ON QUANTUM ERGODICITY FOR HIGHER DIMENSIONAL CAT MAPS

PÄR KURLBERG, ALINA OSTAFE, ZEEV RUDNICK,
AND IGOR E. SHPARLINSKI

Abstract. We study eigenfunction localization for higher dimensional cat maps, a popular model of quantum chaos. These maps are given by linear symplectic maps in $\mathrm{Sp}(2g,\mathbb{Z})$, which we take to be ergodic. Under some natural assumptions, we show that there is a density one sequence of integers N so that as N tends to infinity along this sequence, all eigenfunctions of the quantized map at the inverse Planck constant N are uniformly distributed. For the two-dimensional case ($g = 1$), this was proved by P. Kurlberg and Z. Rudnick (2001). The higher dimensional case offers several new features and requires a completely different set of tools, including from additive combinatorics, in particular a bound of J. Bourgain (2005) for Mordell sums, and a study of tensor product structures for the cat map.

Contents

Date: November 7, 2024.

1991 *Mathematics Subject Classification.* 11L07, 81Q50.

Key words and phrases. Quantum unique ergodicity, linear map, exponential sums.

1. Introduction

1.1. **Quantum ergodicity and the quantized cat map.**

Eigenfunction localization is one of the central topics of Quantum Chaos. In this paper, we examine this question for an important "toy model", the quantized cat map [12], aiming for higher dimensional maps. Our techniques, after a preliminary reduction, combine analytic number theory and additive combinatorics.

Denote by $\mathrm{Sp}(2g, \mathbb{Z})$ the group of all integer matrices A which preserve the symplectic form

$$\omega(\mathbf{x}, \mathbf{y}) = \mathbf{x}_1 \cdot \mathbf{y}_2 - \mathbf{x}_2 \cdot \mathbf{y}_1, \tag{1.1}$$

with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{R}^g \times \mathbb{R}^g$. Any $A \in \mathrm{Sp}(2g, \mathbb{Z})$ generates a classical dynamical system via its action on the torus $\mathbb{T}^{2g} = \mathbb{R}^{2g}/\mathbb{Z}^{2g}$. We say that this dynamical system is ergodic if for almost all initial positions $\mathbf{x} \in \mathbb{T}^{2g}$, the orbit $\{A^j \mathbf{x} : \ j \ge 0\}$ is uniformly distributed in $\mathbb{T}^{2g}$. This is equivalent to A having no eigenvalues which are roots of unity, see [11].

Associated to any $A \in \mathrm{Sp}(2g, \mathbb{Z})$ is a quantum mechanical system. We briefly recall the key definitions: One constructs for each integer $N \ge 1$ (the inverse Planck constant, necessarily an integer here) a

Hilbert space of states $\mathcal{H}_N = L^2((\mathbb{Z}/N\mathbb{Z})^g)$ equipped with the scalar product

$$\langle \varphi_1, \varphi_2 \rangle = \frac{1}{N^g} \sum_{\mathbf{u} \in (\mathbb{Z}/N\mathbb{Z})^g} \varphi_1(\mathbf{u}) \overline{\varphi_2(\mathbf{u})}, \qquad \varphi_1, \varphi_2 \in \mathcal{H}_N.$$

The basic observables are given by the unitary operators

$$\mathrm{T}_N(\mathbf{n}) : \mathcal{H}_N \to \mathcal{H}_N, \qquad \mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbb{Z}^g \times \mathbb{Z}^g = \mathbb{Z}^{2g},$$

as follows

$$(\mathrm{T}_N(\mathbf{n})\varphi)(\mathbf{Q}) = \mathbf{e}_{2N}(\mathbf{n}_1 \cdot \mathbf{n}_2)\, \mathbf{e}_N(\mathbf{n}_2 \cdot \mathbf{Q}) \varphi(\mathbf{Q} + \mathbf{n}_1), \tag{1.2}$$

where hereafter we always follow the convention that integer arguments of functions on $\mathbb{Z}/N\mathbb{Z}$ are reduced modulo N (that is, $\varphi(\mathbf{Q} + \mathbf{n}_1) = \varphi(\mathbf{Q} + (\mathbf{n}_1 \bmod N))$). It is also easy to verify that (1.2) implies

$$\mathrm{T}_N(\mathbf{m})\, \mathrm{T}_N(\mathbf{n}) = \mathbf{e}_{2N}(\omega(\mathbf{m}, \mathbf{n}))\, \mathrm{T}_N(\mathbf{m} + \mathbf{n}),$$

where $\omega(\mathbf{m}, \mathbf{n})$ is defined by (1.1) and

$$\mathbf{e}(z) = \exp(2\pi i z), \qquad \mathbf{e}_k(z) = \mathbf{e}(z/k),$$

see also [19, Equation (2.6)].

For each real-valued function $f \in C^\infty(\mathbb{T}^{2g})$ (an "observable"), one associates a self-adjoint operator $\mathrm{Op}_N(f)$ on $\mathcal{H}_N$, analogous to a pseudo-differential operator with symbol f, defined by

$$\mathrm{Op}_N(f) = \sum_{\mathbf{n} \in \mathbb{Z}^{2g}} \widehat{f}(\mathbf{n})\, \mathrm{T}_N(\mathbf{n}), \tag{1.3}$$

where

$$f(\mathbf{x}) = \sum_{\mathbf{n} \in \mathbb{Z}^{2g}} \hat{f}(\mathbf{n})\, \mathbf{e}(\mathbf{n} \cdot \mathbf{x}). \tag{1.4}$$

Assuming $A = I \bmod 2$ (this condition can be weakened, see, for example, the definition of the subgroup $\mathrm{Sp}_\vartheta(2g, \mathbb{Z})$ of $\mathrm{Sp}(2g, \mathbb{Z})$ as in [16, p. 817], with d instead of g), for each value of the inverse Planck constant $N \geq 1$, there is a unitary operator $U_N(A)$ on $\mathcal{H}_N$, unique up to scalar multiples, which generates the quantum evolution, in the sense that for every observable $f \in C^\infty(\mathbb{T}^{2g})$, we have the exact Egorov property

$$U_N(A)^*\, \mathrm{Op}_N(f) U_N(A) = \mathrm{Op}_N(f \circ A), \tag{1.5}$$

where $U_N(A)^* = \overline{U_N(A)}^{\,t}$, we refer to [18, 23] for a detailed exposition in the case $g = 1$ and [16] for higher dimensions.

The stationary states of the system are the eigenfunctions of $U_N(A)$ and one of the main goals is to study their localization properties. In particular, given any normalized sequence of eigenfunctions $\psi_N \in \mathcal{H}_N$, we ask if the expected values of observables in these eigenfunctions converge, as $N \to \infty$, to the classical average (see § 2.1 for precise definitions), that is, that

$$\lim_{N\to\infty} \langle \mathrm{Op}_N(f)\psi_N, \psi_N\rangle = \int_{\mathbb{T}^{2g}} f(\mathbf{x})d\mathbf{x} \tag{1.6}$$

for all $f \in C^\infty(\mathbb{T}^{2g})$, in which case we say that the sequence of eigenfunctions $\{\psi_N\}$ is uniformly distributed.

A fundamental result is the Quantum Ergodicity Theorem [4,25,28], valid in great generality, which in our setting asserts that if A is ergodic, then for any **orthonormal basis** $\varPsi_N = \{\psi_{j,N} : \ j = 1,\dots,N^g\}$ of eigenfunctions of $U_N(A)$ in $\mathcal{H}_N$, there is a subset $\mathcal{S} \subseteq \{1,\dots,N^g\}$ with asymptotic density one (that is, $\sharp\mathcal{S}/N^g \to 1$, where $\sharp\mathcal{S}$ denotes the cardinality of $\mathcal{S}$) so that $\psi_{j,N}$ are uniformly distributed for all $j \in \mathcal{S}$, see [3]. If <u>all</u> eigenfunctions are uniformly distributed, the system is said to exhibit Quantum Unique Ergodicity [24]. In fact, more generally, setting

$$\Delta_A(f,N) = \max_{\psi_N,\psi'_N} \left| \langle \mathrm{Op}_N(f)\psi_N, \psi'_N\rangle - \langle \psi_N, \psi'_N\rangle \int_{\mathbb{T}^{2g}} f(\mathbf{x})d\mathbf{x} \right|,$$

the maximum taken over all pairs of normalized eigenfunctions ψ_N, ψ'_N of $U_N(A)$, we ask if for all $f \in C^\infty(\mathbb{T}^{2g})$,

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{N}}} \Delta_A(f,N) = 0, \tag{1.7}$$

where $\mathcal{N}$ is a set of integers of asymptotic density 1 (that is, $\mathcal{N} \cap [1,x] = x + o(x)$ as $x \to \infty$).

Remark 1.1. *It is interesting to note that even if we are mainly interested in scarring (that is, decay of diagonal matrix coefficients corresponding to $\psi'_N = \psi_N$ in the above definition of $\Delta_A(f,N)$ and establishing (1.6)), for the full argument we still need estimates for off-diagonals coefficients of the "nontrivial" tensor component in § 6.4.*

The two-dimensional ($g = 1$) cat map is where the first counterexamples ("scars") to QUE have been proved to exist [9], associated with the N, where the period $\mathrm{ord}(A,N)$ of the classical map reduced modulo N was almost minimally small, about $2\log N/\log\lambda$, where $\lambda > 1$ is the largest eigenvalue of A. We note that the relevance of the classical period to the quantum system was recognized early on in the theory [5,12,15]. In [19], it was shown that if $\mathrm{ord}(A,N)$ was somewhat larger than $N^{1/2}$ (and N satisfies a further genericity condition), then all eigenfunctions in $\mathcal{H}_N$ are uniformly distributed. Note that the condition holds for almost all primes [7]. Separately, it was shown that $\operatorname{ord}(A, N)$ is sufficiently large for almost all integers N.

A breakthrough was made by Bourgain [2], who showed that when $N = p$ is prime (that, and the prime power, cases are the basic building block for the theory since the quantization with respect to composite moduli arise as tensor products of quantizations with respect to prime power moduli), for all eigenfunctions to be uniformly distributed it suffices to take $\operatorname{ord}(A, p) > p^{\varepsilon}$, for some $\varepsilon > 0$, a condition that is much easier to establish than a bound bigger than $p^{1/2}$. This allowed Bourgain [2] to give a polynomial rate of convergence for a version of (1.7) over a sequence of almost all integers: for some $\delta > 0$, for almost all N we have $\Delta_A(f, N) \leqslant N^{-\delta}$. Using a different approach, in [20] it is shown that one can take any $\delta < 1/60$.

1.2. **Higher dimensional cat maps.**

Higher dimensional cat maps offer several more challenges. In particular, we address the analogue of [19], namely all eigenfunctions in $\mathcal{H}_N$ being uniformly distributed for almost all integers N. We do not discuss other aspects of localizations, such as entropy bounds [8, 22] and showing that all semiclassical measures have full support [6, 17, 26].

In higher dimensions (that is, for $g > 1$), there is a significant change. Kelmer [16] has shown that if A has nontrivial invariant rational istropic subspaces, then for all N, uniform distribution (1.7) fails – there are so-called scars.

So we assume that there are no nontrivial invariant rational isotropic subspaces. We want to find a full density sequence $\mathcal{N}$ of integers N for which all eigenfunctions of $U_N(A)$ are uniformly distributed, that is, if we fix $f \in C^{\infty}(\mathbb{T}^{2g})$ then we have (1.7). If this holds for all f then we say that A *satisfies QUE for the subsequence* $\mathcal{N}$.

Recall that we assume ergodicity, equivalently, that the eigenvalues of $A \in \mathrm{Sp}(2g, \mathbb{Z})$ are not roots of unity. For our results, we need to impose a further condition on A, that no ratio of distinct eigenvalues is a root of unity. In addition, we assume that the characteristic polynomial $f_A(x) = \det(xI - A) \in \mathbb{Z}[x]$ is separable (that is, has no multiple roots)

Our main result establishes (1.7) for almost all integers under the above conditions on A:

Theorem 1.2. *Let* $A \in \mathrm{Sp}(2g, \mathbb{Z})$*, with a separable characteristic polynomial, be such that no ratio of distinct eigenvalues is a root of* *unity. Assume further that there are no nontrivial A-invariant rational isotropic subspaces. Then A satisfies QUE as in (1.7) for some set $\mathcal{N}$ of asymptotic density* 1.

One can show that if the characteristic polynomial of A is irreducible, then there are no nontrivial A-invariant rational subspaces.

1.3. **Plan of the proof.**

We establish Theorem 1.2 via the following sequences of steps.

(i) To prove (1.7), it suffices to show it for the basic observables (translation operators) $\mathrm{T}_N(\mathbf{n}) = \mathrm{Op}_N(f)$, $f(\mathbf{x}) = \mathbf{e}(\mathbf{x}\cdot\mathbf{n})$ (see also (1.3)), with frequency $\mathbf{n}$ growing slowly with N.

Assume that the characteristic polynomial $f_A(x) = \det(xI - A)$ is irreducible over the rationals. Then we reduce the problem of estimating high powers $|\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle|^{4\nu}$ of the matrix elements for all normalized eigenfunctions ψ, ψ', to a problem of estimating the number of solutions to the matrix congruence

$$A^{k_1} + \ldots + A^{k_{2\nu}} - A^{\ell_1} - \ldots - A^{\ell_{2\nu}} \equiv O \bmod N,$$

for the zero matrix O with $1 \leqslant k_i, \ell_i \leqslant \mathrm{ord}(A, N)$, $i = 1, \ldots, 2\nu$, see Lemmas 4.1 and 4.3 (since indeed $f_A(x)$ being irreducible implies there is no nontrivial zero-divisor in $\mathbb{Q}^{2g}$).

