarnak for discussions on his work with Wenzhi Luo [12], and Dubi Kelmer for his comments.

2. Background

The full details on the cat map and its quantization can be found in [10]. For the reader's convenience we briefly recall the setup: The classical dynamics are given by a hyperbolic linear map $A \in \mathrm{SL}(2,\mathbb{Z})$ so that $x = \binom{p}{q} \in \mathbb{T}^2 \mapsto Ax$ is a symplectic map of the torus. Given an observable $f \in C^\infty(\mathbb{T}^2)$, the classical evolution defined by A is $f \mapsto f \circ A$, where $(f\circ A)(x) = f(Ax)$.

For doing quantum mechanics on the torus, one takes Planck's constant to be $1/N$ and as the Hilbert space of states one takes $\mathcal{H}_N := L^2(\mathbb{Z}/N\mathbb{Z})$, where the inner product is given by

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

The basic observables are given by the operators $T_N(n)$, $n \in \mathbb{Z}^2$, acting on $\psi \in L^2(\mathbb{Z}/N\mathbb{Z})$ via:

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

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

For any smooth classical observable $f \in C^\infty(\mathbb{T}^2)$ with Fourier expansion $f(x) = \sum_{n \in \mathbb{Z}^2} \widehat{f}(n) e(nx)$, its quantization is given by

$$\mathrm{Op}_N(f) := \sum_{n \in \mathbb{Z}^2} \widehat{f}(n) T_N(n) .$$

2.1. *Quantum dynamics.* For A which satisfies a certain parity condition, we can assign unitary operators $U_N(A)$, acting on $L^2(\mathbb{Z}/N\mathbb{Z})$, having the following important properties:

- "Exact Egorov": For all observables $f \in C^\infty(\mathbb{T}^2)$
$$U_N(A)^{-1} \mathrm{Op}_N(f) U_N(A) = \mathrm{Op}_N(f \circ A).$$
- The quantization depends only on A modulo $2N$: If $A \equiv B \mod 2N$ then $U_N(A) = U_N(B)$.
- The quantization is multiplicative: if A, B are congruent to the identity matrix modulo 4 (resp., 2) if N is even (resp., odd), then [10], [13]
$$U_N(AB) = U_N(A) U_N(B).$$

2.2. *Hecke eigenfunctions.* Let α, α^{-1} be the eigenvalues of A. Since A is hyperbolic, α is a unit in the real quadratic field $K = \mathbb{Q}(\alpha)$. Let $\mathfrak{O} = \mathbb{Z}[\alpha]$, which is an order of K. Let $v = (v_1, v_2) \in \mathfrak{O}^2$ be a vector such that $vA = \alpha v$. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we may take $v = (c, \alpha - a)$. Let $I := \mathbb{Z}[v_1, v_2] = \mathbb{Z}[c, \alpha - a] \subset \mathfrak{O}$. Then I is an $\mathfrak{O}$-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 $I \cong \mathbb{Z}^2$ and the action of $\mathfrak{O}$ on the ideal I by multiplication gives a ring homomorphism

$$\iota : \mathfrak{O} \to \mathrm{Mat}_2(\mathbb{Z})$$

with the property that $\det(\iota(\beta)) = \mathcal{N}(\beta)$, where $\mathcal{N} : \mathbb{Q}(\alpha) \to \mathbb{Q}$ is the norm map.

Let $C(2N)$ be the elements of $\mathfrak{O}/2N\mathfrak{O}$ with norm congruent to $1 \mod 2N$, and which congruent to 1 modulo $4\mathfrak{O}$ (resp., $2\mathfrak{O}$) if N is even (resp.,odd). Reducing ι modulo $2N$ gives a map

$$\iota_{2N} : C(2N) \to \mathrm{SL}_2(\mathbb{Z}/2N\mathbb{Z}).$$

Since $C(2N)$ is commutative, the multiplicativity of our quantization implies 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 $U_N(\iota_{2N}(\alpha)) = U_N(A)$.

