Analysis & PDE

Volume 3 No. 1 2010

mathematical sciences publishers

Analysis & PDE

pjm.math.berkeley.edu/apde

See inside back cover or pjm.math.berkeley.edu/apde for submission instructions.

The subscription price for 2010 is US $120/year for the electronic version, and $180/year for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Analysis & PDE, at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

The handling of papers in APDE is managed by the editorial system EditFLOW from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
http://www.mathscipub.org
A NON-PROFIT CORPORATION
Typeset in LaTeX

ANALYSIS AND PDE
Vol. 3, No. 1, 2010

THE INVERSE CONJECTURE FOR THE GOWERS NORM OVER FINITE FIELDS VIA THE CORRESPONDENCE PRINCIPLE

TERENCE TAO AND TAMAR ZIEGLER

The inverse conjecture for the Gowers norms $U^d(V)$ for finite-dimensional vector spaces V over a finite field $\mathbb{F}$ asserts, roughly speaking, that a bounded function f has large Gowers norm $\|f\|_{U^d(V)}$ if and only if it correlates with a phase polynomial $\phi = e_{\mathbb{F}}(P)$ of degree at most $d-1$, thus $P : V \to \mathbb{F}$ is a polynomial of degree at most $d-1$. In this paper, we develop a variant of the Furstenberg correspondence principle which allows us to establish this conjecture in the large characteristic case $\operatorname{char} \mathbb{F} \geqslant d$ from an ergodic theory counterpart, which was recently established by Bergelson, Tao and Ziegler. In low characteristic we obtain a partial result, in which the phase polynomial ϕ is allowed to be of some larger degree $C(d)$. The full inverse conjecture remains open in low characteristic; the counterexamples found so far in this setting can be avoided by a slight reformulation of the conjecture.

1. Introduction

1.1. *The combinatorial inverse conjecture in finite characteristic.* Let $\mathbb{F}$ be a finite field of prime order. Throughout this paper, $\mathbb{F}$ will be considered fixed (for example, $\mathbb{F} = \mathbb{F}_2$ or $\mathbb{F} = \mathbb{F}_3$), and the term *vector space* will be shorthand for *vector space over* $\mathbb{F}$, and more generally any linear algebra term (span, independence, basis, subspace, linear transformation, etc.) will be understood to be over the field $\mathbb{F}$.

If V is a vector space, $f : V \to \mathbb{C}$ is a function, and $h \in V$ is a shift, we define the (multiplicative) derivative $\Delta_h f : V \to \mathbb{C}$ of f by the formula

$$\Delta_h f := (T_h f)\overline{f},$$

where the shift operator T_h with shift h is defined by $T_h f(x) := f(x+h)$. An important special case arises when f takes the form $f = e_{\mathbb{F}}(P)$, where $P : V \to \mathbb{F}$ is a function, and $e_{\mathbb{F}} : \mathbb{F} \to \mathbb{C}$ is the standard character $e_{\mathbb{F}}(j) := e^{2\pi i j/|\mathbb{F}|}$ for $j = 0, \dots, |\mathbb{F}|-1$. In that case we see that $\Delta_h f = e_{\mathbb{F}}(\Delta_h P)$, where $\Delta_h P : V \to \mathbb{F}$ is the (additive) derivative of P, defined as

$$\Delta_h P = T_h P - P.$$

Given an integer $d \geqslant 0$, we say that a function $P : V \to \mathbb{F}$ is a *polynomial of degree at most* d if we have $\Delta_{h_1} \dots \Delta_{h_{d+1}} P = 0$ for all $h_1, \dots, h_{d+1} \in V$, and write $\operatorname{Poly}_d V$ for the set of all polynomials on V of degree at most d; thus for instance $\operatorname{Poly}_0 V$ is the set of constants, $\operatorname{Poly}_1 V$ is the set of linear

MSC2000: 11T06, 37A15.
Keywords: Gowers uniformity norm, Furstenberg correspondence principle, characteristic factor, polynomials over finite fields.
The first author is supported by a grant from the MacArthur Foundation and by NSF grant CCF-0649473. The second author is supported by ISF grant 557/08, by a Landau fellowship (supported by the Taub Foundation), and by an Alon fellowship.

polynomials on V, $\mathrm{Poly}_2 V$ is the set of quadratic polynomials, and so forth. It is easy to see that $\mathrm{Poly}_d V$ is a vector space, and if $V = \mathbb{F}^n = \{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{F}\}$ is the standard n-dimensional vector space, then $\mathrm{Poly}_d V$ has the monomials $x_1^{i_1} \dots x_n^{i_n}$ for $0 \leqslant i_1, \dots, i_n < |\mathbb{F}|$ and $i_1 + \dots + i_n \leqslant d$ as a basis[1].

We shall say that a function $f : V \to \mathbb{C}$ is a *phase polynomial of degree at most d* if all $(d+1)$-th multiplicative derivatives $\Delta_{h_1} \dots \Delta_{h_{d+1}} f$ are identically 1, and write $\mathcal{P}_d(V)$ for the space of all phase polynomials of degree at most d. We have the following equivalence between polynomials and phase polynomials in the high characteristic case:

Lemma 1.2 (phase polynomials are exponentials of polynomials). *Suppose that* $0 \leqslant d < \operatorname{char} \mathbb{F}$, *and* $f : V \to \mathbb{C}$. *Then the following are equivalent*:

(i) $f \in \mathcal{P}_d(V)$.

(ii) $f = e^{2\pi i \theta} e_{\mathbb{F}}(P)$ *for some* $\theta \in \mathbb{R}/\mathbb{Z}$ *and* $P \in \mathrm{Poly}_d V$.

Proof. See [Bergelson et al. 2009, Lemma D.5]. □

Remark 1.3. The lemma fails in the low characteristic case $d \geqslant \operatorname{char} \mathbb{F}$; consider for instance the function $f : \mathbb{F}_2 \to \mathbb{C}$ defined by $f(1) := i$ and $f(0) := 1$. This function lies in $\mathcal{P}_2(\mathbb{F}_2)$ but does not arise from a polynomial in $\mathrm{Poly}_2 \mathbb{F}_2$.

Definition 1.4 (expectation notation). If A is a finite nonempty set and $f : A \to \mathbb{C}$ is a function, we write $|A|$ for the cardinality of A, and $\mathbb{E}_A f$, $\int_A f$, or $\mathbb{E}_{x \in A} f(x)$ for the average $(1/|A|) \sum_{x \in A} f(x)$.

Definition 1.5 (Gowers uniformity norm [Gowers 1998; 2001]). Let V be a finite vector space, let $f : V \to \mathbb{C}$ be a function, and let $d \geqslant 1$ be an integer. We then define the *Gowers norm* $\|f\|_{U^d(V)}$ of f to be the quantity

$$\|f\|_{U^d(V)} := \left| \mathbb{E}_{h_1, \dots, h_d} \int_V \Delta_{h_1} \dots \Delta_{h_d} f \right|^{1/2^d},$$

thus $\|f\|_{U^{d+1}(V)}$ measures the average bias in d-th multiplicative derivatives of f. We also define the *weak Gowers norm* $\|f\|_{u^d(V)}$ of f to be the quantity

$$\|f\|_{u^d(V)} := \sup_{\phi \in \mathcal{P}_{d-1}(V)} \left| \int_V f \overline{\phi} \right|, \tag{1-1}$$

thus $\|f\|_{u^d(V)}$ measures the extent to which f can correlate with a phase polynomial of degree at most $d - 1$.

Remark 1.6. It can in fact be shown that the Gowers and weak Gowers norm are in fact norms for $d \geqslant 2$ (and seminorms for $d = 1$); see [Gowers 2001; Tao and Vu 2006]. Further discussion of these two norms can be found in [Green and Tao 2008]. In view of Lemma 1.2, in the high characteristic case $\operatorname{char} \mathbb{F} \geqslant d$ one can replace the phase polynomial $\phi \in \mathcal{P}_{d-1}(V)$ in (1-1) by the exponential $e_{\mathbb{F}}(P)$ of a polynomial $P \in \mathrm{Poly}_{d-1} V$. However, this is not the case in low characteristic. For instance, let $\mathbb{F} = \mathbb{F}_2$, $V = \mathbb{F}_2^n$, and consider the symmetric function $S_4 : V \to \mathbb{F}_2$ defined by

$$S_4(x_1, \dots, x_n) := \sum_{1 \leqslant i < j < k < l \leqslant n} x_i x_j x_k x_l.$$

[1]The restriction $i_1, \dots, i_n < |\mathbb{F}|$ arises of course from the identity $x^{|\mathbb{F}|} = x$ for all $x \in \mathbb{F}$.

Then the function $f := (-1)^{S_4}$ has low correlation with any exponential $e_{\mathbb{F}}(P) = (-1)^P$ of a cubic polynomial $P \in \mathrm{Poly}_3 V$ in the sense that $\mathbb{E}_{x\in V} f e_{\mathbb{F}}(-P) = o_{n\to\infty}(1)$ [Lovett et al. 2007; Green and Tao 2009a]; on the other hand, it is not hard to verify that the function

$$g(x_1, \ldots, x_n) := e^{2\pi i|x|/8},$$

where $|x|$ denotes the number of indices $1 \leqslant j \leqslant n$ for which $x_j = 1$, lies in $\mathcal{P}_3(V)$ and has a large inner product with f; indeed, since $f(x) = +1$ when $|x| = 0, 1, 2, 3 \mod 8$ and -1 otherwise, we easily check that

$$\begin{aligned}\mathbb{E}_{x\in V} f\bar{g} &= \tfrac{1}{8}\big(1+e^{-2\pi i/8}+e^{-4\pi i/8}+e^{-6\pi i/8}-e^{-8\pi i/8}-e^{-10\pi i/8}-e^{-12\pi i/8}-e^{-14\pi i/8}\big)+o_{n\to\infty}(1)\\ &= \frac{1-i-\sqrt{2}i}{4}+o_{n\to\infty}(1).\end{aligned}$$

We thus see that $\|(-1)^{S_4}\|_{u^4(V)}$ is bounded from below by a positive absolute constant for large n.

Let $\mathcal{D} := \{z \in \mathbb{C} : |z| \leqslant 1\}$ be the compact unit disk. This paper is concerned with the following conjecture.

Conjecture 1.7 (inverse conjecture for the Gowers norm). *Let $\mathbb{F}$ be a finite field and let $d \geqslant 1$ be an integer. Then for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\|f\|_{u^d(V)} \geqslant \varepsilon$ for every finite vector space V and every function $f : V \to \mathcal{D}$ such that $\|f\|_{U^d(V)} \geqslant \delta$.*

Remark 1.8. This result is trivial for $d = 1$, and follows easily from Plancherel's theorem for $d = 2$. The result was established for $d = 3$ in [Green and Tao 2008] (for odd characteristic) and [Samorodnitsky 2007] (for even characteristic), and a formulation of Theorem 1.9 was then conjectured in both papers, in which the phase polynomials were constrained to be (char $\mathbb{F}$)-th roots of unity. This formulation of the conjecture turned out to fail in the low characteristic regime $\mathrm{char}\,\mathbb{F}+1 < d$ [Green and Tao 2009a; Lovett et al. 2007]; however, the counterexamples given there do not rule out the conjecture as formulated above in this case, basically because of the discussion in Remark 1.6.

The case when δ was sufficiently close to 1 (depending on d) was treated in [Alon et al. 2003; 2005], while the case when char $\mathbb{F}$ is large compared to d and δ was established in [Sudan et al. 2001]. In [Green and Tao 2009a], Theorem 1.9 was also established in the case when f was a phase polynomial of degree less than char $\mathbb{F}$. These results have applications to solving linear systems of equations (and in particular, in finding arithmetic progressions) in subsets of vector spaces [Green and Tao 2009b; Gowers and Wolf 2007] and also to polynomiality testing [Samorodnitsky 2007; Bogdanov and Viola 2007]. Conjecture 1.7 is also the finite field analogue of a corresponding inverse conjecture for the Gowers norm in cyclic groups $\mathbb{Z}/N\mathbb{Z}$, which is of importance in solving linear systems of equations in sets of integers such as the primes; see [Green and Tao 2006; Frantzikinakis et al. 2007] for further discussion.

The main result of this paper is to establish this conjecture in the high characteristic case.

Theorem 1.9 (inverse conjecture for the Gowers norm in high characteristic). *Conjecture 1.7 holds whenever* char $\mathbb{F} \geq d$.

In the low characteristic case we have a partial result.

Theorem 1.10 (partial inverse conjecture for the Gowers norm). *Let $\mathbb{F}$ be a finite field and let $d \geqslant 1$ be an integer. Then for every $\delta > 0$ there exists $\varepsilon > 0$ such that $\|f\|_{u^k(V)} \geqslant \varepsilon$ for every finite vector space V and every function $f : V \to \mathfrak{D}$ such that $\|f\|_{U^d(V)} \geqslant \delta$, where $k = C(d)$ depends only on d.*

Remark 1.11. One could in principle make the quantity $k = C(d)$ in Theorem 1.10 explicit, but this would require analyzing the arguments in [Bergelson et al. 2009] in careful detail. One should however be able to obtain reasonable values of k for small d (e.g., $d = 4$).

The proofs of Theorems 1.9, 1.10 rely on four additional ingredients:

- an ergodic inverse theorem for the Gowers norm for $\mathbb{F}^\omega$-systems (Theorems 1.19, 1.20), established in [Bergelson et al. 2009];
- the Furstenberg correspondence principle [Furstenberg 1977], combined with the random averaging trick of Varnavides [1959];
- a statistical sampling lemma (Proposition 3.13); and
- local testability of phase polynomials (Lemma 4.5), essentially established in [Alon et al. 2003; 2005].

Of these ingredients, the ergodic inverse theorem is the most crucial, and we now pause to describe it in detail.

1.12. ***The ergodic inverse conjecture in finite characteristic.*** Let $\mathbb{F}^\omega := \bigcup_{n=0}^\infty \mathbb{F}^n$ be the inverse limit of the finite-dimensional vector spaces $\mathbb{F}^n$, where each $\mathbb{F}^n$ is included in the next space $\mathbb{F}^{n+1}$ in the obvious manner; equivalently, $\mathbb{F}^\omega$ is the space of sequences $(x_i)_{i=1}^\infty$ with $x_i \in \mathbb{F}$, and all but finitely many of the x_i nonzero. This is a countably infinite vector space over $\mathbb{F}$.

Definition 1.13 ($\mathbb{F}^\omega$-system). A $\mathbb{F}^\omega$-*system* is a quadruplet $X = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{F}^\omega})$, where $(X, \mathcal{B}, \mu)$ is a probability space, and $T : h \mapsto T_h$ is a measure-preserving action of $\mathbb{F}^\omega$ on X, thus for each $h \in \mathbb{F}^\omega$, $T_h : X \to X$ is a measure-preserving bijection such that $T_h \circ T_k = T_{h+k}$ for all $h, k \in \mathbb{F}^\omega$. Given any measurable $\phi : X \to \mathbb{C}$ and $h \in \mathbb{F}^\omega$, we define $T_h\phi : X \to \mathbb{C}$ to be the function $T_h\phi := \phi \circ T_h$, and $\Delta_h\phi : X \to \mathbb{C}$ to be the function $\Delta_h\phi := T_h\phi \cdot \overline{\phi}$. We also define the inner product $\langle f, g\rangle := \int_X f\bar{g}\, d\mu$ for all $f, g \in L^2(X)$, where the Lebesgue spaces $L^p(X) = L^p(X, \mathcal{B}, \mu)$ are defined in the usual manner. We say that the system is *ergodic* if the only $\mathbb{F}^\omega$-invariant functions on $L^2(X)$ are the constants.

Definition 1.14 (phase polynomial). Let $X = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{F}^\omega})$ be an $\mathbb{F}^\omega$-system, and let $d \geqslant 0$. We say that a function $\phi \in L^\infty(X)$ is a *phase polynomial of degree at most* d if we have $\Delta_{h_1} \dots \Delta_{h_{d+1}}\phi = 1$ μ-a.e. for all $h_1, \dots, h_{d+1} \in \mathbb{F}^\omega$. We let $\mathcal{P}_d(X)$ denote the space of all phase polynomials.

Remark 1.15. By setting $h_1 = \dots = h_{d+1} = 0$ we see that every phase polynomial $\phi \in \mathcal{P}_d(X)$ has unit magnitude: $|\phi| = 1$ μ-a.e.

Definition 1.16 (Gowers–Host–Kra seminorms [Host and Kra 2005]). Let $X = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{F}^\omega})$ be a $\mathbb{F}^\omega$-system, and let $\phi \in L^\infty(X)$. We define the *Gowers–Host–Kra seminorms* $\|\phi\|_{U^d(X)}$ for $d \geqslant 1$ recursively as follows:

- If $d = 1$, then $\|\phi\|_{U^1(X)} := \limsup_{n\to\infty}\big(\|\mathbb{E}_{h\in\mathbb{F}^n} T_h\phi\|^2_{L^2(X,\mu)}\big)^{1/2}$;
- If $d > 1$, then $\|\phi\|_{U^d(X)} := \limsup_{n\to\infty}\big(\|\Delta_h\phi\|^{2^{d-1}}_{U^{d-1}(X,\mu,T)}\big)^{1/2^d}$.

We also define the *weak Gowers–Host–Kra seminorm* $\|\phi\|_{u^d(X)}$ as

$$\|\phi\|_{u^d(X)} := \sup_{\psi \in \mathcal{P}_{d-1}(X)} |\langle \phi, \psi \rangle|.$$

Example 1.17. If $\phi \in \mathcal{P}_{d-1}(X)$ is a phase polynomial of degree at most $d-1$, then

$$\|\phi\|_{U^d(X)} = \|\phi\|_{u^d(X)} = 1.$$

Remark 1.18. One can use the ergodic theorem to show that the limits here in fact converge, but we will not need this. The U^d are indeed seminorms, but we will not need this either.

In [Bergelson et al. 2009, Corollaries 1.26, 1.27], the following ergodic theory analogues of Theorems 1.9, 1.10 was shown:

Theorem 1.19 (inverse conjecture for the Gowers–Host–Kra seminorm for high characteristic). *Let $X = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be an ergodic $\mathbb{F}^\omega$-system, let* $\operatorname{char} \mathbb{F} \geqslant d \geqslant 1$, *and let $\phi \in L^\infty(X)$ be such that $\|\phi\|_{U^d(X)} > 0$. Then $\|\phi\|_{u^d(X)} > 0$.*

Theorem 1.20 (partial inverse conjecture for the Gowers–Host–Kra seminorm for general characteristic). *Let $X = (X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{F}^\omega})$ be an ergodic $\mathbb{F}^\omega$-system, let $d \geqslant 1$, and let $\phi \in L^\infty(X)$ be such that $\|\phi\|_{U^d(X)} > 0$. Then $\|\phi\|_{u^k(X)} > 0$ for some $k = C(d)$ depending only on d.*

Remark 1.21. The *if* part of this theorem follows easily from van der Corput's lemma; the important part of the theorem for us is the *only if* part. These results can be viewed as a finite field analogue of the results in [Host and Kra 2005] in high characteristic (and a partial analogue in the low characteristic case), and indeed draws heavily on the tools developed in that paper; see [Bergelson et al. 2009] for further discussion. It is quite possible that k can in fact be taken to equal d in Theorem 1.20 (or equivalently, that the condition $\operatorname{char} \mathbb{F} \geqslant d$ can be dropped in Theorem 1.19); this would imply Conjecture 1.7 in full generality.

We will use Theorem 1.20 as a black box, and it will be the primary ingredient in our proof of Theorem 1.10, in much the same way that the Furstenberg recurrence theorem is the primary ingredient in Furstenberg's proof of Szemerédi's theorem in [Furstenberg 1977]. Theorem 1.19 plays a similar role for Theorem 1.9.

As with any other argument using a Furstenberg-type correspondence principle, our bounds are *ineffective*, in that we do not obtain an explicit value of ε in terms of d and δ. In principle, one could finitise the arguments in [Bergelson et al. 2009] (in the spirit of [Tao 2006]) to obtain such an explicit value, but this would be extremely tedious (and not entirely straightforward), and would lead to an extremely poor dependence (such as iterated tower-exponential or worse). We will not pursue this matter here.

2. Notation

We will rely heavily on asymptotic notation. Given any parameters $x_1, \ldots, x_k$, we use $O_{x_1,\ldots,x_k}(X)$ to denote any quantity bounded in magnitude by $C_{x_1,\ldots,x_k} X$ for some finite quantity $C_{x_1,\ldots,x_k}$ depending only on $x_1, \ldots, x_k$. We also write $Y \ll_{x_1,\ldots,x_k} X$ or $X \gg_{x_1,\ldots,x_k} Y$ for $Y = O_{x_1,\ldots,x_k}(X)$. Furthermore, given an asymptotic parameter n that can go to infinity, we use $o_{n\to\infty;x_1,\ldots,x_k}(X)$ to denote any quantity bounded in magnitude by $c_{x_1,\ldots,x_k}(n)X$, where $c_{x_1,\ldots,x_k}(n)$ is a quantity which goes to zero as $n \to \infty$ for fixed $x_1, \ldots, x_k$. Thus for instance, if $r_2 > r_1 > 1$, then $\exp r_1 / \log r_2 = o_{r_2\to\infty;r_1}(1)$.

3. Statistical sampling

It is well known that the *global average* $\mathbb{E}_{h\in V} f(h)$ of a bounded function $f : V \to \mathfrak{D}$ can be accurately estimated (with high probability) by randomly selecting a number of points $x_1, \ldots, x_N \in V$ and computing the empirical Monte Carlo average (or *local average*) $\mathbb{E}_{1\leqslant n\leqslant N} f(x_n)$. Indeed, it is not hard to show (by the second moment method) that with probability $o_{N\to\infty}(1)$, one has

$$\mathbb{E}_{1\leqslant n\leqslant N} f(x_n) = \mathbb{E}_{h\in V} f(h) + o_{N\to\infty}(1).$$

The point here is that the error term is uniform in the choice of f and V.

We now record some variants of this standard *random local averages approximate global averages* fact, in which we perform more exotic empirical averages. We begin with averages along random subspaces of V.

Lemma 3.1 (random sampling for integrals). *Let $v_1, \ldots, v_m$ be points chosen independently at random in a finite-dimensional vector space V, and let $f : V \to \mathfrak{D}$ be a function. With probability $1 - o_{m\to\infty}(1)$, we have*

$$\mathbb{E}_{\vec{a}\in\mathbb{F}^m} f(\vec{a}\cdot\vec{v}) = \mathbb{E}_{h\in V} f(h) + o_{m\to\infty}(1),$$

where $\vec{v} := (v_1, \ldots, v_m)$ and $\vec{a}\cdot\vec{v} := a_1v_1 + \cdots + a_mv_m$.

Remark 3.2. One can easily make the $o_{m\to\infty}(1)$ terms more explicit, but we will not need to do so here.

Proof. We use the second moment method. Note that

$$\mathbb{E}\mathbb{E}_{\vec{a}\in\mathbb{F}^m} f(\vec{a}\cdot\vec{v}) = \mathbb{E}_{h\in V} f(h) + o_{m\to\infty}(1)$$

(the $o_{m\to\infty}(1)$ error arising from the $a = 0$ contribution) so by Chebyshev's inequality it suffices to show that

$$\mathbb{E}|\mathbb{E}_{\vec{a}\in\mathbb{F}^m} f(\vec{a}\cdot\vec{v})|^2 = |\mathbb{E}_{h\in V} f(h)|^2 + o_{m\to\infty}(1).$$

The left side can be rearranged as

$$\mathbb{E}_{\vec{a},\vec{b}\in\mathbb{F}^m} \mathbb{E} f(\vec{a}\cdot\vec{v})\bar{f}(\vec{b}\cdot\vec{v}).$$

It is easy to see that the inner expectation is $|\mathbb{E}_{h\in V} f(h)|^2$ unless $\vec{a} = c\vec{b}$, for some $c \in \mathbb{F}$ in which case it is $O(1)$. The claim follows. □

In the above lemma, f was deterministic and thus independent of the v_i. But we can easily extend the result to the case where f depends on a bounded number of the v_i.

Corollary 3.3 (random sampling for integrals, II). *Let V be a finite-dimensional vector space, let $m \geqslant m_0 \geqslant 0$, let $v_1, \ldots, v_m \in V$ be chosen independently at random, and let $f_{v_1,\ldots,v_{m_0}} : V \to \mathfrak{D}$ be a function that depends on $v_1, \ldots, v_{m_0}$ but is independent of $v_{m_0+1}, \ldots, v_m$. Then with probability $1-o_{m\to\infty;m_0}(1)$, we have*

$$\mathbb{E}_{\vec{a}\in\mathbb{F}^m} f_{v_1,\ldots,v_{m_0}}(\vec{a}\cdot\vec{v}) = \mathbb{E}_{h\in V} f_{v_1,\ldots,v_{m_0}}(h) + o_{m\to\infty;m_0}(1).$$

Proof. We write $\vec{a} = (\vec{a}_0, \vec{a}_1) \in \mathbb{F}^{m_0} \times \mathbb{F}^{m-m_0}$ and $\vec{v} = (\vec{v}_0, \vec{v}_1) \in V^{m_0} \times V^{m-m_0}$. If we condition $\vec{v}_0 = (v_1, \ldots, v_{m_0})$ to be fixed, we see from applying Lemma 3.1 to the remaining random vectors $\vec{v}_1$ that for fixed $\vec{a}_0$, we have

$$\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m-m_0}} f_{v_1,\ldots,v_{m_0}}(\vec{a}\cdot\vec{v}) = \mathbb{E}_{h\in V} f_{v_1,\ldots,v_{m_0}}(\vec{a}_0\cdot\vec{v}_0 + h) + o_{m-m_0\to\infty}(1),$$

with probability $1 - o_{m-m_0\to\infty}(1)$ conditioning on $\vec{v}_0$; integrating this we see that the same is true without the conditioning. We can shift h by $\vec{a}_0 \cdot \vec{v}_0$, move the h average onto the other side, and take expectations to conclude that

$$\mathbb{E}|\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m-m_0}} f_{v_1,\ldots,v_{m_0}}(\vec{a}\cdot\vec{v}) - \mathbb{E}_{h\in V} f_{v_1,\ldots,v_{m_0}}(h)| = o_{m-m_0\to\infty}(1)$$

for each $\vec{a}_0$; averaging over $\vec{a}_0$ by the triangle inequality we obtain the claim. □

Remark 3.4. It is with this corollary that we are implicitly exploiting the highly transitive nature of the symmetry group $GL(V)$ available to us. In the setting of the cyclic group $\mathbb{Z}/N\mathbb{Z}$, the analogue of Lemma 3.1 is still true, namely that one can approximate a global average $\int_{\mathbb{Z}/N\mathbb{Z}} f$ by a local average on random arithmetic progressions of medium length, but this approximation no longer holds if f is allowed to depend on the first few values of that progression, since this of course determines the rest of the progression; this is related to the fact that (for N prime, say), the affine group of $\mathbb{Z}/N\mathbb{Z}$ (which is analogous to $GL(V)$) is 2-transitive but no stronger. In contrast, in the finite field setting, a small subspace of a medium-dimensional subspace does not determine the whole subspace.

We will need to generalise these results further by considering more exotic averages along cubes. A typical result we will need can be stated informally as

$$\mathbb{E}_{\vec{a}_2\in\mathbb{F}^{m_2}}\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m_1}}\int_V f(T_{\vec{a}_1\cdot\vec{v}_1}\bar{f})(T_{\vec{a}_2\cdot\vec{v}_2}\bar{f})(T_{\vec{a}_1\cdot\vec{v}_1+\vec{a}_2\cdot\vec{v}_2}f) \approx \mathbb{E}_{h_1,h_2\in V}\int_V f(T_{h_1}\bar{f})(T_{h_2}\bar{f})T_{h_1+h_2}f \tag{3-1}$$

when m_1 is large, m_2 is large compared with m_1, and $\vec{v}$ is random (see Lemma 3.9 for the formal version of this type of estimate). Such results follow (heuristically, at least), by iterating the previous results. For instance, from Corollary 3.3 we heuristically have

$$\mathbb{E}_{\vec{a}_2\in\mathbb{F}^{m_2}}\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m_1}}\int_V f(T_{\vec{a}_1\cdot\vec{v}_1}\bar{f})(T_{\vec{a}_2\cdot\vec{v}_2}\bar{f})(T_{\vec{a}_1\cdot\vec{v}_1+\vec{a}_2\cdot\vec{v}_2}f)(x) \approx \mathbb{E}_{h_2\in V}\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m_1}}\int_V f(T_{\vec{a}_1\cdot\vec{v}_1}\bar{f})(T_{h_2}\bar{f})(T_{\vec{a}_1\cdot\vec{v}_1+h_2}f)$$

when m_2 is large compared to m_1 and then interchanging the expectations and applying Lemma 3.1 heuristically yields

$$\mathbb{E}_{h_2\in V}\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{m_1}}\int_V f(T_{\vec{a}_1\cdot\vec{v}_1}\bar{f})(T_{h_2}\bar{f})(T_{\vec{a}_1\cdot\vec{v}_1+h_2}f) \approx \mathbb{E}_{h_1\in V}\mathbb{E}_{h_2\in V}\int_V f(T_{h_1}\bar{f})(T_{h_2}\bar{f})(T_{h_1+h_2}f),$$

when m_1 is large, thus giving (3-1).

We will formalise the precise statement along these lines that we need later in this section. We begin with some key definitions.

Definition 3.5 (Lipschitz norm). If $G : \mathcal{D}^n \to \mathbb{C}$ is a function on a polydisk $\mathcal{D}^n$, we define the *Lipschitz norm* $\|G\|_{\mathrm{Lip}}$ of G to be the quantity

$$\|G\|_{\mathrm{Lip}} := \sup_{z\in\mathcal{D}^n}|G(z)| + \sup_{z,w\in\mathcal{D}^n:z\neq w}\frac{|G(z)-G(w)|}{d(z,w)},$$

where we use the metric

$$d\big((z_1,\ldots,z_n),(w_1,\ldots,w_n)\big) := |z_1-w_1| + \cdots + |z_n-w_n|.$$

Definition 3.6 (accurate sampling sequence). Let $k \geqslant 1$, let V be a finite-dimensional vector space, let $f : V \to \mathfrak{D}$ be a bounded function, and let

$$0 = H_0 < H_1 < H_2 < H_3 < \cdots$$

be a sequence of integers (or "scales"). We define an *accurate sampling sequence* for f of degree k and at scales $H_1, H_2, \ldots$ to be an infinite sequence of vectors

$$v_1, v_2, v_3, \ldots \in V$$

such that for every sequence

$$0 \leqslant w_0 < r_1 < \cdots < r_k$$

of scales and every Lipschitz function $G : \mathfrak{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \to \mathbb{C}$, we have

$$\int_V |G_{f,r_0,r_1,\ldots,r_k} - G_{f,r_0}| \leqslant \frac{\|G\|_{\mathrm{Lip}}}{r_1}, \tag{3-2}$$

where

$$G_{f,r_0,r_1,\ldots,r_k}(x) := \mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}},\ldots,\vec{a}_k \in \mathbb{F}^{H_{r_k}}} G\big((f(x + \omega \cdot \mathbf{u} + \vec{b} \cdot \vec{v}_0))_{\omega \in \{0,1\}^k, \vec{b} \in \mathbb{F}^{H_{r_0}}}\big),$$

where

$$\mathbf{u} = (\vec{a}_1 \cdot \vec{v}_1, \ldots, \vec{a}_k \cdot \vec{v}_k); \quad \vec{v}_j = (v_1, \ldots, v_{H_{r_j}}),\ j = 0, \ldots, k,$$

and

$$G_{f,r_0}(x) := \mathbb{E}_{h_1 \in V,\ldots,h_k \in V} G\big((f(x + \omega \cdot \mathbf{h} + \vec{b} \cdot \vec{v}_0))_{\omega \in \{0,1\}^k, \vec{b} \in \mathbb{F}^{H_{r_0}}}\big),$$

where $\mathbf{h} = (h_1, \ldots, h_k)$.

Remark 3.7. The denominator r_1 in (3-2) could be replaced by any other fixed function of r_1 that went to infinity as $r_1 \to \infty$ if desired here.

Remark 3.8. We make the trivial but useful remark that an accurate sampling sequence of degree k is also an accurate sampling sequence of degree k' for any $1 \leqslant k' \leqslant k$. Indeed, to verify (3-2) for a function $G' : \mathfrak{D}^{\{0,1\}^{k'} \times \mathbb{F}^{H_{r_0}}} \to \mathfrak{D}$ and some scales $r_{k'} > \cdots > r_0 \geqslant 0$, one simply adds some dummy scales $r_{k'+1}, \ldots, r_k$ above $r_{k'}$ and extends G' to a function $G : \mathfrak{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \to \mathfrak{D}$ by composing with the obvious restriction map from $\mathfrak{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}}$ to $\mathfrak{D}^{\{0,1\}^{k'} \times \mathbb{F}^{H_{r_0}}}$.