(ii) In turn, this number can be treated by exponential sums. However this reduction does not work directly due to the lack of nontrivial bounds on such sums except when $N = p$ is a prime, modulo which the characteristic polynomial of A splits completely, in which case we can apply a striking result of Bourgain [1] on short “Mordell sums”. The result, roughly speaking, is that there is some $\gamma > 0$ so that for almost all split primes p,

$$\max_{\psi,\psi'} |\langle \mathrm{T}_p(\mathbf{n})\psi, \psi'\rangle| \leqslant p^{-\gamma}. \tag{1.8}$$

(iii) To take advantage of the bound (1.8) for split primes, we prove that the operators $\mathrm{T}_N(\mathbf{n})$ have a tensor product structure with respect to the Chinese Remainder Theorem, however with some losses depending on certain greatest common divisors. Thus we deal with the operators $\mathrm{T}_p(\mathbf{n})$ via exponential sums and use the trivial bound

$$|\langle \mathrm{T}_M(\mathbf{n})\psi, \psi'\rangle| \leqslant 1,$$

where M is the largest divisor of N without split prime factors.

(iv) Finally, using some results from the *anatomy of integers* (§ 7) we show that for almost all integers N, the saving we obtain from the split primes $p \mid N$, exceeds the losses we incur in our version of the Chinese Remainder Theorem.

(v) When the characteristic polynomial $f_A(x)$ of A is reducible, but separable, we require extra consideration, as the reduction to counting solutions of matrix congruences fails when $\mathbf{n}$ is a non-trivial zero-divisor. We make use of an additional tensor structure to reduce to the setting of congruences for a smaller dimensional case, see § 6 for details.

1.4. **Notation.** Throughout the paper, the notations

$$X = O(Y), \qquad X \ll Y, \qquad X \gg Y$$

are all equivalent to the statement that the inequality $|X| \leqslant cY$ holds with some constant $c > 0$, which may depend on the matrix A, and occasionally, where obvious also on the real parameter ε.

We recall that the additive character with period 1 is denoted by

$$z \in \mathbb{R} \ \mapsto \ \mathbf{e}(z) = \exp(2\pi i z).$$

For an integer $k \geqslant 1$ it is also convenient to define

$$\mathbf{e}_k(z) = \mathbf{e}(z/k).$$

The letter p, with or without indices, always denotes prime numbers.

Given an algebraic number γ we denote by $\mathrm{ord}(\gamma, N)$ its order modulo N (assuming that the ideals generated by γ and N are relatively prime in an appropriate number field). In particular, for an element $\lambda \in \mathbb{F}_{p^s}$, $\mathrm{ord}(\lambda, p)$ represents the order of λ in $\mathbb{F}_{p^s}$.

Similarly, we use $\mathrm{ord}(A, N)$ to denote the order of A modulo N (which always exists if $\gcd(\det A, N) = 1$ and in particular for $A \in \mathrm{Sp}(2g, \mathbb{Z})$).

For a finite set $\mathcal{S}$ we use $\sharp\mathcal{S}$ to denote its cardinality.

As usual, we say that a certain property holds for *almost all* elements of a sequence s_n, $n = 1, 2, \ldots$, if it fails for $o(x)$ terms with $n \leqslant x$, as $x \to \infty$. In particular, we say that it holds for *almost all* primes p and positive integers N if for $x \to \infty$, it fails for $o(x/\log x)$ primes $p \leqslant x$ and $o(x)$ positive integers integers $N \leqslant x$, respectively.

Similarly, we say that a certain property holds for *a positive proportion* of primes p or, equivalently for a set of *positive density*, if for some constant $c > 0$, which throughout this work may depend on the matrix A, for all sufficiently large x it holds for at least $cx/\log x$ primes $p \leqslant x$.

1.5. **Acknowledgement.** During the preparation of this work, P.K. was partially supported by the Swedish Research Council (2020-04036) and the Knut and Alice Wallenberg foundation (KAW 2019.0503).

A.O. was supported by the Australian Research Council, DP230100530 and she gratefully acknowledges the hospitality and support of the Max Planck Institute for Mathematics in Bonn, where parts of this work has been carried out. Z.R. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 786758), and by the ISF-NSFC joint research program (grant No. 3109/23). I.S. was supported by the Australian Research Council, DP230100534.

2. A Chinese Remainder Theorem for the operators $\mathrm{T}_N(\mathbf{n})$

2.1. **Observables.**

We begin by defining the mixed translation operators. Given $r \in \mathbb{Z}$, coprime to N, and $\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbb{Z}^g \times \mathbb{Z}^g = \mathbb{Z}^{2g}$, we define a unitary operator $\mathrm{T}_N^{(r)}(\mathbf{n}) : \mathcal{H}_N \to \mathcal{H}_N$ by

$$\left(\mathrm{T}_N^{(r)}(\mathbf{n})\varphi\right)(\mathbf{Q}) = \mathbf{e}_{2N}(r\mathbf{n}_1 \cdot \mathbf{n}_2)\, \mathbf{e}_N(r\mathbf{n}_2 \cdot \mathbf{Q})\varphi(\mathbf{Q} + \mathbf{n}_1).$$

We have

$$\mathrm{T}_N^{(r)}(\mathbf{m})\, \mathrm{T}_N^{(r)}(\mathbf{n}) = \mathbf{e}_{2N}\left(r\omega\left(\mathbf{m}, \mathbf{n}\right)\right) \mathrm{T}_N^{(r)}(\mathbf{m} + \mathbf{n}), \tag{2.1}$$

where $\omega\left(\mathbf{m}, \mathbf{n}\right)$ is defined by (1.1). In particular, taking powers gives

$$\left(\mathrm{T}_N^{(r)}\left(\mathbf{n}\right)\right)^k = \mathrm{T}_N^{(r)}(k\mathbf{n}).$$

The canonical commutation relations can be encapsulated in the relations

$$\mathrm{T}_N^{(r)}(\mathbf{n})\, \mathrm{T}_N^{(r)}(\mathbf{m}) = \mathbf{e}_N\left(r\omega\left(\mathbf{n}, \mathbf{m}\right)\right) \mathrm{T}_N^{(r)}(\mathbf{m})\, \mathrm{T}_N^{(r)}(\mathbf{n})$$

and

$$(\mathrm{T}_N^{(r)}(\mathbf{n}))^N = \mathrm{T}_N^{(r)}(N\mathbf{n}) = (-1)^{rN\mathbf{n}_1\cdot\mathbf{n}_2} I, \quad \mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2).$$

For each function $f \in C^\infty(\mathbb{T}^{2g})$ on the classical phase space (an "observable"), one associates an operator $\mathrm{Op}_{N,r}(f)$ on $\mathcal{H}_N$, analogous to a pseudo-differential operator with symbol f, by

$$\mathrm{Op}_{N,r}(f) = \sum_{\mathbf{n}\in\mathbb{Z}^{2g}} \widehat{f}(\mathbf{n})\, \mathrm{T}_N^{(r)}(\mathbf{n}),$$

where $\widehat{f}(\mathbf{n})$ are defined by (1.4). If f is real valued, then $\mathrm{Op}_{N,r}(f)$ is self-adjoint.

When $r = 1$, we recover the definitions of $\mathrm{T}_N = \mathrm{T}_N^{(1)}$ and $\mathrm{Op}_N = \mathrm{Op}_{N,1}$ in (1.2) and (1.3), respectively.

Let $A \in \mathrm{Sp}(2g, \mathbb{Z})$, satisfying the parity condition $A = I \bmod 2$. Fix $N \geq 1$ and r coprime to N. Then there is a unitary operator

$U_{N,r}(A): \mathcal{H}_N \to \mathcal{H}_N$, unique up to a scalar multiple, so that we have the exact Egorov property

$$U_{N,r}(A)^* \operatorname{T}_N^{(r)}(\mathbf{n}) U_{N,r}(A) = \operatorname{T}_N^{(r)}(\mathbf{n}A), \tag{2.2}$$

fo all $\mathbf{n} \in \mathbb{Z}^{2g}$, which is a full analogue of (1.5).

2.2. **The Chinese Remainder Theorem and a tensor product structure.** Assume that the inverse Planck constant N factors as $N = N_1 \cdot N_2$ with N_1, N_2 coprime. We then use the Chinese Remainder Theorem $\iota : \mathbb{Z}/N\mathbb{Z} \cong \mathbb{Z}/N_1\mathbb{Z} \oplus \mathbb{Z}/N_2\mathbb{Z}$ to get an isomorphism

$$\iota^* : L^2((\mathbb{Z}/N_1\mathbb{Z})^g) \otimes L^2((\mathbb{Z}/N_2\mathbb{Z})^g) = \mathcal{H}_{N_1} \otimes \mathcal{H}_{N_2} \cong \mathcal{H}_N = L^2((\mathbb{Z}/N\mathbb{Z})^g)$$

so that

$$\iota^*(\varphi_1 \otimes \varphi_2)(\mathbf{Q}) = \varphi_1(\mathbf{Q} \bmod N_1) \cdot \varphi_2(\mathbf{Q} \bmod N_2). \tag{2.3}$$

The tensor product $\mathcal{H}_{N_1} \otimes \mathcal{H}_{N_2}$ carries the inner product

$$\|\varphi_1 \otimes \varphi_2\| = \|\varphi_1\| \cdot \|\varphi_2\|$$

and ι^* is actually an isometry, because it maps the orthonormal basis of tensor products of normalized delta functions to normalized delta functions:

$$\iota^* \left(N_1^{g/2} \delta_{\mathbf{u}} \otimes N_2^{g/2} \delta_{\mathbf{v}} \right) = N^{g/2} \delta_{\mathbf{w}}$$

where $\mathbf{w} = \mathbf{u} \bmod N_1$, $\mathbf{w} = \mathbf{v} \bmod N_2$.

Assume $N = N_1 \cdot N_2$ with $N_1 > 1$, $N_2 > 1$ coprime. Fix nonzero $r_1, r_2 \in \mathbb{Z}$ so that

$$N_2 r_2 + N_1 r_1 = 1. \tag{2.4}$$

Necessarily r_2 is coprime to N_1 and r_1 is coprime to N_2.

Lemma 2.1. *For $\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2) \in \mathbb{Z}^g \times \mathbb{Z}^g$, the mixed translation operator $\operatorname{T}_N(\mathbf{n}) = \operatorname{T}_N^{(1)}(\mathbf{n})$ is mapped, via the isomorphism ι^*, to $\operatorname{T}_{N_1}^{(r_2)}(\mathbf{n}) \otimes \operatorname{T}_{N_2}^{(r_1)}(\mathbf{n})$:*

$$\operatorname{T}_N(\mathbf{n}) \iota^* (\varphi_1 \otimes \varphi_2) = \iota^* \left(\left(\operatorname{T}_{N_1}^{(r_2)}(\mathbf{n}) \varphi_1 \right) \otimes \left(\operatorname{T}_{N_2}^{(r_1)}(\mathbf{n}) \varphi_2 \right) \right).$$

Proof. Inserting (2.4) gives

$$\begin{aligned} \mathbf{e}_{2N}(\mathbf{n}_1 \cdot \mathbf{n}_2) &= \mathbf{e}\left(\frac{(N_2 r_2 + N_1 r_1)\mathbf{n}_1 \cdot \mathbf{n}_2}{2N_1 N_2} \right) \\ &= \mathbf{e}\left(\frac{r_2 \mathbf{n}_1 \cdot \mathbf{n}_2}{2N_1} \right) \mathbf{e}\left(\frac{r_1 \mathbf{n}_1 \cdot \mathbf{n}_2}{2N_2} \right) \end{aligned}$$

and for $\mathbf{y} \in (\mathbb{Z}/N\mathbb{Z})^g$,

$$\begin{aligned}\mathbf{e}_N(\mathbf{n}_2 \cdot \mathbf{y}) &= \mathbf{e}\left(\frac{(N_2 r_2 + N_1 r_1)\mathbf{n}_2 \cdot \mathbf{y}}{N_1 N_2}\right)\\ &= \mathbf{e}\left(\frac{r_2 \mathbf{n}_2 \cdot \mathbf{y}}{N_1}\right) \cdot \mathbf{e}\left(\frac{r_1 \mathbf{n}_2 \cdot \mathbf{y}}{N_2}\right).\end{aligned}$$