The matrix elements are invariant under the Hecke operators:

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

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

$$\langle T_N(n)\psi_j, \psi_j\rangle = \langle T_N(nB)\psi_j, \psi_j\rangle\ . \tag{2.2}$$

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

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

This, up to sign, is the quadratic form $\mathcal{N}(xc + y(\alpha - a))/\mathcal{N}(I)$ induced by the norm form on the ideal $I = \mathbb{Z}[c, \alpha - a]$ described in Section 2.2, where $\mathcal{N}(I) = \#\mathfrak{O}/I$. Indeed, since $I = \mathbb{Z}[c, \alpha - a]$ and $\mathfrak{O} = \mathbb{Z}[1, \alpha]$ we have $\mathcal{N}(I) = |c|$. A computation shows that the norm form is then $\mathrm{sign}(c)Q(x, y)$.

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

$$\{B \in \mathrm{SO}(Q, \mathbb{Z}/2N\mathbb{Z}) : B \equiv I \mod 2\}.$$

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

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

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 nondegenerate over the field $\mathbb{Z}/N\mathbb{Z}$. If $N > |Q(m)|$ then $Q(m) \neq 0 \mod N$ and then the assertion reduces to the fact that if Q is a nondegenerate binary quadratic form over the finite field $\mathbb{Z}/N\mathbb{Z}$ ($N \neq 2$ prime) then the special orthogonal group $\mathrm{SO}(Q, \mathbb{Z}/N\mathbb{Z})$ acts transitively on the hyperbolas $\{Q(n) = \nu\}$, $\nu \neq 0 \mod N$. ∎

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

$$(-1)^{n_1 n_2} \langle T_N(n)\psi, \psi\rangle = (-1)^{m_1 m_2} \langle T_N(m)\psi, \psi\rangle.$$

Proof. For ease of notation, set $\varepsilon(n) := (-1)^{n_1 n_2}$. By Lemma 4 it suffices to show that if $m \equiv nB \mod N$ for some $B \in \mathrm{SO}(Q, \mathbb{Z}/N\mathbb{Z})$ then $\varepsilon(n)\langle T_N(n)\psi, \psi\rangle = \varepsilon(m)\langle T_N(m)\psi, \psi\rangle$.

By the Chinese Remainder Theorem,

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

(recall N is odd) and so

$$C(2N) \simeq \{B \in \mathrm{SO}(Q\mathbb{Z}/2N\mathbb{Z}) : B \equiv I \mod 2\} \simeq \mathrm{SO}(Q, \mathbb{Z}/N\mathbb{Z}) \times \{I\}.$$

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

We note that $\varepsilon(n)T_N(n)$ has period N, rather than merely $2N$ for $T_N(n)$ as would follow from (2.1). Then since $m = n\tilde{B} \mod N$,

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

(recall that $\tilde{B} \in C(2N)$ preserves parity: $n\tilde{B} \equiv n \mod 2$, so $\varepsilon(n\tilde{B}) = \varepsilon(n)$). Thus for ψ a Hecke eigenfunction,

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

the last equality by (2.2). □

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

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

and

$$V_\nu(\psi) := \sqrt{N}(-1)^{n_1 n_2} \langle T_N(n)\psi, \psi\rangle, \tag{2.3}$$

where $n \in \mathbb{Z}^2$ 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*

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

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

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

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 $1 \le i,\ j \le 2$.

Proof. If N is inert, then the Weil representation is multiplicity free when restricted to $C(2N)$ (see Lemma 4 in [9].) If N is split, then $C(2N)$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^*$ and the trivial character occurs with multiplicity one, the quadratic character occurs with multiplicity two, and all other characters occur with multiplicity one (see [11, §4.1]). This explains the shape of $\tilde{D}$.

As for the bound on in the split case, it suffices to take $f(x,y) = e(\frac{n_1 x + n_2 y}{N})$ for some $n_1, n_2 \in \mathbb{Z}$. We may give an explicit construction of the Hecke eigenfunctions as follows (see [11, §4] for more details): there exists $M \in \mathrm{SL}_2(\mathbb{Z}/2N\mathbb{Z})$ such that the eigenfunctions ψ_1, ψ_2 can be written as

$$\psi_1 = \sqrt{N} \cdot U_N(M)\delta_0, \quad \psi_2 = \sqrt{\frac{N}{N-1}} \cdot U_N(M)(1-\delta_0)$$

where $\delta_0(x) = 1$ if $x \equiv 0 \mod N$, and $\delta_0(x) = 0$ otherwise. Setting $\phi_1 = \sqrt{N}\delta_0$ and $\phi_2 = \sqrt{\frac{N}{N-1}}(1-\delta_0)$, exact Egorov gives