Roughly speaking, an accurate sampling sequence will allow us to estimate all the global averages that we need for the combinatorial inverse conjecture for the Gowers norm by local averages which are suitable for lifting to the ergodic setting via the correspondence principle. We illustrate the use of such sequences by describing the three special cases of (3-2) that we will actually need in our arguments.

Lemma 3.9 (global Gowers norm can be approximated by local Gowers norm). *Let $d \geqslant 1$, let V be a finite-dimensional vector space, let $f : V \to \mathfrak{D}$ be a bounded function, and let $v_1, v_2, \ldots \in V$ be an accurate sampling sequence for f of degree d and at scales $H_1, H_2, \ldots$. Then for every sequence of scales*

$$0 < r_1 < r_2 < \cdots < r_d,$$

we have

$$\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H_{r_1}},\ldots,\, \vec{a}_d \in \mathbb{F}^{H_{r_d}}} \int_V \Delta_{\vec{a}_1 \cdot \vec{v}_{r_1}} \cdots \Delta_{\vec{a}_d \cdot \vec{v}_{r_d}} f = \|f\|_{U^d(V)}^{2^d} + o_{r_1 \to \infty; d}(1).$$

Remark 3.10. As with all other estimates in this section, the point is that the error term is uniform over all choices of f and V. Note that the $d = 2$ case of this lemma is a formalisation of (3-1).

Proof. We apply (3-2) with $r_0 = 0$, and $G : \mathfrak{D}^{\{0,1\}^d} \to \mathbb{C}$ being the function

$$G\left((z(\omega))_{\omega\in\{0,1\}^d}\right) := \prod_{\omega\in\{0,1\}^d} \mathcal{C}^{\omega_1+\dots+\omega_d} z(\omega),$$

where $\mathcal{C} : z \mapsto \bar{z}$ is the complex conjugation operator. A routine computation gives the identities

$$G_{f,0,r_1,\dots,r_d}(x) = \mathbb{E}_{\vec{a}_1\in\mathbb{F}^{H_{r_1}},\dots,\vec{a}_d\in\mathbb{F}^{H_{r_d}}} \Delta_{\vec{a}_1\cdot\vec{v}_{r_1}} \dots \Delta_{\vec{a}_d\cdot\vec{v}_{r_d}} f, \qquad \int_V G_{f,0} = \|f\|_{U^d(V)}^{2^d}.$$

Also, it is easy to see that the Lipschitz norm $\|G\|_{\mathrm{Lip}}$ is $O_d(1)$. The claim now follows immediately from (3-2) and the triangle inequality. □

Lemma 3.11 (global averages can be approximated by local averages). *Let V be a finite-dimensional vector space, let $f : V \to \mathfrak{D}$ be a bounded function, and let $v_1, v_2, \dots \in V$ be an accurate sampling sequence for f of degree* 1 *and at scales $H_1, H_2, \dots$. Then for every finite sequence $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ and every continuous function $F : \mathfrak{D}^m \to \mathbb{C}$, we have*

$$\int_V |\mathbb{E}_{\vec{a}\in\mathbb{F}^{H_r}} T_{\vec{a}\cdot\vec{v}} g - \int_V g| = o_{r\to\infty; F,m,\vec{b}_1,\dots,\vec{b}_m}(1),$$

where $g : V \to \mathbb{C}$ is the function

$$g(x) := F(T_{\vec{b}_1\cdot\vec{v}} f(x), \dots, T_{\vec{b}_m\cdot\vec{v}} f(x)). \tag{3-3}$$

Proof. By approximating the continuous function F uniformly by a Lipschitz function, we may assume that F is Lipschitz. By adding dummy vectors to the collection $\vec{b}_1, \dots, \vec{b}_m$ if necessary, we may assume that $\{\vec{b}_1, \dots, \vec{b}_m\} = \mathbb{F}^{H_{r_0}}$ for some $r_0 > 0$ depending on $\vec{b}_1, \dots, \vec{b}_m$, thus F is now a Lipschitz function from $\mathfrak{D}^{\mathbb{F}^{H_{r_0}}}$ to $\mathbb{C}$.

Note that to prove the claim we may without loss of generality restrict to the regime $r > r_0$. We now apply (3-2) with $G : \mathfrak{D}^{\{0,1\}\times\mathbb{F}^{H_{r_0}}} \to \mathbb{C}$ being the function

$$G\big((z(\omega,\vec{b}))_{\omega\in\{0,1\},\vec{b}\in\mathbb{F}^{H_{r_0}}}\big) := F\big((z(1,\vec{b}))_{\vec{b}\in\mathbb{F}^{H_{r_0}}}\big).$$

A routine computation gives the identities

$$G_{f,r_0,r}(x) = \mathbb{E}_{\vec{a}\in\mathbb{F}^{H_r}} T_{\vec{a}\cdot\vec{v}} g(x), \qquad G_{f,r_0}(x) = \mathbb{E}_{h\in V} T_h g(x) = \int_V g.$$

Also, it is clear that G is Lipschitz with norm $O_{F,r_0}(1)$. The claim then follows from (3-2). □

Lemma 3.12 (global polynomiality test can be approximated by local polynomiality test). *Let $k \geqslant 1$, let V be a finite-dimensional vector space, let $f : V \to \mathfrak{D}$ be a bounded function, and let $v_1, v_2, \dots \in V$ be an accurate sampling sequence for f of degree k and at scales $H_1, H_2, \dots$. Then for every finite sequence $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ and every continuous function $F : \mathfrak{D}^m \to \mathbb{C}$, we have*

$$\mathbb{E}_{\vec{a}_1\in\mathbb{F}^{H_{r_1}}} \dots \mathbb{E}_{\vec{a}_k\in\mathbb{F}^{H_{r_k}}} \int_V \left|\Delta_{\vec{a}_1\cdot\vec{v}} \dots \Delta_{\vec{a}_k\cdot\vec{v}} g - 1\right| = \mathbb{E}_{h_1,\dots,h_k\in V} \int_V \left|\Delta_{h_1} \dots \Delta_{h_k} g - 1\right| + o_{r_1\to 0; F,m,\vec{b}_1,\dots,\vec{b}_m,k}(1)$$

for any $1 \leqslant r_1 < r_2 < \cdots < r_k$, *where* $g : V \to \mathbb{C}$ *is the function defined by* (3-3).

Proof. Arguing as in Lemma 3.11, we may assume that $\{\vec{b}_1, \ldots, \vec{b}_m\} = \mathbb{F}^{H_{r_0}}$ for some $r_0 > 0$ depending on $\vec{b}_1, \ldots, \vec{b}_m$, and that $F : \mathfrak{D}^{\mathbb{F}^{H_{r_0}}} \to \mathbb{C}$ is Lipschitz.

Note that to prove the claim we may without loss of generality restrict to the regime $r_1 > r_0$. We now apply (3-2) with $G : \mathfrak{D}^{\{0,1\}^k \times \mathbb{F}^{H_{r_0}}} \to \mathbb{C}$ being the function

$$G\big((z(\omega, \vec{b}))_{\omega\in\{0,1\}^k, \vec{b}\in\mathbb{F}^{H_{r_0}}}\big) := \Big| \prod_{\omega\in\{0,1\}^k} \mathscr{C}^{\omega_1+\ldots+\omega_k} F\big((z(\omega,\vec{b}))_{\vec{b}\in\mathbb{F}^{H_{r_0}}}\big) - 1\Big|,$$

where $\mathscr{C}$ is again the complex conjugation operator. A routine computation gives the identities

$$\begin{aligned} G_{f,r_0,r_1,\ldots,r_k}(x) &= \mathbb{E}_{\vec{a}_1\in\mathbb{F}^{H_{r_1}}} \ldots \mathbb{E}_{\vec{a}_k\in\mathbb{F}^{H_{r_k}}} \big|\Delta_{\vec{a}_1\cdot\vec{v}} \ldots \Delta_{\vec{a}_k\cdot\vec{v}}\, g(x) - 1\big|, \\ G_{f,r_0}(x) &= \mathbb{E}_{h_1,\ldots,h_k\in V}\big|\Delta_{h_1}\ldots\Delta_{h_k} g(x) - 1\big|, \end{aligned}$$

for any $r_0 < r_1 < \cdots < r_k$. Also it is clear that G is Lipschitz with norm $O_{F,r_0,k}(1)$. The claim then follows from (3-2) and the triangle inequality. □

Of course, in order to utilise the above lemmas we need to know that such accurate sampling sequences in fact exist. This is the purpose of the following proposition.

Proposition 3.13 (existence of accurate sampling sequence). *Let* $d \geqslant 1$. *Then there exists a sequence*

$$0 = H_0 < H_1 < H_2 < H_3 < \cdots$$

of integers such that for every finite-dimensional vector space V *and any function* $f : V \to \mathfrak{D}$, *there exists an accurate sampling sequence* $v_1, v_2, v_3, \ldots \in V$ *for* f *of degree* d *at scales* $H_1, H_2, \ldots$.

Remark 3.14. The key point here is that the scales $H_1, H_2, H_3, \ldots$ are *universal*; they depend on d, but otherwise and work for all vector spaces V and functions f.

Proof. We select H_j recursively by the formula $H_{j+1} := F(H_j)$, where $F = F_d : \mathbb{N} \to \mathbb{N}$ is a sufficiently rapidly growing function depending on d that we will choose later.

We use the probabilistic method, choosing $v_1, v_2, \ldots \in V$ uniformly at random, and showing that (if F was sufficiently rapid) the resulting sequence will be an accurate sampling sequence with positive probability.

We begin with observing that in order to verify the condition (3-2), it suffices by the triangle inequality to show that with positive probability, one has

$$\int_V \big|G_{f,r_0,r_1,\ldots,r_{d'}} - G_{f,r_0,r_1,\ldots,r_{d'-1}}\big| \leqslant \frac{\|G\|_{\mathrm{Lip}}}{dr_1} \tag{3-4}$$

for all $1 \leqslant d' \leqslant d$, all $0 \leqslant r_0 < \cdots < r_{d'}$, and every Lipschitz function $G : \mathfrak{D}^{\{0,1\}^d \times \mathbb{F}^{H_{r_0}}} \to \mathbb{C}$, where

$$\begin{aligned} &G_{f,r_0,r_1,\ldots,r_{d'}}(x) \\ &\quad := \mathbb{E}_{\vec{a}_1\in\mathbb{F}^{H_{r_1}},\ldots,\vec{a}_{d'}\in\mathbb{F}^{H_{r_{d'}}}} \mathbb{E}_{h_{d'+1},\ldots,h_d\in V}\, G\bigg(\bigg(f\bigg(x + \sum_{j=1}^{d'} \omega_j \vec{a}_j\cdot\vec{v}_j + \sum_{j=d'+1}^{d} \omega_j h_j + b\cdot\vec{v}_0\bigg)\bigg)_{\substack{(\omega_1,\ldots,\omega_d)\in\{0,1\}^d \\ b\in\mathbb{F}^{H_{r_0}}}}\bigg). \end{aligned}$$

By the union bound, it will suffice to show that for all $1 \leqslant d' \leqslant d$ and all $0 \leqslant r_0 < \cdots < r_{d'}$, with probability $1 - o_{H_{r_{d'}} \to \infty; d, H_{r_0}, \ldots, H_{r_{d'-1}}, r_1}(1)$, (3-4) holds for all Lipschitz functions $G : \mathfrak{D}^{\{0,1\}^d \times \mathbb{F}^{H r_0}} \to \mathbb{C}$, since the total failure probability can be made to be less than 1 by choosing F to be sufficiently rapid.

We can normalise G to have Lipschitz norm 1. By the Arzelà–Ascoli theorem, the space of such functions is compact in the uniform topology. In particular, there exists a collection of functions

$$G : \mathfrak{D}^{\{0,1\}^d \times \mathbb{F}^{H r_0}} \to \mathbb{C}$$

of Lipschitz norm 1, S, of size $O_{d, H_{r_0}, r_1}(1)$, such that any other such Lipschitz function lies within $1/(4dr_1)$, say, of a function $G \in S$ in the uniform metric. Because of this, we see from the union bound again that it will suffice to show that for all $1 \leqslant d' \leqslant d$ and all $0 \leqslant r_0 < \cdots < r_{d'}$, and all functions $G : \mathfrak{D}^{\{0,1\}^d \times \mathbb{F}^{H r_0}} \to \mathbb{C}$ of Lipschitz norm 1 in S,

$$\int_V \left| G_{f, r_0, r_1, \ldots, r_{d'}} - G_{f, r_0, r_1, \ldots, r_{d'-1}} \right| \leqslant \frac{1}{2dr_1} \tag{3-5}$$

of (3-4) holds with probability $1 - o_{H_{r_{d'}} \to \infty; d, H_{r_0}, \ldots, H_{r_{d'-1}}, r_1}(1)$.

Fix $d', r_0, \ldots, r_{d'}, G$. By Markov's inequality, it suffices to show that

$$\mathbb{E} \int_V \left| G_{f, r_0, r_1, \ldots, r_{d'}} - G_{f, r_0, r_1, \ldots, r_{d'-1}} \right| = o_{H_{r_{d'}} \to \infty; d, H_{r_0}, \ldots, H_{r_{d'-1}}}(1);$$

by linearity of expectation it thus suffices to show that

$$\mathbb{E} \left| G_{f, r_0, r_1, \ldots, r_{d'}}(x) - G_{f, r_0, r_1, \ldots, r_{d'-1}}(x) \right| = o_{H_{r_{d'}} \to \infty; d, H_{r_0}, \ldots, H_{r_{d'-1}}}(1)$$

uniformly in $x \in V$.

Fix x. We observe that

$$G_{f, r_0, r_1, \ldots, r_{d'}}(x) = \mathbb{E}_{\vec{a} \in \mathbb{F}^{H r_{d'}}} f_{v_1, \ldots, v_{H_{r_{d'-1}}}}(\vec{a} \cdot \vec{v}_{d'}), \qquad G_{f, r_0, r_1, \ldots, r_{d'-1}}(x) = \mathbb{E}_{h \in V} f_{v_1, \ldots, v_{H_{r_{d'-1}}}}(h),$$

where $f_{v_1, \ldots, v_{H_{r_{d'-1}}}} : V \to \mathfrak{D}$ is the function

$$f_{v_1, \ldots, v_{H_{r_{d'-1}}}}(h) :=$$
$$\mathbb{E}_{\vec{a}_1 \in \mathbb{F}^{H r_1}, \ldots, \vec{a}_{d'-1} \in \mathbb{F}^{H r_{d'-1}}} \mathbb{E}_{h_{d'+1}, \ldots, h_d \in V} G\Bigg(\Bigg(f\Bigg(x + \sum_{j=1}^{d'-1} \omega_j \vec{a}_j \cdot \vec{v}_j + \omega_{d'} h_{d'} + \sum_{j=d'+1}^{d} \omega_j h_j + b \cdot \vec{v}_0 \Bigg) \Bigg)_{\substack{(\omega_1, \ldots, \omega_d) \in \{0,1\}^d \\ b \in \mathbb{F}^{H r_0}}} \Bigg).$$

As the notation suggests, the function $f_{v_1, \ldots, v_{H_{r_{d'-1}}}}$ depends on the values of $v_1, \ldots, v_{H_{r_{d'-1}}}$ but not on higher elements of the sequence. Also, as G has Lipschitz norm 1, f takes values in $\mathfrak{D}$. The claim now follows from Corollary 3.3. $\square$

4. Proof of the main theorems

We are now ready to prove the main theorems. We shall just prove Theorem 1.10 using Theorem 1.20; the deduction of Theorem 1.9 using Theorem 1.19 is exactly analogous (see the brief remarks at the end of this section).

Fix $\mathbb{F}$ and d, and let $k = C(d)$ be the quantity in Theorem 1.20. By increasing k if necessary we may assume $k \geqslant d$. Assume for sake of contradiction that Theorem 1.10 failed for this choice of $\mathbb{F}, d, k$. Then we can find $\delta > 0$ and a sequence $f^{(n)} : V^{(n)} \to \mathfrak{D}$ of functions on finite-dimensional vector spaces $V^{(n)}$ such that

$$\|f^{(n)}\|_{U^d(V^{(n)})} \geqslant \delta \tag{4-1}$$

for all n, but

$$\|f^{(n)}\|_{u^k(V^{(n)})} = o_{n\to\infty}(1). \tag{4-2}$$

We now let $F(x) := x$, and let

$$1 < H_1 < H_2 < \cdots$$

be the sequence in Proposition 3.13; it is important to note that this sequence does not depend on n. From that proposition, we can find an accurate sampling sequence

$$v_1^{(n)}, v_2^{(n)}, \ldots \in V^{(n)}$$

for $f^{(n)}$ of degree k at these scales. We fix such a sequence for each n.

We will use these sampling sequences to lift the functions $f^{(n)}$ on $V^{(n)}$ to a universal dynamical system for $\mathbb{F}^\omega$ by the usual Furstenberg correspondence principle method. We begin by constructing this universal space.

Definition 4.1 (Furstenberg universal space). Let $X := \mathfrak{D}^{\mathbb{F}^\omega}$ be the space of functions $\zeta : \mathbb{F}^\omega \to \mathfrak{D}$. With the product topology, this is a compact metrisable space, with Borel σ-algebra $\mathcal{B}$. It has a continuous action $h \mapsto T_h$ of the additive group $\mathbb{F}^\omega$, defined by the formula

$$T_h\zeta(x) := \zeta(x+h).$$

We let $\Pr(X)^T$ be the space of all Borel probability measures μ on X which are invariant with respect to this action; note that $\boldsymbol{X} = (X, \mathcal{B}, \mu, (T_h)_{h\in\mathbb{F}^\omega})$ is a $\mathbb{F}^\omega$-system for any $\mu \in \Pr(X)^T$. If $\mu^{(n)} \in \Pr(X)^T$ is a sequence of such measures, and $\mu \in \Pr(X)^T$ is another measure, we say that $\mu^{(n)}$ *converges vaguely* to μ if we have

$$\lim_{n\to\infty} \int_X \phi(\zeta)\, d\mu^{(n)}(\zeta) \to \int_X \phi(\zeta)\, d\mu(\zeta)$$

for all continuous functions $\phi : X \to \mathbb{C}$.

Because X is compact metrisable, and the action of T is continuous it is a well known fact that $\Pr(X)^T$ is sequentially compact; thus every sequence of measures in $\Pr(X)^T$ has a vaguely convergent subsequence whose limit is also in $\Pr(X)^T$.

For each n, we define a measure $\mu^{(n)} \in \Pr(X)^T$ on X by the formula

$$\mu^{(n)} = \mathbb{E}_{x\in V^{(n)}} \delta_{\zeta_{n,x}},$$

where δ denotes the Dirac mass and for each $x \in V^{(n)}$, $\zeta_{n,x} \in X$ is the function

$$\zeta_{n,x}(\vec{a}) := T_{\vec{a}\cdot\vec{v}^{(n)}} f^{(n)}(x) = T_{\sum_{m=1}^\infty a_m v_m^{(n)}} f^{(n)}(x)$$

for all $\vec{a} \in \mathbb{F}^\omega$ (note the sum on the right side has only finitely many nonzero terms). Observe that $\mu^{(n)}$ is indeed T-invariant. By passing to a subsequence if necessary, we may assume that $\mu^{(n)}$ converges vaguely to a limit $\mu \in \Pr(X)^T$. We write $\boldsymbol{X} := (X, \mathcal{B}, \mu, (T_h)_{h\in\mathbb{F}^\omega})$.

Let $f : X \to \mathcal{D}$ be the indicator function $f(\zeta) := \zeta(0)$. We observe the key correspondence

$$\int_X G(T_{\vec{a}_1} f, \dots, T_{\vec{a}_k} f)\, d\mu^{(n)}(\zeta) = \int_{V^{(n)}} G(T_{\vec{a}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{a}_k \cdot \vec{v}^{(n)}} f^{(n)}) \tag{4-3}$$

for all $\vec{a}_1, \dots, \vec{a}_k \in \mathbb{F}^\omega$, all n, and all continuous $G : \mathcal{D}^k \to \mathbb{C}$.

We now record the (standard) fact that the countable collection of shifts $T_h f$ for $h \in \mathbb{F}^\omega$ generates $L^\infty(\boldsymbol{X})$:

Lemma 4.2 ($T_h f$ generate $L^\infty(\boldsymbol{X})$). *Given any $\phi \in L^\infty(\boldsymbol{X})$ and $\varepsilon > 0$, there exists a finite number of shifts $\vec{h}_1, \dots, \vec{h}_k \in \mathbb{F}^\omega$ and a continuous function $G : \mathcal{D}^k \to \mathbb{C}$ such that*

$$\int_X \left|\phi - G(T_{\vec{h}_1} f, \dots, T_{\vec{h}_k} f)\right| d\mu \leqslant \varepsilon.$$

Proof. For continuous ϕ, the claim follows easily from the Stone–Weierstrass theorem (and in this case we can upgrade the L^1 approximation to L^∞ approximation). As X is compact metrisable, the Borel measure μ is in fact a Radon measure, and so (by Urysohn's lemma) the continuous functions are dense in $L^\infty(\boldsymbol{X})$ in the $L^1(\boldsymbol{X})$ topology, and the claim follows. □

We can now use the machinery of the previous section to deduce various important facts about $\boldsymbol{X}$ and f. For instance, Lemma 3.11 now implies

Lemma 4.3 (ergodicity). *$\boldsymbol{X}$ is ergodic.*

Proof. By the mean ergodic theorem, it suffices to show that

$$\lim_{r \to \infty} \int_X \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_h g - \int_X g \, d\mu \right| d\mu = 0$$

for all $g \in L^\infty(X)$. By Lemma 4.2 and a standard limiting argument it suffices to show this for g which are functions of finitely many shifts of f, say $g = G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_k} f)$. We will then show that

$$\int_X \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h}} g - \int_X g \, d\mu \right| d\mu = o_{r \to \infty; G, k, \vec{b}_1, \dots, \vec{b}_k}(1).$$

By vague convergence it suffices to show that

$$\int_X \left| \mathbb{E}_{h \in \mathbb{F}^{H_r}} T_{\vec{h}} g - \int_X g \, d\mu^{(n)} \right| d\mu^{(n)} = o_{r \to \infty; G, k, \vec{b}_1, \dots, \vec{b}_k}(1)$$

for all n. By (4-3), we can rewrite the left side as

$$\int_V \left| \mathbb{E}_{\vec{h} \in \mathbb{F}^{H_r}} T_{\vec{h} \cdot \vec{v}_r^{(n)}} g^{(n)} - \int_V g^{(n)} \right|,$$

where

$$g^{(n)} := G(T_{\vec{b}_1 \cdot \vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_k \cdot \vec{v}^{(n)}} f^{(n)}).$$

But the claim now follows from Lemma 3.11 (and Remark 3.8). □

In a similar spirit, Lemma 3.9 implies this:

Lemma 4.4 (f has large Gowers–Host–Kra norm). $\|f\|_{U^d(\boldsymbol{X})} \geqslant \delta.$

Proof. From the mean ergodic theorem we have

$$\|f\|_{U^1(X)}^2 = \limsup_{K_1\to\infty} \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{K_1}} \int_X \Delta_{\vec{h}_1} f \, d\mu,$$

and by induction we have

$$\|f\|_{U^d(X)}^{2^d} = \limsup_{K_d\to\infty} \dots \limsup_{K_1\to\infty} \mathbb{E}_{\vec{h}_d\in\mathbb{F}^{K_d}} \dots \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{K_1}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f \, d\mu.$$

It thus suffices to show that

$$\mathbb{E}_{\vec{h}_d\in\mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f \, d\mu > \delta^{2^d} - o_{r_d\to\infty}(1)$$

whenever $1 \leqslant r_d < \dots < r_1$. By reversing the order of averages, it suffices to show that

$$\mathbb{E}_{\vec{h}_d\in\mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f \, d\mu > \delta^{2^d} - o_{r_1\to\infty}(1)$$

whenever $1 \leqslant r_1 < \dots < r_d$. Fix $r_1, \dots, r_d$. By weak convergence, it suffices to show that

$$\mathbb{E}_{\vec{h}_d\in\mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f \, d\mu^{(n)} > \delta^{2^d} - o_{r_1\to\infty}(1)$$

for all n. By (4-1), it suffices to show that

$$\mathbb{E}_{\vec{h}_d\in\mathbb{F}^{H_{r_d}}} \dots \mathbb{E}_{\vec{h}_1\in\mathbb{F}^{H_{r_1}}} \int_X \Delta_{\vec{h}_1} \dots \Delta_{\vec{h}_d} f \, d\mu^{(n)} > \|f^{(n)}\|_{U^d(V^{(n)})}^{2^d} - o_{r_1\to\infty}(1).$$

By (4-3), left side can be rephrased as

$$\int_V \mathbb{E}_{\vec{a}_1\in\mathbb{F}^{H_{r_1}},\dots,\vec{a}_d\in\mathbb{F}^{H_{r_d}}} \Delta_{\vec{a}_1\cdot\vec{v}_{r_1}^{(n)}} \dots \Delta_{\vec{a}_d\cdot\vec{v}_{r_d}^{(n)}} f^{(n)},$$

and the claim now follows from Lemma 3.9 (and Remark 3.8). □

We have now verified all the hypotheses of Theorem 1.19. Applying that theorem, we conclude that $\|f\|_{u^k(X)} > c$ for some $c > 0$ (which could be very small, but positive). Thus we can find a phase polynomial $\phi \in \mathcal{P}_{k-1}(X)$ of degree $k-1$ such that

$$\Big|\int_X f\overline{\phi} \, d\mu\Big| > c.$$

Let $\varepsilon > 0$ be a small number (depending on d, k, c) to be chosen later. By Lemma 4.2, we can find $\vec{b}_1, \dots, \vec{b}_m \in \mathbb{F}^\omega$ (with m potentially quite large, but finite) and a continuous $G : \mathcal{D}^m \to \mathbb{C}$ such that

$$\int_X \big|\phi - G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f)\big| \leqslant \varepsilon. \tag{4-4}$$

Since ϕ takes values in $\mathcal{D}$, we may assume without loss of generality that G does also. If ε is small enough depending on c, we thus have

$$\Big|\int_X f\overline{G(T_{\vec{b}_1} f, \dots, T_{\vec{b}_m} f)} \, d\mu\Big| > c/2.$$

By vague convergence, we thus have

$$\Big|\int_X f\overline{G(T_{\vec b_1}f,\dots,T_{\vec b_m}f)}\,d\mu^{(n)}\Big| > c/4$$

for all sufficiently large n (depending on G, m, c). Using (4-3), we rearrange this as

$$\Big|\int_V f^{(n)}\overline{G(T_{\vec b_1\cdot\vec v^{(n)}}f^{(n)},\dots,T_{\vec b_m\cdot\vec v^{(n)}}f^{(n)})}\Big| > c/4. \tag{4-5}$$

Now let r_1 be a large integer depending on the $\vec b_1,\dots,\vec b_m,\varepsilon$, and let $r_j := r_1+(j-1)$ for $j=2,\dots,d$. Since ϕ is a phase polynomial of degree $k-1$, we have

$$\int_X \big|\Delta_{\vec a_1}\dots\Delta_{\vec a_k}\phi - 1\big|\,d\mu = 0$$

for all $\vec a_1\in\mathbb{F}^{H_{r_1}},\dots,\vec a_k\in\mathbb{F}^{H_{r_k}}$. From many applications of (4-4), the triangle inequality, and the boundedness of ϕ, G, we conclude that

$$\int_X \big|\Delta_{\vec a_1}\dots\Delta_{\vec a_k}G(T_{\vec b_1}f,\dots,T_{\vec b_m}f) - 1\big|\,d\mu \ll_k \varepsilon$$

for all $\vec a_1\in\mathbb{F}^{H_{r_1}},\dots,\vec a_k\in\mathbb{F}^{H_{r_k}}$. By vague convergence, this implies that

$$\int_X \big|\Delta_{\vec a_1}\dots\Delta_{\vec a_k}G(T_{\vec b_1}f,\dots,T_{\vec b_m}f) - 1\big|\,d\mu^{(n)} \ll_k \varepsilon$$

for all sufficiently large n (depending on $\varepsilon, H_{r_1},\dots,H_{r_k}$). Using (4-3), we can rearrange the left side as

$$\int_{V^{(n)}} \big|\Delta_{\vec a_1\cdot\vec v^{(n)}}\dots\Delta_{\vec a_k\cdot\vec v^{(n)}}G(T_{\vec b_1\cdot\vec v^{(n)}}f^{(n)},\dots,T_{\vec b_m\cdot\vec v^{(n)}}f^{(n)}) - 1\big|,$$

and so on averaging we obtain

$$\mathbb{E}_{\vec a_1\in\mathbb{F}^{H_{r_1}},\dots,\vec a_k\in\mathbb{F}^{H_{r_k}}}\int_{V^{(n)}} \big|\Delta_{\vec a_1\cdot\vec v^{(n)}}\dots\Delta_{\vec a_k\cdot\vec v^{(n)}}G(T_{\vec b_1\cdot\vec v^{(n)}}f^{(n)},\dots,T_{\vec b_m\cdot\vec v^{(n)}}f^{(n)}) - 1\big| \ll_k \varepsilon.$$

Applying Lemma 3.12 we conclude (if r_1 is sufficiently large depending on $\vec b_1,\dots,\vec b_m,\varepsilon$) that

$$\mathbb{E}_{h_1,\dots,h_k\in V^{(n)}}\int_{V^{(n)}} \big|\Delta_{h_1}\dots\Delta_{h_k}G(T_{\vec b_1\cdot\vec v^{(n)}}f^{(n)},\dots,T_{\vec b_m\cdot\vec v^{(n)}}f^{(n)}) - 1\big| \ll_k \varepsilon.$$

Now we invoke a local testability lemma:

Lemma 4.5 (polynomiality is locally testable). *Let V be a finite-dimensional vector space, let $k\geqslant 1$, let $g : V\to\mathcal{D}$ be a bounded function, and suppose that*

$$\mathbb{E}_{h_1,\dots,h_k\in V}\int_V \big|\Delta_{h_1}\dots\Delta_{h_k}g - 1\big| \leqslant \varepsilon \tag{4-6}$$

for some $\varepsilon > 0$. Then there exists a phase polynomial $\phi\in\mathcal{P}_{k-1}(V)$ such that

$$\int_V |g-\phi| \leqslant o_{\varepsilon\to 0;d}(1).$$

For $\mathbb{F} = \mathbb{F}_2$, this result is essentially in [Alon et al. 2003; 2005] or [Tao 2007, Proposition 4.6], but for the convenience of the reader, and in view of the subtle difference between phase polynomials and polynomials (see Remark 1.3), we give a full proof of this lemma in Appendix A.

Applying this lemma, we conclude that there exists $\phi^{(n)} \in \mathcal{P}_{k-1}(V^{(n)})$ such that

$$\int_V \left|G(T_{\vec{b}_1\cdot\vec{v}^{(n)}} f^{(n)}, \dots, T_{\vec{b}_m\cdot\vec{v}^{(n)}} f^{(n)}) - \phi^{(n)}\right| \leqslant o_{\varepsilon\to 0;k}(1).$$

Inserting this into (4-5) we conclude that

$$\left|\int_V f^{(n)}\overline{\phi^{(n)}}\right| > c/8$$

if ε is sufficiently small depending on c, k. But this contradicts (4-2). The proof of Theorem 1.10 is complete.

The proof of Theorem 1.9 is identical, but with k now set equal to d, and Theorem 1.19 used instead of Theorem 1.20. We leave the details to the reader.

Remark 4.6. It is tempting to try to adapt these arguments to the cyclic setting $\mathbb{Z}/N\mathbb{Z}$, in which the role of polynomials is replaced by that of a *nilsequence* (see [2006; 2008] for further discussion), thus establishing the *inverse conjecture for the Gowers norm* for $\mathbb{Z}/N\mathbb{Z}$ that was formulated in those papers. The analogue of Theorem 1.19 is known; see [Host and Kra 2005]. However, two obstructions remain before one can carry out this program. The first is to compensate for the rigidity of arithmetic progressions that seems to prevent a counterpart of Corollary 3.3 from holding in the cyclic group setting (see Remark 3.4). The second is that whereas polynomiality is locally testable thanks to Lemma 4.5, it is unclear whether the property of being a nilsequence is similarly testable.