By definition (1.2),

$$\{\mathrm{T}_N(\mathbf{n})\iota^*(\varphi_1 \otimes \varphi_2)\}(\mathbf{y}) = \mathbf{e}_{2N}(\mathbf{n}_1 \cdot \mathbf{n}_2)\, \mathbf{e}_N(\mathbf{n}_2 \cdot \mathbf{y})\iota^*(\varphi_1 \otimes \varphi_2)(\mathbf{y} + \mathbf{n}_1)$$

so that, using (2.3), we obtain

$$\begin{aligned}\{\mathrm{T}_N(\mathbf{n})\iota^*(\varphi_1 \otimes \varphi_2)\}(\mathbf{y}) &= \mathbf{e}_{2N}(\mathbf{n}_1 \cdot \mathbf{n}_2)\, \mathbf{e}_N(\mathbf{n}_2 \cdot \mathbf{y})\iota^*(\varphi_1 \otimes \varphi_2)(\mathbf{y} + \mathbf{n}_1)\\ &= \mathbf{e}\left(\frac{r_2 \mathbf{n}_1 \cdot \mathbf{n}_2}{2N_1}\right) e\left(\frac{r_2 \mathbf{n}_2 \cdot \mathbf{y}}{N_1}\right) \varphi_1(\mathbf{y} + \mathbf{n}_1)\\ &\qquad \cdot \mathbf{e}\left(\frac{r_1 \mathbf{n}_1 \cdot \mathbf{n}_2}{2N_2}\right) \mathbf{e}\left(\frac{r_1 \mathbf{n}_2 \cdot \mathbf{y}}{N_2}\right) \varphi_2(\mathbf{y} + \mathbf{n}_1)\\ &= \left(\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n})\varphi_1\right)(\mathbf{y}) \cdot \left(\mathrm{T}_{N_2}^{(r_1)}(\mathbf{n})\varphi_2\right)(\mathbf{y})\\ &= \iota^*\left(\left(\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}) \otimes \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n})\right)(\varphi_1 \otimes \varphi_2)\right)(\mathbf{y})\end{aligned}$$

as claimed. □

2.3. **Factorization of the quantized map.**

We continue to assume a factorization of $N = N_1 \cdot N_2$ with $N_1 > 1$, $N_2 > 1$ coprime, and such that r_1, r_2 satisfy (2.4). Then the Chinese Remainder Theorem induces an isometry

$$\mathcal{H}_N \simeq \mathcal{H}_{N_1} \otimes \mathcal{H}_{N_2},$$

which is respected by the translation operators (see Lemma 2.1). Furthermore, from now on, we identify the spaces $\mathcal{H}_N$ and $\mathcal{H}_{N_1} \otimes \mathcal{H}_{N_2}$ and thus we do not use the isomorphism map ι^* anymore. We argue that we get a corresponding factorization of the quantized map $U_N(A) = U_{N,1}(A)$ defined by (1.5) as a tensor product:

Lemma 2.2. *There is some $\zeta \in \mathbb{C}$, with $|\zeta| = 1$, such that we have a factorization*

$$U_N(A) = \zeta U_{N_1,r_2}(A) \otimes U_{N_2,r_1}(A).$$

Proof. We saw in Lemma 2.1 that

$$\mathrm{T}_N(\mathbf{n}A) = \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}A) \otimes \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}A).$$

By (2.2), we have

$$\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}A) = U_{N_1,r_2}(A)^*\, \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}) U_{N_1,r_2}(A)$$

and

$$\mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}A) = U_{N_2,r_1}(A)^* \, \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}) U_{N_2,r_1}(A).$$

Hence $\widetilde{U} = U_{N_1,r_2}(A) \otimes U_{N_2,r_1}(A)$ satisfies

$$\begin{aligned} \widetilde{U}^* \, \mathrm{T}_N(\mathbf{n}) \widetilde{U} &= \widetilde{U}^* \left(\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}) \otimes \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}) \right) \widetilde{U} \\ &= \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}A) \otimes \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}A) = \mathrm{T}_N(\mathbf{n}A) \end{aligned}$$

for all $\mathbf{n} \in \mathbb{Z}^{2g}$. Since $U_N(A)$ is the unique (up to a scalar multiple) unitary operator satisfying this relation, we must have that $\widetilde{U}$ is a scalar multiple, of absolute value one, of $U_N(A)$. □

3. Bounding T_N via the tensor product structure

3.1. **Basic properties of tensor products.** We first recall a useful identity regarding operator norms of tensor products. Let V, W be finite dimensional inner product spaces, let $\mathrm{T}_V : V \to V$ and $\mathrm{T}_W : W \to W$ be linear maps, and let $\| \mathrm{T}_V \|$ and $\| \mathrm{T}_W \|$ denote the operator norms of T_V and T_W, respectively. There is a natural inner product on the tensor product $V \otimes W$ — given orthonormal bases $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$ for V, W, respectively, declare $\{v_i \otimes w_j\}_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m}$ to be an orthonormal basis for $V \otimes W$. We then have the relation

$$\| \mathrm{T}_V \otimes \mathrm{T}_W \| = \| \mathrm{T}_V \| \cdot \| \mathrm{T}_W \|, \tag{3.1}$$

between the operator norms (cf. [21, Page 299, Proposition]).

3.2. **Eigenspace decomposition in the coprime case.** Assume that T_V and T_W are both diagonalizable, with eigenvalues being roots of unity (in particular there exists, say minimal, integers $t_1, t_2 > 0$ such that $\mathrm{T}_V^{t_1} = I_V$ and $\mathrm{T}_W^{t_2} = I_W$, where I_V and I_W are the corresponding identity operators; note that we do not assume that T_V and T_W have the same dimensions).

The eigenspaces of $\mathrm{T}_V \otimes \mathrm{T}_W$ are particularly easy to describe in terms of the eigenspaces of T_V and T_W when $\gcd(t_1, t_2) = 1$. Namely, let $\{V_i\}_i$ denote the eigenspaces of T_V, and let $\{W_j\}_j$ denote the eigenspaces of T_W; here we allow both $\mathrm{T}_V, \mathrm{T}_W$ to have eigenvalues with multiplicities. The eigenspaces of $\mathrm{T}_V \otimes \mathrm{T}_W$ are then given by $\{V_i \otimes W_j\}_{i,j}$. Further, if the eigenvalue associated with V_i is denoted μ_i and the eigenvalue associated with W_j is denoted by ν_j, all eigenvalues of $\mathrm{T}_V \otimes \mathrm{T}_W$ are of the form $\lambda_{i,j} = \mu_i \nu_j$. In particular, if $\gcd(t_1, t_2) = 1$ we find that $\mu_{i_1} \nu_{j_1} = \mu_{i_2} \nu_{j_2}$ implies that $i_1 = i_2$ and $j_1 = j_2$. Informally, all multiplicities arise by combining multiplicities from the V_i, W_j eigenspaces (note that this is not true if $\gcd(t_1, t_2) > 1$).

3.3. **Bounds using the tensor product structure.** We now bound matrix coefficients of the special form $\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle$, where, as in § 1.3,

$$\mathrm{T}_N(\mathbf{n}) = \mathrm{Op}_N(\,\mathbf{e}\,(\mathbf{x}\cdot\mathbf{n}))$$

and ψ, ψ' are eigenfunctions of $U_N(A)$.

Let B be an element of $\mathrm{Sp}(2g, \mathbb{Z})$. Assume that $N = N_1N_2$ with coprime integers $N_1 > 1$, $N_2 > 1$. Let t_i denote the order of $B \mod N_i$, $i = 1, 2$. Further assume that

$$\gcd(t_1, t_2) = 1. \tag{3.2}$$

Let $r_1, r_2 \in \mathbb{Z}$ satisfy (2.4). Taking $A = B$ in Lemmas 2.1 and 2.2, we find that

$$\mathrm{T}_N(\mathbf{n}) = \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}) \otimes \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n}) \quad \text{and} \quad U_N(B) = \zeta U_{N_1,r_2}(B) \otimes U_{N_2,r_1}(B)$$

for some $\zeta \in \mathbb{C}^*$ with $|\zeta| = 1$.

Lemma 3.1. *Let ψ, ψ' denote norm one eigenfunctions of $U_N(B)$. With assumptions as above, we then have*

$$|\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle| \leqslant \max_{\varphi,\varphi' \in \Phi_{N_1,r_2}} |\langle \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n})\varphi, \varphi'\rangle|,$$

where Φ_{N_1,r_2} denotes the set of all eigenfunctions of $U_{N_1,r_2}(B)$ of norm one.

Proof. Let E, E' denote the eigenspaces of $U_N(B)$ containing ψ, ψ', and let λ, λ' denote the corresponding eigenvalues. By the discussion in § 3.2, we have

$$E = V_1 \otimes V_2, \tag{3.3}$$

where V_1 and V_2 are eigenspaces of $U_{N_1,r_2}(B)$ and $U_{N_2,r_1}(B)$, respectively, and similarly we have $E' = V_1' \otimes V_2'$. Thus, if we let $S = P_{E'}\,\mathrm{T}_N(\mathbf{n})P_E$, with $P_E : \mathcal{H}_N \to E$ denoting the orthogonal projection onto E (and similarly for $P_{E'}$), we have

$$\max_{\substack{\psi\in E,\, \psi'\in E',\\ \|\psi\|=\|\psi'\|=1}} |\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle| = \|S\|. \tag{3.4}$$

Now, the decomposition (3.3) and Lemma 2.1, after a simple calculation, give that

$$S = S_1 \otimes S_2 \tag{3.5}$$

where $S_1 = P_{V_1'}\,\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n})P_{V_1}$ and $S_2 = P_{V_2'}\,\mathrm{T}_{N_2}^{(r_1)}(\mathbf{n})P_{V_2}$. Since both S_1, S_2 arise as compositions of the unitary maps $\mathrm{T}_{N_1}^{(r_2)}(\mathbf{n}), \mathrm{T}_{N_2}^{(r_1)}(\mathbf{n})$ with orthogonal projections, they are both sub-unitary, and we have the trivial bounds $\|S_1\|, \|S_2\| \leqslant 1$. Thus, by the operator norm identity of tensor products (3.1), we see from (3.5) that

$$\|S\| = \|S_1\| \cdot \|S_2\| \leqslant \|S_1\|.$$

Using this, together with

$$\|S_1\| = \max_{\substack{\varphi \in V_1, \varphi' \in V_1' \\ \|\varphi\| = \|\varphi'\| = 1}} |\langle \mathrm{T}_{N_1}^{(r_2)}(\mathbf{n})\varphi, \varphi'\rangle|,$$

and recalling (3.4), the result now follows. □

We now consider a more general case when instead of the coprimality condition (3.2) we have

$$\gcd(t_1, t_2) = d.$$

Since any eigenfunction of $U_N(A)$ is an eigenfunction for $U_N\left(A^d\right)$ for any integer $d > 0$, we have

$$\max_{\psi, \psi' \in \Psi_N} |\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle| \leqslant \max_{\widetilde{\psi}, \widetilde{\psi}' \in \widetilde{\Psi}_{N,d}} |\langle \mathrm{T}_N(\mathbf{n})\widetilde{\psi}, \widetilde{\psi}'\rangle|, \tag{3.6}$$

where Ψ_N and $\widetilde{\Psi}_{N,d}$ denote the set of normalized eigenfunctions of $U_N(A)$ and $U_N(A^d)$, respectively.

4. Congruences and exponential sums

4.1. Reduction to a counting problem. For a (row) vector $\mathbf{n} \in \mathbb{Z}^{2g}$, $\mathbf{n} \neq \mathbf{0} \bmod N$, we denote by $Q_{2\nu}(N; \mathbf{n})$ the number of solutions of the congruence

$$\mathbf{n}\left(A^{k_1} + \ldots + A^{k_{2\nu}} - A^{\ell_1} - \ldots - A^{\ell_{2\nu}}\right) \equiv \mathbf{0} \bmod N, \tag{4.1}$$

with $1 \leqslant k_i, \ell_i \leqslant \mathrm{ord}(A, N)$, $i = 1, \ldots, 2\nu$.

The key inequality below connects the 4ν-th moment associated to the basic observables $\mathrm{T}_N^{(r)}(\mathbf{n})$ with the number of solutions to the system (4.1). This kind of inequality (for $\nu = 1$) underlies the argument of [19], and also the argument of [2].

Lemma 4.1. *Let $\mathbf{0} \neq \mathbf{n} \in \mathbb{Z}^{2g}$ and let r be an integer coprime to N. Then*

$$\max_{\psi, \psi'} \left|\langle \mathrm{T}_N^{(r)}(\mathbf{n})\psi, \psi'\rangle\right|^{4\nu} \leqslant N^g \frac{Q_{2\nu}(N; \mathbf{n})}{\mathrm{ord}(A, N)^{4\nu}}, \tag{4.2}$$

where the maximum is taken over all pairs of normalized eigenfunctions of $U_{N,r}(A)$.