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

where $(n_1', n_2') \equiv (n_1, n_2)M \mod N$. Since we may assume n not to be an eigenvector of A modulo N, we have $n_1' \not\equiv 0 \mod N$ and $n_2' \not\equiv 0 \mod N$. Hence

$$D_{11} = \langle T_N((n_1', n_2'))\phi_1, \phi_1 \rangle = e\left(\frac{n_1' n_2'}{2N}\right) \delta_0(0 + n_1') = 0$$

since $n_1' \not\equiv 0 \mod N$. The other estimates are analogous. □

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 $k, l, m, n \in \mathbb{Z}^2$. Then*

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

Moreover,

$$\sum_{i=1}^N \langle T_N(k)\psi_i, \psi_i \rangle \overline{\langle T_N(l)\psi_i, \psi_i \rangle} \langle T_N(m)\psi_i, \psi_i \rangle \overline{\langle T_N(n)\psi_i, \psi_i \rangle}$$
$$= \operatorname{tr}\left(D(k)D^*(l)D(m)D^*(n)\right) + O(N^{-2}).$$

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,

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

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

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

The proof of the second assertion is similar. □

4. Proof of Theorem 2

In order to prove Theorem 2 it suffices, by Proposition 6, to show that as $N \to \infty$,

$$\frac{1}{N}\sum_{j=1}^{N} V_\nu(\psi_j)\overline{V_\mu(\psi_j)} \to \mathbb{E}\big(\operatorname{tr} U_\nu \operatorname{tr} U_\mu\big) = \begin{cases} 1 & \text{if } \mu = \nu, \\ 0 & \text{if } \mu \neq \nu, \end{cases}$$

where $U_\mu, U_\nu \in \mathrm{SU}_2$ are random matrices in SU_2, independent if $\nu \neq \mu$.

Proposition 9. *Let $\{\psi_i\}_{i=1}^N$ be a Hecke basis of $\mathcal{H}_N$. If $N \geq N_0(\mu,\nu)$ is prime and $\mu,\nu \not\equiv 0 \mod N$, 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 $m, n \in \mathbb{Z}^2$ such that $Q(m) = \mu$ and $Q(n) = \nu$. By (2.3) and Lemma 8 we find that

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

By definition of $D(n)$ we have

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

We now take the trace of both sides and apply the following easily checked identity (see (2.1)), valid for odd N and $B_1, B_2 \in C(2N)$:

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

We get

$$\begin{aligned} &\frac{1}{N}\sum_{j=1}^{N} V_\nu(\psi_j)\overline{V_\mu(\psi_j)} \\ &\quad= \frac{(-1)^{m_1m_2+n_1n_2}}{|C(2N)|^2}\sum_{\substack{B_1,B_2\in C(2N)\\ nB_1\equiv mB_2 \mod N}} (-1)^{m_1m_2+n_1n_2}N + O(N^{-1}) \\ &\quad= \frac{N}{|C(2N)|}\cdot|\{B\in C(2N) : n \equiv mB \mod N\}| + O(N^{-1}), \end{aligned} \tag{4.1}$$

which, since $|C(2N)| = N \pm 1$, equals $1 + O(N^{-1})$ if there exists $B \in C(2N)$ such that $n \equiv mB \mod N$, and $O(N^{-1})$ otherwise. Finally, for N sufficiently large (i.e., $N \geq N_0(\mu, \nu)$), Lemma 4 gives that $n \equiv mB \mod N$ 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

$$\frac{1}{N}\sum_{j=1}^{N} V_\kappa(\psi_j)\overline{V_\lambda(\psi_j)}V_\mu(\psi_j)\overline{V_\nu(\psi_j)} \to \mathbb{E}\big(\operatorname{tr} U_\kappa \operatorname{tr} U_\lambda \operatorname{tr} U_\mu \operatorname{tr} U_\nu\big) \tag{5.1}$$