Appendix: Proof of Lemma 4.5

In this appendix we give a proof of Lemma 4.5, following the arguments in [Alon et al. 2003; 2005] and [Tao 2007, Proposition 4.6]. We begin with a variant of Lemma 1.2:

Lemma A.1 (discreteness). *Let $k \geqslant 0$, let V be a finite-dimensional vector space, and $\phi \in \mathcal{P}_k(V)$. Then there exists $\theta \in \mathbb{R}/\mathbb{Z}$ and an integer $K \geqslant 1$ depending only on $\mathbb{F}$ such that $\phi(x)$ is equal to $e^{2\pi i\theta}$ times a K-th root of unity for every $x \in V$.*

Proof. See [Bergelson et al. 2009, Lemma D.5], (which gives the explicit value $K = p^{\lfloor k/p\rfloor+1}$, where p is the characteristic of $\mathbb{F}$). □

Lemma A.2 (rigidity). *Let $k \geqslant 0$, let V be a finite-dimensional vector space, and take $\phi \in \mathcal{P}_k(V)$. Suppose that $\int_V |\phi-1| \leqslant \varepsilon$ for some $\varepsilon > 0$. If ε is sufficiently small depending on k, $\mathbb{F}$, then ϕ is constant.*

Proof. We induct on k. For $k = 0$ the claim is obvious, and for $k = 1$ ϕ is a linear character (times a phase) and the claim can be worked out by hand. Now suppose $k \geqslant 2$ and the claim has already been shown for smaller values of k. Since ϕ is a phase polynomial, we have $\Delta_0 \dots \Delta_0\phi = 1$, and thus ϕ has unit magnitude. Observe that if $\int_V |\phi - 1| \leqslant \varepsilon$, then $\int_V |T_h\phi - 1| \leqslant \varepsilon$ for every $h \in V$. Using the elementary estimate

$$|\Delta_h\phi - 1| \leqslant |\phi - 1| + |T_h\phi - 1|$$

(using the fact that ϕ has unit magnitude) we conclude that

$$\int_V |\Delta_h \phi - 1| \leqslant 2\varepsilon,$$

for every $h \in V$. On the other hand, $\Delta_h \phi \in \mathcal{P}_{k-1}(V)$, so by induction hypothesis (if ε is small enough) we conclude that $\Delta_h \phi$ is constant for all $h \in V$. Thus $\phi \in \mathcal{P}_1(V)$, but then the claim follows from the case of $k-1$. □

We now prove Lemma 4.5. The case $k = 1$ is easy, so suppose that $k \geqslant 2$ and the claim has already been established for $k-1$. To abbreviate the notation we shall write $o(1)$ for $o_{\varepsilon \to 0;k}(1)$. We say that a statement $P(x)$ holds for *most* $x \in V$ if it holds for $(1 - o(1))|V|$ elements of v.

We fix k, V, f. We may assume that ε is small depending on d, as the claim is trivial otherwise. From (4-6) and Markov's inequality we see that

$$\mathbb{E}_{h_1,\dots,h_{k-1} \in V} \int_V \left| \Delta_{h_1} \dots \Delta_{h_{k-1}} \Delta_h f - 1 \right| = o(1) \tag{A-1}$$

for most $h \in V$. Let us call h *good* if (A-1) holds. Applying the induction hypothesis, we conclude that for any good h there exists[2] $\phi_h \in \mathcal{P}_{k-2}(V)$ such that

$$\int_V |\Delta_h f - \phi_h| \leqslant o(1).$$

In particular, this implies (by Markov's inequality) that for all good h, we have

$$f(x+h)\overline{f(x)} = \phi_h(x) + o(1)$$

for most V. Since f is bounded in magnitude by 1, this implies that

$$|f(x)| = 1 - o(1)$$

for most x, and for all good h we have

$$f(x+h) = \phi_h(x) f(x) + o(1) \tag{A-2}$$

for most x.

We now pause to perform a discretisation trick. Write $p := \operatorname{char} \mathbb{F}$. From repeated applications of (A-2) we see that

$$f(x) = f(x + ph) = \phi_h(x)\phi_h(x+h) \dots \phi_h(x + (p-1)h) f(x) + o(1)$$

for most x, and thus

$$\phi_h(x)\phi_h(x+h) \dots \phi_h(x + (p-1)h) = 1 + o(1)$$

for at least one x. On the other hand, from Lemma A.1 ϕ_h takes values in $e^{2\pi i \theta}$ times K-th roots of unity for some fixed K depending only on d, p. Thus $e^{2\pi i p \theta}$ times a K-th root of unity is within $o(1)$ of 1,

[2] This quantity plays the same role that *cocycles* do in ergodic theory.

and so $e^{2\pi i\theta}$ lies within $o(1)$ of a pK-th root of unity. Rotating ϕ_h by $o(1)$ if necessary we may assume that $e^{2\pi i\theta}$ is *exactly* a pK-th root of unity, and in particular we have

$$\phi_h^{pK} \equiv 1 \tag{A-3}$$

whenever h is good.

Now suppose that h_1, h_2, h_3, h_4 are good and form an *additive quadruple* in the sense that $h_1 + h_2 = h_3 + h_4$. Then from (A-2) we see that

$$f(x + h_1 + h_2) = f(x)\phi_{h_1}(x)\phi_{h_2}(x + h_1) + o(1) \tag{A-4}$$

for most x, and similarly

$$f(x + h_3 + h_4) = f(x)\phi_{h_3}(x)\phi_{h_4}(x + h_3) + o(1)$$

for most x. Since $|f(x)| = 1 + o(1)$ for most x, we conclude the approximate cocycle relationship

$$\phi_{h_1}(x)\phi_{h_2}(x + h_1)\overline{\phi_{h_3}(x)\phi_{h_4}(x + h_3)} = 1 + o(1)$$

for most x. In particular, the average of the left side in x is $1 - o(1)$. Applying Lemma A.2 (and assuming ε small enough), we conclude that the left side is *constant* in x; using the discretisation (A-3), we conclude (again for ε small enough) that it is in fact 1. Thus

$$\phi_{h_1}(x)\phi_{h_2}(x + h_1) = \phi_{h_3}(x)\phi_{h_4}(x + h_3) \tag{A-5}$$

for *all* x and any good additive quadruple h_1, h_2, h_3, h_4.

Now for any $k \in V$, define the quantity $\psi(k) \in \mathbb{C}$ by the formula

$$\psi(k) := \phi_{h_1}(0)\phi_{h_2}(h_1) \tag{A-6}$$

whenever h_1, h_2, h_1+h_2 are simultaneously good. Note that the existence of such an h_1, h_2 is guaranteed since most h are good, and (A-5) ensures that the right side of (A-6) does not depend on the exact choice of h_1, h_2 and so ψ is well-defined. From (A-3) we see that ψ takes values in the pK-th roots of unity, and in particular only has $O(1)$ possible values.

Now let $x \in V$ and h be good. Then, since most elements of V are good, we can find good r_1, r_2, s_1, s_2 such that $r_1 + r_2 = x$ and $s_1 + s_2 = x + h$. From (A-4) we see that

$$\begin{aligned} f(y + x) &= f(y + r_1 + r_2) = f(y)\phi_{r_1}(y)\phi_{r_2}(y + r_1) + o(1), \\ f(y + x + h) &= f(y + s_1 + s_2) = f(y)\phi_{s_1}(y)\phi_{s_2}(y + s_1) + o(1), \end{aligned}$$

for most y. Also from (A-2) we have

$$f(y + x + h) = f(y + x)\phi_h(y + x) + o(1)$$

for most y. Combining these (and the fact that $|f(y)| = 1 + o(1)$ for most y) we see that

$$\phi_{s_1}(y)\phi_{s_2}(y + s_1)\overline{\phi_{r_1}(y)\phi_{r_2}(y + r_1)\phi_h(y + x)} = 1 + o(1)$$

for most y. Taking expectations and applying Lemma A.2 and (A-3) as before, we conclude that

$$\phi_{s_1}(y)\phi_{s_2}(y + s_1)\overline{\phi_{r_1}(y)\phi_{r_2}(y + r_1)\phi_h(y + x)} = 1$$

for *all* y. Specialising to $y = 0$ and applying (A-6) we conclude that

$$\phi_h(x) = \psi(x+h)\overline{\psi(x)} = \Delta_h \psi(x) \tag{A-7}$$

for all $x \in V$ and good h; thus we have successfully "integrated" ϕ_h. We can then extend $\phi_h(x)$ to all $h \in V$ (not just good h) by viewing (A-7) as a *definition*. Observe that if $h \in V$, then $h = h_1 + h_2$ for some good h_1, h_2, and from (A-7) we have

$$\phi_h(x) = \phi_{h_1}(x)\phi_{h_2}(x+h_1).$$

In particular, since the right side lies in $\mathcal{P}_{k-2}(V)$, the left side does also. Thus we see that $\Delta_h \psi \in \mathcal{P}_{k-2}(V)$ for all $h \in V$, and thus $Q \in \mathcal{P}_{k-1}(V)$. If we then set $g(x) := f(x)\overline{\psi}(x)$, then from (A-2), (A-7) we see that for every $h \in H$ we have

$$g(x+h) = g(x) + o(1)$$

for most x. From Fubini's theorem, we thus conclude that there exists an x such that $g(x+h) = g(x) + o(1)$ for most h, thus g is almost constant. Since $|g(x)| = 1 + o(1)$ for most x, we thus conclude the existence of a phase $\theta \in \mathbb{R}/\mathbb{Z}$ such that $g(x) = e^{2\pi i\theta} + o(1)$ for most x. We conclude that

$$f(x) = e^{2\pi i\theta}\psi(x) + o(1)$$

for most x, and Lemma 4.5 then follows.

Acknowledgments

The authors are greatly indebted to Ben Green for helpful conversations and to Vitaly Bergelson for encouragement.

References

[Alon et al. 2003] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron, "Testing low-degree polynomials over GF(2)", pp. 188–199 in *Approximation, randomization, and combinatorial optimization*, edited by S. Arora et al., Lecture Notes in Comput. Sci. **2764**, Springer, Berlin, 2003. MR 2005c:68285

[Alon et al. 2005] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron, "Testing Reed–Muller codes", *IEEE Transactions on Information Theory* **51** (2005), 4032–4039.

[Bergelson et al. 2009] V. Bergelson, T. Tao, and T. Ziegler, "An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$", preprint, University of California, Los Angeles, 2009, Available at http://tinyurl.com/yhdz5a5.

[Bogdanov and Viola 2007] A. Bogdanov and E. Viola, "Pseudorandom bits for polynomials", pp. 41–51 in *FOCS '07: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society, Washington, DC, USA, 2007.

[Frantzikinakis et al. 2007] N. Frantzikinakis, B. Host, and B. Kra, "Multiple recurrence and convergence for sequences related to the prime numbers", *J. Reine Angew. Math.* **611** (2007), 131–144. MR 2009c:11023 Zbl 1126.37005

[Furstenberg 1977] H. Furstenberg, "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions", *J. Analyse Math.* **31** (1977), 204–256. MR 58 #16583 Zbl 0347.28016

[Gowers 1998] W. T. Gowers, "A new proof of Szemerédi's theorem for arithmetic progressions of length four", *Geom. Funct. Anal.* **8**:3 (1998), 529–551. MR 2000d:11019 Zbl 0907.11005

[Gowers 2001] W. T. Gowers, "A new proof of Szemerédi's theorem", *Geom. Funct. Anal.* **11**:3 (2001), 465–588. MR 2002k:11014 Zbl 1028.11005

[Gowers and Wolf 2007] T. Gowers and J. Wolf, "The true complexity of a system of linear equations", preprint, 2007. arXiv 0711.0185v1

[Green and Tao 2006] B. Green and T. Tao, "Linear equations in primes", preprint, 2006. To appear in *Annals of Mathematics*. arXiv 0606088v2

[Green and Tao 2008] B. Green and T. Tao, "An inverse theorem for the Gowers $U^3(G)$ norm", *Proc. Edinb. Math. Soc.* (2) **51**:1 (2008), 73–153. MR 2009g:11012 Zbl 05249333

[Green and Tao 2009a] B. Green and T. Tao, "The distribution of polynomials over finite fields, with applications to the Gowers norms", *Contributions to Discrete Mathematics* **4**:2 (2009), 1–36.

[Green and Tao 2009b] B. Green and T. Tao, "New bounds for Szemerédi's theorem, I: Progressions of length 4 in finite field geometries", *Proc. Lond. Math. Soc.* (3) **98**:2 (2009), 365–392. MR 2010a:11020

[Host and Kra 2005] B. Host and B. Kra, "Nonconventional ergodic averages and nilmanifolds", *Ann. of Math.* (2) **161**:1 (2005), 397–488. MR 2007b:37004 Zbl 1077.37002

[Lovett et al. 2007] S. Lovett, R. Meshulam, and A. Samorodnitsky, "Inverse conjecture for the Gowers norm is false", preprint, 2007. arXiv 0711.3388v3

[Samorodnitsky 2007] A. Samorodnitsky, "Low-degree tests at large distances", pp. 506–515 in *STOC'07. Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (San Diego, CA), edited by U. Feige, ACM, New York, 2007. MR 2009f:68077 Zbl 05485486

[Sudan et al. 2001] M. Sudan, L. Trevisan, and S. Vadhan, "Pseudorandom generators without the XOR lemma", *J. Comput. System Sci.* **62**:2 (2001), 236–266. MR 2002f:68040 Zbl 1005.65006

[Tao 2006] T. Tao, "A quantitative ergodic theory proof of Szemerédi's theorem", *Electron. J. Combin.* **13**:1 (2006), 1–49. MR 2007i:37016 Zbl 1127.11011

[Tao 2007] T. Tao, "Structure and randomness in combinatorics", pp. 3–15 in *FOCS '07: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society, Washington, DC, USA, 2007.

[Tao and Vu 2006] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, 2006. MR 2008a:11002 Zbl 1127.11002

[Varnavides 1959] P. Varnavides, "On certain sets of positive density", *J. London Math. Soc.* **34** (1959), 358–360. MR 21 #5595 Zbl 0088.25702

Received 30 Oct 2008. Revised 29 Oct 2009. Accepted 10 Dec 2009.

TERENCE TAO: tao@math.ucla.edu
University of California, Mathematics Department, Los Angeles, CA 90095-1555, United States
http://www.math.ucla.edu/~tao/

TAMAR ZIEGLER: tamarzr@tx.technion.ac.il
Department of Mathematics, Technion, 32000 Haifa, Israel

ANALYSIS AND PDE
Vol. 3, No. 1, 2010

BILINEAR FORMS ON THE DIRICHLET SPACE

NICOLA ARCOZZI, RICHARD ROCHBERG, ERIC SAWYER AND BRETT D. WICK

We show that the bilinear form $B_b(f, g) = \langle fg, b\rangle$ is bounded on the Dirichlet space of holomorphic functions on the unit disk if and only if $|b'|^2\,dx\,dy$ is a Carleson measure for the Dirichlet space. This is completely analogous to the results for boundedness of Hankel forms on the Hardy and Bergman spaces, but the proof is quite different, relying heavily on potential-theoretic constructions.

1. Introduction

A Hankel form is a bilinear form B on a space of holomorphic functions with the characteristic property that for any f, g, $B(f, g)$ is a linear function of fg. These forms have been studied extensively on Hardy spaces and on Bergman type spaces; some references are mentioned below. Here we consider boundedness of Hankel forms on the Dirichlet space. In contrast to Hardy and Bergman spaces, the Dirichlet space is a potential space and hence, not surprisingly, capacity estimates play a central role in the analysis. Thus, although our main results are strongly analogous to earlier work, the techniques are quite different.

Overview. Let $\mathcal{D}$ be the classical Dirichlet space, the Hilbert space of holomorphic functions on the disk with inner product

$$\langle f, g\rangle_{\mathcal{D}} = f(0)\overline{g(0)} + \int_{\mathbb{D}} f'(z)\overline{g'(z)}\,dA$$

and normed by $\|f\|_{\mathcal{D}}^2 = \langle f, f\rangle_{\mathcal{D}}$. Given a holomorphic *symbol function* b we define the associated Hankel type bilinear form, initially for $f, g \in \mathcal{P}(\mathbb{D})$, the space of polynomials, by

$$T_b(f, g) := \langle fg, b\rangle_{\mathcal{D}}.$$

The norm of T_b is

$$\|T_b\|_{\mathcal{D}\times\mathcal{D}} := \sup\big\{|T_b(f, g)| : \|f\|_{\mathcal{D}} = \|g\|_{\mathcal{D}} = 1\big\}.$$

We say a positive measure μ on the disk is a *Carleson measure for* $\mathcal{D}$ if

$$\|\mu\|_{CM(\mathcal{D})} := \sup\left\{\int_{\mathbb{D}} |f|^2 d\mu : \|f\|_{\mathcal{D}} = 1\right\} < \infty,$$

MSC2000: 30C85, 47B35, 31C25.
Keywords: Dirichlet space, Hankel form.
Arcozzi's work is partially supported by the COFIN project Analisi Armonica, funded by the Italian Minister for Research, Rochberg's work is supported by the National Science Foundation under Grant 0070642. Sawyer's work supported by the National Science and Engineering Council of Canada. Wick's work is supported by the National Science Foundation under Grant 0752703.

and that a function b is in the space $\mathscr{X}$ if the measure $d\mu_b := |b'(z)|^2 dA$ is a Carleson measure. We norm $\mathscr{X}$ by

$$\|b\|_{\mathscr{X}} := |b(0)| + \left\| |b'(z)|^2 dA \right\|_{CM(\mathscr{D})}^{1/2}$$

and denote by $\mathscr{X}_0$ the norm closure in $\mathscr{X}$ of the space of polynomials.

Our main result is this:

Theorem 1.1. (1) *T_b is bounded if and only if $b \in \mathscr{X}$. In that case*

$$\|T_b\|_{\mathscr{D}\times\mathscr{D}} \approx \|b\|_{\mathscr{X}}.$$

(2) *T_b is compact if and only if $b \in \mathscr{X}_0$.*

This result is part of an intriguing pattern of results involving boundedness of Hankel forms on Hardy spaces in one and several variables and boundedness of Schrödinger operators on the Sobolev space. We recall some of those results in the next subsection.

Boundedness criteria for bilinear forms can be recast as weak factorization of function spaces. We present details and related earlier results later in this introduction. In particular we will see that the first statement in Theorem 1.1 is equivalent to a weak factorization of the predual of $\mathscr{X}$; in notation we introduce below

$$(\mathscr{D} \odot \mathscr{D})^* = \mathscr{X}. \tag{1-1}$$

At the end of the introduction (page 25) we describe the relation between Theorem 1.1 and classical results about Hankel matrices.

The proof of Theorem 1.1 comes in Sections 2 and 3. It is easy to see that $\|T_b\|_{\mathscr{D}\times\mathscr{D}} \le C\|b\|_{\mathscr{X}}$. To obtain the other inequality we must use the boundedness of T_b to show $|b'|^2 dA$ is a Carleson measure. Analysis of the capacity-theoretic characterization of Carleson measures due to Stegenga allows us to focus attention on a certain set V in $\mathbb{D}$ and the relative sizes of $\int_V |b'|^2$ and the capacity of the set $\overline{V} \cap \partial\overline{\mathbb{D}}$. To compare these quantities we construct $V_{\exp}$, an expanded version of the set V which satisfies two conflicting conditions. First, $V_{\exp}$ is not much larger than V, either when measured by $\int_{V_{\exp}} |b'|^2$ or by the capacity of the $\overline{V}_{\exp} \cap \partial\overline{\mathbb{D}}$. Second, $\mathbb{D} \setminus V_{\exp}$ is well separated from V in a way that allows the interaction of quantities supported on the two sets to be controlled. Once this is done we can construct a function $\Phi_V \in \mathscr{D}$ which is approximately one on V and which has Φ_V' approximately supported on $\mathbb{D} \setminus V_{\exp}$. Using Φ_V we build functions f and g with the property that

$$|T_b(f, g)| = \int_V |b'|^2 + \text{error}.$$

The technical estimates on Φ_V allow us to show that the error term is small and the boundedness of T_b then gives the required control of $\int_V |b'|^2$.

Once the first part of the theorem is established, the second follows rather directly.

Other bilinear forms. The Hardy space of the unit disk, $H^2(\mathbb{D})$, can be defined as the space of holomorphic functions on the disk with inner product

$$\langle f, g\rangle_{H^2(\mathbb{D})} = f(0)\overline{g(0)} + \int_{\mathbb{D}} f'(z)\overline{g'(z)}\,(1-|z|^2)\,dA$$

and normed by $\|f\|^2_{H^2(\mathbb{D})} = \langle f, f\rangle_{H^2(\mathbb{D})}$. Given a holomorphic *symbol function* b the Hankel form with symbol b is the bilinear form

$$T_b^{H^2(\mathbb{D})}(f, g) := \langle fg, b\rangle_{H^2(\mathbb{D})}. \tag{1-2}$$

The boundedness criteria for such forms was given by Nehari [1957]. He used the fact that functions in the Hardy space H^1 can be written as the product of functions in H^2 and showed $T_b^{H^2(\mathbb{D})}$ will be bounded if and only if b is in the dual space of H^1. Using Ch. Fefferman's identification of the dual of H^1 we can reformulate this in the language of Carleson measures. We say a positive measure μ on the disk is a *Carleson measure for* $H^2(\mathbb{D})$ if

$$\|\mu\|_{CM(H^2(\mathbb{D}))} := \sup\left\{\int_{\mathbb{D}} |f|^2 d\mu \, : \, \|f\|_{H^2(\mathbb{D})} = 1\right\} < \infty.$$

The form $T_b^{H^2(\mathbb{D})}$ is bounded if and only if b is in the function space BMO or, equivalently, if and only if

$$|b'(z)|^2(1 - |z|^2)\, dA \in CM(H^2(\mathbb{D})).$$

Later, in [Coifman et al. 1976], Nehari's theorem was viewed as a result about Calderón–Zygmund singular integrals on spaces of homogenous type and an analogous result was proved for $H^2(\partial\mathbb{B}^n)$, the Hardy space of the sphere in complex n-space. In that context the Hankel form is defined similarly

$$T_b^{H^2(\partial\mathbb{B}^n)}(f, g) := \langle fg, b\rangle_{H^2(\partial\mathbb{B}^n)}.$$

That form is bounded if and only if b is in $BMO(\partial\mathbb{B}^n)$ or, equivalently, if and only if, with ∇ denoting the invariant gradient on the ball,

$$|\nabla b(z)|^2 dV \in CM(H^2(\partial\mathbb{B}^n)).$$

The approach in [Coifman et al. 1976] is not well suited for analysis on the Hardy space of the polydisk, $H^2(\mathbb{D}^n)$. However Ferguson, Lacey, and Terwilleger were able to extend methods of multivariable harmonic analysis and obtain a result for $H^2(\mathbb{D}^n)$ [Ferguson and Lacey 2002; Lacey and Terwilleger 2009]. They showed that a Hankel form on $H^2(\mathbb{D}^n)$, again defined as a form whose value only depends on the product of its arguments, is bounded if and only if the symbol function b lies in $BMO(\mathbb{D}^n)$ or, equivalently, if and only if derivatives of b can be used to generate a Carleson measure for $H^2(\mathbb{D}^n)$.

Maz'ya and Verbitsky [2002] presented a boundedness criterion for a bilinear form associated to the Schrödinger operator. Although their viewpoint and proof techniques were quite different from those used for Hankel forms, their result is formally very similar. We change their formulation slightly to make the analogy more visible, our b is related to their V by $b = -\Delta^{-1}V$. Let $\mathring{L}^1_2(\mathbb{R}^n)$ be the energy space (homogenous Sobolev space) obtained by completing $C_0^\infty(\mathbb{R}^n)$ with respect to the quasinorm induced by the Dirichlet inner product

$$\langle f, g\rangle_{\mathrm{Dir}} = \int_{\mathbb{R}^n} \nabla f \cdot \overline{\nabla g}\, dx.$$

Given b, a bilinear Schrödinger form on $\mathring{L}^1_2(\mathbb{R}^n) \times \mathring{L}^1_2(\mathbb{R}^n)$ is defined by

$$S_b(f, g) = \langle fg, b\rangle_{\mathrm{Dir}}.$$

Although the relevant class of measures in this context was first studied by Maz'ya we will use a notation which emphasizes the analogy with the previous situations. We will write $\mu \in CM(\mathring{L}_2^1(\mathbb{R}^n))$ if

$$\|\mu\|_{CM(\mathring{L}_2^1(\mathbb{R}^n))} := \sup\left\{\int_{\mathbb{R}^n} |f|^2 d\mu : \|f\|_{\mathring{L}_2^1(\mathbb{R}^n)} = 1\right\} < \infty.$$

Corollary 2 of [Maz'ya and Verbitsky 2002] is that S_b is bounded if and only if

$$\left|(-\Delta)^{1/2} b\right|^2 dx \in CM(\mathring{L}_2^1(\mathbb{R}^n)).$$

It would be very satisfying to know an underlying reason for the similarity of these various results to each other and to Theorem 1.1.

Reformulation in terms of weak factorization. In his proof Nehari used the fact that any function $f \in H^1(\mathbb{D})$ could be factored as $f = gh$ with $g, h \in H^2(\mathbb{D})$, $\|f\|_{H^1(\mathbb{D})} = \|g\|_{H^2(\mathbb{D})}\|h\|_{H^2(\mathbb{D})}$. In [Coifman et al. 1976] the authors develop a weak substitute for this. For two Banach spaces of functions, $\mathcal{A}$ and $\mathcal{B}$, defined on the same domain, define the weakly factored space $\mathcal{A} \odot \mathcal{B}$ to be the completion of finite sums $f = \sum a_i b_i$; $\{a_i\} \subset \mathcal{A}$, $\{b_i\} \subset \mathcal{B}$ using the norm

$$\|f\|_{\mathcal{A}\odot\mathcal{B}} = \inf\left\{\sum \|a_i\|_{\mathcal{A}}\|b_i\|_{\mathcal{B}} : f = \sum a_i b_i\right\}.$$

It is shown in [Coifman et al. 1976] that $H^2(\partial\mathbb{B}^n)\odot H^2(\partial\mathbb{B}^n) = H^1(\partial\mathbb{B}^n)$ and consequentially

$$\left(H^2(\partial\mathbb{B}^n)\odot H^2(\partial\mathbb{B}^n)\right)^* = BMO(\partial\mathbb{B}^n). \tag{1-3}$$

(In this context, by $=$ we mean equality of the function spaces and equivalence of the norms.) Based on the analogy between (1-1) and (1-3) we think of $\mathcal{D} \odot \mathcal{D}$ as a type of H^1 space and of $\mathcal{X}$ as a type of BMO space. That viewpoint is developed further in [Arcozzi et al. 2008].

The precise formulation of (1-1) is the following corollary.

Corollary 1.2. *For $b \in \mathcal{X}$ set $\Lambda_b h = T_b(h, 1)$, then $\Lambda_b \in (\mathcal{D}\odot\mathcal{D})^*$. Conversely, if $\Lambda \in (\mathcal{D}\odot\mathcal{D})^*$ there is a unique $b \in \mathcal{X}$ so that for all $h \in \mathcal{P}(\mathbb{D})$ we have $\Lambda h = T_b(h, 1) = \Lambda_b h$. In both cases $\|\Lambda_b\|_{(\mathcal{D}\odot\mathcal{D})^*} \approx \|b\|_{\mathcal{X}}$.*

Proof. *If $b \in \mathcal{X}$ and $f \in \mathcal{D} \odot \mathcal{D}$, say $f = \sum g_i h_i$ with $\sum \|g_i\|_{\mathcal{D}}\|h_i\|_{\mathcal{D}} \le \|f\|_{\mathcal{D}\odot\mathcal{D}} + \varepsilon$, then*

$$|\Lambda_b f| = \left|\sum_{i=1}^{\infty}\langle g_i h_i, b\rangle_{\mathcal{D}}\right| = \left|\sum_{i=1}^{\infty} T_b(g_i, h_i)\right| \le \|T_b\| \sum_{i=1}^{\infty} \|g_i\|_{\mathcal{D}}\|h_i\|_{\mathcal{D}} \le \|T_b\|(\|f\|_{\mathcal{D}\odot\mathcal{D}} + \varepsilon).$$

It follows that $\Lambda_b f = \langle f, b\rangle_{\mathcal{D}}$ defines a continuous linear functional on $\mathcal{D} \odot \mathcal{D}$ with $\|\Lambda_b\| \le \|T_b\|$.

Conversely, if $\Lambda \in (\mathcal{D} \odot \mathcal{D})^*$with norm $\|\Lambda\|$, then for all $f \in \mathcal{D}$

$$|\Lambda f| = |\Lambda(f \cdot 1)| \le \|\Lambda\|\|f\|_{\mathcal{D}}\|1\|_{\mathcal{D}} = \|\Lambda\|\|f\|_{\mathcal{D}}.$$

Hence there is a unique $b \in \mathcal{D}$ such that $\Lambda f = \Lambda_b f$ for $f \in \mathcal{D}$. Finally, if $f = gh$ with $g, h \in \mathcal{D}$ we have

$$|T_b(g, h)| = |\langle gh, b\rangle_{\mathcal{D}}| = |\Lambda_b f| = |\Lambda f| \le \|\Lambda\|\|f\|_{\mathcal{D}\odot\mathcal{D}} \le \|\Lambda\|\|g\|_{\mathcal{D}}\|h\|_{\mathcal{D}},$$

which shows that T_b extends to a continuous bilinear form on $\mathcal{D} \odot \mathcal{D}$ with $\|T_b\| \le \|\Lambda\|$. By Theorem 1.1 we conclude $b \in \mathcal{X}$ and collecting the estimates that $\|\Lambda\| = \|\Lambda_b\|_{(\mathcal{D}\odot\mathcal{D})^*} \approx \|T_b\| \approx \|b\|_{\mathcal{X}}$. □

There is a bilinear form related to T_b which was studied earlier and which is also related to a weak factorization statement. Define K_b by $K_b(f, g) = \int_{\mathbb{D}} f' g \bar{b}' \, dV$. It was shown independently in [Coifman and Murai 1988; Tolokonnikov 1991; Rochberg and Wu 1993] that K_b is bounded if and only if $b \in \mathscr{X}$. (In fact the work reported in the last of these papers began as an attempt to prove Theorem 1.1.) Define the space $\partial^{-1}(\partial\mathscr{D} \odot \mathscr{D})$ to be the completion of the space of functions f which have $f' = \sum_{i=1}^N g_i' h_i$ (and thus $f = \partial^{-1} \sum (\partial g_i) h_i)$) using the norm

$$\|f\|_{\partial^{-1}(\partial\mathscr{D}\odot\mathscr{D})} = \inf \left\{ \sum \|g_i\|_{\mathscr{D}} \|h_i\|_{\mathscr{D}} : f' = \sum_{i=1}^N g_i' h_i \right\}.$$

Theorem 1.3 [Coifman and Murai 1988; Tolokonnikov 1991; Rochberg and Wu 1993]. *K_b is bounded if and only if $b \in \mathscr{X}$, equivalently,*

$$(\partial^{-1}(\partial\mathscr{D} \odot \mathscr{D}))^* = \mathscr{X}.$$

In fact this follows from Theorem 1.1. In proving that if $b \in \mathscr{X}$ then T_b is bounded we actually show directly that K_b is bounded and then note that

$$T_b(f, g) = K_b(f, g) + K_b(g, f) + (fg\bar{b})(0). \tag{1-4}$$

In the other direction, if K_b is bounded then the same relation shows T_b is bounded and we can then appeal to Theorem 1.1.

The representation (1-4) gives an insight into why Theorem 1.1 seems to be more difficult than those earlier results. The proofs of Theorem 1.3 in the three papers cited give, explicitly or implicitly, estimates from below for $|K_b(f, g)|$. In proving Theorem 1.1 we need to estimate $|T_b(f, g)|$ from below. Although the formula (1-4) invites using that representation as a starting point for analysis of T_b. It was unclear to us how to analyze the potential cancellation between terms on the right hand side of (1-4) and that potential cancellation appears to be a basic issue here.

Combining the previous two results we have, with the obvious notation:

Corollary 1.4.
$$\partial(\mathscr{D} \odot \mathscr{D}) = \partial\mathscr{D} \odot \mathscr{D}.$$

In contrast

$$\partial(\mathscr{D} \odot \mathscr{D}) \neq \partial^{1/2}\mathscr{D} \odot \partial^{1/2}\mathscr{D}.$$

To see this note that $\partial^{1/2}\mathscr{D} \odot \partial^{1/2}\mathscr{D} = H^2(\mathbb{D}) \odot H^2(\mathbb{D}) = H^1(\mathbb{D})$ and that $f(z) = (\log(1-z))^{3/2}$ satisfies $f' \in \partial(\mathscr{D} \odot \mathscr{D})$, $f' \notin H^1$.

Reformulation in terms of matrices. If T_b is given by (1-2) with $b(z) = \sum b_n z^n$ then the matrix representation of T_b with respect to the monomial basis is $(\bar{b}_{i+j})$. Nehari's theorem gives a boundedness condition for such Hankel matrices; matrices $(a_{i,j})$ for which $a_{i,j}$ is a function of $i + j$. There are analogous results for Hankel forms on Bergman spaces. Those forms have matrices

$$\big((i+1)^{\alpha}(j+1)^{\beta}(i+j+1)^{\gamma}\bar{b}(i+j)\big) \tag{1-5}$$

with $\alpha, \beta > 0$ and are bounded if and only if $b(z)$ is in the Bloch space. The criteria for (1-5) to belong to the Schatten–von Neumann classes is known if $\min\{\alpha, \beta\} > -1/2$ and it is known that those results do not extend to $\min\{\alpha, \beta\} \leq -1/2$. For all of this see [Peller 2003, Chapter 6.8].