Proof. We abbreviate $\tau = \operatorname{ord}(A,N)$. Given a pair (ψ,ψ') of normalized eigenfunctions of $U_{N,r}(A)$ with eigenvalues λ,λ', put $\mu = \lambda'/\lambda$ and note that μ is a root of unity (since λ and λ' are). Let

$$D(\mathbf{n}) = \frac{1}{\tau}\sum_{i=1}^{\tau} U_{N,r}(A)^{-i}\,\mathrm{T}_N^{(r)}(\mathbf{n})U_{N,r}(A)^i\mu^i = \frac{1}{\tau}\sum_{i=1}^{\tau}\mathrm{T}_N^{(r)}(\mathbf{n}A^i)\mu^i$$

be the μ-twisted time averaged observable, where the last equality comes from (2.2). Then for any pair of eigenfunctions (ψ,ψ') of $U_{N,r}(A)$, with eigenvalues λ and λ', we have

$$\langle \mathrm{T}_N^{(r)}(\mathbf{n})\psi,\psi'\rangle = \langle D(\mathbf{n})\psi,\psi'\rangle,$$

see also the proof of [19, Proposition 4]. Put $H(\mathbf{n}) = D(\mathbf{n})^*D(\mathbf{n})$; note that $H(\mathbf{n})$ is Hermitian. Clearly

$$|\langle D(\mathbf{n})\psi,\psi'\rangle| \leqslant \|D(\mathbf{n})\| = \|H(\mathbf{n})\|^{1/2},$$

where $\|H(\mathbf{n})\|$ denotes the operator norm of $H(\mathbf{n})$. Therefore for any $\nu \geq 1$,

$$|\langle \mathrm{T}_N^{(r)}(\mathbf{n})\psi,\psi'\rangle|^{4\nu} \leqslant \|H(\mathbf{n})\|^{2\nu} = \|H(\mathbf{n})^\nu\|^2.$$

We bound the operator norm by the *Hilbert–Schmidt norm* and obtain

$$\|H(\mathbf{n})^\nu\|^2 \leqslant \|H(\mathbf{n})^\nu\|_{HS}^2 = \operatorname{tr}((H(\mathbf{n})^\nu)^*H(\mathbf{n})^\nu) = \operatorname{tr}(H(\mathbf{n})^{2\nu}),$$

where tr denotes the operator trace. Finally, compute

$$\begin{aligned} H(\mathbf{n})^{2\nu} = \frac{1}{\tau^{4\nu}} &\sum_{k_1,\ldots,k_{2\nu},\ell_1\ldots,\ell_{2\nu}=1}^{\tau} \\ &\qquad\times\prod_{j=1}^{2\nu}\left(\mathrm{T}_N^{(r)}\left(\mathbf{n}A^{k_j}\right)\mathrm{T}_N^{(r)}\left(-\mathbf{n}A^{\ell_j}\right)\right)\mu^{\sum_{j=1}^{2\nu}(k_j-\ell_j)} \\ = \frac{1}{\tau^{4\nu}} &\sum_{k_1,\ldots,k_{2\nu},\ell_1\ldots,\ell_{2\nu}=1}^{\tau}\gamma(\mathbf{k},\boldsymbol{\ell},\mu)\,\mathrm{T}_N^{(r)}\left(\mathbf{n}\sum_{j=1}^{2\nu}\left(A^{k_j}-A^{\ell_j}\right)\right)\end{aligned}$$

with some complex coefficients $\gamma(\mathbf{k},\boldsymbol{\ell},\mu)$, satisfying $|\gamma(\mathbf{k},\boldsymbol{\ell},\mu)| = 1$, where $\mathbf{k} = (k_1,\ldots,k_{2\nu})$ and $\boldsymbol{\ell} = (\ell_1\ldots,\ell_{2\nu})$, and where the last equality comes from (2.1). Taking the trace and using

$$|\operatorname{tr}\mathrm{T}_N^{(r)}(\mathbf{m})| = \begin{cases} N^g & \text{if } \mathbf{m} = \mathbf{0} \bmod N,\\ 0 & \text{otherwise,}\end{cases}$$

we find

$$\left|\langle \mathrm{T}_N^{(r)}(\mathbf{n})\psi,\psi'\rangle\right|^{4\nu} \leqslant \operatorname{tr}\left(H(\mathbf{n})^{2\nu}\right) \leqslant \frac{N^g}{\tau^{4\nu}}Q_{2\nu}(N;\mathbf{n})$$

which concludes the proof. □

Remark 4.2. *If the right hand side of (4.2) tends to zero, then (1.7) is satisfied, that is, all eigenfunctions of $U_N(A)$ are uniformly distributed, and more generally, all off-diagonal matrix coefficients tend to zero. Thus Lemma 4.1 reduces the problem (1.7) to a purely arithmetic issue.*

4.2. **Linear independence of matrix powers.** The primal goal of this section is to show that if the characteristic polynomial f_A of A is separable over $\mathbb{Q}$ we can essentially eliminate the dependence on the vector $\mathbf{n}$ in our argument, except in some special cases. In particular, instead of $Q_{2\nu}(N;\mathbf{n})$ we can consider the number of solutions of the congruence

$$A^{k_1} + \ldots + A^{k_{2\nu}} \equiv A^{\ell_1} + \ldots + A^{\ell_{2\nu}} \bmod N, \tag{4.3}$$

with $1 \leqslant k_i, \ell_i \leqslant \operatorname{ord}(A, N)$, $i = 1, \ldots, 2\nu$. This is based on the following result which is also used in our bounds on exponential sums. However, we first need to introduce the notion of zero-divisors amongst the row vectors $\mathbf{n} \in \mathbb{Z}^{2g}$. For this we first identify $\mathbb{Q}^{2g} \cong \mathbb{Q}[X]/(f_A(X))$ as a $\mathbb{Q}[A]$ module. We say that $\mathbf{n}$ is a *zero-divisor*, if its image $\widetilde{\mathbf{n}} \in \mathbb{Q}[X]/(f_A(X))$ is a zero-divisor in this module (we follow the convention that zero is also a zero divisor, call all other non-zero divisors *nontrivial.*)

Lemma 4.3. *Let $A \in \mathrm{Sp}(2g, \mathbb{Z})$ have a separable characteristic polynomial. Then for any row vector $\mathbf{n} \in \mathbb{Z}^{2g}$, which is not a zero-divisor, we have:*

(i) *the vectors $\mathbf{n}, \mathbf{n}A, \ldots, \mathbf{n}A^{2g-1}$ are linearly independent;*
(ii) *there exists some $p_0(A)$, depending only on A, such that for all primes $p > p_0(A)\|\mathbf{n}\|_2^{2g}$, the vectors $\mathbf{n}, \mathbf{n}A, \ldots, \mathbf{n}A^{2g-1}$ are linearly independent modulo p.*

Proof. In the case when the characteristic polynomial is irreducible, Part (i) is proved in [19, page 210] (for $n = 2$) and in the proof of [20, Theorem 2.5] (which is done over a finite field but it remains valid over any field).

In our more general case of separability, assume that the vectors $\mathbf{n}A^i$, $i = 0, \ldots, 2g-1$, are linearly dependent over $\mathbb{Q}$, that is, there is a linear relation

$$\sum_{i=0}^{2g-1} c_i \mathbf{n} A^i = \mathbf{n} \left(\sum_{i=0}^{2g-1} c_i A^i \right) = \mathbf{0}$$

for some $c_i \in \mathbb{Q}$ not all zero. Since $\mathbf{n}$ is not a zero-divisor, we obtain

$$\sum_{i=0}^{2g-1} c_i A^i = \mathbf{0}.$$

However, this shows that the minimal polynomial of A has degree at most $2g-1$, which contradicts the fact that the minimal polynomial is f_A since it is separable. This concludes the proof of Part (i).

To show Part (ii) one considers the determinant of the matrix having rows $\mathbf{n}, \ldots, \mathbf{n}A^{2g-1}$ whose vanishing is equivalent to linear independence; it is an integer, nonzero by Part (i), hence for all primes p not dividing it we have linear independence mod p. □

4.3. **Reduction to a system of exponential equations.** We now consider (4.1) for a prime $N = p$.

Let $A \in \mathrm{Sp}(2g, \mathbb{Z})$ have separable characteristic polynomial $f_A \in \mathbb{Z}[X]$. Assume that p is large enough so that f_A is separable modulo p, which also implies that A is diagonalisable over $\overline{\mathbb{F}}_p$.

Next, let

$$f_A(X) = h_1(X) \cdots h_t(X) \bmod p$$

be the factorization of f_A into irreducible factors $h_i \in \mathbb{F}_p[X]$ of degrees $d_i = \deg h_i$, $i = 1 \ldots, t$. In particular, any root of h_i belongs to $\mathbb{F}_{p^{d_i}}$, $i = 1 \ldots, t$. For each h_i we fix a root $\lambda_i \in \mathbb{F}_{p^{d_i}}$ and consider the system of equations (in the algebraic closure of $\mathbb{F}_p$)

$$\lambda_i^{k_1} + \cdots + \lambda_i^{k_{2\nu}} = \lambda_i^{\ell_1} + \cdots + \lambda_i^{\ell_{2\nu}}, \qquad i = 1 \ldots, t, \tag{4.4}$$

with $1 \leqslant k_j, \ell_j \leqslant \mathrm{ord}(A, p)$, $j = 1, \ldots, 2\nu$. It is easy to see that for all choices of roots $\lambda_1, \ldots, \lambda_t$ we get equivalent systems.

Next, we reduce counting the number of solutions to (4.3) to counting the number of solutions to (4.4).

In fact our treatment depends only on the degrees $d_1, \ldots, d_t$ and so we denote the number of solutions to (4.4) by $R_{2\nu}(d_1, \ldots, d_t; p)$.

Lemma 4.4. *Under the above assumptions, there exists some $p_0(A)$, depending only on A, such that for any vector $\mathbf{n} \in \mathbb{Z}^{2g}$, which is not a zero divisor, and $p > p_0(A)\|\mathbf{n}\|_2^{2g}$ we have*

$$Q_{2\nu}(p; \mathbf{n}) = R_{2\nu}(d_1, \ldots, d_t; p).$$

Proof. Let us denote

$$B = A^{k_1} + \cdots + A^{k_{2\nu}} - A^{\ell_1} - \cdots - A^{\ell_{2\nu}}.$$

Multiplying (4.1) by powers of A, we conclude that

$$\left(\mathbf{n}A^i\right) B \equiv \mathbf{0} \bmod p, \qquad i = 0, \ldots, 2g-1,$$

which is equivalent to

$$\begin{pmatrix} \mathbf{n} \\ \mathbf{n}A \\ \dots \\ \mathbf{n}A^{2g-1} \end{pmatrix} B \equiv \mathbf{0} \bmod p.$$

From Lemma 4.3 (ii), there exists some $p_0(A)$, depending only on A, such that for $p > p_0(A)\|\mathbf{n}\|_2^{2g}$ all rows $\mathbf{n}, \mathbf{n}A, \dots, \mathbf{n}A^{2g-1}$ are linearly independent modulo p, and thus from the above we conclude that B vanishes over $\mathbb{F}_p$.

Since A is diagonalisable over $\overline{\mathbb{F}}_p$, the equation above is equivalent to

$$\Lambda^{k_1} + \ldots + \Lambda^{k_{2\nu}} \equiv \Lambda^{\ell_1} + \ldots + \Lambda^{\ell_{2\nu}} \bmod p,$$

where Λ is a diagonal matrix with elements on the diagonal all the roots of h_i, $i = 1, \dots, t$. Since for each irreducible factor of f modulo p, all roots are conjugate (that is, the roots of h_i in $\mathbb{F}_{p^{d_i}}$ are $\lambda_i, \lambda_i^p, \dots, \lambda_i^{p^{d_i-1}}$), we conclude the proof. □

5. Multiplicative orders and exponential sums

5.1. **Ergodicity and the order modulo p.** It is natural that our argument, as in [2, 19, 20], rests on various results on multiplicative orders.

We begin by showing that the multiplicative orders of the eigenvalues of $A \in \mathrm{Sp}(2g, \mathbb{Z})$, and their ratios, are sufficiently large for almost all primes. The argument is a modification of that of Hooley [13].

We recall the definition of $\mathrm{ord}(\lambda, p)$ in § 1.4 and also that we say that p is split prime if the characteristic polynomial of the matrix A splits completely modulo p.

Lemma 5.1. *Assume that $A \in \mathrm{Sp}(2g, \mathbb{Z})$ has separable characteristic polynomial and that no eigenvalue or ratio of distinct eigenvalues is a root of unity. Let $\lambda_1, \dots, \lambda_{2g}$ be the eigenvalues of A. Then for almost all split primes p we have*

$$\mathrm{ord}(\lambda_i, p), \mathrm{ord}(\lambda_i/\lambda_j, p) > p^{1/2}/\log p, \quad 1 \leqslant i \neq j \leqslant 2g.$$

Proof. For a sufficiently large $Y \geqslant 2$, let

$$A(Y) = \prod_{n \leqslant Y} \prod_{1 \leqslant i \leqslant 2g} \mathrm{Nm}_{K/\mathbb{Q}}(\lambda_i^n - 1) \prod_{1 \leqslant j < h \leqslant 2g} \mathrm{Nm}_{K/\mathbb{Q}}(\lambda_j^n - \lambda_h^n),$$

where $\mathrm{Nm}_{K/\mathbb{Q}}(\zeta)$ is the *norm* of $\zeta \in K = Q(\lambda_1, \dots, \lambda_n)$ in $\mathbb{Q}$. Note that $A(Y) \neq 0$ because of the condition on the avoidance of roots of unity among the eigenvalues and their ratios, and $A(Y) \in \mathbb{Z}$ since all eigenvalues are algebraic integers. Since

$$\mathrm{Nm}_{K/\mathbb{Q}}\left(\lambda_i^n - 1\right) = \prod_{\sigma \in \mathrm{Gal}(K/\mathbb{Q})} \left(\sigma(\lambda_i)^n - 1\right)$$

and

$$\mathrm{Nm}_{K/\mathbb{Q}}\left(\lambda_j^n - \lambda_h^n\right) = \prod_{\sigma \in \mathrm{Gal}(K/\mathbb{Q})} \left(\sigma(\lambda_j)^n - \sigma(\lambda_h)^n\right),$$

where both products are over all automorphisms σ from the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ of K over $\mathbb{Q}$, and thus

$$\log \mathrm{Nm}_{K/\mathbb{Q}}(\lambda_i^n - 1), \log \mathrm{Nm}_{K/\mathbb{Q}}(\lambda_j^n - \lambda_h^n) \ll n,$$

we see that

$$\log |A(Y)| \ll Y^2. \tag{5.1}$$

Let $\mathcal{P}(Y)$ be the set of primes for which

$$\min_{1 \leqslant i \leqslant 2g} \min_{1 \leqslant j < h \leqslant 2g} \{\mathrm{ord}(\lambda_i, p), \mathrm{ord}(\lambda_j/\lambda_h, p)\} \leqslant Y.$$