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

Let $S \subset \mathbb{Z}^4$ 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 10. *Let $\{\psi_i\}_{i=1}^N$ be a Hecke basis of $\mathcal{H}_N$ and let $\kappa, \lambda, \mu, \nu \in \mathbb{Z}$. If N is a sufficiently large prime, then*

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

Given Proposition 10 it is straightforward to deduce (5.1), we need only to note that $\mathbb{E}\big((\operatorname{tr} U)^4\big) = 2$, $\mathbb{E}\big((\operatorname{tr} U)^2\big) = 1$, and $\mathbb{E}\big(\operatorname{tr} U\big) = 0$.

The proof of Proposition 10 will occupy the remainder of this section. For the reader's convenience, here is a brief outline:

(1) Express the left-hand side of (5.1) an exponential sum.

(2) 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.

(3) Determine the number of solutions.

5.2. *Ergodic averaging.*

LEMMA 11. *Choose $k, l, m, n \in \mathbb{Z}^2$ such that $Q(k) = \kappa$, $Q(l) = \lambda$, $Q(m) = \mu$, and $Q(n) = \nu$. Then*

$$\frac{1}{N}\sum_{j=1}^{N} V_\kappa(\psi_j)\overline{V_\lambda(\psi_j)}V_\mu(\psi_j)\overline{V_\nu(\psi_j)} = \frac{N^2}{|C(2N)|^4} \cdot \cdot \sum_{\substack{B_1,B_2,B_3,B_4 \in C(N) \\ kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \mod N}} e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right). \tag{5.2}$$

The proof of Lemma 11 is an extension of the arguments proving the analogous (4.1) in the proof of Proposition 9 and is left to the reader.

5.3. *Exponential sums over curves.* In order to show that there is quite a bit of cancellation in (5.2) 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.

THEOREM 12 (Bombieri [1, Th. 6]). *If $d_1 d_2 < p$ and R is not constant on any component Γ of X then*

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

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

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

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

Proof. Taking norms, we obtain $1 = \mathcal{N}(a) + \mathcal{N}(b) + \operatorname{tr}(\overline{a}b\gamma_2)$ and hence $\operatorname{tr}(\overline{a}b\gamma_2)$ is constant. If $\overline{a}b \neq 0$, 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 $D \in \mathbb{F}_N$.) $\square$

LEMMA 14. *Fix $k, l, m, n \in \mathbb{Z}^2$ and let X be the set of solutions to*

$$k - lB_2 + mB_3 - nB_4 \equiv 0 \mod N, \; B_2, B_3, B_4 \in C(N).$$

If $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \mod N$, 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

$$k - lB_2 + mB_3 - nB_4 \equiv 0 \mod N$$

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 $Q(l), Q(m) \not\equiv 0 \mod N$ implies that λ, μ are nonzero[3], 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

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

Lemma 13 gives (note that $\kappa \neq 0$ since $Q(k) \not\equiv 0 \mod N$) 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.

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

LEMMA 15. *Assume that* $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \mod N$, *and let* $\mathrm{Sol}(k, l, m, n)$ *be the number of solutions to the equations*

$$kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \mod N \tag{5.3}$$

$$\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4) \equiv -C \mod N \tag{5.4}$$

where $B_i \in C(N)$. *If* $C \equiv 0 \mod N$ *and* N *is sufficiently large, then*

$$\mathrm{Sol}(k, l, m, n) = \begin{cases} 2|C(N)|^2 & \text{if } Q(k) = Q(l) = Q(m) = Q(n), \\ |C(N)|^2 + O(|C(N)|) & \text{if } (Q(k), Q(l), Q(m), Q(n)) \in S, \\ O(|C(N)|) & \text{otherwise.} \end{cases} \tag{5.5}$$

On the other hand, if $C \not\equiv 0 \mod N$ *then*

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

Proof. For simplicity[4], we will assume that N is inert. It will be convenient to use the language of algebraic number theory; we identify $(\mathbb{Z}/N\mathbb{Z})^2$ 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

$$\omega(m, n) = xw - zy = \mathrm{Im}(\overline{(x + y\sqrt{D})}(z + w\sqrt{D}))$$

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

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

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

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

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

and hence

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

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

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

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

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

[4]The 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.