The matrix representations of the forms T_b and K_b with respect to the basis of normalized monomials of $\mathcal{D}$ are of the form (1-5) with (α, β) equal to $\left(-\frac{1}{2}, -\frac{1}{2}\right)$ in the first case and $\left(-\frac{1}{2}, \frac{1}{2}\right)$ in the second.

2. Preliminary steps in the proof of Theorem 1.1

Proof of (2) given (1). Suppose T_b is compact. For any holomorphic function $k(z)$ on $\mathbb{D}$ and r, $0 < r < 1$, set $S_r k(z) = k(rz)$. A computation with monomials verifies that

$$T_{S_r b}(f, g) = T_b(S_r f, S_r g).$$

As $r \to 1$, S_r converges strongly to I. Using this and that T_b is compact we obtain $\lim \|T_{S_r b} - T_b\| = 0$. Hence, by the first part of the theorem $\lim \|S_r b - b\|_{\mathcal{X}} = 0$. The Taylor coefficients of $S_r b$ decay geometrically, hence $S_r b \in \mathcal{X}_0$ and thus $b \in \mathcal{X}_0$.

In the other direction note that if b is a polynomial then T_b is finite rank and hence compact. If $\{b_n\} \subset \mathcal{P}(\mathbb{D})$ is a sequence of polynomials which converge in norm to $b \in \mathcal{X}_0$ then, by the first part of the theorem, T_b is the norm limit of the T_{b_n} and hence is also compact. □

Proof of the easy direction of (1). Suppose that μ_b is a Carleson measure for $\mathcal{D}$. For $f, g \in \mathcal{P}(\mathbb{D})$ we have

$$\begin{aligned}
|T_b(f, g)| &= \left| f(0)g(0)\overline{b(0)} + \int_{\mathbb{D}} \big(f'(z)g(z) + f(z)g'(z)\big)\overline{b'(z)}\, dA \right| \\
&\le \big|f(0)g(0)b(0)\big| + \int_{\mathbb{D}} \big|f'(z)g(z)b'(z)\big|\, dA + \int_{\mathbb{D}} \big|f(z)g'(z)b'(z)\big|\, dA \\
&\le \big|(fgb)(0)\big| + \|f\|_{\mathcal{D}} \left(\int_{\mathbb{D}} |g|^2\, d\mu_b \right)^{1/2} + \|g\|_{\mathcal{D}} \left(\int_{\mathbb{D}} |f|^2\, d\mu_b \right)^{1/2} \\
&\le C\big(|b(0)| + \|\mu_b\|_{CM(\mathcal{D})}\big) \|f\|_{\mathcal{D}} \|g\|_{\mathcal{D}} \\
&= C\|b\|_{\mathcal{X}} \|f\|_{\mathcal{D}} \|g\|_{\mathcal{D}}.
\end{aligned}$$

Thus T_b has a bounded extension to $\mathcal{D} \times \mathcal{D}$ with $\|T_b\| \le C\|b\|_{\mathcal{X}}$. □

We note for later that if T_b extends to a bounded bilinear form on $\mathcal{D}$ then $b \in \mathcal{D}$, equivalently, $d\mu_b$ is a finite measure. To see this note that for all $f \in \mathcal{P}(\mathbb{D})$, $|\langle f, b\rangle_{\mathcal{D}}| = |T_b(f, 1)| \le \|T_b\| \|f\|_{\mathcal{D}} \|1\|_{\mathcal{D}}$. Thus $b \in \mathcal{D}$ and

$$\|b\|_{\mathcal{D}} \le C\|T_b\|. \tag{2-1}$$

Disk capacity and disk blow-ups. To complete the proof of Theorem 1.1 we must show that if T_b is bounded then $\mu_{b:} = |b'|^2 dA$ is a $\mathcal{D}$-Carleson measure. We will do this by showing that μ_b satisfies a capacitary condition introduced by Stegenga [1980].

For an interval I in the circle we let I_m be its midpoint and $z(I) = (1 - |I|/2\pi) I_m$ be the associated index point in the disk. In the other direction let $I(z)$ be the interval such that $z(I(z)) = z$. Let $T(I)$ be the tent over I, the convex hull of I and $z(I)$ and let $T(z) = T(z(I)) := T(I)$. More generally, for any open subset H of the circle $\mathbb{T}$, we define $T(H)$, the *tent region* of H in the disk $\mathbb{D}$, by

$$T(H) = \bigcup_{I \subset H} T(I).$$

For G in the circle $\mathbb{T}$ define the capacity of G by

$$\operatorname{Cap}_{\mathbb{D}} G = \inf\left\{\|\psi\|_{\mathcal{D}}^2 : \psi(0) = 0, \operatorname{Re}\psi(z) \geq 1 \text{ for } z \in G\right\}. \tag{2-2}$$

Stegenga [1980] has shown that μ is a $\mathcal{D}$-Carleson measure exactly if for any finite collection of disjoint arcs $\{I_j\}_{j=1}^N$ in the circle $\mathbb{T}$ we have

$$\mu\Big(\bigcup_{j=1}^N T(I_j)\Big) \leq C \operatorname{Cap}_{\mathbb{D}}\Big(\bigcup_{j=1}^N I_j\Big). \tag{2-3}$$

We will need to understand how the capacity of a set changes if we expand it in certain ways. For I an open arc and $0 < \rho \leq 1$, let I^ρ be the arc concentric with I having length $|I|^\rho$.

Definition 2.1 (disk blowup). For G open in $\mathbb{T}$ we call

$$G_{\mathbb{D}}^\rho = \bigcup_{I\subset G} T(I^\rho)$$

the *disk blowup* (of order ρ) of G.

The important feature of the disk blowup is that it achieves a good geometric separation between $\mathbb{D}\setminus G_{\mathbb{D}}^\rho$ and $G_{\mathbb{D}}^1 = T(G)$. This plays a crucial role in using Schur's test to estimate an integral later, as well as in estimating an error term near the end of the paper.

Lemma 2.2. *Let G be an open subset of the circle $\mathbb{T}$. If $w \in G_{\mathbb{D}}^1 = T(G)$ and $z \notin G_{\mathbb{D}}^\rho$ then $|z-w| \geq (1-|w|^2)^\rho$.*

Proof. The inequality follows from the definition of $G_{\mathbb{D}}^\rho$ and the inclusion

$$T(I^\rho) \subset \left\{z : |z - z(I)| < 2(1-|z(I)|)^{2\rho}\right\}. \qquad \square$$

It would be useful to us if we knew there were constants C_ρ, for each $0 < \rho < 1$, such that

$$\operatorname{Cap}_{\mathbb{D}} \bigcup_{I\subset G} I^\rho \leq C_\rho \operatorname{Cap}_{\mathbb{D}} G. \tag{2-4}$$

and

$$\lim_{\rho\to 1^-} C_\rho = 1. \tag{2-5}$$

Bishop [1994] proved (2-4) but did not obtain (2-5). In a short while we will obtain Lemma 2.8, an analog of (2-4) and (2-5) in a tree model, and that will play an important role in the proof. After we show that tree and disk are comparable (Corollary 2.12) we will also have the tree result (2-4), which will likewise be used in the proof. It remains an open question whether the disk result (2-5) holds.

Tree capacity and tree blow-ups. In our study of capacities and approximate extremals it will sometimes be convenient to transfer our arguments to and from the Bergman tree $\mathcal{T}$ and to work with the associated tree capacities. We now recall the notation associated to $\mathcal{T}$. Further properties of $\mathcal{T}$ are in the Appendix and a more extensive investigation with other applications is in [Arcozzi et al. 2007].

Let $\mathcal{T}$ be the standard Bergman tree in the unit disk $\mathbb{D}$. That is $\mathcal{T} = \{x\}$ is the index set for the subsets $\{B_x\}$ of $\mathbb{D}$ obtained by decomposing $\mathbb{D}$, first with the circles $C_k = \{z : |z| = 1-2^{-k}\}$, $k = 1, 2, \ldots$ and then for each k making 2^k radial cuts in the ring bounded by C_k and C_{k+1}. We refer to the $\{B_x\}$ as boxes and we emphasize the standard bijection between the boxes and the intervals on the circle $\{I(B_x)\}$ obtained by radial projection of the boxes. This also induces a bijection with the point set $\{z(I(B_x))\}$ in the disk; furthermore, $z(I(B_x)) \in B_x$. At times we will use the label x to denote the point $z(I(B_x))$.

$\mathcal{T}$ is a rooted dyadic tree with root $\{0\}$, which we denote by o. For a vertex x of $\mathcal{T}$ we denote its immediate predecessor by x^{-1} and its two immediate successors by x_+ and x_-. We let $d(x)$ equal the number of nodes on the geodesic $[o, x]$. The successor set of x is $S(x) = \{y \in \mathcal{T} : y \geq x\}$.

We say that $S \subset \mathcal{T}$ is a *stopping time* if no pair of distinct points in S are comparable in $\mathcal{T}$. Given stopping times $E, F \subset \mathcal{T}$ we say that $F \succ E$ if for every $x \in F$ there is $y \in E$ above x, that is, with $x > y$. For stopping times $F \succ E$ denote by $\mathcal{G}(E, F)$ the union of all those geodesics connecting a point of $x \in F$ to the point $y \in E$ above it.

The bijections between $\{B_x\}$, $\{I(B_x)\}$, and $\{z(I(B_x))\}$ induce bijections between other sets. We will be particularly interested in three types of sets:

- *stopping times* W in the tree $\mathcal{T}$,
- $\mathcal{T}$*-open subsets* G of the circle $\mathbb{T}$,
- $\mathcal{T}$*-tent regions* Γ of the disk $\mathbb{D}$.

The bijections are given as follows. For W a *stopping time* in $\mathcal{T}$, its associated $\mathcal{T}$*-open set* in $\mathbb{T}$ is the $\mathcal{T}$-shadow $S_{\mathcal{T}}(W) = \bigcup\{I(x) : x \in W\}$ of W on the circle (this also *defines* the collection of $\mathcal{T}$-open sets). The associated $\mathcal{T}$*-tent region* in $\mathbb{D}$ is $T_{\mathcal{T}}(W) = \bigcup\big\{T(I(\kappa)) : \kappa \in W\big\}$ (this also *defines* the collection of $\mathcal{T}$-tent regions).

At times we will identify a stopping time $W = W_{\mathcal{T}}$ in a tree $\mathcal{T}$ with its associated $\mathcal{T}$-shadow on the circle and its $\mathcal{T}$-tent region in the disk and will use W or $W_{\mathcal{T}}$ to denote any of them. When we do this the exact interpretation will be clear from the context.

Note that for any open subset E of the circle $\mathbb{T}$, there is a unique $\mathcal{T}$-open set $G \subset E$ such that $E \setminus G$ is at most countable. We often informally identify the open sets E and G.

For a functions k, K defined on $\mathcal{T}$ set

$$Ik(x) = \sum_{y \in [o,x]} k(y), \qquad \Delta K(x) = K(x) - K(x^-)$$

with the convention that $K(o^-) = 0$.

For $\Omega \subseteq \mathcal{T}$ a point $x \in \mathcal{T}$ is in the interior of Ω if $x, x^{-1}, x_+, x_- \in \Omega$. A function H is *harmonic* in Ω if

$$H(x) = \tfrac{1}{3}\big[H(x^{-1}) + H(x_+) + H(x_-)\big] \tag{2-6}$$

for every point x which is interior in Ω. If $H = Ih$ is harmonic then for all x in the interior of Ω

$$h(x) = h(x_+) + h(x_-). \tag{2-7}$$

Let $\mathrm{Cap}_{\mathcal{T}}$ be the tree capacity associated with $\mathcal{T}$:

$$\mathrm{Cap}_{\mathcal{T}}(E) = \inf\big\{\|f\|^2_{\ell^2(\mathcal{T})} : If \geq 1 \text{ on } E\big\}. \tag{2-8}$$

More generally, if $E, F \subset \mathcal{T}$ are disjoint stopping times with $F \succ E$, the capacity of the pair (E, F), commonly known as a condenser, is given by

$$\operatorname{Cap}_{\mathcal{T}}(E, F) = \inf\Big\{\|f\|^2_{\ell^2(\mathcal{T})} : If \geq 1 \text{ on } F,\ \operatorname{supp}(f) \subset \bigcup_{e \in E} S(e)\Big\}. \tag{2-9}$$

Let $\mathcal{T}_\theta$ be the rotation of the tree $\mathcal{T}$ by the angle θ, and let $\operatorname{Cap}_{\mathcal{T}_\theta}$ be the tree capacity associated with $\mathcal{T}_\theta$ as in (2-8), and extend the definition to open subsets G of the circle $\mathbb{T}$ by

$$\operatorname{Cap}_{\mathcal{T}_\theta}(G) = \inf\bigg\{\sum_{\kappa \in \mathcal{T}_\theta} f(\kappa)^2 : If(\beta) \geq 1 \text{ for } \beta \in \mathcal{T}_\theta,\ I(\beta) \subset G\bigg\}.$$

This is consistent with the definition of tree capacity of a stopping time W in $\mathcal{T}_\theta$; that is, if

$$G = \bigcup\big\{I(\kappa) : \kappa \in W\big\},$$

we have

$$\operatorname{Cap}_{\mathcal{T}_\theta}(W) = \operatorname{Cap}_{\mathcal{T}_\theta}(\{o\}, W) = \operatorname{Cap}_{\mathcal{T}_\theta}(G).$$

When the angle θ is not important, we will simply write $\mathcal{T}$ with the understanding that all results have analogues with $\mathcal{T}_\theta$ in place of $\mathcal{T}$.

We will use functions on the disk which are approximate extremals for measuring capacity, that is functions for which the equality in (2-2) is approximately attained. A tool in doing that is an analysis of the model problems on a tree. The following result about tree capacities and extremals is proved in the Appendix.

Proposition 2.3. *Suppose $E, F \subset \mathcal{T}$ are disjoint stopping times with $F \succ E$.*

(1) *There is an extremal function $H = Ih$ such that* $\operatorname{Cap}(E, F) = \|h\|^2_{\ell^2}$.

(2) *The function H is harmonic on $\mathcal{T} \setminus (E \cup F)$.*

(3) *If S is a stopping time in $\mathcal{T}$, then $\sum_{\kappa \in S} |h(\kappa)| \leq 2Cap(E, F)$.*

(4) *The function h is positive on $\mathcal{G}(E, F)$ and zero elsewhere.*

Definition 2.4 (stopping time blowup). Given $0 \leq \rho \leq 1$ and a stopping time W in a tree $\mathcal{T}$, define the *stopping time blowup* $W^\rho_{\mathcal{T}}$ of W in $\mathcal{T}$ as the set of minimal tree elements in $\{R^\rho\kappa : \kappa \in \mathcal{T}_\theta\}$, where $R^\rho\kappa$ denotes the unique element in the tree $\mathcal{T}$ satisfying

$$o \leq R^\rho\kappa \leq \kappa, \qquad \rho d(\kappa) \leq d(R^\rho\kappa) < \rho d(\kappa) + 1. \tag{2-10}$$

Clearly $W^\rho_{\mathcal{T}}$ is a stopping time in $\mathcal{T}$. Note that $R^1\kappa = \kappa$. The element $R^\rho\kappa$ can be thought of as the ρ-th root of κ, since $|R^\rho\kappa| = 2^{-d(R^\rho\kappa)} \approx 2^{-\rho d(\kappa)} = |\kappa|^\rho$.

If W is a stopping time for $\mathcal{T}$ and $W^\rho_{\mathcal{T}}$ is the stopping time blowup of W, then there is a good estimate for the tree capacity of $W^\rho_{\mathcal{T}}$ given in Lemma 2.8 below: $\operatorname{Cap}_{\mathcal{T}}(\{o\}, W^\rho_{\mathcal{T}}) \leq \rho^{-2} \operatorname{Cap}_{\mathcal{T}}(\{o\}, W)$. Unfortunately there is not a good condenser estimate of the form $\operatorname{Cap}_{\mathcal{T}}(W^\rho_{\mathcal{T}}, W) \leq C_\rho \operatorname{Cap}_{\mathcal{T}}(\{o\}, W)$; the left side can be infinite when the right side is finite. We now introduce another type of blowup, a tree analog of the disk blowup, for which we do have an effective condenser estimate. We do this using a capacitary extremal function and a comparison principle. Let W be a stopping time in $\mathcal{T}$. By Proposition 2.3, there is a unique extremal function $H = Ih$ such that

$$Ih(x) = H(x) = 1 \ \text{ for } x \in W \qquad \text{and} \qquad \operatorname{Cap}_{\mathcal{T}} W = \|h\|_{\ell^2}^2. \tag{2-11}$$

Definition 2.5 (capacitary blowup). Given a stopping time W in $\mathcal{T}$, the corresponding extremal H satisfying (2-11), and $0 < \rho < 1$, define the *capacitary blowup* $\widehat{W_{\mathcal{T}}^{\rho}}$ of W by

$$\widehat{W_{\mathcal{T}}^{\rho}} = \big\{ t \in \mathcal{G}(\{o\}, W) : H(t) \geq \rho \quad \text{and} \quad H(x) \leq \rho \text{ for } x < t \big\}.$$

Clearly $\widehat{W_{\mathcal{T}}^{\rho}}$ is a stopping time in $\mathcal{T}$.

Lemma 2.6. $$\operatorname{Cap}_{\mathcal{T}} \widehat{W_{\mathcal{T}}^{\rho}} \leq \rho^{-2} \operatorname{Cap}_{\mathcal{T}} W.$$

Proof. Let H be the extremal for W in (2-11) and set $h = \Delta H$, $h^{\rho} = h/\rho$ and $H^{\rho} = H/\rho$. Then H^{ρ} is a candidate for the infimum in the definition of capacity of $\widehat{W_{\mathcal{T}}^{\rho}}$, and hence, by the comparison principle,

$$\operatorname{Cap}_{\mathcal{T}} \widehat{W_{\mathcal{T}}^{\rho}} \leq \|h^{\rho}\|_{\ell^2}^2 = \left(\frac{1}{\rho}\right)^2 \|h\|_{\ell^2}^2 = \rho^{-2} \operatorname{Cap}_{\mathcal{T}} W. \qquad \square$$

The next lemma is used in the proof of our main estimate, (3-1). It requires an upper bound on $\operatorname{Cap}_{\mathbb{D}}(G)$. However, (3-1) is straightforward if $\operatorname{Cap}_{\mathbb{D}}(G)$ bounded away from zero so that restriction is not a problem. In fact, moving forward we will assume, at times implicitly, that $\operatorname{Cap}_{\mathbb{D}}(G)$ is not large.

Lemma 2.7. $\operatorname{Cap}_{\mathcal{T}}(W, \widehat{W_{\mathcal{T}}^{\rho}}) \leq \dfrac{4}{(1-\rho)^2} \operatorname{Cap}_{\mathcal{T}} W$ *provided* $\operatorname{Cap}_{\mathcal{T}} W \leq (1-\rho)^2/4$.

Proof. Let H be the extremal for W in (2-11). For $t \in \widehat{W_{\mathcal{T}}^{\rho}}$ we have by our assumption,

$$h(t) \leq \|h\|_{\ell^2} \leq \sqrt{\operatorname{Cap}_{\mathcal{T}} W} \leq \tfrac{1}{2}(1-\rho),$$

and so

$$H(t) = H(t^{-}) + h(t) \leq \rho + \tfrac{1}{2}(1-\rho) = \tfrac{1}{2}(1+\rho).$$

If we define $\widetilde{H}(t) = 2/(1-\rho)\big(H(t) - \frac{1}{2}(1+\rho)\big)$, then $\widetilde{H} \leq 0$ on $\widehat{W_{\mathcal{T}}^{\rho}}$ and $\widetilde{H} = 1$ on W. Thus $\widetilde{H}$ is a candidate for the capacity of the condenser and so, by the comparison principle,

$$\operatorname{Cap}_{\mathcal{T}}(W, \widehat{W_{\mathcal{T}}^{\rho}}) \leq \|\triangle \widetilde{H}\|_{\ell^2(\mathcal{G}(W_{\mathcal{T}}^{\rho}, W))}^2 \leq \|\triangle \widetilde{H}\|_{\ell^2(\mathcal{T})}^2 = \left(\frac{2}{1-\rho}\right)^2 \|h\|_{\ell^2(\mathcal{T})}^2 = \frac{4}{(1-\rho)^2} \operatorname{Cap}_{\mathcal{T}} W. \quad \square$$

We also have good *tree* separation inherited from the stopping time blowup $W_{\mathcal{T}}^{\rho}$. This gives our substitute for (2-4) and (2-5).

Lemma 2.8. $W_{\mathcal{T}}^{\rho} \subset \widehat{W_{\mathcal{T}}^{\rho}}$ *as open subsets of the circle or, equivalently, as $\mathcal{T}$-tent regions in the disk. Consequently* $\operatorname{Cap}_{\mathcal{T}} W_{\mathcal{T}}^{\rho} \leq \rho^{-2} \operatorname{Cap}_{\mathcal{T}} W$.

Proof. The restriction of H to a geodesic is a concave function of distance from the root, and so if $o < z < w \in W$, then

$$H(z) \geq \left(1 - \frac{d(z)}{d(w)}\right) H(o) + \frac{d(z)}{d(w)} H(w) = \frac{d(z)}{d(w)} \geq \rho, \quad z \in \widehat{W_{\mathcal{T}}^{\rho}},$$

and this proves $W_{\mathcal{T}}^{\rho} \subset \widehat{W_{\mathcal{T}}^{\rho}}$. The inequality now follows from Lemma 2.6. $\square$

Holomorphic approximate extremals and capacity estimates. We now define a holomorphic approximation Φ to the extremal function $H = Ih$ on $\mathcal{T}$ constructed in Proposition 2.3. We will use a parameter s. We always suppose $s > -1$ and additional specific assumptions will be made at various places. Define

$$\begin{aligned}\varphi_\kappa(z) &= \left(\frac{1-|\kappa|^2}{1-\overline{\kappa}z}\right)^{1+s},\\ \Phi(z) &= \sum_{\kappa\in\mathcal{T}} h(\kappa)\varphi_\kappa(z) = \sum_{\kappa\in\mathcal{T}} h(\kappa)\left(\frac{1-|\kappa|^2}{1-\overline{\kappa}z}\right)^{1+s}.\end{aligned} \tag{2-12}$$

Note that for $\tau \in \mathcal{T}$

$$\sum_{\kappa\in\mathcal{T}} h(\kappa) I\delta_\kappa(\tau) = I\left(\sum_{\kappa\in\mathcal{T}} h(\kappa)\delta_\kappa\right)(\tau) = Ih(\tau) = H(\tau),$$

and so

$$\Phi(z) - H(z) = \sum_{\kappa\in\mathcal{T}} h(\kappa)\{\varphi_\kappa - I\delta_\kappa\}(z). \tag{2-13}$$

Define Γ_s by

$$\Gamma_s h(z) = \int_{\mathbb{D}} h(\zeta)\frac{(1-|\zeta|^2)^s}{(1-\overline{\zeta}z)^{1+s}}\, dA, \tag{2-14}$$

and recall that for appropriate constant c_s, $c_s\Gamma_s$ is a projection onto holomorphic functions [Zhu 2005, Thm 2.11]. For notational convenience we absorb the constant c_s into the measure dA. Thus for $h \in \mathcal{P}(\mathbb{D})$,

$$\Gamma_s h(z) = h(z). \tag{2-15}$$

We then have $\Phi = \Gamma_s g$ where

$$g(\zeta) = \sum_{\kappa\in\mathcal{T}} h(\kappa)\frac{1}{|B_\kappa|}\frac{(1-\overline{\zeta}\kappa)^{1+s}}{(1-|\zeta|^2)^s}\chi_{B_\kappa}(\zeta), \tag{2-16}$$

and B_κ is the Euclidean ball centered at κ with radius $c(1-|\kappa|)$ where c is a small positive constant to be chosen later. The function Φ satisfies the following estimates.

Proposition 2.9. *Set $F = \widehat{E^\rho_{\mathcal{T}}}$ and write $E = \{w_k\}_k$. Suppose $z \in \mathbb{D}$ and $s > -1$. Then*

$$\begin{cases} |\Phi(z) - \Phi(w_k)| \le C\,\mathrm{Cap}_{\mathcal{T}}(E, F), & z \in T(w_k),\\ \operatorname{Re}\Phi(w_k) \ge c > 0, & k \ge 1,\\ |\Phi(w_k)| \le C, & k \ge 1,\\ |\Phi(z)| \le C\,\mathrm{Cap}_{\mathcal{T}}(E, F), & z \notin F.\end{cases} \tag{2-17}$$

Corollary 2.10. *Let the situation be as in the proposition. If $s > -\frac{1}{2}$ then $\Phi = \Gamma_s g$, where g satisfies*

$$\int_{\mathbb{D}} |g(\zeta)|^2 dA \le C\,\mathrm{Cap}_{\mathcal{T}}(E, F); \tag{2-18}$$

and if $s > \frac{1}{2}$ then

$$\|\Phi\|^2_{\mathcal{D}} \le \int_{\mathbb{D}} |g(\zeta)|^2 dA \le C\,\mathrm{Cap}_{\mathcal{T}}(E, F). \tag{2-19}$$

Proof. From (2-13) we have

$$|\Phi(z)-H(z)| \le \sum_{\kappa\in[o,z]} \left|h(\kappa)\{\varphi_\kappa(z)-1\}\right| + \sum_{\kappa\notin[o,z]} |h(\kappa)\varphi_\kappa(z)| = \mathrm{I}(z)+\mathrm{II}(z).$$

Also, h is nonnegative and supported in $V_G^\gamma \setminus V_G^\alpha$. We first show that

$$\mathrm{II}(z) \le \sum_{\kappa\notin[o,z]} h(\kappa)\left|\frac{1-|\kappa|^2}{1-\bar{\kappa}z}\right|^{1+s} \le C\,\mathrm{Cap}(E,F).$$

For $A>1$ let

$$\Omega_j = \left\{\kappa \in \mathcal{T} : A^{-j-1} < \left|\frac{1-|\kappa|^2}{1-\bar{\kappa}z}\right| \le A^{-j}\right\}.$$

Lemma 2.11. *For every j the set Ω_j is a union of two stopping times for $\mathcal{T}$.*

Proof. Let Ω_j^1 be the subset of Ω_j of points whose distance from the root is odd and set $\Omega_j^2 = \Omega_j \setminus \Omega_j^1$. We will show both are stopping times; that is, if for $r=1,2$, $\kappa\in\Omega_j^r$, $\lambda\in\mathcal{T}$, and $\kappa\in[o,\lambda)$, then $\lambda\notin\Omega_j^r$.

Set $\delta\kappa = \lambda-\kappa$. We have

$$\begin{aligned}\left|\frac{1-\bar{\lambda}z}{1-|\lambda|^2}\right| &= \frac{1-|\kappa|^2}{1-|\lambda|^2}\left|\frac{1-(\overline{\kappa+\delta\kappa})z}{1-|\kappa|^2}\right| \\ &= \frac{1-|\kappa|^2}{1-|\lambda|^2}\left|\frac{1-\bar{\kappa}z}{1-|\kappa|^2}-\frac{\overline{\delta\kappa}z}{1-|\kappa|^2}\right| \ge \frac{1-|\kappa|^2}{1-|\lambda|^2}\left(\left|\frac{1-\bar{\kappa}z}{1-|\kappa|^2}\right|-\frac{|\overline{\delta\kappa}z|}{1-|\kappa|^2}\right).\end{aligned} \tag{2-20}$$

By the construction of the tree $(1-|\kappa|^2)\sim 2^s(1-|\lambda|^2)$ for some positive integer s, and if κ and λ are in the same Ω_j^r then $s\ge 2$. Also, by the construction of $\mathcal{T}$, we have

$$\frac{|\overline{\delta\kappa}z|}{1-|\kappa|^2} \le \frac{\sqrt{2}(1-|\kappa|)|z|}{1-|\kappa|^2} \lesssim \frac{\sqrt{2}}{2},$$

and hence we continue with

$$\left|\frac{1-\bar{\lambda}z}{1-|\lambda|^2}\right| \ge 4\left(A^j-\frac{\sqrt{2}}{2}\right).$$

We are done if $A^{j+1}\le 4(A^j-\sqrt{2}/2)$ for each j. That holds if $A\le 4(1-\sqrt{2}/2)<1.17$. □

Now by the stopping time property, item 3 in Proposition 2.3, we have

$$\sum_{\kappa\in\Omega_j} h(\kappa) \le C\,\mathrm{Cap}_{\mathcal{T}}(E,F),\ \ j\ge 0.$$

Altogether we then have

$$\mathrm{II}(z) \le \sum_{j=0}^{\infty}\sum_{\kappa\in\Omega_j} h(\kappa)A^{-j(1+s)} \le C_s\,\mathrm{Cap}_{\mathcal{T}}(E,F).$$

If $z\in\mathbb{D}\setminus F$ then $\mathrm{I}(z)=0$ and $H(z)=0$ and we have

$$|\Phi(z)| = |\Phi(z)-H(z)| \le \mathrm{II}(z) \le C_s\,\mathrm{Cap}_{\mathcal{T}}(E,F),$$

which is the fourth line in (2-17).

If $z \in T(w_j)$, then for $\kappa \notin [o, w_j]$ we have $|\varphi_\kappa(w_j)| \le C|\varphi_\kappa(z)|$, and for $\kappa \in [o, z]$ we have

$$|\varphi_\kappa(z) - \varphi_\kappa(w_j)| = \left| \left(\frac{1-|\kappa|^2}{1-\overline{\kappa} z} \right)^{1+s} - \left(\frac{1-|\kappa|^2}{1-\overline{\kappa} w_j} \right)^{1+s} \right| \le C_s \frac{|z-w_j|}{1-|\kappa|^2}.$$

Thus for $z \in T(w_j^\alpha)$,

$$\begin{aligned} |\Phi(z) - \Phi(w_j)| &\le \sum_{\kappa \in [o, w_j^\alpha]} h(\kappa) |\varphi_\kappa(z) - \varphi_\kappa(w_j)| + C \sum_{\kappa \notin [o,z]} h(\kappa) |\varphi_\kappa(z)| \\ &\le C_s \sum_{\kappa \in [o, w_j^\alpha]} h(\kappa) \frac{|z-w_j|}{1-|\kappa|^2} + C\, \mathrm{II}(z) \le C_s \operatorname{Cap}_{\mathcal{T}}(E, F), \end{aligned}$$

since $h(\kappa) \le C \operatorname{Cap}_{\mathcal{T}}(E, F)$ and $\sum_{\kappa \in [o, w_j]} \frac{1}{1-|\kappa|^2} \approx \frac{1}{1-|w_j|^2}$. This proves the first line in (2-17).

Moreover, we note that for $s = 0$ and $\kappa \in [o, w_j]$,

$$\operatorname{Re} \varphi_\kappa(w_j) = \operatorname{Re} \frac{1-|\kappa|^2}{1-\overline{\kappa} w_j} = \operatorname{Re} \frac{1-|\kappa|^2}{|1-\overline{\kappa} w_j|^2} (1 - \kappa \overline{w_j}) \ge c > 0.$$

A similar result holds for $s > -1$ provided the Bergman tree $\mathcal{T}$ is constructed sufficiently thin depending on s. It then follows from $\sum_{\kappa \in [o, w_j]} h(\kappa) = 1$ that

$$\begin{aligned} \operatorname{Re} \Phi(w_j) &= \sum_{\kappa \in [o, w_j]} h(\kappa) \operatorname{Re} \varphi_\kappa(w_j) + \sum_{\kappa \notin [o, w_j]} h(\kappa) \operatorname{Re} \varphi_\kappa(w_j) \\ &\ge c \sum_{\kappa \in [o, w_j]} h(\kappa) - C \operatorname{Cap}_{\mathcal{T}}(E, F) \ge c' > 0. \end{aligned}$$

We trivially have

$$|\Phi(w_j)| \le \mathrm{I}(z) + \mathrm{II}(z) \le C \sum_{\kappa \in [o, w_j]} h(\kappa) + C \operatorname{Cap}_{\mathcal{T}}(E, F) \le C,$$

and this completes the proof of (2-17).

Now we prove (2-18). From property 1 of Proposition 2.3 we obtain

$$\begin{aligned} \int_{\mathbb{D}} |g(\zeta)|^2 dA &= \int_{\mathbb{D}} \left| \sum_{\kappa \in \mathcal{T}} h(\kappa) \frac{1}{|B_\kappa|} \frac{(1-\overline{\zeta}\kappa)^{1+s}}{(1-|\zeta|^2)^s} \chi_{B_\kappa}(\zeta) \right|^2 dA \\ &= \sum_{\kappa \in \mathcal{T}} |h(\kappa)|^2 \frac{1}{|B_\kappa|^2} \int_{B_\kappa} \frac{|1-\overline{\zeta}\kappa|^{2+2s}}{(1-|\zeta|^2)^{2s}} \, dA \approx \sum_{\kappa \in \mathcal{T}} |h(\kappa)|^2 \approx \operatorname{Cap}_{\mathcal{T}}(E, F). \end{aligned}$$

Finally (2-19) follows from (2-18) and [Böe 2002, Lemma 2.4]. □

Corollary 2.12. *Let G be a finite union of arcs in the circle $\mathbb{T}$. Then*

$$\operatorname{Cap}_{\mathbb{D}}(G) \approx \operatorname{Cap}_{\mathcal{T}}(G), \tag{2-21}$$

where $\operatorname{Cap}_{\mathbb{D}}$ denotes Stegenga's capacity on the circle $\mathbb{T}$.