We observe that for $p \in \mathcal{P}(Y)$, we must have $p \mid A(Y)$, and hence

$$\sharp\mathcal{P}(Y) \leqslant \omega\left(A(Y)\right), \tag{5.2}$$

where, as usual, $\omega(k)$ denotes the number of prime divisors of the integer $k \geqslant 1$. From the trivial observation that $\omega(k)! \leqslant k$ and the Stirling formula, we derive

$$\omega(k) \ll \frac{\log k}{\log\log(k+2)}, \qquad k \geqslant 1. \tag{5.3}$$

Putting together (5.1), (5.2) and (5.3), we see that

$$\sharp\mathcal{P}(Y) \ll Y^2/\log Y.$$

Since the number of primes $p \leqslant X$ is $\pi(X) \sim X/\log X$, we can take $Y = \sqrt{X}/\log X$ to assure that for all but $o\left(\pi(X)\right)$ primes $p \leqslant X$, we have

$$\mathrm{ord}(\lambda_i, p), \mathrm{ord}(\lambda_i/\lambda_j, p) > \sqrt{X}/\log X \geq \sqrt{p}/\log p, \quad 1 \leqslant i \neq j \leqslant 2g.$$

Since splitting fields of polynomials are Galois extensions, by the Chebotarev Density Theorem, see [14, Theorem 21.2], for a positive proportion of primes p, see our convention in § 1.4, the characteristic polynomial of the matrix A splits modulo p. This concludes the proof. □

5.2. **Relation with short exponential sums.** As discussed in § 4.1, one relates the uniform distribution of the eigenfunctions of the operator $U_N(A)$, as well as the decay of off-diagonal matrix elements, to bounding the number of solutions $Q_{2\nu}(N;\mathbf{n})$ for $\mathbf{n}\in\mathbb{Z}^{2g}$ to the matrix congruence (4.1), see Lemma 4.1.

Following the discussion after Theorem 1.2 and Lemma 4.1, we thus reduce the problem to showing that

$$Q_{2\nu}(p;\mathbf{n}) = o\left(\frac{\operatorname{ord}(A,p)^{4\nu}}{p^g}\right)$$

for a set of 'good' primes p for which the characteristic polynomial of A splits completely over $\mathbb{F}_p$, with eigenvalues $\lambda_i\in\mathbb{F}_p^*$, $i=1,\dots,2g$.

In turn, using the orthogonality of exponential sums, this leads us to a problem of obtaining nontrivial cancellation in exponential sums of the form

$$\sum_{j=1}^{\operatorname{ord}(A,p)} \mathbf{e}_p\left(\alpha_1\lambda_1^j+\ldots+\alpha_{2g}\lambda_{2g}^j\right)$$

for $(\alpha_1,\dots,\alpha_{2g})\in\mathbb{F}_p^{2g}$.

These exponential sums are not treatable by algebro-geometric methods of Weil and Deligne, but fortunately they can be treated by methods from additive combinatorics. In particular, we make use of the bounds of Bourgain [1, Corollary] on Mordell type sums over prime fields.

Lemma 5.2. *For every $\varepsilon>0$ there exists some $\delta>0$ such that the following holds. Let $\alpha_1,\dots,\alpha_s\in\mathbb{F}_p$, not all zero, and $\lambda_1,\dots,\lambda_s\in\mathbb{F}_p^*$ be such that*

$$\operatorname{ord}(\lambda_i,p),\quad \operatorname{ord}(\lambda_i/\lambda_j,p)\geqslant p^{\varepsilon},\qquad 1\leqslant i,j\leqslant s,\ i\neq j.$$

Then

$$\left|\sum_{x=1}^{T}\mathbf{e}_p\left(\alpha_1\lambda_1^x+\ldots+\alpha_s\lambda_s^x\right)\right|\ll Tp^{-\delta},$$

where T is the order of the subgroup of $\mathbb{F}_p^*$ generated by $\lambda_1,\dots,\lambda_s$.

According to Lemma 4.4, in the split case, that is, for $\lambda_i\in\mathbb{F}_p^*$, $i=1,\dots,2g$, the number of solutions to the system (4.4) is given by

$Q_{2\nu}(p;\mathbf{n}) = R_{2\nu}(1,\dots,1;p)$ (with $t = 2g$ therein). Using the orthogonality of exponential functions we obtain

$$\begin{aligned} Q_{2\nu}(p;\mathbf{n}) &= R_{2\nu}(1,\dots,1;p)\\ &= \frac{1}{p^{2g}} \sum_{\alpha\in\mathbb{F}_p^{2g}} \left|\sum_{t=1}^{T} \mathbf{e}_p\left(\alpha_1\lambda_1^t+\ldots+\alpha_{2g}\lambda_{2g}^t\right)\right|^{4\nu}\\ &< \frac{T^{4\nu}}{p^{2g}} + \max_{\mathbf{0}\neq\alpha\in\mathbb{F}_p^{2g}} \left|\sum_{t=1}^{T} \mathbf{e}_p\left(\alpha_1\lambda_1^t+\ldots+\alpha_{2g}\lambda_{2g}^t\right)\right|^{4\nu}. \end{aligned}$$

Inserting Lemma 5.2, exactly as in [2], we derive

$$Q_{2\nu}(p;\mathbf{n}) \ll \frac{T^{4\nu}}{p^{2g}} + T^{4\nu}p^{-4\nu\delta} \leqslant 2\frac{T^{4\nu}}{p^{2g}}$$

for $\nu \geq g/(2\delta)$. Hence we find:

Corollary 5.3. *Let $A \in \mathrm{Sp}(2g,\mathbb{Z})$ have separable characteristic polynomial. For every $\varepsilon > 0$ there exists some integer $\nu_0 > 0$ such that the following holds. For a prime p so that A splits modulo p, let the eigenvalues of A be*

$$\lambda_1,\dots,\lambda_{2g} \in \mathbb{F}_p^*.$$

Assume that

$$\mathrm{ord}(\lambda_i,p), \quad \mathrm{ord}(\lambda_i/\lambda_j,p) \geqslant p^{\varepsilon}, \qquad 1 \leqslant i,j \leqslant 2g,\ i \neq j.$$

Then, for any vector $\mathbf{n} \in \mathbb{Z}^{2g}$, which is not a zero-divisor and such that for $p > p_0(A)\|\mathbf{n}\|_2^{2g}$, where $p_0(A)$ is as in Lemma 4.4, and all $\nu \geq \nu_0$, we have

$$Q_{2\nu}(p;\mathbf{n}) \ll \frac{\mathrm{ord}(A,p)^{4\nu}}{p^{2g}}.$$

5.3. Bounding $\langle \mathrm{T}_p^{(r)}(\mathbf{n})\psi,\psi'\rangle$ for a positive proportion of primes.

We remark that the assumptions of Lemma 5.2 and Corollary 5.3 hold for a positive proportion of primes p (in fact, for a full density subset of the set of primes p for which the characteristic polynomial of A splits completely, see Lemma 7.3). Hence, combining Lemma 4.1 and Corollary 5.3, we obtain the desired estimate (1.8) on $\langle \mathrm{T}_p^{(r)}(\mathbf{n})\psi,\psi'\rangle$ when $\mathbf{n}$ is not a zero-divisor.

Corollary 5.4. *Let $A \in \mathrm{Sp}(2g,\mathbb{Z})$ have separable characteristic polynomial. There exists some constant $\gamma > 0$, depending only on A, such that for a positive proportion of primes p the following holds: For all* *integers r coprime to p, and for any $\mathbf{n} \in \mathbb{Z}^{2g}$, which is not a zero-divisor and with $p > p_0(A)\|\mathbf{n}\|_2^{2g}$, where $p_0(A)$ is as in Lemma 4.4, we have*

$$\max_{\psi,\psi'} \left| \langle \mathrm{T}_p^{(r)}(\mathbf{n})\psi, \psi' \rangle \right| \leqslant p^{-\gamma},$$

the maximum over all pairs of normalized eigenvectors of $U_{p,r}(A)$.

6. Treatment of zero-divisors

6.1. **Preliminaries.** We remark that if f_A is irreducible then there are no nontrivial zero-divisors, and thus the results of § 4.1 allow us to complete the proof. However in the case when f_A is separable but not irreducible we need additional considerations to treat vectors $\mathbf{n} \in \mathbb{Z}^{2g}$ which are zero-divisors, as defined in § 4.2. Thus this section is not needed if one is only interested in the case of matrices $A \in \mathrm{Sp}(2g, \mathbb{Z})$ with irreducible characteristic polynomials.

6.2. **Remarks on symplectic spaces.** We next record some basic facts regarding symplectic vector spaces. Let W be a symplectic space, that is, a vector space with a non-degenerate alternating bilinear form, which we denote $\langle \cdot, \cdot \rangle$. We note that a subspace $U \subseteq V$ is symplectic, that is, the restriction of the symplectic form to U is non-degenerate, if and only if $U \cap U^{\perp} = \{\mathbf{0}\}$.

Lemma 6.1. *Let $A \in \mathrm{Sp}(V)$ be a symplectic matrix over V. Assume that $U \subseteq V$ is an A-invariant subspace on which A acts irreducibly, and assume that U is not isotropic. Then U is symplectic, and its orthogonal complement $U^{\perp}$ is also A-invariant and symplectic.*

Proof. Assume for contradiction that the restriction of the above bilinear form $\langle \cdot, \cdot \rangle$ to U is degenerate. Then there exist nonzero $\mathbf{u}_0 \in U$ such that $\langle \mathbf{u}, \mathbf{u}_0 \rangle = 0$ for all $\mathbf{u} \in U$, and hence $\langle A^i\mathbf{u}, A^i\mathbf{u}_0 \rangle = 0$ for all $\mathbf{u} \in U$ and all integers $i \geq 0$ (note that here we follow the usual convention of groups acting on the left.)

Since A is symplectic it is invertible, and so is the restriction to U, hence $\langle \mathbf{u}, A^i\mathbf{u}_0 \rangle = 0$ for all $\mathbf{u} \in U$. Since the span of $A^i\mathbf{u}_0$, $i = 0, 1, \ldots$, equals U we find that $U \subseteq U^{\perp}$, contradicting that U is not isotropic.

The argument for the first part of second assertion is similar. If $\mathbf{w} \in U^{\perp}$ then $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ for all $\mathbf{u} \in U$, and thus $\langle A\mathbf{u}, A\mathbf{w} \rangle = 0$ for all $\mathbf{u} \in U$ and hence, again using that $A|_U$ (that is, the map induced by A on U) is invertible, we have $\langle \mathbf{u}, A\mathbf{w} \rangle = 0$ for all $\mathbf{u} \in U$ and thus $U^{\perp}$ is A-invariant. Since U is symplectic we have $U \cap U^{\perp} = \{0\}$ and thus $W = U \oplus U^{\perp}$ (since $\dim(U) + \dim(U^{\perp}) = \dim(W)$ always holds).

Now, if the restriction of the form to $U^{\perp}$ is degenerate there exists $\mathbf{v} \in U^{\perp}$ with $\langle \mathbf{v}, \mathbf{u}^{\perp} \rangle = 0$ for all $\mathbf{u}^{\perp} \in U^{\perp}$, and since $\langle \mathbf{v}, \mathbf{u} \rangle = 0$ for all

$\mathbf{u} \in U$, we find that $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w} \in W$, which contradicts W being symplectic. □

A simple consequence of Lemma 6.1 is that if W splits into irreducible A-invariant subspaces, then each such subspace is either symplectic or isotropic. If there exist an invariant isotropic subspace, there is scarring as shown by Kelmer [16, Theorem 1]. Otherwise, we can decompose W into smaller invariant symplectic subspaces and use a certain tensor product structure to reduce the dimension, and this allows us to treat the problem of small zero-divisors.

6.3. **Quantized cat maps and tensor products revisited.**

Let $A \in \mathrm{Sp}(2g, \mathbb{Z})$ have separable characteristic polynomial and let $N = p$ be a prime. Let us consider an element $\mathbf{n} \in \mathbb{Z}^{2g}$ for which the reduction modulo p in $\mathbb{Z}^{2g}/(p\mathbb{Z}^{2g}) \simeq \mathbb{F}_p^{2g}$ is not a zero-divisor in the sense defined in § 4.2, where we identify $\mathbb{F}_p^{2g} \simeq \mathbb{F}_p[x]/(f_A(x))$. In order to bound the matrix coefficient $\langle \mathrm{T}_N^{(r)}(\mathbf{n})\psi, \psi' \rangle$ we need some further properties of the quantization related to invariant symplectic subspaces and an associated tensor product structure; these properties are consequences of $U_{p,r}(A)$ being implicitly defined via the Weil (or oscillator) representation of $\mathrm{Sp}(2g, \mathbb{F}_p)$. We briefly outline the construction below, for more details see [10, 16].