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 (5.3) and (5.4) 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 13 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 are 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

$$\mu_1 = \mu_2, \quad \mu_4 = \mu_3. \tag{5.6}$$

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

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

$$\mu_2 = \mu_3, \quad \mu_1 = \mu_4. \tag{5.7}$$

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

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

which in turn gives that

$$\begin{aligned}(5.8)\quad \mu_1 &= \mu_2 + \mu_4 - \mu_3 = \mu_2 + \frac{\mu_2 - \mu_3}{\overline{\mu_3} - \overline{\mu_2}}\overline{\mu_2} - \mu_3 \\ &= \frac{\mu_2 - \mu_3}{\overline{\mu_3} - \overline{\mu_2}}(\overline{\mu_3} - \overline{\mu_2}) + \frac{\mu_2 - \mu_3}{\overline{\mu_3} - \overline{\mu_2}}\overline{\mu_2} = \frac{\mu_2 - \mu_3}{\overline{\mu_3} - \overline{\mu_2}}\overline{\mu_3} = \frac{\mu_2\overline{\mu_3} - N_3}{\mu_2\overline{\mu_3} - N_2}\mu_2.\end{aligned}$$

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

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

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, $C = \mathrm{Im}(K) = \mathrm{Im}(\overline{\mu_1}\mu_2 + \overline{\mu_3}\mu_4) = 0$. 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

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

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 $N_3N_4 = N_1N_2$, 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.

Now Lemma 4 gives 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 (5.2).

PROPOSITION 16. *If* $Q(k), Q(l), Q(m), Q(n) \not\equiv 0 \mod N$ *then, for* N *sufficiently large, we have*

$$\sum_{\substack{B_1,B_2,B_3,B_4 \in C(N) \\ kB_1 - lB_2 + mB_3 - nB_4 \equiv 0 \mod N}} e\left(\frac{t(\omega(kB_1, -lB_2) + \omega(mB_3, -nB_4))}{N}\right) \tag{5.9}$$

$$= \begin{cases} 2|C(N)|^2 + O(|C(N)|) & \text{if } Q(k) = Q(l) = Q(m) = Q(n), \\ |C(N)|^2 + O(|C(N)|) & \text{if } (Q(k), Q(l), Q(m), Q(n)) \in S, \\ O(|C(N)|^{3/2}) & \text{otherwise.} \end{cases}$$

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

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

for $B' \in C(N)$, we may rewrite the left hand side of (5.9) as $|C(N)|$ times

$$\sum_{\substack{B_2,B_3,B_4 \in C(N) \\ k - lB_2 + mB_3 - nB_4 \equiv 0 \mod N}} e\left(\frac{t(\omega(k, -lB_2) + \omega(mB_3, -nB_4))}{N}\right). \tag{5.10}$$

Let X be the set of solutions to

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

By Lemma 14, 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 15 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 that (5.10) is $O(N^{1/2}) = O(|C(N)|^{1/2})$.

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 15 gives the following: $C \equiv 0 \mod N$, and (5.10) equals $\mathrm{Sol}(k,l,m,n)$, which in turn equals $|C(N)|^2$ or $2|C(N)|^2$ depending on whether $Q(k) \equiv Q(l) \equiv Q(m) \equiv Q(n) \mod N$ or not. □

Proposition 10 now follows from Lemma 11 and Proposition 16 on recalling 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

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

where $f_0 = f - \int_{\mathbb{T}^2} f(y)dy$. Using the Fourier expansion and collecting together frequencies n lying in the same A-orbit this equals

$$\sum_{t=-\infty}^{\infty} \sum_{0\neq n\in\mathbb{Z}^2} \widehat{f}(n)\overline{\widehat{f}(nA^t)} = \sum_{m\in(\mathbb{Z}^2-0)/\langle A\rangle} \left| \sum_{n\in m\langle A\rangle} \widehat{f}(n)\right|^2$$

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

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

The comparison with with our answer $\sum_{\nu\neq 0}\left|\sum_{Q(n)=\nu}(-1)^{n_1n_2}\widehat{f}(n)\right|^2$ in (1.1), is now clear: Both expressions would coincide if each hyperbola $\{n \in \mathbb{Z}^2 : Q(n) = \nu\}$ 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.* There is yet another analogy with the modular domain, pointed out to us by Peter Sarnak: We define a differential operator L on $C^\infty(\mathbb{T}^2)$ 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) $B(f, g) = \mathbb{E}(X_f X_g)$ 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 [12] (cf. Appendix 5 of Sarnak's survey [14]).