Proof. To prove the inequality $\lesssim$ in (2-21) we use Proposition 2.9 to obtain a test function for estimating the Stegenga capacity of G. We take $F = \{o\}$ and $E = G$ in Proposition 2.9. Let c, C be the constants in Proposition 2.9, and suppose that $\operatorname{Cap}(E, F) \le c/(3C)$. Set $\Psi(z) = \frac{3}{c}(\Phi(z) - \Phi(0))$. Then $\Psi(0) = 0$ and

$$\operatorname{Re}\Psi(z) = \frac{3}{c}\big\{\operatorname{Re}\Phi(z) - \operatorname{Re}\Phi(0)\big\} \ge \frac{3}{c}\big\{c - 2C\operatorname{Cap}_{\mathcal{T}}(E, F)\big\} \ge 1, \quad z \in G.$$

By definition (2-2) and (2-19) we have, for $G \subset \mathbb{T}$,

$$\operatorname{Cap}_{\mathbb{D}}(G) \le \|\Psi\|_{\mathcal{D}}^2 = \left(\frac{3}{c}\right)^2 \|\Phi\|_{\mathcal{D}}^2 \le \left(\frac{3}{c}\right)^2 C\operatorname{Cap}_{\mathcal{T}}(E, F) \le C\operatorname{Cap}_{\mathcal{T}} E = C\operatorname{Cap}_{\mathcal{T}} G.$$

To obtain the opposite inequality we use $\psi \in \mathcal{D}$, an extremal function for computing $\operatorname{Cap}_{\mathbb{D}} G$. For $R > 0$, $z \in \mathbb{D}$ let $B(z, R)$ be the hyperbolic disk of radius R centered at z. Pick R large enough so that for all $\kappa \in \mathcal{T} \setminus \{o\}$ we have $B(\kappa, R) \supset \operatorname{convexhull}(B_\kappa \cup B_{\kappa^{-1}})$. Our candidate for estimating $\operatorname{Cap}_{\mathcal{T}}$ is given by setting $h(o) = 0$ and

$$h(\kappa) = (1 - |\kappa|^2)\sup\big\{|\psi'(z)| : z \in B(\kappa, R)\big\}; \ \kappa \in \mathcal{T} \setminus \{o\}.$$

We have the pointwise estimate

$$\begin{aligned}\operatorname{Re}\psi(\beta) \le |\psi(\beta)| &\le \sum_{\kappa\in[o,\beta]} \big|\psi(\kappa) - \psi(\kappa^{-1})\big| \\ &\le \sum_{\kappa\in[o,\beta]} |\kappa - \kappa^{-1}| \sup\big\{|\psi'(z)| : z \in \operatorname{segment}(\kappa, \kappa^{-1})\big\} \le C \sum_{\kappa\in[o,\beta]} h(\kappa) = CIh(\beta).\end{aligned}$$

We have the norm estimate, with $z(\kappa)$ denoting the appropriate point in $B(\kappa, R)$,

$$\begin{aligned}\|h\|_{\ell^2(\mathcal{T})}^2 &= \sum_{\kappa\in\mathcal{T}} (1 - |\kappa^2|)^2 |\psi'(z(\kappa))|^2 \le C \sum_{\kappa\in\mathcal{T}} \frac{(1 - |\kappa^2|)^2}{|B(\kappa, R)|} \int_{B(\kappa,R)} |\psi'(z)|^2\, dA \\ &\le C \sum_{\kappa\in\mathcal{T}} \int_{B(\kappa,R)} |\psi'(z)|^2\, dA \le C \int_{\mathbb{D}} |\psi'(z)|^2\, dA \le C\|\psi\|_{\mathcal{D}}^2.\end{aligned}$$

Here the first inequality uses the submean value property for the subharmonic function $|\psi'(z)|^2$, the second uses straightforward estimates for $|B(\kappa, R)|$, and the next estimate holds because the $B(\kappa, R)$ are approximately disjoint; $\sum \chi_{B(\kappa,R)}(z) \le C$. Recalling definition (2-8) we find

$$\operatorname{Cap}_{\mathcal{T}} G \le C\left\|\frac{1}{c}\psi\right\|_{\mathcal{D}}^2 = \frac{C}{c^2}\operatorname{Cap}_{\mathbb{D}} G. \qquad \square$$

Abbreviate $\operatorname{Cap}_{\mathcal{T}_\theta}$ by $\operatorname{Cap}_\theta$, and let $T_\theta(E)$ be the $\mathcal{T}_\theta$-tent region corresponding to an open subset E of the circle $\mathbb{T}$. Recall that $T(E) = \bigcup_{I\subset E} T(I)$. Now define M by

$$M := \sup_{E \text{ open} \subset \mathbb{T}} \frac{\int_{\mathbb{T}} \mu_b(T_\theta(E))\, d\theta}{\int_{\mathbb{T}} \operatorname{Cap}_\theta(E)\, d\theta}. \tag{2-22}$$

Corollary 2.13. $$\|\mu_b\|_{CM(\mathcal{D})}^2 \approx M.$$

Proof. Using Corollary 2.12 and $T_\theta(E) \subset T(E)$, we have

$$M \leq C \sup_{E \text{ open} \subset \mathbb{T}} \frac{\int_{\mathbb{T}} \mu_b(T(E))\, d\theta}{\int_{\mathbb{T}} \operatorname{Cap}_{\mathbb{D}}(E)\, d\theta} = C \sup_{E \text{ open} \subset \mathbb{T}} \frac{\mu_b(T(E))}{\operatorname{Cap}_{\mathbb{D}}(E)} \approx \|\mu_b\|^2_{CM(\mathfrak{D})},$$

where the final comparison is Stegenga's theorem. Conversely, one can verify using an argument in the style of the one in (2-25) below that for $0 < \rho < 1$,

$$\mu_b(E) \leq C \int_{\mathbb{T}} \mu_b(T_\theta(E^\rho_{\mathbb{D}}))\, d\theta \leq CM \int_{\mathbb{T}} \operatorname{Cap}_\theta(E^\rho_{\mathbb{D}}) d\theta \approx CM \operatorname{Cap}_{\mathbb{D}}(E^\rho_{\mathbb{D}}) \leq CM \operatorname{Cap}_{\mathbb{D}}(E).$$

Here the third line uses (2-21) with $E^\rho_{\mathbb{D}}$ and $\mathcal{T}(\theta)$ in place of G and $\mathcal{T}$, and the final inequality follows from (2-4). Thus from Stegenga's theorem we obtain

$$\|\mu_b\|^2_{CM(\mathfrak{D})} \approx \sup_{E \text{ open} \subset \mathbb{T}} \frac{\mu_b(E)}{\operatorname{Cap}_{\mathbb{D}}(E)} \leq CM. \qquad \square$$

Given $0 < \delta < 1$, let G be an open set in $\mathbb{T}$ such that

$$\frac{\int_{\mathbb{T}} \mu_b(T_\theta(G)) d\theta}{\int_{\mathbb{T}} \operatorname{Cap}_\theta(G)\, d\theta} \geq \delta M. \tag{2-23}$$

We need to know that $\mu_b(V^\beta_G \setminus V_G)$ is small compared to $\mu_b(V_G)$. This crucial step of the proof is where we use the asymptotic capacity estimate Lemma 2.8.

Proposition 2.14. *Given $\varepsilon > 0$ we can choose $\delta = \delta(\varepsilon) < 1$ in* (2-23) *and $\beta = \beta(\varepsilon) < 1$ so that, for any G satisfying* (2-23), *we have*

$$\mu_b(V^\beta_G \setminus V_G) \leq \varepsilon \mu_b(V_G), \tag{2-24}$$

where $V^\beta_G = G^\beta_{\mathbb{D}}$ and $V_G = G^1_{\mathbb{D}} = T(G)$.

Proof. Let $G^\rho(\theta) = G^\rho_{\mathcal{T}_\theta}$. Lemma 2.8 shows that $\operatorname{Cap}_\theta(G^\rho(\theta)) \leq \rho^{-2} \operatorname{Cap}_\theta(G)$ for $0 \leq \theta < 2\pi$, $0 < \rho < 1$, and if we integrate on $\mathbb{T}$ we obtain

$$\int_{\mathbb{T}} \operatorname{Cap}_\theta(G^\rho(\theta))\, d\theta \leq \rho^{-2} \int_{\mathbb{T}} \operatorname{Cap}_\theta(G)\, d\theta.$$

From (2-22) and (2-23) we thus have

$$\int_{\mathbb{T}} \mu_b(T_\theta(G^\rho(\theta)))\, d\theta \leq M \int_{\mathbb{T}} \operatorname{Cap}_\theta(G^\rho(\theta))\, d\theta \leq M\rho^{-2} \int_{\mathbb{T}} \operatorname{Cap}_\theta(G)\, d\theta \leq \frac{1}{\delta\rho^2} \int_{\mathbb{T}} \mu_b(T_\theta(G))\, d\theta.$$

It follows that

$$\begin{aligned} \int_{\mathbb{T}} \mu_b(T_\theta(G^\rho(\theta)) \setminus T_\theta(G))\, d\theta &= \int_{\mathbb{T}} \mu_b(T_\theta(G^\rho(\theta)))\, d\theta - \int_{\mathbb{T}} \mu_b(T_\theta(G))\, d\theta \\ &\leq \Big(\frac{1}{\delta\rho^2} - 1\Big) \int_{\mathbb{T}} \mu_b(T_\theta(G))\, d\theta. \end{aligned}$$

Now, with $\eta = \frac{1}{2}(\rho+1)$,

$$\begin{aligned}\int_{\mathbb{T}} \mu_b\big(T_\theta(G^\rho(\theta)) \setminus T_\theta(G)\big)\,d\theta &= \int_{\mathbb{T}}\int_{T_\theta(G^\rho(\theta))\setminus T_\theta(G)} d\mu_b(z)\,d\theta \geq \int_{\mathbb{T}}\int_{T_\theta(G^\rho(\theta))\setminus T(G)} d\mu_b(z)\,d\theta \\ &\geq \int_{\mathbb{T}}\int_{T_\theta(G^\rho(\theta))\setminus T(G)} d\mu_b(z)\,d\theta \qquad (2\text{-}25)\\ &= \int_{\mathbb{D}} \frac{1}{2\pi}\int_{\{\theta: z\in T_\theta(G^\rho(\theta))\setminus T(G)\}} d\theta\,d\mu_b(z) \geq \frac{1}{2}\int_{T(G^\eta_{\mathbb{D}})\setminus T(G)} d\mu_b(z),\end{aligned}$$

since every $z \in T(G^\eta_{\mathbb{D}})$ lies in $T_\theta(G^\rho(\theta))$ for at least half of the θ's in $[0, 2\pi)$. Here we may assume that the components of $G^\rho_{\mathbb{D}}$ have small length since otherwise we trivially have $\int_{\mathbb{T}} \mathrm{Cap}_{\mathcal{T}(\theta)}(G)\,d\theta \geq c > 0$. We continue with

$$M \leq \frac{1}{c}\int d\mu_b \leq \frac{1}{c}\|b\|^2_{\mathcal{D}} \leq \frac{C}{c}\|T_b\|^2. \tag{2-26}$$

Combining the inequalities above, using $\rho = 2\eta - 1$, $1/2 \leq \rho < 1$, and choosing $\delta = \eta$, we obtain

$$\begin{aligned}\mu_b(T(G^\eta_{\mathbb{D}}) \setminus T(G)) &\leq 2\left(\frac{1}{\delta\rho^2} - 1\right)\int_{\mathbb{T}} \mu_b(T_\theta(G))\,d\theta \\ &= 2\left(\frac{1}{\eta(2\eta-1)^2} - 1\right)\int_{\mathbb{T}} \mu_b(T_\theta(G))\,d\theta \leq C(1-\eta)\int_{\mathbb{T}} \mu_b(T_\theta(G))\,d\theta,\end{aligned}$$

for $\frac{3}{4} \leq \eta < 1$. Recalling that $V^\eta_G = T(G^\eta_{\mathbb{D}})$ and that for all θ we have $T_\theta(G) \subset T(G) = V_G$ this becomes

$$\mu_b(V^\eta_G \setminus V_G) \leq C(1-\eta)\int_{\mathbb{T}} \mu_b(T_\theta(G))\,d\theta \leq C(1-\eta)\mu_b(V_G), \quad \tfrac{3}{4} \leq \eta < 1,$$

Hence given $\varepsilon > 0$ it is possible to select δ and β so that (2-24) holds. □

Schur estimates and a bilinear operator on trees. We begin with a bilinear version of Schur's well known theorem.

Proposition 2.15. *Let (X, μ), (Y, ν) and (Z, ω) be measure spaces and $H(x, y, z)$ be a nonnegative measurable function on $X \times Y \times Z$. Define, initially for nonnegative functions f, g,*

$$T(f, g)(x) = \int_{Y\times Z} H(x, y, z) f(y)\,d\nu(y) g(z)\,d\omega(z), \quad x \in X,$$

For $1 < p < \infty$, suppose there are positive functions h, k, and m on X, Y, and Z respectively such that

$$\int_{Y\times Z} H(x, y, z)k(y)^{p'} m(z)^{p'} d\nu(y)\,d\omega(z) \leq (Ah(x))^{p'},$$

for μ-a.e. $x \in X$, and

$$\int_X H(x, y, z)h(x)^p\,d\mu(x) \leq (Bk(y)m(z))^p,$$

for $\nu \times \omega$-a.e. $(y, z) \in Y \times Z$. Then T is bounded from $L^p(\nu) \times L^p(\omega)$ to $L^p(\mu)$ and $\|T\| \leq AB$.

Proof. We have

$$\begin{aligned}\int_X |Tf(x)|^p \, d\mu(x) &\le \int_X \biggl(\int_{Y\times Z} H(x,y,z) k(y)^{p'} m(z)^{p'} \, d\nu(y)\, d\omega(z) \biggr)^{p/p'} \\ &\qquad \times \biggl(\int_{Y\times Z} H(x,y,z) \Bigl(\frac{f(y)}{k(y)}\Bigr)^p d\nu(y) \Bigl(\frac{g(z)}{m(z)}\Bigr)^p d\omega(z) \biggr) \, d\mu(x) \\ &\le A^p \int_{Y\times Z} \biggl(\int_X H(x,y,z) h(x)^p \, d\mu(x) \biggr) \Bigl(\frac{f(y)}{k(y)}\Bigr)^p d\nu(y) \Bigl(\frac{g(z)}{m(z)}\Bigr)^p d\omega(z) \\ &\le A^p B^p \int_{Y\times Z} k(y)^p m(z)^p \Bigl(\frac{f(y)}{k(y)}\Bigr)^p d\nu(y) \Bigl(\frac{g(z)}{m(z)}\Bigr)^p d\omega(z) \\ &= (AB)^p \int_Y f(y)^p \, d\nu(y) \int_Z g(z)^p \, d\omega(z). \qquad \square\end{aligned}$$

This proposition can be used, along with the estimates

$$\int_{\mathbb{D}} \frac{(1-|w|^2)^t}{|1-\overline{w}z|^{2+t+c}} \, dw \approx \begin{cases} C_t & \text{if } c<0,\ t>-1, \\ -C_t \log(1-|z|^2) & \text{if } c=0,\ t>-1, \\ C_t(1-|z|^2)^{-c} & \text{if } c>0,\ t>-1, \end{cases} \tag{2-27}$$

to prove a corollary we will use later [Zhu 2005, Thm 2.10].

Corollary 2.16. *Define*

$$Tf(z) = (1-|z|^2)^a \int_{\mathbb{D}} \frac{(1-|w|^2)^b}{(1-\overline{w}z)^{2+a+b}} f(w)\, dw, \quad Sf(z) = (1-|z|^2)^a \int_{\mathbb{D}} \frac{(1-|w|^2)^b}{|1-\overline{w}z|^{2+a+b}} f(w)\, dw.$$

Suppose $t \in \mathbb{R}$ and $1 \le p < \infty$. Then T is bounded on $L^p(\mathbb{D}, (1-|z|^2)^t dA)$ if and only if S is bounded on $L^p(\mathbb{D}, (1-|z|^2)^t dA)$ if and only if

$$-pa < t+1 < p(b+1). \tag{2-28}$$

We now use Proposition 2.15 to show that if $\mathcal{A}, \mathcal{B} \subset \mathcal{T}$ are well separated then a certain bilinear operator mapping on $\ell^2(\mathcal{A}) \times \ell^2(\mathcal{B})$ maps boundedly into $L^2(\mathbb{D})$.

Lemma 2.17. *Suppose $\mathcal{A}$ and $\mathcal{B}$ are subsets of T, $h \in \ell^2(\mathcal{A})$ and $k \in \ell^2(\mathcal{B})$, and $1/2 < \alpha < 1$. Suppose further that $\mathcal{A}$ and $\mathcal{B}$ satisfy the separation condition, $\forall \kappa \in \mathcal{A}$, $\gamma \in \mathcal{B}$, then we have*

$$|\kappa - \gamma| \ge (1-|\gamma|^2)^\alpha. \tag{2-29}$$

Then the bilinear map of (h,k) to functions on the disk given by

$$T(h,b)(z) = \biggl(\sum_{\kappa\in\mathcal{A}} h(\kappa) \frac{(1-|\kappa|^2)^{1+s}}{|1-\overline{\kappa}z|^{2+s}} \biggr) \biggl(\sum_{\gamma\in\mathcal{B}} b(\gamma) \frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}} \biggr)$$

is bounded from $\ell^2(\mathcal{A}) \times \ell^2(\mathcal{B})$ to $L^2(\mathbb{D})$.

Remark 2.18. For $h \in \ell^2(\mathcal{A})$ and $b \in \ell^2(\mathcal{B})$ set

$$H(z) = \sum_{\kappa\in\mathcal{A}} h(\kappa) \frac{(1-|\kappa|^2)^{1+s}}{(1-\overline{\kappa}z)^{2+s}}, \; B(z) = \sum_{\gamma\in\mathcal{B}} b(\gamma) \frac{(1-|\gamma|^2)^{1+s}}{(1-\overline{\gamma}z)^{1+s}}.$$

By [Zhu 2005, Thm 2.30] $H \in L^2(\mathbb{D})$ and $B \in \mathcal{D}$. There are unbounded functions in $\mathcal{D}$; hence these facts do not ensure that $HB \in L^2(\mathbb{D})$. The lemma shows that if $\mathcal{A}$ and $\mathcal{B}$ are separated then $HB \in L^2(\mathbb{D})$.

Proof of Lemma 2.17. We will verify the hypotheses of the previous proposition. The kernel function here is

$$H(z,\kappa,\gamma) = \frac{(1-|\kappa|^2)^{1+s}}{|1-\overline{\kappa}z|^{2+s}} \frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}}, \quad z \in \mathbb{D}, \kappa \in \mathcal{A}, \gamma \in \mathcal{B},$$

with Lebesgue measure on $\mathbb{D}$ and counting measure on $\mathcal{A}$ and $\mathcal{B}$. We will take as Schur functions

$$h(z) = (1-|z|^2)^{-1/4}, \quad k(\kappa) = (1-|\kappa|^2)^{1/4}, \quad \text{and} \quad m(\gamma) = (1-|\gamma|^2)^{\varepsilon/2},$$

on $\mathbb{D}$, $\mathcal{A}$ and $\mathcal{B}$ respectively, where $\varepsilon = \varepsilon(\alpha, s) > 0$ will be chosen sufficiently small later. We must then verify

$$\sum_{\kappa\in\mathcal{A}} \sum_{\gamma\in\mathcal{B}} \frac{(1-|\kappa|^2)^{3/2+s}}{|1-\overline{\kappa}z|^{2+s}} \frac{(1-|\gamma|^2)^{1+\varepsilon+s}}{|1-\overline{\gamma}z|^{1+s}} \le A^2(1-|z|^2)^{-1/2} \tag{2-30}$$

for $z \in \mathbb{D}$, and

$$\int_{\mathbb{D}} \frac{(1-|\kappa|^2)^{1+s}}{|1-\overline{\kappa}z|^{2+s}} \frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}} (1-|z|^2)^{-1/2} dA \le B^2(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{\varepsilon} \tag{2-31}$$

for $\kappa \in \mathcal{A}$ and $\gamma \in \mathcal{B}$.

To prove (2-30) we write

$$\sum_{\kappa\in\mathcal{A}} \sum_{\gamma\in\mathcal{B}} \frac{(1-|\kappa|^2)^{3/2+s}}{|1-\overline{\kappa}z|^{2+s}} \frac{(1-|\gamma|^2)^{1+\varepsilon+s}}{|1-\overline{\gamma}z|^{1+s}} = \left(\sum_{\kappa\in\mathcal{A}} \frac{(1-|\kappa|^2)^{3/2+s}}{|1-\overline{\kappa}z|^{2+s}}\right)\left(\sum_{\gamma\in\mathcal{B}} \frac{(1-|\gamma|^2)^{1+\varepsilon+s}}{|1-\overline{\gamma}z|^{1+s}}\right).$$

Then from (2-27) we obtain

$$\sum_{\kappa\in\mathcal{A}} \frac{(1-|\kappa|^2)^{3/2+s}}{|1-\overline{\kappa}z|^{2+s}} \le C \int_{\mathbb{D}} \frac{(1-|w|^2)^{-1/2+s}}{|1-\overline{w}z|^{2+s}}\, dw \le C(1-|z|^2)^{-1/2}$$

and

$$\sum_{\gamma\in\mathcal{B}} \frac{(1-|\gamma|^2)^{1+\varepsilon+s}}{|1-\overline{\gamma}z|^{1+s}} \le C \int_{\zeta\in V_G} \frac{(1-|\zeta|^2)^{-1+\varepsilon+s}}{|1-\overline{\zeta}z|^{1+s}}\, dA \le C,$$

which yields (2-30).

We now prove (2-31). We will make repeated use of (2-29) as well as the following consequence of it (via the triangle inequality):

$$(1-|\kappa|^2) \le C|\kappa-\gamma| \quad \text{for all } \kappa \in \mathcal{A},\ \gamma \in \mathcal{B}.$$

We set $\kappa^* = \dfrac{\kappa}{|\kappa|}$, $\gamma^* = \dfrac{\gamma}{|\gamma|}$, and we express the integral

$$\int_{\mathbb{D}} \frac{(1-|\kappa|^2)^{1+s}}{|1-\overline{\kappa}z|^{2+s}} \frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}} (1-|z|^2)^{-1/2}\,dA =: \mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V}$$

as a sum of integrals over five regions:

$$\begin{aligned}
\mathrm{I} &\text{ over } \left\{|z-\gamma^*| \le 1-|\gamma|^2\right\},\\
\mathrm{II} &\text{ over } \left\{1-|\gamma|^2 \le |z-\gamma^*| \le \tfrac{1}{2}|\kappa-\gamma|\right\},\\
\mathrm{III} &\text{ over } \left\{|z-\kappa^*| \le 1-|\kappa|^2\right\},\\
\mathrm{IV} &\text{ over } \left\{1-|\kappa|^2 \le |z-\kappa^*| \le \tfrac{1}{2}|\kappa-\gamma|\right\},\\
\mathrm{V} &\text{ over } \left\{|z-\gamma^*|, |z-\kappa^*| \ge |\kappa-\gamma|\right\}.
\end{aligned}$$

We have

$$\begin{aligned}
\mathrm{I} &\approx \frac{(1-|\kappa|^2)^{1+s}}{|\kappa-\gamma|^{2+s}} \int_{|z-\gamma^*|\le 1-|\gamma|^2} (1-|z|^2)^{-1/2}\,dA\\
&\approx \frac{(1-|\kappa|^2)^{1+s}(1-|\gamma|^2)^{3/2}}{|\kappa-\gamma|^{2+s}} \le C(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{3(1-\alpha)/2},\\
\mathrm{II} &\approx \frac{(1-|\kappa|^2)^{1+s}(1-|\gamma|^2)^{1+s}}{|\kappa-\gamma|^{2+s}} \int_{1-|\gamma|^2\le|z-\gamma^*|\le\frac{1}{2}|\kappa-\gamma|} \frac{(1-|z|^2)^{-1/2}}{|z-\gamma^*|^{1+s}}\,dA\\
&\approx \frac{(1-|\kappa|^2)^{1+s}(1-|\gamma|^2)^{1+s}}{|\kappa-\gamma|^{2+s}} (1-|\gamma|^2)^{1/2-s}\\
&= \frac{(1-|\kappa|^2)^{1+s}(1-|\gamma|^2)^{3/2}}{|\kappa-\gamma|^{2+s}} \le C(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{3(1-\alpha)/2},\\
\mathrm{III} &\approx \frac{(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{1+s}}{|\kappa-\gamma|^{1+s}} \le C(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{(1+s)(1-\alpha)},\\
\mathrm{IV} &\le C(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{\varepsilon} \qquad \text{for some } \varepsilon > 0,\\
\mathrm{V} &\approx \int_{|z-\gamma^*|,|z-\kappa^*|\ge|\kappa-\gamma|} \frac{(1-|\kappa|^2)^{1+s}}{|z-\kappa^*|^{2+s}} \frac{(1-|\gamma|^2)^{1+s}}{|z-\gamma^*|^{1+s}} (1-|z|^2)^{-1/2}\,dA\\
&\approx \frac{(1-|\kappa|^2)^{1+s}(1-|\gamma|^2)^{1+s}}{|\kappa-\gamma|^{3/2+2s}} \le C(1-|\kappa|^2)^{1/2}(1-|\gamma|^2)^{(1+s)(1-\alpha)}.
\end{aligned}$$

□

3. The main bilinear estimate

To complete the proof we will show that μ_b is a $\mathcal{D}$-Carleson measure by verifying Stegenga's condition (2-3); that is, we will show that for any finite collection of disjoint arcs $\{I_j\}_{j=1}^N$ in the circle $\mathbb{T}$ we have

$$\mu_b\Big(\bigcup_{j=1}^N T(I_j)\Big) \le C\,\mathrm{Cap}_{\mathbb{D}}\Big(\bigcup_{j=1}^N I_j\Big).$$

In fact we will see that it suffices to verify this for the sets $G = \bigcup_{j=1}^N I_j$ described in (2-23) that are almost extremal for (2-22). We will prove the inequality

$$\mu_b(V_G) \le C\|T_b\|^2 \operatorname{Cap}_{\mathbb{D}}(G). \tag{3-1}$$

Once we have this, Corollary 2.12 yields

$$M = \frac{\int_{\mathbb{T}} \mu_b(T_\theta(G))\, d\theta}{\int_{\mathbb{T}} \operatorname{Cap}_\theta(G)\, d\theta} \le \frac{\mu_b(V_G)}{\int_{\mathbb{T}} \operatorname{Cap}_\theta(G)\, d\theta} \le C\|T_b\|^2.$$

By Corollary 2.13 $\|\mu_b\|^2_{CM(\mathcal{D})} \approx M$ which then completes the proof of Theorem 1.1.

We now turn to (3-1). Let $\frac{1}{2} < \beta < \beta_1 < \gamma < \alpha < 1$, with additional constraints to be added later. Suppose G (2-23) with $\varepsilon > 0$ to be chosen. We define in succession the following regions in the disk:

$$V_G = T_{\mathcal{T}}(G), \qquad V_G^\alpha = G_{\mathbb{D}}^\alpha, \qquad V_G^\gamma = \widehat{(V_G^\alpha)_{\mathcal{T}}^{\gamma/\alpha}}, \qquad V_G^\beta = (V_G^\gamma)_{\mathbb{D}}^{\beta/\gamma}.$$

Thus V_G is the $\mathcal{T}$-tent associated with G, V_G^α is a disk blowup of G, V_G^γ is a $\mathcal{T}$-capacitary blowup of V_G^α, and V_G^β is a disk blowup of V_G^γ. Using the natural bijections described earlier, we write

$$V_G = \{w_k\}_k, \qquad V_G^\alpha = \{w_k^\alpha\}_k, \qquad V_G^\gamma = \{w_k^\gamma\}_k, \qquad V_G^\beta = \{w_k^\beta\}_k, \tag{3-2}$$

with $w_k, w_k^\alpha, w_k^\gamma, w_k^\beta \in \mathcal{T}$. Following earlier notation we write $E = V_G^\alpha$ and $F = V_G^\gamma$.