Hereafter, to simplify the notation in this section we regard p as a fixed prime, and suppress the dependence on p and $\mathbf{n}$ in most places. Let W be a symplectic vector space over $\mathbb{F}_p$, and assume that W splits into a direct sum of symplectic subspaces, that is, $W = W_1 \oplus W_2$ where $W_1 \perp W_2$ (that is, $W_2 = W_1^{\perp}$), and the restrictions of the symplectic form to W_1 and W_2 are both non-degenerate. We emphasise that in our application, W_1, W_2 depend not only on p but on $\mathbf{n}$ as well: we write $\mathbb{F}_p^{2g} \simeq W_1 \oplus W_2$, where the image of $\mathbf{n}$ in W_2 is zero, whereas the image in W_1 does not correspond to a zero-divisor.

With $V_i \subseteq W_i$, $i = 1, 2$, denoting maximal isotropic subspaces, we note that $V = V_1 \oplus V_2 \subseteq W$ is a maximal isotropic subspace. We may define the Heisenberg group

$$H(W) = \{(f, \mathbf{w}) : \ f \in \mathbb{F}_p, \ \mathbf{w} \in W\}$$

with the group law given by

$$(f, \mathbf{w}) \cdot (f', \mathbf{w}') = (f + f' + \langle \mathbf{w}, \mathbf{w}' \rangle, \mathbf{w} + \mathbf{w}')$$

where $\langle \cdot, \cdot \rangle$ denotes the symplectic form on W (and similarly $H(W_i)$ for $i = 1, 2$).

Let $Z \subseteq H(W_1) \times H(W_2)$ denote the subgroup

$$Z = \{(f, \mathbf{0}) \times (-f, \mathbf{0}) : \ f \in \mathbb{F}_p\}.$$

We find that the surjection $H(W_1) \times H(W_2) \to H(W)$, given by

$$(f_1, \mathbf{w}_1) \times (f_2, \mathbf{w}_2) \to (f_1 + f_2, \mathbf{w}_1 + \mathbf{w}_2)$$

factors through Z, and that we have the isomorphism

$$(H(W_1) \times H(W_2))/Z \cong H(W).$$

The irreducible non-abelian representations of $H(W)$ arise in the following way. Given a non-trivial additive character $\chi : \mathbb{F}_p \to \mathbb{C}$, let $K = \mathbb{F}_p \times V \subseteq H(W)$ denote a maximal abelian subgroup of $H(W)$ and extend χ to K (say, by letting $\chi(f, v) = \chi(f)$). We remark that the character χ depends on r present in the definition of our observables $T_p^{(r)}$, but the precise dependence is not important; we only need that $\gcd(r, p) = 1$ implies that χ is non-trivial. By inducing the extended character χ from K to $H(W)$, we obtain an irreducible representation $\rho : \ H(W) \to GL(L^2(V))$, and similarly irreducible representations

$$\rho_\nu : \ H(W_\nu) \to GL(L^2(V_\nu)), \qquad \nu = 1, 2.$$

Now, as $V = V_1 \times V_2$ we have $L^2(V) = L^2(V_1) \otimes L^2(V_2)$. Since the action of Z is trivial, we find that $H(W_1) \times H(W_2)$, and thus $H(W)$, acts in a natural way on $L^2(V_1) \otimes L^2(V_2)$.

Briefly, the Weil representation π of $\mathrm{Sp}(2g, \mathbb{F}_p) = \mathrm{Sp}(W)$ is then defined as follows: $\mathrm{Sp}(W)$ acts on $H(W)$, and this induces an action on the set of irreducible representations of $H(W)$. The action preserves the central character, and since irreducible representations of $H(W)$ are determined by their central characters (this holds since $H(W)$ is a two step nilpotent group), the action on the set of irreducible representations is, up to intertwining operators, trivial. In particular, for each $g \in \mathrm{Sp}(W)$, define ρ^g by $\rho^g(h) = \rho(g(h))$ (for $h \in H(W)$); we then find that $\rho \simeq \rho^g$, that is, there exists an intertwining operator (only defined up to a scalar; it turns out that this gives projective representation of $\mathrm{Sp}(W)$; for p odd a non-trivial fact is that it is possible to choose scalars to obtain a true representation) $\pi(g)$ acting on $L^2(V) = L^2(V_1) \otimes L^2(V_2)$ so that $\pi(g)\rho^g = \rho\pi(g)$. Further, we similarly obtain "smaller" Weil representations ρ_ν of $\mathrm{Sp}(W_\nu)$ acting on $L^2(V_\nu)$, for $\nu = 1, 2$; to fix compatible central characters it is convenient to use the maps $H(W_\nu) \to H(W_1) \times H(W_2) \to H(W)$ to obtain the action of $\mathrm{Sp}(W_\nu)$ on $L^2(V_\nu)$.

The product $\mathrm{Sp}(W_1) \times \mathrm{Sp}(W_2)$, under the inclusion

$$\mathrm{Sp}(W_1) \times \mathrm{Sp}(W_2) \subseteq \mathrm{Sp}(W),$$

then acts componentwise on the tensor product $L^2(V_1) \otimes L^2(V_2)$. In particular, if $A \in \mathrm{Sp}(W)$ leaves both W_1 and W_2 invariant, let $A_\nu \in \mathrm{Sp}(W_\nu)$ denote the corresponding restrictions of A to W_ν, for $\nu = 1, 2$. We now note that letting $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ denote the reduction of $\mathbf{n}$ modulo p, we can write,

$$\begin{aligned} U_{p,r}(A) &= U_1(A_1) \otimes U_2(A_2), \\ \mathrm{T}_p^{(r)}(\mathbf{n}) &= \rho((0,\mathbf{w})) = \rho_1((0,\mathbf{w}_1)) \otimes \rho_2((0,\mathbf{w}_2)), \end{aligned} \tag{6.1}$$

where $U_{p,r}(A) = \pi(A)$ and $U_\nu(A_\nu) = \pi_\nu(A_\nu)$ for $\nu = 1, 2$.

6.4. **Eigenfunctions of tensor products.** We next describe eigenfunctions of $U_{p,r}(A)$ in terms of the tensor product structure. With W, W_1, W_2 and V, V_1, V_2 as in § 6.3, for $\nu = 1, 2$, we may decompose $L^2(V_\nu)$ into $U_\nu(A_\nu)$-eigenspaces

$$E_{\nu,\lambda} = \ker(U_\nu(A_\nu) - \lambda I), \qquad \lambda \in \Lambda_\nu,$$

(possibly with multiplicities), where λ ranges over the set of eigenvalues Λ_ν of $U_\nu(A_\nu)$.

Further, for $\nu = 1, 2$ we may find bases of orthonormal eigenfunctions $\psi_{\nu,\lambda,i} \in E_{\nu,\lambda}$, $i = 1, \ldots, I_{\nu,\lambda}$, for some positive integers $I_{\nu,\lambda} = \dim(E_{\nu,\lambda})$ with

$$\sum_{\lambda \in \Lambda_1} I_{1,\lambda} + \sum_{\lambda \in \Lambda_2} I_{2,\lambda} = 2g,$$

which follows from the separability of the characteristic polynomial of A. That is,

$$U_\nu(A_\nu)\psi_{\nu,\lambda,i} = \lambda_{\nu,i}\psi_{\nu,\lambda,i}, \qquad \nu = 1, 2, \ i = 1, \ldots, I_{\nu,\lambda},$$

where $\lambda_{\nu,i}$ ranges over the whole set Λ_ν. We further note that the set

$$\{\psi_{1,\lambda_1,i_1} \otimes \psi_{2,\lambda_2,i_2} : \ \lambda_\nu \in \Lambda_\nu, \ i_\nu = 1, \ldots, I_{\nu,\lambda}, \ \nu = 1, 2\}$$

gives an orthonormal eigenbasis of $L^2(V) = L^2(V_1) \otimes L^2(V_2)$. In particular, the eigenvalues of $U_{p,r}(A) = U_1(A_1) \otimes U_2(A_2)$ are given by

$$\Lambda = \{\lambda_1\lambda_2 : \ \lambda_1 \in \Lambda_1, \lambda_2 \in \Lambda_2\},$$

and for $\mu \in \Lambda$, an eigenbasis for $E_\mu = \ker(U_{p,r}(A) - \mu I)$ is given by

$$\{\psi_{1,\lambda,i} \otimes \psi_{2,\mu/\lambda,j} : \ \lambda \in \Lambda_1, \ i = 1, \ldots, I_{1,\lambda}, \ j = 1, \ldots, I_{2,\mu/\lambda}\}.$$

Note that the quantizations $U_{p,r}(A), U_1(A_1)$, and $U_2(A_2)$ are only defined up to scalars, but once we have chosen scalars for $U_1(A_1)$, and $U_2(A_2)$ we may chose the scalar for $U_{p,r}(A)$ so that multiplicativity of eigenvalues hold.

We can now bound matrix coefficients corresponding to observables having zero-divisors.

Lemma 6.2. *Let $A \in \mathrm{Sp}(2g,\mathbb{Z})$ with a separable characteristic polynomial, such that there are no A-invariant rational istropic subspaces. There exists some constant $\gamma > 0$, depending only on A, such that for a positive proportion of primes p the following holds: Let $\psi \in E_\mu$ and $\psi' \in E_{\mu'}$ denote two eigenfunctions of $U_{p,r}(A)$, and let $\mathbf{w}$ denote a non-trivial zero-divisor. Then for $p > p_0(A)\|\mathbf{w}\|_2^{2g}$, where $p_0(A)$ is as in Lemma 4.4, we have*

$$|\langle \mathrm{T}_p^{(r)}(\mathbf{w})\psi, \psi'\rangle| \ll p^{-\gamma}\|\psi\|_2 \cdot \|\psi'\|_2.$$

Proof. Let $\mathbf{0} \neq \mathbf{w} \in \mathbb{Z}^{2g}$, which is a zero-divisor. Then there is an A-stable rational subspace W_1, necessarily symplectic by Lemma 6.1, so that with respect to the decomposition $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2) \in W_1 \oplus W_2$, where $W_2 = W_1^\perp$, the component $\mathbf{w}_1 \in W_1$ of $\mathbf{w}$ is not a zero divisor, while the component $\mathbf{w}_2$ in $W_2 = W_1^\perp$ is zero.

For $\mu, \mu' \in \Lambda$, write

$$\psi = \sum_{(\lambda,i,j)\in\Omega} \alpha_{\lambda,i,j}\psi_{1,\lambda,i} \otimes \psi_{2,\mu/\lambda,j},$$

where

$$\Omega = \{(\lambda, i, j) :\ \lambda \in \Lambda_1,\ i = 1, \dots, I_{1,\lambda},\ j = 1, \dots, I_{2,\mu/\lambda}\},$$

and

$$\psi' = \sum_{(\lambda',i',j')\in\Omega'} \beta_{\lambda',i',j'}\psi_{1,\lambda',i'} \otimes \psi_{2,\mu'/\lambda',j'},$$

where

$$\Omega' = \{(\lambda', i', j') :\ \lambda' \in \Lambda_1,\ i' = 1, \dots, I_{2,\lambda'},\ j' = 2, \dots, I_{2,\mu'/\lambda'}\},$$

with complex coefficients $\alpha_{\lambda,i,j}, \beta_{\lambda',i',j'} \in \mathbb{C}$.

Since $\mathbf{w}_2 = \mathbf{0}$, by (6.1), we have

$$\mathrm{T}_p^{(r)}(\mathbf{w}) = \rho((0,\mathbf{w})) = \rho_1((0,\mathbf{w}_1)) \otimes \rho_2((0,\mathbf{w}_2)) = \rho_1((0,\mathbf{w}_1)) \otimes \mathrm{Id},$$

and thus

$$\begin{aligned}\langle \rho((0,\mathbf{w}))\psi, \psi'\rangle = \sum_{(\lambda,i,j)\in\Omega}\ \sum_{(\lambda',i',j')\in\Omega'} & \overline{\alpha_{\lambda,i,j}}\beta_{\lambda',i',j'} \\ & \cdot\langle \rho_1((0,\mathbf{w}_1))\psi_{1,\lambda,i}, \psi_{1,\lambda',i'}\rangle\langle \psi_{2,\mu/\lambda,j}, \psi_{2,\mu'/\lambda',j'}\rangle.\end{aligned}$$

Now, since

$$\langle \psi_{2,\mu/\lambda,j}, \psi_{2,\mu'/\lambda',j'}\rangle = \begin{cases} 1 & \text{if } j = j' \text{ and } \mu/\lambda = \mu'/\lambda', \\ 0 & \text{otherwise,}\end{cases}$$

only terms for which $j = j'$ and for which $\lambda' = \eta(\lambda)$ for the bijection $\eta : \Lambda_1 \to \Lambda_1$ contribute (more precisely, we have $\eta(\lambda) = (\lambda\mu')/\mu$. Hence

$$\begin{aligned}&\langle \rho((0,\mathbf{w}))\psi, \psi'\rangle \\ &= \sum_{(\lambda,i,j)\in\Omega} \sum_{i'=1}^{I_{1,\eta(\lambda)}} \overline{\alpha_{\lambda,i,j}}\beta_{\eta(\lambda),i',j}\langle \rho_1((0,\mathbf{w}_1))\psi_{1,\lambda,i}, \psi_{1,\eta(\lambda),i'}\rangle.\end{aligned} \tag{6.2}$$