6.3. *Connection with character sums.* Conjecture 1 is related to the value distributions of certain character sums, at least in the case of *split* primes, that is primes N for which the cat map A is diagonalizable modulo N. Let $M \in \mathrm{SL}_2(\mathbb{Z}/2N\mathbb{Z})$ be such that $A = MDM^{-1} \mod 2N$. In [11] we explained that in that case, all but one of the normalized Hecke eigenfunctions are given in terms of the Dirichlet characters χ modulo N as $\psi_\chi := \sqrt{\frac{N}{N-1}} U_N(M)\chi$. We can then write the matrix elements $\langle T_N(n)\psi_\chi, \psi_\chi \rangle$ as characters sums: Setting $(m_1, m_2) = nM$, we have

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

and Conjecture 1 gives a prediction for the value distribution of these sums as χ varies.

Royal Institute of Technology, Stockholm, Sweden
E-mail address: kurlberg@math.kth.se
URL: www.math.kth.se/~kurlberg

Tel Aviv University, Tel Aviv 69978, Israel
E-mail address: rudnick@post.tau.ac.il

References

[1] E. Bombieri, On exponential sums in finite fields, *Amer. J. Math.* **88** (1966), 71–105.

[2] Y. Colin de Verdière, Ergodicité et fonctions propres du laplacien, *Comm. Math. Phys.* **102** (1985), 497–502.

[3] M. Degli Esposti, Quantization of the orientation preserving automorphisms of the torus, *Ann. Inst. H. Poincaré Phys. Théor.* **58** (1993), 323–341.

[4] M. Degli Esposti, S. Graffi, and S. Isola, Classical limit of the quantized hyperbolic toral automorphisms, *Comm. Math. Phys.* **167** (1995), 471–507.

[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** (1995), 5893–5903.

[6] F. Faure, S. Nonnenmacher, and S. De Bièvre, Scarred eigenstates for quantum cat maps of minimal periods, *Comm. Math. Phys.* **29** (2003), 449–492.

[7] M. Feingold and A. Peres, Distribution of matrix elements of chaotic systems, *Phys. Rev. A* **34** (1986), 591–595.

[8] J. H. Hannay and M. V. Berry, Quantization of linear maps on a torus-Fresnel diffraction by a periodic grating, *Phys. D* **1** (1980), 267–290.

[9] P. Kurlberg, A local Riemann hypothesis. II, *Math. Z.* **233** (2000), 21–37.

[10] P. Kurlberg and Z. Rudnick, Hecke theory and equidistribution for the quantization of linear maps of the torus, *Duke Math. J.* **103** (2000), 47–77.

[11] ———, Value distribution for eigenfunctions of desymmetrized quantum maps, *Internat. Math. Res. Not.* **(2001)**, No. 18 985–1002.

[12] W. Z. Luo and P. Sarnak, Quantum invariance for Hecke eigenforms, *Ann. Sci. École Norm. Sup.* (4) **37** (2004), 769–799.

[13] F. Mezzadri, On the multiplicativity of quantum cat maps, *Nonlinearity* **15** (2002), 905–922.

[14] P. Sarnak, Spectra of hyperbolic surfaces, *Bull. Amer. Math. Soc.* **40** (2003) 441–478 (electronic).

[15] A. I. Schnirelman, Ergodic properties of eigenfunctions. *Uspkehi Mat. Nauk* **29** (1974), 181–182.

[16] T. Watson, Rankin triple products and quantum chaos, Ph.D. thesis, Princeton University, 2003.

[17] S. Zelditch, Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.* **55** (1987), 919–941.

[18] ———, Quantum ergodicity of C^* dynamical systems, *Comm. Math. Phys.* **177** (1996), 507–528.

(Received March 26, 2003)