We proceed by estimating $T_b(f, g)$ for well chosen f and g in $\mathcal{D}$. Let Φ be as in (2-12); we then have the estimates in Proposition 2.9 and Corollary 2.10. Set $g = \Phi^2$; then g is approximately equal to χ_{V_G}. The function f will be, approximately, $b'\chi_{V_G}$;

$$f(z) = \Gamma_s\Big(\frac{1}{(1+s)\overline{\zeta}}\chi_{V_G} b'(\zeta)\Big)(z) = \int_{V_G} \frac{b'(\zeta)(1-|\zeta|^2)^s}{(1-\overline{\zeta}z)^{1+s}} \frac{dA}{(1+s)\overline{\zeta}}. \tag{3-3}$$

We now analyze $T_b(f, g)$. From (3-3) and (2-15) we have

$$f'(z) = \int_{V_G} \frac{b'(\zeta)(1-|\zeta|^2)^s}{(1-\overline{\zeta}z)^{2+s}}\, dA = b'(z) - \int_{\mathbb{D}\setminus V_G} \frac{b'(\zeta)(1-|\zeta|^2)^s}{(1-\overline{\zeta}z)^{2+s}}\, dA = b'(z) + \Lambda b'(z),$$

where the last term is defined by

$$\Lambda b'(z) = -\int_{\mathbb{D}\setminus V_G} \frac{b'(\zeta)(1-|\zeta|)^s}{(1-\overline{\zeta}z)^{2+s}}\, dA. \tag{3-4}$$

We have

$$T_b(f, g) = (f\Phi^2\bar{b})(0) + \int_{\mathbb{D}} \big\{f'(z)\Phi(z) + 2f(z)\Phi'(z)\big\}\Phi(z)\overline{b'(z)}\, dA =: (1) + (2) + (3) + (4), \tag{3-5}$$

with

$$(1) = (f\Phi^2\bar{b})(0), \qquad (3) = 2\int_{\mathbb{D}} \Phi(z)\Phi'(z) f(z)\,\overline{b'(z)}\, dA,$$

$$(2) = \int_{\mathbb{D}} |b'(z)|^2\Phi(z)^2\, dA, \qquad (4) = \int_{\mathbb{D}} \Lambda b'(z)\,\overline{b'(z)}\,\Phi(z)^2\, dA.$$

Now we write

$$(2)=\int_{\mathbb{D}}|b'(z)|^2\Phi(z)^2dA=\left\{\int_{V_G}+\int_{V_G^\beta\setminus V_G}+\int_{\mathbb{D}\setminus V_G^\beta}\right\}|b'(z)|^2\Phi(z)^2\,dA=:(2_A)+(2_B)+(2_C). \quad (3\text{-}6)$$

The main term is (2_A). By (2-17) and (2-1) it satisfies

$$(2_A)=\mu_b(V_G)+\int_{V_G}|b'(z)|^2(\Phi(z)^2-1)\,dA=\mu_b(V_G)+O\bigl(\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)\bigr), \quad (3\text{-}7)$$

Rearranging this and using (3-5) and (3-6) we find

$$\mu_b(V_G)\le C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)+|T_b(f,g)|+|(1)|+(2_B)+(2_C)+|(3)|+|(4)|. \quad (3\text{-}8)$$

Using the boundedness of T_b and Corollary 2.10 we have

$$|T_b(f,g)|=|T_b(f,\Phi^2)|=|T_b(f\Phi,\Phi)|\le\|T_b\|\|f\Phi\|_{\mathfrak{D}}\|\Phi\|_{\mathfrak{D}}\le C\|T_b\|\|f\Phi\|_{\mathfrak{D}}\sqrt{\operatorname{Cap}_{\mathcal{T}}(E,F)}. \quad (3\text{-}9)$$

For (1) we use the elementary estimate

$$|(1)|\le C\|b\|_{\mathfrak{D}}^2\operatorname{Cap}_{\mathcal{T}}(E,F)\le C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F).$$

For (2_B) we use (2-24) to obtain

$$(2_B)\le C\mu_b(V_G^\beta\setminus V_G)\le C\varepsilon\mu_b(V_G). \quad (3\text{-}10)$$

Using (2-17) once more, we see that (2_C) satisfies

$$(2_C)\le\int_{\mathbb{D}\setminus V_G^\beta}|b'(z)|^2(C_{\alpha,\beta,\rho}\operatorname{Cap}_{\mathcal{T}}(E,F))^2\,dA\le C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F). \quad (3\text{-}11)$$

Putting these estimates into (3-8) we obtain

$$\mu_b(V_G)\le C(\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)+\|T_b\|\|f\Phi\|_{\mathfrak{D}}\sqrt{\operatorname{Cap}_{\mathcal{T}}(E,F)}+|(3)|+|(4)|). \quad (3\text{-}12)$$

For small positive ε we estimate (3) using Cauchy–Schwarz as follows:

$$\begin{aligned}|(3)|&\le 2\int_{\mathbb{D}}|\Phi(z)b'(z)||\Phi'(z)f(z)|\,dA\\&\le\varepsilon\int_{\mathbb{D}}|\Phi(z)b'(z)|^2dA+\frac{C}{\varepsilon}\int_{\mathbb{D}}|\Phi'(z)f(z)|^2\,dA\\&=(3_A)+(3_B).\end{aligned}$$

Using the decomposition and the argument surrounding term (2) we obtain

$$(3_A)\le\varepsilon\{\int_{V_G}+\int_{V_G^\beta\setminus V_G}+\int_{\mathbb{D}\setminus V_G^\beta}\}|\Phi(z)b'(z)|^2dA\le C\varepsilon(\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)). \quad (3\text{-}13)$$

To estimate term (3_B) we use

$$\begin{aligned}|f(z)| &\le \left|\Gamma_s\Big(\frac{1}{(1+s)\overline{\zeta}}\chi_{V_G}b'(\zeta)\Big)(z)\right| \\ &\le \int_{V_G}\frac{(1-|\zeta|^2)^s}{|1-\overline{\zeta}z|^{1+s}}|b'(\zeta)|\,dA \\ &\approx \sum_{\gamma\in\mathcal{T}\cap V_G}\frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}}\int_{B_\gamma}|b'(\zeta)|(1-|\zeta|^2)\,d\lambda(\zeta) \\ &= \sum_{\gamma\in\mathcal{T}\cap V_G}\frac{(1-|\gamma|^2)^{1+s}}{|1-\overline{\gamma}z|^{1+s}}b(\gamma),\end{aligned}$$

where

$$\sum_{\gamma\in\mathcal{T}\cap V_G} b(\gamma)^2 \approx \sum_{\gamma\in\mathcal{T}\cap V_G}\int_{B_\gamma}|b'(\zeta)|^2(1-|\zeta|^2)^2\,d\lambda(\zeta) = \int_{V_G}|b'(\zeta)|^2\,dA.$$

We now use the separation of $\mathbb{D}\setminus V_G^\alpha$ and V_G. The facts that $\mathcal{A}=supp(h)\subset\mathbb{D}\setminus V_G^\alpha$ and $\mathcal{B}=\mathcal{T}\cap V_G\subset V_G$, together with Lemma 2.2, ensure that (2-29) is satisfied and hence we can use Lemma 2.17 and the representation of Φ in (2-12) to continue with

$$(3_B) = \int_{\mathbb{D}}|\Phi'(z)f(z)|^2 dA \le C\Big(\sum_{\kappa\in\mathcal{A}}h(\kappa)^2\Big)\Big(\sum_{\gamma\in\mathcal{B}}b(\gamma)^2\Big).$$

We also have from (2-1) and Corollary 2.10 that

$$\Big(\sum_{\kappa\in\mathcal{A}}h(\kappa)^2\Big)\Big(\sum_{\gamma\in\mathcal{B}}b(\gamma)^2\Big) \le C\,\mathrm{Cap}_{\mathcal{T}}(E,F)\|T_b\|^2.$$

Altogether we then have

$$(3_B) \le C\,\mathrm{Cap}_{\mathcal{T}}(E,F)\|T_b\|^2, \tag{3-14}$$

and thus also

$$|(3)| \le \varepsilon\mu_b(V_G) + C\|T_b\|^2\,\mathrm{Cap}_{\mathcal{T}}(E,F). \tag{3-15}$$

We begin our estimate of term (4) by

$$|(4)| = \left|\int_{\mathbb{D}}\Lambda b'(z)\,\overline{b'(z)}\,\Phi(z)^2\,dA\right| \le \sqrt{\int_{\mathbb{D}}|b'(z)\Phi(z)|^2\,dA}\sqrt{\int_{\mathbb{D}}|\Lambda b'(z)\Phi(z)|^2\,dA}, \tag{3-16}$$

where the first factor is $\sqrt{(3_A)/\varepsilon}$. We claim the following estimate for the second factor,

$$\sqrt{(4_A)} := \|\Phi\Lambda b'\|_{L^2(\mathbb{D})}:$$

Lemma 3.1. $$(4_A) = \int_{\mathbb{D}}|\Phi(z)\Lambda b'(z)|^2 dA \le C\mu_b(V_G^\beta\setminus V_G) + C\|T_b\|^2\,\mathrm{Cap}_{\mathcal{T}}(E,F).$$

Proof. From (3-4) we obtain

$$
\begin{aligned}
(4_A) &= \int_{\mathbb{D}} |\Phi(z)|^2 \left| \left\{ \int_{V_G^\beta \setminus V_G} + \int_{\mathbb{D}\setminus V_G^\beta} \right\} \frac{b'(\zeta)(1-|\zeta|)^s}{(1-\overline{\zeta}z)^{2+s}} dA \right|^2 dA \\
&\le C \int_{\mathbb{D}} |\Phi(z)|^2 \left(\int_{V_G^\beta \setminus V_G} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}z|^{2+s}}\, dA \right)^2 dA + C \int_{\mathbb{D}} |\Phi(z)|^2 \left| \int_{\mathbb{D}\setminus V_G^\beta} \frac{b'(\zeta)(1-|\zeta|)^s}{(1-\overline{\zeta}z)^{2+s}} dA \right|^2 dA \\
&=: (4_{AA}) + (4_{AB}).
\end{aligned}
$$

Corollary 2.16 shows that

$$
|(4_{AA})| \le \int_{\mathbb{D}} \left(\int_{V_G^\beta \setminus V_G} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}z|^{2+s}}\, dA \right)^2 dA \le C \int_{V_G^\beta \setminus V_G} |b'(\zeta)|^2\, dA = C\mu_b(V_G^\beta \setminus V_G).
$$

We write the second integral as

$$
(4_{AB}) = \left\{ \int_{V_G^\gamma} + \int_{\mathbb{D}\setminus V_G^\gamma} \right\} |\Phi(z)|^2 \left| \int_{\mathbb{D}\setminus V_G^\beta} \frac{b'(\zeta)(1-|\zeta|)^s}{(1-\overline{\zeta}z)^{2+s}}\, dA \right|^2 dA =: (4_{ABA}) + (4_{ABB}),
$$

where, by Corollary 2.16 again,

$$
(4_{ABB}) \le C\, \mathrm{Cap}_{\mathcal{T}}(E, F)^2 \int_{\mathbb{D}} |b'(\zeta)|^2 dA \le C\|T_b\|^2\, \mathrm{Cap}_{\mathcal{T}}(E, F)^2 \le C\|T_b\|^2\, \mathrm{Cap}_{\mathcal{T}}(E, F),
$$

where the final estimate, $\mathrm{Cap}_{\mathcal{T}}(E, F) \le C$, follows from our assumption that $\mathrm{Cap}_{\mathbb{D}}(G)$ is small. Indeed, (2-4) then shows that $\mathrm{Cap}_{\mathbb{D}}(E)$ is small and hence $\mathrm{Cap}_{\mathcal{T}}(E)$ is small as well by Corollary 2.12. Lemma 2.7 then shows that $\mathrm{Cap}_{\mathcal{T}}(E, F)$ is small, and in particular bounded.

Finally, with $\beta < \beta_1 < \gamma < \alpha < 1$, Corollary 2.16 shows that the term (4_{ABA}) satisfies the following estimate. Recall that $V_G^\gamma = \bigcup J_k^\gamma$ and $w_j^\gamma = z(J_k^\gamma)$. We set $A_\ell = \{k : J_k^\gamma \subset J_\ell^{\beta_1}\}$ and define $\ell(k)$ by the condition $k \in A_{\ell(k)}$. From Lemma 2.2 we have $\mathrm{sidelength}(J_k^\gamma) \le \mathrm{sidelength}(J_\ell^{\beta_1})^{1/\rho}$, with $\rho = \beta_1/\gamma$. Hence

$$
\begin{aligned}
(4_{ABA}) &\le C \int_{V_G^\gamma} \left(\int_{\mathbb{D}\setminus V_G^\beta} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}z|^{2+s}} d\zeta \right)^2 dA \\
&\approx C \sum_k \int_{J_k^\gamma} |J_k^\gamma| \left(\int_{\mathbb{D}\setminus V_G^\beta} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}w_k^\gamma|^{2+s}}\, d\zeta \right)^2 dA \\
&= C \sum_k \frac{|J_k^\gamma|}{|J_{\ell(k)}^{\beta_1}|} |J_{\ell(k)}^{\beta_1}| \left(\int_{\mathbb{D}\setminus V_G^\beta} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}w_k^\gamma|^{2+s}}\, d\zeta \right)^2 \\
&\approx C \sum_\ell \frac{\sum_{k\in A_\ell} |J_k^\gamma|}{|J_\ell^{\beta_1}|} \int_{J_\ell^{\beta_1}} \left(\int_{\mathbb{D}\setminus V_G^\beta} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}z|^{2+s}}\, d\zeta \right)^2 dA \\
&\le C|V_G^{\beta_1}|^{\varepsilon(\gamma-\beta_1)} \int_{V_G^{\beta_1}} \left(\int_{\mathbb{D}\setminus V_G^\beta} \frac{|b'(\zeta)|(1-|\zeta|)^s}{|1-\overline{\zeta}z|^{2+s}}\, d\zeta \right)^2 dA \\
&\le C|V_G^{\beta_1}|^{\varepsilon(\gamma-\beta_1)} \|b\|_{\mathcal{D}}^2 \le C\|T_b\|^2\, \mathrm{Cap}_{\mathcal{T}}(E, F).
\end{aligned}
$$

We continue from ((3-16)). We know that $|(4)| \leq \sqrt{(3_A)/\varepsilon}\sqrt{(4_A)}$ We estimate (3_A) using (3-13) and (4_A) using Lemma 3.1. After that we continue by using (2-24) so

$$\begin{aligned} |(4)| &\leq \sqrt{C\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)} \times \sqrt{C\mu_b(V_G^\beta \setminus V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)} \\ &\leq \sqrt{C\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)} \times \sqrt{\varepsilon\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)} \\ &\leq \sqrt{\varepsilon}\mu_b(V_G)+C\sqrt{\mu_b(V_G)}\sqrt{\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)}+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F). \end{aligned} \tag{3-17}$$

Now, recalling that $f' = b' + \Lambda b'$,

$$\begin{aligned} \|\Phi f\|_{\mathcal{D}}^2 &\leq C\int |\Phi'(z)f(z)|^2 dA + C\int |\Phi(z)(b'(z)+\Lambda b'(z))|^2 dA \\ &\leq C(3_B) + C\frac{1}{\varepsilon}(3_A) + C(4_A). \\ &\leq C\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F), \end{aligned} \tag{3-18}$$

by Lemma 3.1 and the estimates (3-13) and (3-14) for (3_A) and (3_B). □

Using Proposition 2.14 and the estimates (3-15), (3-17), and (3-18) in (3-12) we obtain

$$\begin{aligned} \mu_b(V_G) &\leq \sqrt{\varepsilon}\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)+C\sqrt{\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F)}\sqrt{\mu_b(V_G)} \\ &\leq \sqrt{\varepsilon}\mu_b(V_G)+C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F). \end{aligned}$$

We absorb the first term into the right side. Now using Lemma 2.7, Lemma 2.8 again, and Corollary 2.12 we obtain

$$\operatorname{Cap}_{\mathcal{T}}(E,F) \leq C\operatorname{Cap}_{\mathbb{D}} G.$$

Finally we have

$$\mu_b(V_G) \leq C\|T_b\|^2\operatorname{Cap}_{\mathcal{T}}(E,F) \leq C\|T_b\|^2\operatorname{Cap}_{\mathbb{D}} G,$$

which is (3-1).

Appendix: Tree extremals

Let E be a stopping time in $\mathcal{T}$. Recall that

$$\operatorname{Cap}_{\mathcal{T}}(E) = \inf\{\|h\|_{\ell^2}^2 : Ih \geq 1 \text{ on } E\}. \tag{A-1}$$

We call functions which can be used in computing the infimum *admissible*.

Much of the following proposition as well as Proposition 2.3 could be extracted from general capacity theory such as presented in, for instance, [Adams and Hedberg 1996]. Statement (3) is the discrete analog of the fact that continuous capacity can be interpreted as the derivative at infinity of a Green function.

Proposition A.2. *Suppose $E \subset \mathcal{T}$ is given.*

(1) *There is a function h such that the infimum in the definition of* $\operatorname{Cap}_{\mathcal{T}}(E)$ *is achieved.*

(2) *If $x \notin E$,*

$$h(x) = h(x_+) + h(x_-). \tag{A-2}$$

(3) $h(o) = \|h\|_{\ell^2}^2$.

(4) *h is strictly positive on* $\mathcal{G}(o, E)$ *and zero elsewhere.*

(5) $Ih|_E = 1$.

Proof. Consider first the case when E is a finite subset of $\mathcal{T}$. Multiplying an admissible function by the characteristic function of $\mathcal{G}(o, E)$ leaves it admissible and reduces the ℓ^2 norm. Hence we need only consider functions supported on the finite set of vertices in $\mathcal{G}(o, E)$. In that context it is easy to see that an extremal exists, call it h. Now consider (2). Suppose $x \in \mathcal{T} \setminus E$ and consider the competing function h^* which takes the same values as h except possible at x, x_+, and x_- and whose values at those points are determined by

(i) $h^*(x) + h^*(x_+) = h(x) + h(x_+)$ and $h^*(x) + h^*(x_-) = h(x) + h(x_-)$,

(ii) $h^*(x)^2 + h^*(x_+)^2 + h^*(x_-)^2$ is minimal subject to (i).

Then h^* is admissible, $\|h^*\|_{\ell^2}^2 \leq \|h\|_{\ell^2}^2$, and, doing the calculus problem, h^* satisfies (A-2). Hence h must satisfy (A-2).

If $h(x) < 0$ at some point, replacing its value by zero leaves the function admissible while reducing the ℓ^2 norm, hence $h \geq 0$. To complete the proof of (4) we must show that we cannot have an $x \in \mathcal{G}(o, E)$ at which $h(x) = 0$. Suppose we had such a point. By (A-2) and the fact that $h \geq 0$, we have $h \equiv 0$ on $S_{\mathcal{T}}(x)$. Hence by admissibility $Ih(x^{-1}) \geq 1$. Let $y \neq x$ be the point such that $x^{-1} = y^{-1}$. If $h(y) > 0$ then setting $h(y) = 0$ we would decrease the ℓ^2 norm while keeping the function admissible. Thus $h(y) = 0$ and, by (A-2), $h(x^{-1}) = 0$. Continuing in this way we find that $h \equiv 0$ an the geodesic from o to some $e \in E$, an impossibility for an admissible function. Item (5) is a consequence of this. If $Ih(e) > 1$ for some $e \in E$ and $h(e) > 0$ then we could decrease $h(e)$ slightly, reducing the norm of h and still have h admissible thus contradicting the supposition that h is extremal.

It remains to show (3) and we do that by induction on the size of E. If $E = \{e\}$ is a single point having distance $d - 1 \geq 0$ from o then the extremal is $h \equiv 1/d$ on $[o, e]$ and $\|h\|_{\ell^2}^2 = d(1/d)^2 = h(o)$. Given E with more than one point, let z be the uniquely determined branching point in $\mathcal{G}(o, E)$ having the least distance from the root. Consider the rooted trees $\mathcal{T}_\pm = S(z_\pm)$ with roots $z_\pm$. Set $E_\pm = E \cap \mathcal{T}_\pm$ and let $h_\pm$ be the extremal functions for the computation of $\mathrm{Cap}_{\mathcal{T}_\pm}(E_\pm)$. By induction, we have $\|h_\pm\|_{\ell^2}^2 = h_\pm(z_\pm)$. From properties (1)-(5) satisfied by the extremal functions h, h_+ and h it is easy to see that

$$h(x) = \begin{cases} (1 - Ih(z))h_\pm(x) & \text{if } x \in \mathcal{G}(z_\pm), \\ h(o) & \text{if } x \in [o, z], \\ 0 & \text{otherwise.} \end{cases}$$

In particular, $Ih(z) = dh(o)$ if there are d points in $[o, z]$ such that

$$h(o) = h(z) = h(z_+) + h(z_-) = \frac{h_+(z_+) + h_-(z_-)}{1 - Ih(z)} = \frac{h_+(z_+) + h_-(z_-)}{1 - dh(o)}. \tag{A-3}$$

Rescaling and using the induction hypothesis,

$$\begin{aligned}\|h\|_{\ell^2}^2 &= (\|h_+\|_{\ell^2}^2 + \|h_-\|_{\ell^2}^2)(1-dh(o))^2 + dh(o)^2 = (h_+(z_+) + h_-(z_-))(1-dh(o))^2 + dh(o)^2 \\ &= \frac{h(z_+)+h(z_-)}{1-dh(o)}(1-dh(o))^2 + dh(o)^2 = \frac{h(z)}{1-dh(o)}(1-dh(o))^2 + dh(o)^2 \\ &= \frac{h(o)}{1-dh(o)}(1-dh(o))^2 + dh(o)^2 = h(o).\end{aligned}$$

We note in passing that, by (3), formula (A-3) gives a recursive formula for computing tree capacities.

Suppose now that E is infinite. Select a sequence of finite sets $E_n = \{e_1, \ldots, e_n\}$ such that $E_n \nearrow E$. Let h_n be the corresponding extremal functions and $H_n = Ih_n$. We claim that the sequence H_n increases, in the sense specified below. Let $K = H_n - H_{n-1} = I(h_n - h_{n-1}) = Ik_n$. By (A-2), the function K satisfies the mean value property on $\mathcal{G}(o, E_n) \setminus (\{o\} \cup E_n)$:

$$K(x) = \tfrac{1}{3}[K(x_+) + K(x_-) + K(x^{-1})], \quad \text{if } x \in \mathcal{G}(o, E_n) \setminus (\{o\} \cup E_n).$$

Moreover, K vanishes on $\{o\} \cup E_{n-1}$ and it is positive at e_n, since $H_{n-1}(e_n) \leq 1 = H_n(e_n)$, by (3) and (4). By the maximum principle (an easy consequence of the mean value property), $K_n \geq 0$ in $\mathcal{G}(o, E_n)$. Hence, the limit $Ih = H = \lim_n H_n$ exists in $\mathcal{G}(o, E)$ and it is finite because each H_n is bounded above by 1. Since $h(x) = H(x) - H(x^{-1}) = \lim h_n(x)$, h is admissible for E and it satisfies (3), (4) and (5).

Also, $h_n \to h$ as $n \to \infty$, pointwise, and $\|h_n\|_{\ell^2}^2 = h_n(o) \to h(o)$, by dominated convergence, hence,

$$h(o) = \lim_{n\to\infty} \|h_n\|_{\ell^2}^2 = \|h\|_{\ell^2}^2,$$

which is (3) for h.

It remains to prove that h is extremal. Suppose k is another admissible function for E, and let k_n be its restriction to $\mathcal{G}(o, E_n)$, which is clearly admissible for E_n. By the extremal character of the functions h_n, we have

$$\|k\|_{\ell^2}^2 = \lim_{n\to\infty} \|k_n\|_{\ell^2}^2 \leq \lim_{n\to\infty} \|h_n\|_{\ell^2}^2 = \lim_{n\to\infty} h_n(o) = h(o) = \|h\|_{\ell^2}^2.$$

Hence, h is extremal among the admissible functions for E. □

Proof of Proposition 2.3. Consider each $e \in E$ as the root of the tree $\mathcal{T}_e = S(e)$. Set $F_e = F \cap S(e)$ and let h_e be the extremal function (from the previous proposition) for computing $\mathrm{Cap}_{\mathcal{T}_e}(F_e)$. Using the previous proposition it is straightforward to check that $h = \sum h_e$ is the required extremal function and has the required properties. □

Acknowledgments

This work was begun while Rochberg, Sawyer and Wick were visiting the Fields Institute. We thank the institute for its hospitality and its excellent working conditions. While there we had interesting discussions with Michael Lacey. He suggested an approach to the theorem that was quite different from what we had been envisioning and his comments helped shape our approach. In particular they led us to suspect that a good estimate for the collar error term, Proposition 2.14, could play a crucial role. We thank him for his involvement.

References

[Adams and Hedberg 1996] D. R. Adams and L. I. Hedberg, *Function spaces and potential theory*, Grundlehren der Math. Wissenschaften **314**, Springer, Berlin, 1996. MR 97j:46024 Zbl 0834.46021

[Arcozzi et al. 2007] N. Arcozzi, R. Rochberg, and E. Sawyer, "Onto interpolating sequences for the Dirichlet space", preprint, 2007, Available at http://amsacta.cib.unibo.it/2480/01/ontoamsacta.pdf.

[Arcozzi et al. 2008] N. Arcozzi, R. Rochberg, E. Sawyer, and B. Wick, "Function Spaces Related to the Dirichlet Space", preprint, 2008. arXiv 0812.3422

[Bishop 1994] C. J. Bishop, "Interpolating sequences for the Dirichlet space and its multipliers", preprint, 1994, Available at http://www.math.sunysb.edu/~bishop/papers/mult.pdf.

[Böe 2002] B. Böe, "Interpolating sequences for Besov spaces", *J. Funct. Anal.* **192**:2 (2002), 319–341. MR 2003h:46044 Zbl 1018.46018

[Coifman and Murai 1988] R. R. Coifman and T. Murai, "Commutators on the potential-theoretic energy spaces", *Tohoku Math. J.* (2) **40**:3 (1988), 397–407. MR 89i:42029 Zbl 0679.42009

[Coifman et al. 1976] R. R. Coifman, R. Rochberg, and G. Weiss, "Factorization theorems for Hardy spaces in several variables", *Ann. of Math.* (2) **103**:3 (1976), 611–635. MR 54 #843 Zbl 0326.32011

[Ferguson and Lacey 2002] S. H. Ferguson and M. T. Lacey, "A characterization of product BMO by commutators", *Acta Math.* **189**:2 (2002), 143–160. MR 2004e:42026 Zbl 1039.47022

[Lacey and Terwilleger 2009] M. Lacey and E. Terwilleger, "Hankel operators in several complex variables and product BMO", *Houston J. Math.* **35**:1 (2009), 159–183. MR 2491875 Zbl 1163.47024

[Maz'ya and Verbitsky 2002] V. G. Maz'ya and I. E. Verbitsky, "The Schrödinger operator on the energy space: boundedness and compactness criteria", *Acta Math.* **188**:2 (2002), 263–302. MR 2004b:35050 Zbl 1013.35021

[Nehari 1957] Z. Nehari, "On bounded bilinear forms", *Ann. of Math.* (2) **65** (1957), 153–162. MR 18,633f Zbl 0077.10605

[Peller 2003] V. V. Peller, *Hankel operators and their applications*, Springer, New York, 2003. MR 2004e:47040 Zbl 1030.47002

[Rochberg and Wu 1993] R. Rochberg and Z. J. Wu, "A new characterization of Dirichlet type spaces and applications", *Illinois J. Math.* **37**:1 (1993), 101–122. MR 93j:30039 Zbl 0789.31009

[Stegenga 1980] D. A. Stegenga, "Multipliers of the Dirichlet space", *Illinois J. Math.* **24**:1 (1980), 113–139. MR 81a:30027 Zbl 0432.30016

[Tolokonnikov 1991] V. A. Tolokonnikov, "Carleson's Blaschke products and Douglas algebras", *Algebra i Analiz* **3**:4 (1991), 186–197. In Russian; translated in *St. Petersburg Math. J.* **3**:4 (1992), 881–892. MR 93c:46098 Zbl 0763.46051

[Zhu 2005] K. Zhu, *Spaces of holomorphic functions in the unit ball*, Graduate Texts in Mathematics **226**, Springer, New York, 2005. MR 2006d:46035 Zbl 1067.32005

Received 8 Jul 2009. Revised 99 2999. Accepted 5 Aug 2009.

NICOLA ARCOZZI: `arcozzi@dm.unibo.it`
Dipartimento di Matematica, Università di Bologna, 40127 Bologna, Italy

RICHARD ROCHBERG: `rr@math.wustl.edu`
Department of Mathematics, Washington University, St. Louis, MO 63130, United States

ERIC SAWYER: `erictsawyer@gmail.com`
Department of Mathematics and Statistics, McMaster University, Hamilton, ON L8S 4K1, Canada

BRETT D. WICK: `brett.wick@gmail.com`
Department of Mathematics, LeConte College, 1523 Greene Street, University of South Carolina, Columbia, SC 29208, United States

ANALYSIS AND PDE
Vol. 3, No. 1, 2010

POISSON STATISTICS FOR EIGENVALUES OF CONTINUUM RANDOM SCHRÖDINGER OPERATORS

JEAN-MICHEL COMBES, FRANÇOIS GERMINET AND ABEL KLEIN

We show absence of energy levels repulsion for the eigenvalues of random Schrödinger operators in the continuum. We prove that, in the localization region at the bottom of the spectrum, the properly rescaled eigenvalues of a continuum Anderson Hamiltonian are distributed as a Poisson point process with intensity measure given by the density of states. In addition, we prove that in this localization region the eigenvalues are simple.

These results rely on a Minami estimate for continuum Anderson Hamiltonians. We also give a simple, transparent proof of Minami's estimate for the (discrete) Anderson model.

1. Introduction

Local fluctuations of eigenvalues of random operators are believed to distinguish between localized and delocalized regimes, indicating an Anderson metal-insulator transition. Exponential decay of eigenfunctions implies that disjoint regions of space are uncorrelated and create almost independent eigenvalues, leading to the absence of energy levels repulsion, which is mathematically translated in terms of a Poisson point process. On the other hand, extended states imply that distant regions have mutual influence, and thus create some repulsion between energy levels.

Local fluctuations of eigenvalues have been studied within the context of random matrix theory, in particular Wigner matrices and GUE matrices [Bellissard 2004; Disertori et al. 2002; Erdős et al. 2009b; 2009a; Johansson 1998; 2001; Schenker and Schulz-Baldes 2007]. It is challenging to understand random hermitian band matrices from the perspective of their eigenvalues fluctuations, by proving a transition between Poisson statistics and a semi-circle law for the density of states (a signature of energy levels repulsion), and relate this to the (discrete) Anderson model [Bellissard 2004; Disertori et al. 2002]. CMV matrices are another class of random matrices for which Poisson statistics and a transition to energy levels repulsion have been proved been proved [Killip and Stoiciu 2009; Stoiciu 2006; 2007].

For random Schrödinger operators, Poisson statistics for eigenvalues were first proved by Molchanov [1980/81] for the same one-dimensional continuum random Schrödinger operator for which Anderson localization was first rigorously established [Gol'dsheĭd et al. 1977]. Molchanov's proof was based on a detailed analysis of localization in finite intervals for this particular random Schrödinger operator [Molchanov 1978].

Poisson statistics for eigenvalues of the Anderson model was established in [Minami 1996]. The Anderson model, a random Schrödinger operator on $\ell^2(\mathbb{Z}^d)$, is the discrete analogue of the Anderson

MSC2000: primary 82B44; secondary 47B80, 60H25.
Keywords: Anderson localization, Poisson statistics of eigenvalues, Minami estimate, level statistics.
Klein was supported in part by NSF Grant DMS-0457474.

Hamiltonian. A crucial ingredient in Minami's proof is an estimate of the probability of two or more eigenvalues in an interval. The key step in the proof of this estimate, namely [Minami 1996, Lemma 2], estimates the average of a determinant whose entries are matrix elements of the imaginary part of the resolvent. The more recent proofs of Minami's estimate by Bellissard et al. [2007] and Graf and Vaghi [2007] are variants of Minami's. Since those arguments do not seem to extend to the continuum, a Minami-type estimate and Poisson statistics for the eigenvalues have until now been challenging questions for continuum Anderson Hamiltonians.

Here we introduce a fundamentally new approach to Minami's estimate. Unlike the previous approach, ours relies on averaging spectral projections, a technique that does extend to the continuum. Combined with a property of rank-one perturbations, it provides a simple and transparent proof of Minami's estimate for the Anderson model, valid for single-site probability distributions with compact support and no atoms, which is presented here as an illustration of the method. On the continuum, our proof of Minami's estimate circumvents the unavailability of that rank-one property by averaging the spectral shift function, using refined bounds on the density of states not previously available.

Once we have Minami's estimate in the continuum, we prove Poisson statistics for eigenvalues of the Anderson Hamiltonian. We start by approximating the point process defined by the rescaled eigenvalues by superpositions of independent point processes, as in [Molchanov 1980/81; Minami 1996]. But our proof that these superpositions converge weakly to the desired Poisson point process differs from Minami's for the Anderson model, since his way of identifying the intensity measure of the Poisson process, which relies on complex analysis, is not readily applicable in the continuum. We identify this intensity measure using methods of real analysis.

Klein and Molchanov [2006] showed that Minami's estimate implies simplicity of eigenvalues for the Anderson model, a result previously obtained by Simon [1994] by different methods. Their arguments can also be applied in the continuum, so we also obtain simplicity of eigenvalues in the continuum. Previous results [Combes and Hislop 1994; Germinet and Klein 2006] proved only finite multiplicity of the eigenvalues in the localization region.

2. Main results

To state our results we introduce the following notation. We write

$$\Lambda_L(x) := x + \left[-\tfrac{L}{2}, \tfrac{L}{2}\right[^d \tag{2-1}$$

for the (half-open, half-closed) box of side $L > 0$ centered at $x \in \mathbb{R}^d$. By Λ_L we denote a box $\Lambda_L(x)$ for some $x \in \mathbb{R}^d$. Given a box $\Lambda = \Lambda_L(x)$, we set $\widetilde{\Lambda} = \Lambda \cap \mathbb{Z}^d$. If B is a set, we write χ_B for its characteristic function. We set $\chi_x^{(L)} := \chi_{\Lambda_L(x)}$. The Lebesgue measure of a Borel set $B \subset \mathbb{R}$ will be denoted by $|B|$. If $r > 0$, we denote by $[r]$ the largest integer less than equal to r, and by $[\![r]\!]$ the smallest integer bigger than r. By a constant we will always mean a finite constant. Constants such as $C_{a,b,\ldots}$ will be finite and depending only on the parameters or quantities $a, b, \ldots$; they will be independent of other parameters or quantities in the equation.

We consider random Schrödinger operators on $\mathrm{L}^2(\mathbb{R}^d)$ of the type

$$H_\omega := -\Delta + V_{\mathrm{per}} + V_\omega, \tag{2-2}$$

where Δ is the d-dimensional Laplacian operator, V_{per} is a bounded $\mathbb{Z}^d$-periodic potentia,; and $V_{\boldsymbol{\omega}}$ is an Anderson-type random potential, given by

$$V_{\boldsymbol{\omega}}(x) := \sum_{j\in\mathbb{Z}^d} \omega_j\, u_j(x), \quad \text{with } u_j(x) = u(x-j), \tag{2-3}$$

where the single-site potential u is a nonnegative bounded measurable function on $\mathbb{R}^d$ with compact support, uniformly bounded away from zero in a neighborhood of the origin, and $\boldsymbol{\omega} = \{\omega_j\}_{j\in\mathbb{Z}^d}$ is a family of independent identically distributed random variables, whose common probability distribution μ is nondegenerate with a bounded density ρ with compact support.