We now apply Corollary 5.4 with respect to the matrix A_1 in the decomposition (6.1), which applies since $\mathbf{w}_1$ is not a zero-divisor. Then, by the Cauchy inequality, for every λ and j fixed, we have

$$\begin{aligned}&\left|\sum_{i=1}^{I_{1,\lambda}} \sum_{i'=1}^{I_{1,\eta(\lambda)}} \overline{\alpha_{\lambda,i,j}}\beta_{\eta(\lambda),i',j}\langle \rho_1((0,\mathbf{w}_1))\psi_{1,\lambda,i}, \psi_{1,\eta(\lambda),i'}\rangle\right| \\ &\qquad \ll p^{-\gamma}\left(\sum_{i=1}^{I_{1,\lambda}} |\alpha_{\lambda,i,j}|^2\right)^{1/2}\left(\sum_{i'=1}^{I_{1,\eta(\lambda)}} |\beta_{\eta(\lambda),i',j}|^2\right)^{1/2}\end{aligned} \tag{6.3}$$

Finally, using the Cauchy inequality again, and then recalling that η is a bijection on Λ, we derive

$$\begin{aligned}&\sum_{\lambda\in\Lambda}\sum_{j=1}^{I_{2,\mu/\lambda}}\left(\sum_{i=1}^{I_{1,\lambda}} |\alpha_{\lambda,i,j}|^2\right)^{1/2}\left(\sum_{i'=1}^{I_{1,\eta(\lambda)}} |\beta_{\eta(\lambda),i',j}|^2\right)^{1/2} \\ &\quad \ll \left(\sum_{\lambda\in\Lambda}\sum_{j=1}^{I_{2,\mu/\lambda}}\sum_{i=1}^{I_{1,\lambda}} |\alpha_{\lambda,i,j}|^2\right)^{1/2}\left(\sum_{\lambda\in\Lambda}\sum_{j=1}^{I_{2,\mu/\lambda}}\sum_{i'=1}^{I_{1,\eta(\lambda)}} |\beta_{\eta(\lambda),i',j}|^2\right)^{1/2} \\ &\qquad = \|\psi\|_2\cdot\|\psi'\|_2\end{aligned}$$

and recalling (6.2) and (6.3), we conclude the proof. □

7. Anatomy of integers

7.1. **Some sums and products over primes.** It is convenient to denote by $\log_k x$ the k-fold iterated logarithm, that is, for $x \geqslant 1$ we set

$$\log_1 x = \log x \qquad \text{and} \qquad \log_k = \log_{k-1}\max\{\log x, 2\}, \quad k = 2, 3, \ldots.$$

We begin by recording an upper bound for Mertens type sums over primes in progressions, together with a simple consequence.

Lemma 7.1. *Let q be a prime and let $j \geqslant 1$ be an integer. We have*

$$\sum_{\substack{p\leqslant x\\ q\mid p^j-1}} \frac{1}{p} \ll q^{-1/j} + \frac{\log_2 x}{q},$$

where the implied constant depends only on j.

Proof. For an integer $k \geq 0$ define the dyadic interval $I_k = [2^k q, 2^{k+1} q]$, and note that $q \mid p^j - 1$ implies that p must lie in a progression $p \equiv a \mod q$, where $0 \leqslant a < q$ ranges over over at most j possible values. For any a, the Brun–Titchmarsh inequality, see, for example, [14, Theorem 6.6] or [27, Chapter I, Theorem 4.16], implies that

$$\sum_{\substack{p \in I_k \\ p \equiv a \mod q}} 1/p \ll \frac{2^{k+1} q}{q \log(2^{k+1} q/q)} \cdot \frac{1}{2^k q} \ll \frac{1}{q(k+1)}.$$

If $2^k q \leqslant x$ we have $k \ll \log x$, and summing over such k we find that the contribution from primes $p \geqslant q$ is $O\left(q^{-1} \log_2 x\right)$. Since there are at most j primes $p < q$ for which $q \mid p^j - 1$, and each such prime satisfies $p > q^{1/j}$ we find that the contribution from $p < q$ is $O\left(q^{-1/j}\right)$, and the proof is concluded. □

We remark that for $j = 1$ the bound of Lemma 7.1 simplifies as

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod q}} \frac{1}{p} \ll \frac{\log_2 x}{q}. \tag{7.1}$$

We control the contribution from small prime divisors of $p - 1$ as follows. For a prime q and positive integer k, we define $v_q(k)$ to be the positive integer ℓ such that

$$q^\ell \mid k \qquad \text{and} \qquad q^{\ell+1} \nmid k.$$

We fix some $z > 0$ and let

$$s_z(N) = \prod_{p \mid N} \prod_{q \leqslant z} q^{v_q(p-1)} = \prod_{p \mid N} \prod_{\substack{q \leqslant z \\ q^\ell \| p-1}} q^\ell, \tag{7.2}$$

that is, $s_z(N)$ is the product of the z-smooth parts of $p-1$, as p ranges over all prime divisors of N.

Lemma 7.2. *Let*

$$Z = \exp\left((\log_2 x)(\log_3 x)^{3/2}\right) \qquad \textit{and} \qquad z = (\log_2 x)^{O(1)}.$$

For all but $o(x)$ integers $N \leqslant x$ we have $s_z(N) \leqslant Z$.

Proof. From the definition of $s_z(N)$ in (7.2), extending over all powers $q^\ell \leqslant x$, $q \leqslant z$, such that $q^\ell \mid (p-1)$, we have

$$\sum_{N \leqslant x} \log s_z(N) \ll \sum_{\substack{q^\ell \leqslant x \\ q \leqslant z, \text{ prime}}} \log(q^\ell) \sum_{p \equiv 1 \mod q^\ell} \lfloor x/p \rfloor = S_1 + S_{\geqslant 2},$$

where S_1 is the contribution from the terms corresponding to $\ell = 1$ and $S_{\geqslant 2}$ is the contribution from the terms with $\ell \geqslant 2$.

For S_1, we have

$$S_1 \ll x \sum_{q \leqslant z,\ \text{prime}} \log q \sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod q}} \frac{1}{p}.$$

Using (7.1) applied to the inner sum, we now derive

$$\begin{aligned} S_1 &\ll x \sum_{q \leqslant z} \log q \frac{\log_2 x}{q} \\ &\ll x(\log_2 x) \sum_{q \leqslant z} \frac{\log q}{q} \ll x(\log_2 x)(\log z). \end{aligned} \tag{7.3}$$

The sum $S_{\geqslant 2}$ is estimated trivially by discarding the primality conditions on p and thus using that

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \bmod q^\ell}} \frac{1}{p} \leqslant \sum_{1 \leqslant k \leqslant x/q^\ell} \frac{1}{1 + kq^\ell} \ll \frac{\log x}{q^\ell},$$

which implies, after we abandon the condition of primality on q and the inequality $q \leqslant z$,

$$\begin{aligned} S_{\geqslant 2} &\ll x(\log x) \sum_{2 \leqslant \ell \leqslant \log x/\log 2} \sum_{1 \leqslant m \leqslant x^{1/\ell}} \frac{\log(m^\ell)}{m^\ell} \\ &\ll x(\log x)^2 \sum_{2 \leqslant \ell \leqslant \log x/\log 2} x^{-1+1/\ell} \ll x^{1/2}(\log x)^2. \end{aligned} \tag{7.4}$$

Clearly the bound on S_1 in (7.3) dominates the bound on $S_{\geqslant 2}$ in (7.4). Hence,

$$\sum_{N \leqslant x} \log s_z(N) \ll x(\log_2 x)(\log z) \ll x(\log_2 x)(\log_3 x).$$

Therefore we have $s_z(N) \geqslant Z = \exp\left((\log_2 x)(\log_3 x)^{3/2}\right)$ for at most

$$O\left(x(\log_2 x)(\log_3 x)(\log Z)^{-1}\right) = O\left(x\left(\log_3 x\right)^{-1/2}\right)$$

positive integers $N \leqslant x$. □

7.2. **Good primes and integers.** We recall that $A \in \mathrm{Sp}(2g, \mathbb{Z})$.

We say that a prime p is *good* if the following two conditions are satisfied:

- the characteristic polynomial of A is separable and splits completely modulo p;

- for the roots $\lambda_1, \dots, \lambda_{2g}$ of the characteristic polynomial of A modulo p we have
$$\mathrm{ord}(\lambda_i, p),\ \mathrm{ord}(\lambda_i/\lambda_j, p) \geqslant p^{1/3}, \qquad 1 \leqslant i, j \leqslant s,\ i \neq j.$$

We note that the exponent $1/3$ is somewhat arbitrary and can be replaced by any $\gamma < 1/2$.

Let $\mathcal{P}_{\mathrm{good}}$ denote the set of good primes. Applying Lemma 5.1, we now derive

Lemma 7.3. *The set $\mathcal{P}_{\mathrm{good}}$ is of positive density.*

Next, given integers $U \geqslant V \geqslant 1$ we define

$$\mathcal{P}_{\mathrm{good}}(V, U) = \mathcal{P}_{\mathrm{good}} \cap [V, U].$$

We now set

$$\begin{aligned} D(x) &= (\log x)^{(\log_3 x)^2}, \\ V(x) &= \exp\left(\exp\left(\sqrt{\log_2 x}\right)\right), \\ W(x) &= x^{\log_3 x / \log_2 x}, \end{aligned} \tag{7.5}$$

and define the following set $\mathcal{N}_{\mathrm{good}}(x)$ of *good* integers

$$\begin{aligned} \mathcal{N}_{\mathrm{good}} = \{N :\ \exists p \in \mathcal{P}_{\mathrm{good}}(V(N), W(N)) \\ \text{with } N = pM,\ M \in \mathbb{Z},\ \gcd(p, M) = 1, \\ \gcd(p-1, \mathrm{ord}(A, M)) \leqslant D(N)\}. \end{aligned} \tag{7.6}$$

We then set

$$\mathcal{N}_{\mathrm{good}}(x) = \mathcal{N}_{\mathrm{good}} \cap [1, x].$$

The next statement is our main tool.

Lemma 7.4. *We have*

$$\sharp \mathcal{N}_{\mathrm{good}}(x) = x + o(x).$$

Proof. It is certainly enough to show that

$$\sharp \left(\mathcal{N}_{\mathrm{good}} \cap [x/2, x]\right) = x/2 + o(x).$$

In turn, we set

$$D_0 = D(x/2), \quad V_0 = V(x), \quad W_0 = W(x/2),$$

such that

$$[V_0, W_0] \subseteq [V(N), W(N)],$$

for all $N \in [x/2, x]$ and define the following set $\widetilde{\mathcal{N}}_{\mathrm{good}}(x)$ of *good* integers $N \leqslant x$:

$$\begin{aligned}\widetilde{\mathcal{N}}_{\mathrm{good}}(x) = \{N \leqslant x: \ & \exists p \in \mathcal{P}_{\mathrm{good}}(V_0, W_0) \text{ with } N = pM, \ M \in \mathbb{Z}, \\ & \gcd(p, M) = 1, \ \gcd(p-1, \mathrm{ord}(A, M)) \leqslant D_0\}.\end{aligned}$$

Clearly $\widetilde{\mathcal{N}}_{\mathrm{good}}(x) \subseteq \mathcal{N}_{\mathrm{good}}(x)$, hence it is enough to show that

$$\sharp\widetilde{\mathcal{N}}_{\mathrm{good}}(x) = x + o(x). \tag{7.7}$$

That is, in the above, we first consider integers N in a dyadic interval. This allows us to replace $\mathcal{P}_{\mathrm{good}}(V(N), W(N))$ with $\mathcal{P}_{\mathrm{good}}(V_0, W_0)$. After this is done, we can bring back integers below $x/2$ as well: if the exceptional set is of size $o(x)$ on $[1, x]$ then so it is on $[x/2, x]$ and we are done. Thus indeed we only need to establish (7.7).

First recall that by Lemma 7.3 the set of good primes $\mathcal{P}_{good}$ is of positive density. Therefore, there are some constants $C, c > 0$ (depending on the matrix A) such that for $Z \geqslant 2$ the set $\mathcal{P}_{\mathrm{good}}(Z, CZ)$ contains at least $cZ/\log Z + O(1)$ primes, that is,

$$\sharp\mathcal{P}_{\mathrm{good}}(Z, CZ) \geqslant c\frac{Z}{\log Z} + O(1). \tag{7.8}$$

Taking x sufficiently large such that the interval $[2, W_0]$ contains I non-overlapping intervals of the form $[C^i, C^{i+1})$, $i = 1, \dots, I$, where

$$\log W_0 \ll I \ll \log W_0,$$

we derive

$$\begin{aligned}\sum_{p \in \mathcal{P}_{\mathrm{good}}(V_0, W_0)} 1/p &\geqslant \sum_{p \in \mathcal{P}_{\mathrm{good}}(2, W_0)} 1/p - \sum_{p \leqslant V_0, \ \mathrm{prime}} 1/p \\ &\geqslant \sum_{i=1}^{I} \sum_{p \in \mathcal{P}_{\mathrm{good}}(C^i, C^{i+1})} 1/p - \sum_{p \leqslant V_0, \ \mathrm{prime}} 1/p \\ &\geqslant \sum_{i=1}^{I} C^{-i} \sharp\mathcal{P}_{\mathrm{good}}\left(C^i, C^{i+1}\right) - \sum_{p \leqslant V_0, \ \mathrm{prime}} 1/p.\end{aligned}$$

Next, recalling (7.8) and the Mertens formula (or simply using (7.1) with $q = 1$), we obtain

$$\begin{aligned}\sum_{p\in\mathcal{P}_{\mathrm{good}}(V_0,W_0)} 1/p &\geqslant \sum_{i=1}^{I} C^{-i}\left(c\frac{C^i}{i\log C}+O(1)\right)+O(\log_2 V_0)\\ &\geqslant \frac{c}{\log C}\sum_{i=1}^{I}\frac{1}{i}+O(\log_2 V_0)\geqslant \frac{c}{\log C}\log I+O(\log_2 V_0)\\ &\gg \log_2 W_0+O(\log_2 V_0)\gg \log_2 W_0 \gg \log_2 x.\end{aligned}$$

Therefore, by the classical Brun sieve, see, for example, [27, Chapter I, Theorem 4.4], we conclude that

$$\prod_{p\in\mathcal{P}_{\mathrm{good}}(V_0,W_0)}(1-1/p)\ll \exp\left(-\sum_{p\in\mathcal{P}_{\mathrm{good}}(V_0,W_0)}1/p\right)\leqslant (\log x)^{-\gamma}$$

for some $\gamma > 0$, which depends only on C and c, and thus only on the matrix A. In particular, almost all $N \leqslant x$ are divisible by some $p \in \mathcal{P}_{\mathrm{good}}(V_0, W_0)$.