We normalize $H_{\boldsymbol{\omega}}$ as follows. We first require $\inf \operatorname{supp} \mu = 0$, which can always be realized by changing the periodic potential V_{per}. Next we assume $\|u\|_\infty = 1$, which can achieved by rescaling μ. We then adjust V_{per} by adding a constant so $\inf \sigma(-\Delta + V_{\mathrm{per}}) = 0$, in which case $[0, E_*] \subset \sigma(-\Delta + V_{\mathrm{per}})$ for some $E_* > 0$. Thus, without loss of generality, we will assume that the random Schrödinger operator H_ω given in (2-2)–(2-3) is normalized as follows:

(I) The free Hamiltonian $H_0 := -\Delta + V_{\mathrm{per}}$ has 0 as the bottom of its spectrum:

$$\inf \sigma(H_0) = 0. \tag{2-4}$$

(II) The single-site potential u is a measurable function on $\mathbb{R}^d$ such that

$$\|u\|_\infty = 1 \quad \text{and} \quad u_- \chi_{\Lambda_{\delta_-}(0)} \le u \le \chi_{\Lambda_{\delta_+}(0)} \quad \text{with } u_-, \delta_\pm \in\,]0,\infty[; \tag{2-5}$$

we set

$$U_+ := \big\| \textstyle\sum_{j\in\mathbb{Z}^d} u_j \big\|_\infty \le \max\{1, \delta_+^d\}. \tag{2-6}$$

(III) $\boldsymbol{\omega} = \{\omega_j\}_{j\in\mathbb{Z}^d}$ is a family of independent, identically distributed random variables, whose common probability distribution μ has a density ρ such that

$$\{0, M_\rho\} \in \operatorname{ess\,supp} \rho \subset [0, M_\rho] \quad \text{with } M_\rho \in\,]0,\infty[\text{ and } \rho_+ := \|\rho\|_\infty < \infty. \tag{2-7}$$

A random Schrödinger operator H_ω on $\mathrm{L}^2(\mathbb{R}^d)$ as in (2-2)–(2-3), normalized as in (I)-(III), will be called an *Anderson Hamiltonian*. The common probability distribution μ in (III) is said to be *uniform-like* if its density ρ also satisfies $\rho_- := \operatorname{ess\,inf} \rho\chi_{[0,M_\rho]} > 0$, in which case we have

$$\rho_- \chi_{[0,M_\rho]} \le \rho \le \rho_+ \chi_{[0,M_\rho]} \quad \text{with } \rho_\pm, M_\rho \in\,]0,\infty[. \tag{2-8}$$

An Anderson Hamiltonian $H_{\boldsymbol{\omega}}$ is a $\mathbb{Z}^d$-ergodic family of random self-adjoint operators. It follows from standard results [Klein and Molchanov 2006; Carmona and Lacroix 1990; Pastur and Figotin 1992] that there exist fixed subsets Σ, Σ_{pp}, Σ_{ac} and Σ_{sc} of $\mathbb{R}$ so that the spectrum $\sigma(H_{\boldsymbol{\omega}})$ of $H_{\boldsymbol{\omega}}$, as well as its pure point, absolutely continuous, and singular continuous components, are equal to these fixed sets with probability one. With our normalization, the nonrandom spectrum Σ of an Anderson Hamitonian $H_{\boldsymbol{\omega}}$ satisfies [Kirsch and Martinelli 1982]

$$\sigma(H_0) \subset \Sigma \subset [0,\infty[, \tag{2-9}$$

so $\inf \Sigma = 0$ and $[0, E_*] \subset \Sigma$ for some $E_* = E_*(V_{\mathrm{per}}) > 0$. Note that $\Sigma = \sigma(-\Delta) = [0,\infty[$ if $V_{\mathrm{per}} = 0$.

An Anderson Hamiltonian $H_{\boldsymbol{\omega}}$ exhibits Anderson and dynamical localization at the bottom of the spectrum [Martinelli and Holden 1984; Combes and Hislop 1994; Klopp 1995; Kirsch et al. 1998; Germinet and De Bièvre 1998; Damanik and Stollmann 2001; Germinet and Klein 2001; 2003a; Aizenman et al. 2006]. More precisely, there exists an energy $E_1 > 0$ such that $[0, E_1] \subset \Xi^{\mathrm{CL}}$, where Ξ^{CL} is the region of complete localization for the random operator $H_{\boldsymbol{\omega}}$ [Germinet and Klein 2004; 2006]. (See Appendix A for a discussion of localization. Note that $\mathbb{R} \setminus \Sigma \subset \Xi^{\mathrm{CL}}$ in our definition.) Similarly, given an energy $E_1 > 0$, we have $[0, E_1] \subset \Xi^{\mathrm{CL}}$ if ρ_+ in (2-7) is sufficiently small, corresponding to a large disorder regime.

Finite volume operators will be defined for finite boxes $\Lambda = \Lambda_L(j)$, where $j \in \mathbb{Z}^d$ and $L \in 2\mathbb{N}$, $L > \delta_+$. Given such Λ, we will consider the random Schrödinger operator $H_{\boldsymbol{\omega}}^{(\Lambda)}$ on $\mathrm{L}^2(\Lambda)$ given by the restriction of the Anderson Hamiltonian $H_{\boldsymbol{\omega}}$ to Λ with periodic boundary condition. To do so, we identify Λ with a torus in the usual way by identifying opposite edges, and define finite volume operators

$$H_{\boldsymbol{\omega}}^{(\Lambda)} := H_0^{(\Lambda)} + V_{\boldsymbol{\omega}}^{(\Lambda)} \quad \text{on} \quad \mathrm{L}^2(\Lambda). \tag{2-10}$$

The finite volume free Hamiltonian $H_0^{(\Lambda)}$ is given by

$$H_0^{(\Lambda)} := -\Delta^{(\Lambda)} + V_{\mathrm{per}}^{(\Lambda)} \quad \text{on} \quad \mathrm{L}^2(\Lambda), \tag{2-11}$$

where $\Delta^{(\Lambda)}$ is the Laplacian on Λ with periodic boundary condition and $V_{\mathrm{per}}^{(\Lambda)}$ is the restriction of V_{per} to Λ. The random potential $V_{\boldsymbol{\omega}}^{(\Lambda)}$ is the restriction of $V_{\boldsymbol{\omega}^{(\Lambda)}}$ to Λ, where, given $\boldsymbol{\omega} = \{\omega_i\}_{i \in \mathbb{Z}^d}$, $\boldsymbol{\omega}^{(\Lambda)} = \{\omega_i^{(\Lambda)}\}_{i \in \mathbb{Z}^d}$ is defined as follows:

$$\omega_i^{(\Lambda)} = \begin{cases} \omega_i & \text{if } i \in \Lambda, \\ \omega_k^{(\Lambda)} & \text{if } k - i \in L\mathbb{Z}^d. \end{cases} \tag{2-12}$$

The random finite volume operator $H_{\boldsymbol{\omega}}^{(\Lambda)}$ is covariant with respect to translations in the torus. If $B \subset \mathbb{R}$ is a Borel set, we write $P_{\boldsymbol{\omega}}^{(\Lambda)}(B) := \chi_B(H_{\boldsymbol{\omega}}^{(\Lambda)})$ and $P_{\boldsymbol{\omega}}(B) := \chi_B(H_{\boldsymbol{\omega}})$ for the spectral projections.

The finite volume operator $H_{\boldsymbol{\omega}}^{(\Lambda)}$ has a compact resolvent, and hence its ($\boldsymbol{\omega}$-dependent) spectrum consists of isolated eigenvalues with finite multiplicity. It satisfies a Wegner estimate [Combes and Hislop 1994; Combes et al. 2007a]: Given $E_0 > 0$, there exists a constant K_W, independent of Λ, such that for all intervals $I \subset [0, E_0]$ we have

$$\mathbb{E}\left\{\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\right\} \le K_W\, \rho_+ |I| |\Lambda|. \tag{2-13}$$

The constant K_W given in [Combes and Hislop 1994; Combes et al. 2007a] depends on E_0, d, u, V_{per}, M_ρ, but not on ρ_+.

The integrated density of states (IDS) for $H_{\boldsymbol{\omega}}$ is given, for a.e. $E \in \mathbb{R}$, by

$$N(E) := \lim_{L \to \infty} |\Lambda_L(0)|^{-1} \operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda_L(0))}(]-\infty, E]) \quad \text{for } \mathbb{P}\text{-a.e. } \boldsymbol{\omega}, \tag{2-14}$$

in the sense that the limit exists and is the same for $\mathbb{P}$-a.e. $\boldsymbol{\omega}$ [Carmona and Lacroix 1990; Pastur and Figotin 1992]. It follows from (2-13) that the IDS $N(E)$ is locally Lipschitz, hence continuous, so (2-14) holds for all $E \in \mathbb{R}$. For all $E \in \mathbb{R}$ we have

$$N(E) = \lim_{L \to \infty} \mathbb{E}\left\{|\Lambda_L|^{-1} \operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda_L)}(]-\infty, E])\right\}. \tag{2-15}$$

$N(E)$ is a nondecreasing absolutely continuous function on $\mathbb{R}$, the cumulative distribution function of the density of states measure, given by

$$\eta(B) := \mathbb{E}\,\mathrm{tr}\big\{\chi_0^{(1)} P_\omega(B) \chi_0^{(1)}\big\} \quad \text{for a Borel set } B \subset \mathbb{R}. \tag{2-16}$$

In particular $N(E)$ is differentiable a.e. with respect to Lebesgue measure, with $n(E) := N'(E) \geq 0$ being the density of the measure η, so $n(E) > 0$ for η-a.e. E.

Given an energy $\mathscr{E} \in \Sigma$, using (2-13) we define a point process $\xi_{\mathscr{E},\omega}^{(\Lambda)}$ on the real line by the rescaled spectrum of the finite volume operator $H_\omega^{(\Lambda)}$ near $\mathscr{E}$:

$$\xi_{\mathscr{E},\omega}^{(\Lambda)}(B) := \mathrm{tr}\big\{\chi_B(|\Lambda|(H_\omega^{(\Lambda)} - \mathscr{E}))\big\} = \mathrm{tr}\big\{P_\omega^{(\Lambda)}(\mathscr{E} + |\Lambda|^{-1} B)\big\} \tag{2-17}$$

for a Borel set $B \subset \mathbb{R}$. (We refer to [Daley and Vere-Jones 1988] for definitions and results concerning random measures and point processes.)

Theorem 2.1. *Let H_ω be an Anderson Hamiltonian with $\delta_- \geq 2$ and a uniform-like distribution μ. Then there exists an energy $E_0 > 0$, such that*:

(a) *For all energies $\mathscr{E} \in \Xi^{\mathrm{CL}} \cap [0, E_0[$ such that the IDS $N(E)$ is differentiable at $\mathscr{E}$ with $n(\mathscr{E}) := N'(\mathscr{E})$ positive, the point process $\xi_{\mathscr{E},\omega}^{(\Lambda_L)}$ converges weakly, as $L \to \infty$, to the Poisson point process $\xi_{\mathscr{E}}$ on $\mathbb{R}$ with intensity measure $\nu_{\mathscr{E}}(B) := \mathbb{E}\,\xi_{\mathscr{E}}(B) = n(\mathscr{E})|B|$, that is, $\mathrm{d}\nu_{\mathscr{E}} = n(\mathscr{E})\mathrm{d}E$.*

(b) *With probability one, every eigenvalue of H_ω in $\Xi^{\mathrm{CL}} \cap [0, E_0[$ is simple.*

Similarly, given an energy $E_0 > 0$, (a) and (b) hold if the probability distribution μ in (2-8) *has a density ρ with $(\rho_+/\rho_-)\rho_+^{2^d-1}$ sufficiently small. In fact, there exists a constant $Q_{d,V_{\mathrm{per}}} > 0$, such that (a) and (b) hold whenever*

$$U_+ u_-^{-2^d} \frac{\rho_+}{\rho_-} \rho_+^{2^d-1} \gamma_d(E_0) \min\big\{1, E_0^{2^d-d-1}\big\} \max\big\{1, E_0^{2^{d+2}}\big\} \leq Q_{d,V_{\mathrm{per}}}, \tag{2-18}$$

where we have $\gamma_d(E_0) = 1$ if $d \geq 2$, and $\gamma_1(E_0) = \gamma_{1,V_{\mathrm{per}}}(E_0) \in]0, 1]$ with $\lim_{E_0 \to 0} \gamma_1(E_0) = 0$.

The next theorem gives our Minami estimate for the continuum Anderson Hamiltonian, a crucial ingredient for proving Theorem 2.1.

Theorem 2.2. *Let H_ω be an Anderson Hamiltonian with $\delta_- \geq 2$ and a uniform-like distribution μ. Then there exists a constant $Q_{d,V_{\mathrm{per}}} > 0$, such that whenever* (2-18) *holds for an energy $E_0 > 0$, we have the Minami estimate*

$$\mathbb{E}\big\{(\mathrm{tr}\, P_\omega^{(\Lambda)}(I))(\mathrm{tr}\, P_\omega^{(\Lambda)}(I) - 1)\big\} \leq K_M (\rho_+ |I| |\Lambda|)^2, \tag{2-19}$$

for all intervals $I \subset [0, E_0]$ and $\Lambda = \Lambda_L$ with $L \geq L(E_0)$, with a constant

$$K_M \leq C_{d,V_{\mathrm{per}},M_\rho} (1 + E_0)^{4[\![d/4]\!]}. \tag{2-20}$$

In more detail:

(i) *If H_ω is an Anderson Hamiltonian with $\delta_- \geq 2$, there exists a constant $C_{d,V_{\mathrm{per}}}$ such that, given an energy $E_0 > 0$, the Wegner estimate* (2-13) *holds for all intervals $I \subset [0, E_0]$ with a constant*

$$K_W \leq C_{d,V_{\mathrm{per}}} u_-^{-2^d} \rho_+^{2^d-1} \gamma_d(E_0) \min\big\{1, E_0^{2^d-d-1}\big\} \max\big\{1, E_0^{2^{d+2}}\big\}, \tag{2-21}$$

where we have $\gamma_d(E_0) = 1$ if $d \geq 2$, and $\gamma_1(E_0) = \gamma_{1,V_{\mathrm{per}}}(E_0) \in]0, 1]$ with $\lim_{E_0 \to 0} \gamma_1(E_0) = 0$.

(ii) *If H_ω is an Anderson Hamiltonian with a uniform-like distribution μ, and for a given $E_0 > 0$ the constant K_W in* (2-13) *satisfies*

$$2K_W U_+ \frac{\rho_+}{\rho_-} \le 1, \tag{2-22}$$

then (2-19) *holds for all intervals $I \subset [0, E_0]$ with a constant $K_M = C_{d,V_{\mathrm{per}},u,M_\rho,E_0} K_W$. If in addition $\delta_- \ge 2$, we have* (2-20).

Our approach to Minami's estimate is discussed in Section 3, where it is illustrated by a proof of the estimate for the (discrete) Anderson model (Theorem 3.3). We also comment on the differences between the discrete and the continuum cases.

On the lattice (the Anderson model), the Wegner estimate (2-13) is a simple consequence of spectral averaging ((3-14)), and holds with $K_W = 1$ for all E_0 [Wegner 1981; Fröhlich and Spencer 1983; Carmona et al. 1987; Kirsch 2008]. On the continuum the Wegner estimate, which has not been as simple to prove, comes with an E_0 dependent constant K_W (which also depends on d, V_{per}, and u) [Combes and Hislop 1994; Combes et al. 2007a]. The proof given in [Combes and Hislop 1994] requires the covering condition $\delta_- \ge 1$. It allows estimates of the constant, but the estimates do not go to 0 as either E_0 or ρ_+ go to 0. The proof in [Combes et al. 2007a] does not require a covering condition, but it uses [Combes et al. 2003, Proposition 1.3] (cf. [Combes et al. 2007a, Theorem 2.1]), which relies on the unique continuation principle to show that some constant is strictly positive, giving no control on the constant in (2-13). To prove that (2-22) holds, so we have (2-19), we need suitable control of the constant K_W, as in (2-21). To obtain this control we introduce a double averaging procedure which uses the covering condition $\delta_- \ge 2$.

Note that the estimate (2-21) provides a bound on the differentiated density of states $n(E) := N'(E)$ in the interval $[0, E_0]$, whenever it exists, since it then follows from (2-13) and (2-21) that

$$n(E) \le C_{d,V_{\mathrm{per}}} u_-^{-2^d} \rho_+^{2^d} \gamma_d(E) \min\{1, E^{2^d-d-1}\} \max\{1, E^{2^{d+2}}\}. \tag{2-23}$$

Once we have the Minami estimate (2-19), we may prove Poisson statistics and simplicity of eigenvalues. The next theorem is proven for arbitrary Anderson Hamiltonians.

Theorem 2.3. *Let H_ω be an Anderson Hamiltonian. Suppose there exists an open interval $\mathcal{I}$ such that for all large boxes Λ the estimate* (2-19) *holds for any interval $I \subset \mathcal{I}$ with $|I| \le \delta_0$, for some $\delta_0 > 0$, with some constant K_M.*

(a) *For all energies $\mathcal{E} \in \mathcal{I} \cap \Xi^{\mathrm{CL}}$ such that the IDS $N(E)$ is differentiable at $\mathcal{E}$ with $n(\mathcal{E}) := N'(\mathcal{E}) > 0$, the point process $\xi^{(\Lambda_L)}_{\mathcal{E},\omega}$ converges weakly, as $L \to \infty$, to the Poisson point process $\xi_{\mathcal{E}}$ on $\mathbb{R}$ with intensity measure $\nu_{\mathcal{E}}(B) := \mathbb{E}\,\xi_{\mathcal{E}}(B) = n(\mathcal{E})|B|$, that is, $\mathrm{d}\nu_{\mathcal{E}} = n(\mathcal{E})\mathrm{d}E$.*

(b) *With probability one, every eigenvalue of H_ω in $\mathcal{I} \cap \Xi^{\mathrm{CL}}$ is simple.*

Theorem 2.3(a) is proven by approximating the point process $\xi^{(\Lambda_L)}_{\mathcal{E},\omega}$ by superpositions of independent point processes, as in [Molchanov 1980/81; Minami 1996], which are then shown to converge weakly to the desired Poisson point process. But here our proof diverges from Minami's, who used the connection, valid for the Anderson model, between the Borel transform of the density of states measure η and averages of the matrix elements of the imaginary part of the resolvent, to identify the intensity measure of the limit point process. Instead, we introduce the random measures

$$\theta^{(\Lambda)}_{\mathcal{E},\omega}(B) := \mathrm{tr}\{\chi_\Lambda P_\omega(\mathcal{E} + |\Lambda|^{-1} B)\chi_\Lambda\} \quad \text{for a Borel set } B \subset \mathbb{R}, \tag{2-24}$$

justified by (2-13)–(2-16), which we show to have the same weak limit as the point processes $\xi^{(\Lambda)}_{\mathscr{E},\boldsymbol{\omega}}$, and use them to show that, thanks to the Lebesgue Differentiation Theorem, the intensity measure $\nu_{\mathscr{E}}$ of the limit point process $\xi_{\mathscr{E}}$ satisfies $\mathrm{d}\nu_{\mathscr{E}} = n(\mathscr{E})\mathrm{d}E$.

Theorem 2.1 follows immediately by combining Theorem 2.2 and Theorem 2.3. Theorem 2.2 is proven in Sections 4 and 5. In Section 4 we prove Wegner estimates with control of the constant in Lemma 4.1, and a Wegner estimate with one random variable ω_j fixed in Lemma 4.2. Theorem 2.2(i) follows from Lemma 4.1(i). Section 5 contains the proof of Minami's estimate: Theorem 2.2(ii) is proven in Lemma 5.1(i), completing the proof of Theorem 2.2. Theorem 2.3 is proven in Sections 6 and 7. In Section 6 we prove Theorem 2.3(a), namely the convergence of the rescaled eigenvalues to a Poisson point process. Finally, in Section 7 we discuss how Theorem 2.3(b) follows from the Minami estimate (2-19) and [Klein and Molchanov 2006].

Some comments about our notation: Finite volumes will always be understood to be boxes $\Lambda = \Lambda_L(j_0)$ with $j_0 \in \mathbb{Z}^d$ and $L \in 2\mathbb{N}$, $L > \delta_+$. We will always identify such Λ with the torus $j_0 + \mathbb{R}^d/L\mathbb{Z}^d$. If $j \in \widetilde{\Lambda}$, we will consider subboxes $\Lambda_s^{(\Lambda)}(j)$ of Λ, where $0 < s \le L$, defined by

$$\Lambda_s^{(\Lambda)}(j) := \left\{\textstyle\bigcup_{k \in L\mathbb{Z}^d} \Lambda_s(j+k)\right\} \cap \Lambda,$$

that is, $\chi_{\Lambda_s^{(\Lambda)}(j)} := \chi_\Lambda \sum_{k\in L\mathbb{Z}^d} \chi_{\Lambda_s(j+k)}$. Similarly, we define functions $u_j^{(\Lambda)}$ on the torus Λ by $u_j^{(\Lambda)} := \chi_\Lambda \sum_{k\in L\mathbb{Z}^d} u_{j+k}$, that is, the function u_j will be assumed to have been wrapped around the torus Λ. Note that we then have $V_{\boldsymbol{\omega}}^{(\Lambda)} = \sum_{j\in\widetilde{\Lambda}} \omega_j u_j^{(\Lambda)}$. We will abuse the notation and just write $\Lambda_s(j)$ for $\Lambda_s^{(\Lambda)}(j)$, u_j for $u_j^{(\Lambda)}$, and $V_{\boldsymbol{\omega}}^{(\Lambda)} = \sum_{j\in\widetilde{\Lambda}} \omega_j u_j$. In addition, given $j \in \Upsilon \cap \mathbb{Z}^d$, where $\Upsilon = \Lambda_L(0)$ or $\mathbb{R}^d$, we write $\boldsymbol{\omega} = (\boldsymbol{\omega}_j^\perp, \omega_j)$, and $H^{(\Upsilon)}_{(\omega_j^\perp,\omega_j=s)} = H^{(\Upsilon)}_{(\omega_j^\perp,s)}$, $P^{(\Upsilon)}_{(\omega_j^\perp,\omega_j=s)}(I) = P^{(\Upsilon)}_{(\omega_j^\perp,s)}(I)$ when we want to make explicit that $\omega_j = s$.

3. A new approach to Minami's estimate illustrated by a proof for the (discrete) Anderson Model

The starting point and key idea in our approach is contained in the following simple lemma.

Lemma 3.1. *Consider the self-adjoint operator $H_s = H_0 + sW$ on the Hilbert space $\mathscr{H}$, where H_0 and W are self-adjoint operators on $\mathscr{H}$, with $W \ge 0$ bounded, and $s \ge 0$. Let $P_s(J) = \chi_J(H_s)$ for an interval J, and suppose* $\operatorname{tr} P_s(]-\infty, c]) < \infty$ *for all $c \in \mathbb{R}$ and $s \ge 0$. Then, for all $a, b \in \mathbb{R}$ with $a < b$ we have*

$$\operatorname{tr} P_s(]a,b]) \le \left\{\operatorname{tr} P_0(]-\infty,b]) - \operatorname{tr} P_t(]-\infty,b])\right\} + \operatorname{tr} P_t(]a,b]) \quad \textit{for} \quad 0 \le s \le t. \tag{3-1}$$

Proof. Let $a, b \in \mathbb{R}$ with $a < b$ and $0 \le s \le t$. Then, since $W \ge 0$, we have

$$\begin{aligned} \operatorname{tr} P_s(]a,b]) &= \operatorname{tr} P_s(]-\infty,b]) - \operatorname{tr} P_s(]-\infty,a]) \\ &\le \operatorname{tr} P_0(]-\infty,b]) - \operatorname{tr} P_t(]-\infty,a]) \\ &= \operatorname{tr} P_0(]-\infty,b]) - \operatorname{tr} P_t(]-\infty,b]) + \operatorname{tr} P_t(]a,b]), \end{aligned} \tag{3-2}$$

as required. □

We will also use the basic spectral averaging estimate: Let H_0 and W be self-adjoint operators on a Hilbert space $\mathscr{H}$, with $W \ge 0$ bounded. Consider the random operator $H_\xi := H_0 + \xi W$, where ξ is a random variable with a nondegenerate probability distribution μ with compact support. The basic spectral averaging estimate for such perturbations of self-adjoint operators says that, given $\varphi \in \mathscr{H}$ with

$\|\varphi\| = 1$, then for all bounded intervals $I \subset \mathbb{R}$ we have (see [Combes and Hislop 1994, Corollary 4.2], [Combes et al. 2007a, (3.16)])

$$\mathbb{E}_\xi\left\{\langle\varphi, \sqrt{W}\chi_I(H_\xi)\sqrt{W}\varphi\rangle\right\} := \int \mathrm{d}\mu(\xi)\, \langle\varphi, \sqrt{W}\chi_I(H_\xi)\sqrt{W}\varphi\rangle \le Q_\mu(|I|), \tag{3-3}$$

where

$$Q_\mu(s) := \begin{cases} \rho_\infty s & \text{if } \mu \text{ has a bounded density } \rho \text{ as in (2-7),} \\ 8\sup_{a\in\mathbb{R}} \mu([a, a+s]) & \text{otherwise.} \end{cases} \tag{3-4}$$

As a consequence, given a trace class operator $S \ge 0$ on $\mathcal{H}$, we have

$$\mathbb{E}_\xi\left\{\operatorname{tr}\{\sqrt{W}\chi_I(H_\xi)\sqrt{W}S\}\right\} \le (\operatorname{tr} S)\, Q_\mu(|I|). \tag{3-5}$$

Note that the measure μ has no atoms if and only if $\lim_{s\downarrow 0} Q_\mu(s) = 0$.

Lemma 3.1 will allow the decoupling of random variables for the performance of two spectral averagings.

We will first illustrate our approach to Minami's estimate by giving a simple and transparent proof of the estimate for in the discrete case, that is, for the Anderson model. We will then comment on how to proceed in the continuum case, that is, for the Anderson Hamiltonian.

Minami's estimate for the (discrete) Anderson model. An Anderson model will be a discrete random Schrödinger operator of the form

$$H_{\boldsymbol{\omega}} = H_0 + V_{\boldsymbol{\omega}} \quad \text{on} \quad \ell^2(\mathbb{Z}^d), \tag{3-6}$$

where H_0 is a bounded self-adjoint operator and $V_{\boldsymbol{\omega}}$ is the random potential given by $V_{\boldsymbol{\omega}}(j) = \omega_j$ for $j \in \mathbb{Z}^d$, where $\boldsymbol{\omega} = \{\omega_j\}_{j\in\mathbb{Z}^d}$ is a family of independent, identically distributed random variables with common probability distribution μ. (The usual Anderson model has $H_0 = -\Delta$, where Δ is the discrete Laplacian.) We assume μ has compact support and no atoms. Adjusting H_0 and μ, we may assume

$$\{0, M\} \in \operatorname{supp}\mu \subset [0, M] \quad \text{with } M \in]0, \infty[. \tag{3-7}$$

Restrictions of $H_{\boldsymbol{\omega}}$ to finite volumes $\Lambda \subset \mathbb{Z}^d$ are denoted by $H_{\boldsymbol{\omega}}^{(\Lambda)}$, a self-adjoint operator of the form

$$H_{\boldsymbol{\omega}}^{(\Lambda)} = H_0^{(\Lambda)} + V_{\boldsymbol{\omega}}^{(\Lambda)} \quad \text{on} \quad \ell^2(\Lambda), \tag{3-8}$$

where $H_0^{(\Lambda)}$ is a self-adjoint restriction of H_0 to the finite-dimensional Hilbert space $\ell^2(\Lambda)$, and $V_{\boldsymbol{\omega}}^{(\Lambda)}$ is the restriction of $V_{\boldsymbol{\omega}}$ to Λ. (In the discrete case our results are not sensitive to the choice of $H_{0,\Lambda}$, they hold for any boundary condition.) Given a Borel set $J \subset \mathbb{R}$, we write $P_{\boldsymbol{\omega}}^{(\Lambda)}(J) = P_{H_{\boldsymbol{\omega}}}^{(\Lambda)}(J) = \chi_J(H_{\boldsymbol{\omega}}^{(\Lambda)})$ for the associated spectral projection.

What makes the discrete case much easier than the continuum is that in the discrete case finite volume operators are finite-dimensional and each random variable couples a rank-one perturbation. Given a unit vector φ in a Hilbert space $\mathcal{H}$, we let Π_φ denote the orthogonal projection onto $\mathbb{C}\varphi$, the one-dimensional subspace spanned by φ. With this notation, the potentials in (3-6) and (3-8) are given by sums of rank-one perturbations:

$$V_{\boldsymbol{\omega}} = \sum_{j\in\mathbb{Z}^d} \omega_j \Pi_j \quad \text{and} \quad V_{\boldsymbol{\omega}}^{(\Lambda)} = \sum_{j\in\Lambda} \omega_j \Pi_j, \quad \text{with } \Pi_j = \Pi_{\delta_j}. \tag{3-9}$$

For rank-one perturbations Lemma 3.1 has the following consequence:

Lemma 3.2. *Let H_s be as in Lemma 3.1 with $W = \Pi_\varphi$ for some unit vector $\varphi \in \mathcal{H}$. Then, for all $a, b \in \mathbb{R}$ with $a < b$ we have*

$$\operatorname{tr} P_s(]a, b]) \leq 1 + \operatorname{tr} P_t(]a, b]) \quad \textit{for all } 0 \leq s \leq t. \tag{3-10}$$

Proof. Let $0 \leq s \leq t$. Recall that for any $c \in \mathbb{R}$ we always have

$$0 \leq \operatorname{tr} P_s(]-\infty, c]) - \operatorname{tr} P_t(]-\infty, c]) \leq 1, \tag{3-11}$$

the last inequality being a consequence of the min-max principle applied to rank-one perturbations, for example, [Kirsch 2008, Lemma 5.22]. Thus (3-10) follows immediately from (3-1). □

For rank-one perturbations the fundamental spectral averaging estimate (3-3) may be stated as follows: Consider the random self-adjoint operator

$$H_\xi = H_0 + \xi \Pi_\varphi \quad \text{on } \mathcal{H}, \tag{3-12}$$

where H_0 is a self-adjoint operator on the Hilbert space $\mathcal{H}$, $\varphi \in \mathcal{H}$ with $\|\varphi\| = 1$, and ξ is a random variable with a nondegenerate probability distribution μ with compact support. Let $P_\xi(J) = \chi_J(H_\xi)$ for a Borel set $J \subset \mathbb{R}$. Then for all bounded intervals $I \subset \mathbb{R}$ we have [Wegner 1981; Fröhlich and Spencer 1983; Carmona et al. 1987; Kirsch 2008; Combes and Hislop 1994; Combes et al. 2007a]

$$\mathbb{E}_\xi\big\{\langle \varphi, P_\xi(I)\varphi\rangle\big\} := \int \mathrm{d}\mu(\xi)\, \langle \varphi, P_\xi(I)\varphi\rangle \leq Q_\mu(|I|). \tag{3-13}$$

The Wegner estimate for an Anderson model [Wegner 1981; Fröhlich and Spencer 1983; Carmona et al. 1987; Kirsch 2008] is an immediate consequence of (3-13):

$$\mathbb{E}\big\{\operatorname{tr} P_{H_\omega}^{(\Lambda)}(I)\big\} = \sum_{j \in \Lambda} \mathbb{E}_{\omega_j^\perp}\big\{\mathbb{E}_{\omega_j}\{\langle \delta_j, P_{H_\omega}^{(\Lambda)}(I)\delta_j\rangle\}\big\} \leq Q_\mu(|I|)|\Lambda|. \tag{3-14}$$

We can now prove Minami's estimate for an Anderson model for arbitrary μ with compact support and no atoms, a result previously known only for μ with a bounded density [Minami 1996; Bellissard et al. 2007; Graf and Vaghi 2007].

Theorem 3.3. *Let H_ω be an Anderson model as in* (3-6), *with μ arbitrary except for compact support and no atoms. Let $\Lambda \subset \mathbb{Z}^d$ be a finite volume. For any bounded interval I we have*

$$\mathbb{E}\big\{(\operatorname{tr} P_\omega^{(\Lambda)}(I))(\operatorname{tr} P_\omega^{(\Lambda)}(I) - 1)\big\} \leq (Q_\mu(|I|)|\Lambda|)^2. \tag{3-15}$$

Theorem 3.3 is extended in [Combes et al. 2009], allowing for n arbitrary intervals and arbitrary single-site probability measure μ with no atoms. We also give applications of (3-15), deriving new results about the multiplicity of eigenvalues and Mott's formula for the ac-conductivity when the single-site probability distribution is Hölder continuous.