We now set $z = (\log_2 x)^{2g+1}$ and note that $D_0 > Z$, where Z is as in Lemma 7.2. Thus Lemma 7.2 allows us to discard $o(x)$ positive integers $N \leqslant x$ with

$$s_z(N) \geqslant D_0,$$

where $s_z(N)$ is defined by (7.2). Hence for the remaining integers $N \in [x/2, x]$ we have

$$s_z(N) < D_0 \leqslant D(N).$$

We also discard $O(x/V_0)$ integers $N \leqslant x$ which are divisible by p^2 for some prime $p > V_0$. Hence, for the remaining integers N, for any $p \in \mathcal{P}_{\mathrm{good}}(V_0, W_0)$ with $p \mid N$ we now have $\gcd(p, N/p) = 1$.

Furthermore, for the remaining $N \leqslant x$, we see that if

$$\gcd(p-1, \mathrm{ord}(A, N/p)) > D_0,$$

then, since $s_z(p) < s_z(N) < D_0$, there is a prime $q > z$ with $q \mid p-1$ and another prime $\ell \mid N$, $\ell \neq p$, such that

$$q \mid \mathrm{ord}(A,\ell) \mid \prod_{j=1}^{2g}\left(\ell^j-1\right).$$

Hence to conclude the proof it suffices to show that for every $j = 1, \ldots, 2g$ we have

$$\sum_{q>z,\ \text{prime}} \ \sum_{\substack{p\leqslant x,\ \text{prime} \\ p\equiv 1 \bmod q}} \ \sum_{\substack{\ell\leqslant x/p,\ \text{prime} \\ p\neq \ell \\ q\mid \ell^j-1}} \frac{x}{\ell p} = o(x). \tag{7.9}$$

To establish (7.9), we first discard the condition $\ell \neq p$, and extend the summation over ℓ up to $\ell \leqslant x$. Then we recall Lemma 7.1 (for the sum over ℓ) and its special case (7.1) (for the sum over p) and derive

$$\begin{aligned}
\sum_{q>z,\ \text{prime}} \ \sum_{\substack{p\leqslant x,\ \text{prime} \\ p\equiv 1 \bmod q}} \ \sum_{\substack{\ell\leqslant x/p,\ \text{prime} \\ p\neq \ell \\ q\mid \ell^j-1}} \frac{x}{\ell p} &\leqslant x \sum_{q>z,\ \text{prime}} \ \sum_{\substack{p\leqslant x,\ \text{prime} \\ p\equiv 1 \bmod q}} \frac{1}{p} \sum_{\substack{\ell\leqslant x,\ \text{prime} \\ q\mid \ell^j-1}} \frac{1}{\ell} \\
&\ll x \sum_{q>z,\ \text{prime}} \frac{\log_2 x}{q} \left(q^{-1/j} + \frac{\log_2 x}{q}\right) \\
&\ll x \left(\frac{\log_2 x}{z^{1/j}} + \frac{(\log_2 x)^2}{z^2}\right) \\
&\ll x \left(\frac{\log_2 x}{z^{1/(2g)}} + \frac{(\log_2 x)^2}{z^2}\right).
\end{aligned}$$

Recalling our choice $z = (\log_2 x)^{2g+1}$, we obtain (7.9).

Thus all together we have discarded $o(x)$ integers and all remaining integers $N \leqslant x$ belong to $\widetilde{\mathcal{N}}_{\text{good}}$. Hence we see that (7.7) holds, and the result follows. □

8. Proof of Theorem 1.2

We recall the definition of good integers given by (7.6). We now show that (1.7) holds with $\mathcal{N} = \mathcal{N}_{\text{good}}$, that is,

$$\lim_{\substack{N\to\infty \\ N\in\mathcal{N}_{\text{good}}}} \max_{\psi_N,\psi_N'} \left| \langle \mathrm{Op}_N(f)\psi_N, \psi_N'\rangle - \langle \psi_N, \psi_N'\rangle \int_{\mathbb{T}^{2g}} f(\mathbf{x})d\mathbf{x} \right| = 0,$$

where the maximum is taken over all pairs of normalized eigenfunctions ψ_N, ψ_N' of $U_N(A)$. By Lemma 7.4, the set $\mathcal{N}_{\text{good}}$ is of full density and hence this is sufficient for our goal.

As in [2, 19], using the rapid decay of coefficients of $f \in C^\infty(\mathbb{T}^{2g})$, it suffices to show that

$$\max_{\substack{\mathbf{n}\in\mathbb{Z}^{2g} \\ 0<|\mathbf{n}|\leqslant L(N)}} \max_{\psi_N,\psi_N'} |\langle \mathrm{T}_N(\mathbf{n})\psi_N, \psi_N'\rangle| \to 0$$

as $N \to \infty$, $N \in \mathcal{N}_{\text{good}}$, with ψ_N, ψ'_N running over all normalized eigenfunctions of $U_N(A)$, with a slowly growing function $L(N) \to \infty$.

We recall the definition of the functions $D(x)$, $V(x)$ and $W(x)$ as in (7.5). In particular, we take $L(N)$ to grow sufficiently slowly to guarantee that for any $p \in \mathcal{P}_{\text{good}}(V(N), W(N))$ and for any $\mathbf{n} \in \mathbb{Z}^{2g}$ with $0 < |\mathbf{n}| \leqslant L(N)$ the conditions of Corollary 5.4 and Lemma 6.2 are satisfied provided that N is sufficiently large.

We now fix some $N \in \mathcal{N}_{\text{good}}$ and choose a prime p which satisfies all properties in (7.6).

We set

$$d = \gcd\left(\operatorname{ord}(A,p), \operatorname{ord}(A,M)\right).$$

Clearly

$$d \leqslant \gcd(p-1, \operatorname{ord}(A,M)) \leqslant D(N).$$

Now, applying (3.6), and then Lemma 3.1 (with $N_1 = p$ and with A^d instead of A), we derive

$$|\langle \mathrm{T}_N(\mathbf{n})\psi, \psi'\rangle| \leqslant \max_{\varphi,\varphi' \in \Phi_{p,r}} |\langle \mathrm{T}_p^r(\mathbf{n})\varphi, \varphi'\rangle|, \tag{8.1}$$

where r is some integer coprime to p and φ, φ' range over all normalized eigenfunctions of $U_{p,r}(A^d)$.

We note that the roots of the characteristic polynomial of A^d are $\lambda_1^d, \ldots, \lambda_{2g}^d$, where $\lambda_1, \ldots, \lambda_{2g}$ are the roots of the characteristic polynomial of A modulo p and we also have

$$\operatorname{ord}(\lambda_i^d, p),\ \operatorname{ord}(\lambda_i^d/\lambda_j^d, p) \geqslant p^{1/3}d^{-1} \gg p^{1/4}, \qquad 1 \leqslant i,j \leqslant 2g,\ i \neq j,$$

since obviously for a sufficiently large N we have

$$d \leqslant D(N) \leqslant V(N)^{1/12} \leqslant p^{1/12}.$$

Thus the conditions of Corollary 5.3 are satisfied. Combining Corollary 5.4 and Lemma 6.2 (when $\mathbf{n}$ is a zero divisor) with (8.1) we conclude the proof.

References

[1] J. Bourgain, 'Mordell's exponential sum estimate revisited', *J. Amer. Math. Soc.*, **18** (2005), 477–499. 6, 19

[2] J. Bourgain, 'A remark on quantum ergodicity for CAT maps', *Geometric aspects of functional analysis*, Lecture Notes in Math., v. 1910, Springer, Berlin, 2007, 89–98. 5, 13, 17, 20, 32

[3] A. Bouzouina and S. de Bièvre, 'Equipartition of the eigenfunctions of quantized ergodic maps on the torus', *Commun. Math. Phys.* **178** (1996), 83–105. 4

[4] Y. Colin de Verdière, 'Ergodicité et fonctions propres du laplacien', *Commun. Math. Phys.* **102** (1985), 497–502. 4

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

[6] S. Dyatlov and M. Jézéquel, ‘Semiclassical measures for higher dimensional quantum cat maps’, *Annales Henri Poincaré*, (to appear). 5

[7] P. Erdős and R. Murty,‘On the order of $a \pmod p$’, *Proc. 5th Canadian Number Theory Assoc. Conf.*, Amer. Math. Soc., 1999, 87–97. 5

[8] F. Faure and S. Nonnenmacher, ‘On the maximal scarring for quantum cat map eigenstates.’, *Commun. Math. Phys.* **245** (2004), 201–214. 5

[9] F. Faure, S. Nonnenmacher and S. de Bièvre, ‘Scarred eigenstates for quantum cat maps of minimal periods’, *Commun. Math. Phys.* **239** (2003), 449–492. 4

[10] S. Gurevich and R. Hadani ‘The geometric Weil representation’, *Selecta Math. (N.S.)* **13** (2007), 465–481 22

[11] P. R. Halmos, ‘On automorphisms of compact groups’, *Bull. Amer. Math. Soc.* **49** (1943), 619–624. 2

[12] J. H. Hanna and M. V. Berry, ‘Quantization of linear maps on the torus – Fresnel diffraction by a periodic grating’, *Physica D: Nonlinear Phenomena* **1** (1980), 267–290. 2, 4

[13] C. Hooley, ‘Artin’s conjecture for primitive roots’, *J. Reine Angew. Math.* **225** (1967), 209–220. 17

[14] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc., Providence, RI, 2004. 18, 27

[15] J. P. Keating, ‘The cat maps: quantum mechanics and classical motion’, *Nonlinearity* **4** (1991), 309–341. 4

[16] D. Kelmer, ‘Arithmetic quantum unique ergodicity for symplectic linear maps of the multidimensional torus’, *Annals of Math.* **171** (2010), 815–879. 3, 5, 22

[17] E. Kim, ‘Characterizing the support of semiclassical measures for higher-dimensional cat maps’, with an Appendix by T. C. Anderson and R. J. Lemke Oliver, *Preprint*, 2024, available from `https://arxiv.org/abs/2410.13449`. 5

[18] P. Kurlberg and Z. Rudnick, ‘Hecke theory and equidistribution for the quantization of linear maps of the torus’, *Duke Math. J.* **103** (2000), 47–78. 3

[19] P. Kurlberg and Z. Rudnick, ‘On quantum ergodicity for linear maps of the torus’, *Comm. Math. Phys.*, **222** (2001), 201–227. 3, 4, 5, 13, 14, 15, 17, 32

[20] A. Ostafe, I. E. Shparlinski and J. F. Voloch, ‘Equations and character sums with matrix powers, Kloosterman sums over small subgroups and quantum ergodicity’, Int. Math. Res. Not. IMRN (2023), no. 16, 14196–14238. 5, 15, 17

[21] M. Reed and B. Simon, *Methods of modern mathematical physics, I: Functional analysis*, 2nd edition, Academic Press, Inc., New York, 1980. 11

[22] G. Riviére, ‘Entropy of semiclassical measures for symplectic linear maps of the multidimensional torus’, *Int. Math. Res. Not.* **2011** (2011), 2396–2443. 5

[23] Z. Rudnick, ‘The arithmetic theory of quantum maps. Equidistribution in number theory, an introduction’, *NATO Sci. Ser. II Math. Phys. Chem.*, vol. 237, Springer, Dordrecht, 2007, 331–342, 3

[24] Z. Rudnick and P. Sarnak, ‘The behaviour of eigenstates of arithmetic hyperbolic manifolds’, *Commun. Math. Phys.* **161** (1994), 195–213. 4

[25] A. Schnirelman, ‘Ergodic properties of eigenfunctions’, *Usp. Math. Nauk* **29** (1974), 181–182. 4

[26] N. Schwartz, 'The full delocalization of eigenstates for the quantized cat map', *Preprint*, 2021, available from https://arxiv.org/abs/2103.06633. 5

[27] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Grad. Studies Math., vol. 163, Amer. Math. Soc., 2015. 27, 31

[28] Z. Zelditch, 'Uniform distribution of eigenfunctions on compact hyperbolic surfaces', *Duke Math. J.* **55** (1987), 919–941. 4

Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden
Email address: kurlberg@math.kth.se

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia
Email address: alina.ostafe@unsw.edu.au

School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
Email address: rudnick@tauex.tau.ac.il

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia
Email address: igor.shparlinski@unsw.edu.au