Proof of Theorem 3.3. Fix $\Lambda \subset \mathbb{Z}^d$ and let I be a bounded interval. Since the measure μ has no atoms, it follows from (3-14) that $\mathbb{E}_\omega\big\{\operatorname{tr} P_\omega^{(\Lambda)}(\{c\})\big\} = 0$ for any $c \in \mathbb{R}$. Thus we may take all intervals to be of the form $]a,b]$, and use Lemma 3.2 to decouple the random variable ω_j from the random variables $\boldsymbol{\omega}_j^{\perp}$. In view of (3-7), for all $\tau_j \geq M$, $j \in \mathbb{Z}^d$, we have

$$\begin{aligned}(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1) &= \sum_{j\in\Lambda}\bigl\{\langle\delta_j, P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\delta_j\rangle(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\bigr\}\\ &\leq \sum_{j\in\Lambda}\bigl\{\langle\delta_j, P_{(\boldsymbol{\omega}_j^{\perp},\omega_j)}^{(\Lambda)}(I)\delta_j\rangle(\operatorname{tr} P_{(\boldsymbol{\omega}_j^{\perp},\tau_j)}^{(\Lambda)}(I))\bigr\}. \end{aligned}\tag{3-16}$$

We now average over the random variables $\boldsymbol{\omega} = \{\omega_j\}_{j\in\mathbb{Z}^d}$. Using (3-13), we get

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\omega}}\bigl\{(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\bigr\} &\leq \sum_{j\in\Lambda}\mathbb{E}_{\boldsymbol{\omega}_j^{\perp}}\bigl\{(\operatorname{tr} P_{(\boldsymbol{\omega}_j^{\perp},\tau_j)}^{(\Lambda)}(I))\bigl(\mathbb{E}_{\omega_j}\bigl\{\langle\delta_j, P_{(\boldsymbol{\omega}_j^{\perp},\omega_j)}^{(\Lambda)}(I)\delta_j\rangle\bigr\}\bigr)\bigr\}\\ &\leq Q_\mu(|I|)\sum_{j\in\Lambda}\mathbb{E}_{\boldsymbol{\omega}_j^{\perp}}\bigl\{\operatorname{tr} P_{(\boldsymbol{\omega}_j^{\perp},\tau_j)}^{(\Lambda)}(I)\bigr\}. \end{aligned}\tag{3-17}$$

This holds for all $\tau_j \geq M$, $j \in \mathbb{Z}^d$, so we now take $\tau_j = M + \tilde{\omega}_j$, where $\tilde{\boldsymbol{\omega}} = \{\tilde{\omega}_j\}_{j\in\mathbb{Z}^d}$ and $\boldsymbol{\omega} = \{\omega_j\}_{j\in\mathbb{Z}^d}$ are two independent, identically distributed collections of random variables. Now $\boldsymbol{\tau} = \{\tau_j\}_{j\in\mathbb{Z}^d}$ are independent identically distributed random variables with a common probability distribution $\mu_{\boldsymbol{\tau}}$ such that $Q_{\mu_{\boldsymbol{\tau}}} = Q_\mu$. We get

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\omega}}\bigl\{(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\bigr\} &= \mathbb{E}_{\boldsymbol{\tau}}\bigl\{\mathbb{E}_{\boldsymbol{\omega}}\bigl\{(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\bigr\}\bigr\}\\ &\leq Q_\mu(|I|)\sum_{j\in\Lambda}\mathbb{E}_{(\boldsymbol{\omega}_j^{\perp},\tau_j)}(\operatorname{tr} P_{(\boldsymbol{\omega}_j^{\perp},\tau_j)}^{(\Lambda)}(I)) \leq (Q_\mu(|I|)|\Lambda|)^2, \end{aligned}\tag{3-18}$$

where we used the Wegner estimate (3-14). (More precisely, we estimate as in (3-14); the random variables do not need to be identically distributed.) □

Stepping up to the continuum. Unfortunately things are not so simple for the continuum Anderson Hamiltonian. The main reason is that the random potential $V_{\boldsymbol{\omega}}$ in (2-3) is a sum of independent random perturbations of infinite rank, not of rank one as in the discrete case, and thus the a priori bound in (3-11), and also Lemma 3.2, are not applicable anymore.

To prove Minami's estimate on the continuum we will use the fundamental spectral averaging estimate as in (3-5). The straightforward expansion of the trace in (3-14) and (3-17) cannot be used for the spectral averaging, even with u_j instead of δ_j, and will be replaced by a more sophisticated expansion in terms of trace class operators, as in [Combes and Hislop 1994; Combes et al. 2007a] ((4-1)–(4-5)). Lemma 3.1 will be modified, since the term in brackets in (3-1) does not satisfy an a priori bound as in (3-11) anymore. This term will be estimated using the Birman–Solomyak formula; see (5-3), (5-4). The bound in (3-11) is then replaced by averaging the resulting expression over all the other random variables and using the Wegner estimate (2-13); see (5-9). The resulting bound is useful if the constant K_W in (2-13) is not too big (we have $K_W = 1$ in the lattice, as can be seen in (3-14)). Since previous proofs of the Wegner estimate do not give the desired control of K_W, we must revisit the Wegner estimate. We introduce a double averaging procedure that provides the desired estimates on the constant K_W (Lemma 4.1). In addition, because of the way we use the Birman–Solomyak formula, we do not have freedom in the choice of τ_j as in (3-16), we have to take $\tau_j = M_\rho$. Thus we cannot average in $\boldsymbol{\tau}$ as in (3-18); this argument is replaced by a refinement of the Wegner estimate where one of the random variables is fixed (Lemma 4.2).

4. The Wegner estimate revisited

Let $H_{\boldsymbol{\omega}}$ be the Anderson Hamiltonian, $E_0 > 0$, $I \subset [0, E_0]$ an interval, and Λ a finite box. To prove the Wegner estimate (2-13), it is shown in [Combes and Hislop 1994; Combes et al. 2007a] that

$$\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) \le Q_1 \sum_{j,k\in\widetilde{\Lambda}} \left|\operatorname{tr}\{\sqrt{u_k}\, P_{\boldsymbol{\omega}}^{(\Lambda)}(I) \sqrt{u_j}\, T_{j,k}^{(\Lambda)}\}\right|, \tag{4-1}$$

where $\{T_{j,k}^{(\Lambda)}\}_{j,k\in\widetilde{\Lambda}}$ are (nonrandom) trace class operators in $\mathrm{L}^2(\Lambda)$ such that

$$\max_{j\in\widetilde{\Lambda}} \left\{ \sum_{k\in\widetilde{\Lambda}} \|T_{j,k}^{(\Lambda)}\|_1 \right\} \le Q_2, \tag{4-2}$$

the constants Q_1, Q_2 depending only on $E_0, d, u, V_{\text{per}}, M_\rho$. Letting

$$T_{j,k}^{(\Lambda)} = U_{j,k}^{(\Lambda)} |T_{j,k}^{(\Lambda)}|$$

be the polar decomposition of the operator $T_{j,k}^{(\Lambda)}$, recalling that then $|T_{j,k}^{(\Lambda)*}| = U_{j,k}^{(\Lambda)} T_{j,k}^{(\Lambda)} U_{j,k}^{(\Lambda)*}$, and setting

$$S_j^{(\Lambda)} := \frac{1}{2} \sum_{k\in\widetilde{\Lambda}} \left(|T_{j,k}^{(\Lambda)*}| + |T_{k,j}^{(\Lambda)}|\right) \ge 0 \quad \text{for} \quad j \in \widetilde{\Lambda}, \tag{4-3}$$

we obtain

$$\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) \le Q_1 \sum_{j\in\widetilde{\Lambda}} \operatorname{tr}\left\{\sqrt{u_j}\, P_{\boldsymbol{\omega}}^{(\Lambda)}(I) \sqrt{u_j}\, S_j^{(\Lambda)}\right\}, \tag{4-4}$$

with

$$\max_{j\in\widetilde{\Lambda}} \left\{\operatorname{tr} S_j^{(\Lambda)}\right\} \le Q_2. \tag{4-5}$$

If we now take the expectation in (4-4), use (3-5) and (4-5), we get the Wegner estimate (2-13) with $K_W = Q_1 Q_2$.

We will need control of the constant K_W and a Wegner estimate with one of the random variables, say ω_0, fixed. In the course of obtaining control over K_W we will derive (4-1) with estimates on the constants Q_1 and Q_2 in the case when $\delta_- \ge 1$.

A Wegner estimate with control of the constants.

Lemma 4.1. *Let $H_{\boldsymbol{\omega}}$ be an Anderson Hamiltonian.*

(i) *Assume $\delta_- \ge 2$. Then there exists a constant $C_{d,V_{\text{per}}}$ such that, given an energy $E_0 > 0$,* (2-13) *holds for all intervals $I \subset [0, E_0]$ with a constant*

$$K_W \le C_{d,V_{\text{per}}} \left(\frac{\rho_+}{u_-}\right)^{2^d} \gamma_d(E_0) \min\left\{1, E_0^{2^d-d-1}\right\} \max\left\{1, E_0^{2^{d+2}}\right\}, \tag{4-6}$$

where we have $\gamma_d(E_0) = 1$ if $d \ge 2$, and $\gamma_1(E_0) = \gamma_{1,V_{\text{per}}}(E_0) \in \,]0, 1]$ with $\lim_{E_0\to 0} \gamma_1(E_0) = 0$.

(ii) *Assume $\delta_- \ge 1$. Then, given an energy $E_0 > 0$,* (4-1)–(4-5) *hold for all intervals $I \subset [0, E_0]$ with constants*

$$Q_1 = (1 + E_0)^{2[\![d/4]\!]} \quad \textit{and} \quad Q_2 = C'_{d,V_{\text{per}}}, \tag{4-7}$$

and hence (2-13) *holds for all intervals* $I \subset [0, E_0]$ *with a constant*

$$K_W \le C'_{d,V_{\text{per}}}(1+E_0)^{2[\![d/4]\!]}. \tag{4-8}$$

Proof. Assume $\delta_- \ge m$, where m is either 1 or 2. We set

$$\chi_j^{(m)} = \chi_{\Lambda_m(j)} \quad \text{for } j \in \widetilde{\Upsilon} := \Upsilon \cap \mathbb{Z}^d,$$

where Υ is either $\mathbb{R}^d$ or a finite box Λ (recall that in this case $\chi_{\Lambda_m(j)}$ denotes $\chi^{(\Lambda)}_{\Lambda_m(j)}$, a subbox in the torus). Note that for any $j_0 \in \widetilde{\Upsilon}$ we have

$$\sum_{j\in(j_0+m\mathbb{Z}^d)\cap\Upsilon} \chi_j^{(m)} = 1. \tag{4-9}$$

We also let $\hat{\chi}_j^{(m)} = {u_j}^{-1/2}\chi_j^{(m)}$ on $\Lambda_m(j)$, $\hat{\chi}_j^{(m)} = 0$ otherwise. It follows from (2-5) that

$$\hat{\chi}_j^{(m)} \le u_-^{-1/2}\chi_j^{(m)}.$$

(Recall we write u_j for $u_j^{(\Lambda)}$.)

To prove (i), assume $\delta_- \ge 2$. We write $\boldsymbol{\omega}' = \{\omega_j\}_{j\in 2\mathbb{Z}^d}$, $\boldsymbol{\omega}'' = \{\omega_j\}_{j\notin 2\mathbb{Z}^d}$. We set

$$H_{\boldsymbol{\omega}''} := H_0 + V_{\boldsymbol{\omega}''}, \quad V_{\boldsymbol{\omega}''} := \sum_{j\notin 2\mathbb{Z}^d} \omega_j u_j. \tag{4-10}$$

Note that $H_{\boldsymbol{\omega}''}$ is a $2\mathbb{Z}^d$ ergodic family of random self-adjoint operators, and we have

$$H_{\boldsymbol{\omega}} \ge H_{\boldsymbol{\omega}''} \ge H_0, \quad H_{\boldsymbol{\omega}''} \ge V_{\boldsymbol{\omega}''}. \tag{4-11}$$

Fix an energy $E_0 > 0$, a box Λ, and let $I =]a, b] \subset [0, E_0]$. Set $p = 2^{d+1}$. Given $t > 0$, the function $g_t(x) = (1+tx)^{-2p}$ is convex on the interval $]-1/t, \infty[$. Thus, using (4-11), we can proceed as in [Combes and Hislop 1994] using convexity and Jensen's inequality (see Lemma B.1 in Appendix B), and then (4-9) and (2-5), to get

$$\begin{aligned}
\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) &\le (1+tE_0)^{2p} \operatorname{tr}\big\{P_{\boldsymbol{\omega}}^{(\Lambda)}(I)(1+tH_{\boldsymbol{\omega}}^{(\Lambda)})^{-2p} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\big\} \\
&\le (1+tE_0)^{2p} \operatorname{tr}\big\{P_{\boldsymbol{\omega}}^{(\Lambda)}(I)(1+tH_{\boldsymbol{\omega}''}^{(\Lambda)})^{-2p} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\big\} \\
&= (1+tE_0)^{2p} \operatorname{tr}\big\{P_{\boldsymbol{\omega}}^{(\Lambda)}(I)(1+tH_{\boldsymbol{\omega}''}^{(\Lambda)})^{-2p}\big\} \\
&= (1+tE_0)^{2p} \sum_{j,k\in\Lambda\cap 2\mathbb{Z}^d} \operatorname{tr}\big\{P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\chi_j^{(2)}(1+tH_{\boldsymbol{\omega}''}^{(\Lambda)})^{-2p}\chi_k^{(2)}\big\} \\
&= (1+tE_0)^{2p} \sum_{j,k\in\Lambda\cap 2\mathbb{Z}^d} \operatorname{tr}\big\{\sqrt{u_k}\, P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\sqrt{u_j}\,\hat{\chi}_j^{(2)}(1+tH_{\boldsymbol{\omega}''}^{(\Lambda)})^{-2p}\hat{\chi}_k^{(2)}\big\}.
\end{aligned} \tag{4-12}$$

It then follows from (3-5), proceeding as in (4-1)–(4-4) (see also [Combes et al. 2007a, Lemma 2.1]), that

$$\begin{aligned}\mathbb{E}_{\omega'} \operatorname{tr} P_{\omega}^{(\Lambda)}(I) &\le (1+tE_0)^{2p}\rho_+|I| \sum_{j,k\in\Lambda\cap 2\mathbb{Z}^d} \big\|\hat{\chi}_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-2p}\hat{\chi}_k^{(2)}\big\|_1\\ &\le (1+tE_0)^{2p}u_-^{-1}\rho_+|I| \sum_{j,k\in\Lambda\cap 2\mathbb{Z}^d} \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-2p}\chi_k^{(2)}\big\|_1. \end{aligned} \tag{4-13}$$

We now use several deterministic estimates. First,

$$\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-2p}\chi_k^{(2)}\big\|_1 \le \sum_{r\in\Lambda\cap 2\mathbb{Z}^d} \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_2 \big\|\chi_r^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_k^{(2)}\big\|_2. \tag{4-14}$$

Second,

$$\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_2^2 \le \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\| \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_1. \tag{4-15}$$

Third, we estimate

$$\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|$$

using the Combes–Thomas estimate. We use the precise estimate provided in [Germinet and Klein 2003b, (19) in Theorem 1] (with $\gamma=\frac{1}{2}$), modified for finite volume operators with periodic boundary condition as in [Figotin and Klein 1996, Lemma 18] and [Klein and Koines 2001, Theorem 3.6], plus the fact that we are using boxes of side 2. For $L\ge L_d$ we have, with $d_\Lambda(j,r)$ the distance on the torus Λ,

$$\begin{aligned}\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\| &= t^{-p}\big\|\chi_j^{(2)}(t^{-1}+H_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\| \le t^{-p}\big(\tfrac{4}{3}t\big)^p \exp\frac{1}{2\sqrt{t}}\exp\Big(-\frac{1}{8\sqrt{td}}d_\Lambda(j,r)\Big)\\ &= \big(\tfrac{4}{3}\big)^p \exp\frac{1}{2\sqrt{t}}\exp\Big(-\frac{1}{8\sqrt{td}}d_\Lambda(j,r)\Big).\end{aligned} \tag{4-16}$$

Fourth, note that

$$\begin{aligned}\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_1 &\le \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\big\|_2 \big\|\chi_r^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\big\|_2\\ &= \big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_j^{(2)}\big\|_1^{1/2} \big\|\chi_r^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_1^{1/2}.\end{aligned} \tag{4-17}$$

We now average over ω''. Using (4-14)–(4-17), we have

$$\begin{aligned}&\mathbb{E}_{\omega''}\big\{\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_1^{1/2}\big\|\chi_r^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_k^{(2)}\big\|_1^{1/2}\big\}\\ &\quad\le \mathbb{E}_{\omega''}\big\{\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_j^{(2)}\big\|_1^{1/4}\big\|\chi_r^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_r^{(2)}\big\|_1^{1/2}\times\big\|\chi_k^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_k^{(2)}\big\|_1^{1/4}\big\}\\ &\quad\le \beta_t := \mathbb{E}_{\omega''}\big\{\big\|\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_0^{(2)}\big\|_1\big\},\end{aligned} \tag{4-18}$$

where we used Hölder's inequality plus translation invariance (in the torus) of the expectation.

It now follows from (4-14), (4-15), (4-16), (4-17), and (4-18) that

$$\begin{aligned}
\mathbb{E}_{\omega''}\Big\{\sum_{j,k\in\Lambda\cap 2\mathbb{Z}^d}&\big\|\chi_j^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-2p}\chi_k^{(2)}\big\|_1\Big\}\\
&\le \beta_t\ \exp\frac{1}{2\sqrt{t}}\big(\tfrac{4}{3}\big)^p\sum_{j,k,r\in\Lambda\cap 2\mathbb{Z}^d}\exp\Big(-\frac{1}{16\sqrt{td}}d_\Lambda(j,r)\Big)\exp\Big(-\frac{1}{16\sqrt{td}}d_\Lambda(r,k)\Big)\\
&\le 2^{-d}\beta_t\ \exp\frac{1}{2\sqrt{t}}\big(\tfrac{4}{3}\big)^p|\Lambda|\bigg(\sum_{r\in 2\mathbb{Z}^d}\exp\Big(-\frac{1}{16\sqrt{td}}|r|\Big)\bigg)^2\\
&= 2^{-d}\beta_t\ \exp\frac{1}{2\sqrt{t}}\big(\tfrac{4}{3}\big)^p|\Lambda|\bigg(\sum_{s\in\mathbb{Z}}\exp\Big(-\frac{1}{8d\sqrt{t}}|s|\Big)\bigg)^{2d}\\
&\le 2^{-d}\beta_t\ \exp\frac{1}{2\sqrt{t}}\big(\tfrac{4}{3}\big)^p|\Lambda|\bigg(1+2\int_0^\infty \mathrm{d}s\ \exp\Big(-\frac{1}{8d\sqrt{t}}|s|\Big)\bigg)^{2d}\\
&\le 2^{-d}\beta_t\exp\frac{1}{2\sqrt{t}}\big(\tfrac{4}{3}\big)^p|\Lambda|(1+16d\sqrt{t})^{2d},
\end{aligned}\tag{4-19}$$

so we conclude from (4-13) that

$$\mathbb{E}_\omega\ \mathrm{tr}\ P_\omega^{(\Lambda)}(I)\le\big(\tfrac{4}{3}\big)^p\frac{1}{2u_-}(1+tE_0)^{2p}\beta_t\ \exp\frac{1}{2\sqrt{t}}(1+16d\sqrt{t})^{2d}\rho_+|I||\Lambda|.\tag{4-20}$$

We now estimate β_t. We have, using periodicity, and again Lemma B.1 with $H_{\omega''}^{(\Lambda)}\ge V_{\omega''}$ and (2-5),

$$\begin{aligned}
\beta_t &:= \mathbb{E}_{\omega''}\big\{\mathrm{tr}\{\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p}\chi_0^{(2)}\}\big\}=\frac{2^d}{|\Lambda|}\mathbb{E}_{\omega''}\big\{\mathrm{tr}\{(1+tH_{\omega''}^{(\Lambda)})^{-p}\}\big\}\\
&\le\frac{2^d}{|\Lambda|}\mathbb{E}_{\omega''}\big\{\mathrm{tr}\big\{(1+tH_{\omega''}^{(\Lambda)})^{-p/4}(1+tV_{\omega''})^{-p/2}(1+tH_{\omega''}^{(\Lambda)})^{-p/4}\big\}\big\}\\
&=\frac{2^d}{|\Lambda|}\mathbb{E}_{\omega''}\big\{\mathrm{tr}\big\{(1+tV_{\omega''})^{-p/4}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}(1+tV_{\omega''})^{-p/4}\big\}\big\}\\
&=\mathbb{E}_{\omega''}\big\{\mathrm{tr}\big\{\chi_0^{(2)}(1+tV_{\omega''})^{-p/4}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}(1+tV_{\omega''})^{-p/4}\chi_0^{(2)}\big\}\big\}\\
&=\mathbb{E}_{\omega''}\big\{\mathrm{tr}\big\{(1+tV_{\omega''})^{-p/4}\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\chi_0^{(2)}(1+tV_{\omega''})^{-p/4}\big\}\big\}\\
&\le\mathbb{E}_{\omega''}\big\{(1+tu_-\hat\omega_0)^{-p/2}\ \mathrm{tr}\{\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\chi_0^{(2)}\}\big\},
\end{aligned}\tag{4-21}$$

where we set, with $Q:=\{0,1\}^d\setminus\{0\}\subset\mathbb{Z}^d$,

$$\hat\omega_0=\sum_{q\in Q}\hat\omega_{0,q}\quad\text{ with }\hat\omega_{0,q}:=\min\{\omega_{q+i}:i\in 2\mathbb{Z}^d,\ |q+i|_\infty=1\}.\tag{4-22}$$

Note that $|Q|=2^d-1$, and $(q+2\mathbb{Z}^d)\cap(q'+2\mathbb{Z}^d)=\emptyset$ if $q,q'\in Q$ with $q\ne q'$, so $\{\hat\omega_{0,q}\}_{q\in Q}$ are independent random variables.

Now, with $\Theta:=\max\{-\operatorname{ess\,inf}V_{\mathrm{per}},0\}$,

$$\begin{aligned}
\mathrm{tr}\big\{\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\chi_0^{(2)}\big\}&\le\Big\{\sup_{E\ge 0}\Big(\frac{1+\Theta+E}{1+tE}\Big)^{p/2}\Big\}\ \mathrm{tr}\big\{\chi_0^{(2)}(H_{\omega''}^{(\Lambda)}+1+\Theta)^{-p/2}\chi_0^{(2)}\big\}\\
&\le C_{d,\Theta}\max\{1,t^{-p/2}\},
\end{aligned}\tag{4-23}$$

where (as in the proof of Lemma A.4 of [Germinet and Klein 2004], for example) we used the fact that $\operatorname{tr}\bigl\{\chi_0^{(2)}(H_{\omega''}^{(\Lambda)}+1+\Theta)^{-p/2}\chi_0^{(2)}\bigr\}$ is uniformly bounded, independently of Λ — itself a consequence of the inequality $p = 2^{d+1} \geq 4[\![d/4]\!]$, where $[\![d/4]\!]$ is the smallest integer exceeding $d/4$.

Moreover, since $p = 2^{d+1} > 2(2^d - 1)$,

$$\begin{aligned}
\mathbb{E}_{\omega''}\{(1+tu_-\hat{\omega}_0)^{-p/2}\} &\leq \prod_{q\in Q} \mathbb{E}_{\omega''}\bigl\{(1+tu_-\hat{\omega}_{0,q})^{-p/(2(2^d-1))}\bigr\} \\
&= \prod_{q\in Q} \mathbb{E}_{\omega''}\Bigl\{\max_{\substack{i\in 2\mathbb{Z}^d \\ |q+i|_\infty=1}} (1+tu_-\omega_{q+i})^{-p/(2(2^d-1))}\Bigr\} \\
&\leq \bigl(2d\,\mathbb{E}_{\omega_0}\bigl\{(1+tu_-\omega_0)^{-p/(2(2^d-1))}\bigr\}\bigr)^{2^d-1} \\
&\leq \biggl(2d\,\rho_+\int_0^\infty \mathrm{d}\omega_0\,(1+tu_-\omega_0)^{-p/(2(2^d-1))}\biggr)^{2^d-1} \\
&\leq \biggl(\frac{2d(2^d-1)\,\rho_+}{(2^d-1-\frac{p}{2})tu_-}\biggr)^{2^d-1} = C_d'\Bigl(\frac{\rho_+}{tu_-}\Bigr)^{2^d-1}.
\end{aligned} \tag{4-24}$$

Thus, we have

$$\beta_t \leq C_{d,\Theta}'\max\{1, t^{-2^d}\}\Bigl(\frac{\rho_+}{tu_-}\Bigr)^{2^d-1}, \tag{4-25}$$

so it follows from (4-20) that

$$\mathbb{E}_{\omega}\operatorname{tr}P_{\omega}^{(\Lambda)}(I) \leq \frac{C_{d,\Theta}'}{u_-}(1+tE_0)^{2^{d+2}}\,\exp\frac{1}{2\sqrt{t}}(1+16d\sqrt{t})^{2d}\max\{1,t^{-2^d}\}\Bigl(\frac{\rho_+}{tu_-}\Bigr)^{2^d-1}\rho_+|I||\Lambda|. \tag{4-26}$$

If $E_0 \leq 3$, we choose $t = 1/E_0$, obtaining

$$\mathbb{E}_{\omega}\operatorname{tr}P_{\omega}^{(\Lambda)}(I) \leq C_{d,\Theta}''\Bigl(\frac{\rho_+}{u_-}\Bigr)^{2^d} E_0^{2^d-d-1}|I||\Lambda|. \tag{4-27}$$

If $E_0 > 3$, we take $t = 1$, getting

$$\mathbb{E}_{\omega}\operatorname{tr}P_{\omega}^{(\Lambda)}(I) \leq C_{d,\Theta}'''\Bigl(\frac{\rho_+}{u_-}\Bigr)^{2^d} E_0^{2^{d+2}}|I||\Lambda|. \tag{4-28}$$

Thus, for all $E_0 > 0$ we have

$$\mathbb{E}_{\omega}\operatorname{tr}P_{\omega}^{(\Lambda)}(I) \leq \frac{C_{d,\Theta}}{u_-}\Bigl(\frac{\rho_+}{u_-}\Bigr)^{2^d-1}\min\{1, E_0^{2^d-d-1}\}\max\{1, E_0^{2^{d+2}}\}\rho_+|I||\Lambda|. \tag{4-29}$$

For $d = 1$ we need to do a bit better. In this case we redo (4-23) as follows:

$$\operatorname{tr}\bigl\{\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-p/2}\chi_0^{(2)}\bigr\} \leq \operatorname{tr}\bigl\{\chi_0^{(2)}(1+tH_{\omega''}^{(\Lambda)})^{-1}\chi_0^{(2)}\bigr\} \leq \alpha_t := \operatorname{tr}\bigl\{\chi_0^{(2)}(1+tH_0^{(\Lambda)})^{-1}\chi_0^{(2)}\bigr\}. \tag{4-30}$$

For $d = 1$ the estimate (4-26) now becomes

$$\mathbb{E}_{\omega}\operatorname{tr}P_{\omega}^{(\Lambda)}(I) \leq \frac{C_{1,\theta}}{u_-}(1+tE_0)^8\,\exp\frac{1}{2\sqrt{t}}(1+16\sqrt{t})^2\alpha_t\Bigl(\frac{\rho_+}{tu_-}\Bigr)\rho_+|I||\Lambda|, \tag{4-31}$$

and thus (4-29) becomes

$$\mathbb{E}_{\omega} \operatorname{tr} P_{\omega}^{(\Lambda)}(I) \leq \frac{C_{1,\Theta}}{u_-} \frac{\rho_+}{u_-} \gamma_1(E_0) \max\{1, E_0^8\} \rho_+ |I| |\Lambda|. \tag{4-32}$$

where $\gamma_1(E_0) \leq 1$ and $\lim_{E_0 \to 0} \gamma_1(E_0) = 0$ uniformly in Λ large.

This proves (i). To prove (ii), we now assume $\delta_- \geq 1$. We proceed as in the proof of (i), with $\omega' = \omega$ and $\omega'' = \{\omega_j\}_{j \notin \mathbb{Z}^d} = \varnothing$, that is $V_{\omega''} = 0$ and $H_{\omega''} = H_0$. We also now fix $p = 2[\![d/4]\!]$. Then (4-12) yields (4-1) with $Q_1 = (1 + tE_0)^{2p}$ and $T_{j,k}^{(\Lambda)} = \hat{\chi}_j^{(1)}(1 + tH_0^{(\Lambda)})^{-2p} \hat{\chi}_k^{(1)}$. Proceeding as in (4-14)–(4-19) gives (4-2) with

$$Q_2 = \beta_t^{(0)} \exp \frac{1}{4\sqrt{t}} \left(\tfrac{4}{3}\right)^p (1 + 32d\sqrt{t})^{2d}, \tag{4-33}$$

where, as in (4-23),

$$\beta_t^{(0)} := \left\| \chi_0^{(1)} (1 + tH_0^{(\Lambda)})^{-p} \chi_0^{(1)} \right\|_1 \leq C_{d,\Theta} \max\{1, t^{-p}\} \leq C_{d,\Theta}. \tag{4-34}$$

We now set $t = 1$, obtaining (4-7) and (4-8). □

A Wegner estimate with ω_0 fixed. Let $\Upsilon = \Lambda_L(0)$ or $\mathbb{R}^d$. Given $\tau \in \mathbb{R}$, we consider (recall $u_0 = u$)

$$H_{(\omega^{(0)},\tau)}^{(\Upsilon)} = H_{(\omega^{(0)},\omega_0=\tau)}^{(\Upsilon)} = H_{\omega}^{(\Upsilon)} + (\tau - \omega_0)u. \tag{4-35}$$

Lemma 4.2. *Let H_ω be an Anderson Hamiltonian, $E_0 > 0$. Given $\tau \in \mathbb{R}$, there exists a constant $\widetilde{K}_W = \widetilde{K}_W(d, u, V_{\text{per}}, E_0, M_\rho, \tau)$, such that for any interval $I \subset [0, E_0]$ and finite box $\Lambda = \Lambda_L(0)$ we have*

$$\mathbb{E}_{\omega^{(0)}} \left\{ \operatorname{tr} P_{(\omega^{(0)},\tau)}^{(\Lambda)}(I) \right\} \leq \widetilde{K}_W \rho_+ |I| |\Lambda|. \tag{4-36}$$

Moreover, if $\delta_- \geq 2$, we have

$$\widetilde{K}_W \leq C_{d,V_{\text{per}},\tau} (1 + E_0)^{2[\![\frac{d}{4}]\!]}. \tag{4-37}$$

Proof. We will show that the proof of Theorem 1.3 of [Combes et al. 2007a] can be modified to yield the proposition. *All references of the form (2.N) in this proof will be to that paper unless otherwise stated.*

We introduce the background potential

$$H_1 := H_0 + \tau \sum_{j \in 2\mathbb{Z}^d} u_j = -\Delta + V_{\text{per}}^{(2)}, \tag{4-38}$$

where $V_{\text{per}}^{(2)} = V_{\text{per}} + \tau \sum_{j \in 2\mathbb{Z}^d} u_j$ is a $2\mathbb{Z}^d$-periodic potential. It follows that

$$H_{(\omega^{(0)},\tau)} = H_1 + V_{\omega^{(0)}}(\tau) \quad \text{with } V_{\omega^{(0)}}(\tau) := \sum_{j \in (2\mathbb{Z})^d \setminus \{0\}} (\omega_j - \tau) u_j + \sum_{j \in \mathbb{Z} \setminus (2\mathbb{Z})^d} \omega_j u_j. \tag{4-39}$$

The main point is that the single-site potential $u_0 = u$ does not appear in the sum, but all the other u_j's appear with a random coefficient.

To prove (4-36) with no conditions on δ_-, we proceed as in Section 2 of [Combes et al. 2007a]. We take an interval $I \subset [0, E_0]$, write $\tilde{I} = [0, E_0{+}1]$; I and $\tilde{I}$ replace the intervals Δ and $\tilde{\Delta}$ in that paper. The potential V_Λ in equation (2.7) there is replaced by $V_{\omega^{(0)}}^{(\Lambda)}(\tau)$, which only involves the random variables $\omega^{(0)}$. As a consequence, the sum in (2.10) runs over indices $i, j \in \widetilde{\Lambda} \setminus \{0\}$. The spectral averaging in (2.13) can thus be performed with respect to the random variables $\omega^{(0)}$. Similarly for (2.18), since $\tilde{K}(n)_{i_1,j_n}$

of (2.17) is now constructed only with the single-site potentials u_j's present in $V^{(\Lambda)}_{\omega^{(0)}}(\tau)$, that is, u_j with $j \in \widetilde{\Lambda} \setminus \{0\}$. We thus get the analog of (2.20), with $M_0 = M_\rho + |\tau|$, namely, with $P_1(B) = \chi_B(H_1)$,

$$\mathbb{E}_{\omega^{(0)}}\big\{\operatorname{tr}\big\{P^{(\Lambda)}_{(\omega^{(0)},\tau)}(I)P_1^{(\Lambda)}(\mathbb{R}\setminus\tilde{I})\big\}\big\} \le K_1\rho_+|I||\Lambda|, \tag{4-40}$$

for an appropriate constant K_1.

It remains to bound $\mathbb{E}_{\omega^{(0)}}\big\{\operatorname{tr}\big\{P^{(\Lambda)}_{(\omega^{(0)},\tau)}(I)P_1^{(\Lambda)}(\tilde{I})\big\}\big\}$. For this purpose, we set

$$\tilde{V}_1 = \sum_{j\in(\mathrm{e}_1+2\mathbb{Z}^d)} u_j, \tag{4-41}$$

where $\mathrm{e}_1 = (1, 0, 0, \dots, 0) \notin 2\mathbb{Z}^d$, and we use H_1 and $\tilde{V}_1^{(\Lambda)}$, the restriction of $\tilde{V}_1$ to Λ, instead of H_0 and $\tilde{V}_\Lambda = \sum_{j\in\mathbb{Z}^d\cap\Lambda} u_j$, in the crucial estimate (2.1) of [Combes et al. 2007a]. Since H_1 and $\tilde{V}_1$ are both $2\mathbb{Z}^d$-periodic, we have[1] the equivalent of (2.1),

$$P_1^{(\Lambda)}(\tilde{I})\tilde{V}_1^{(\Lambda)}P_1^{(\Lambda)}(\tilde{I}) \ge C(E_0, u, V_{\mathrm{per}}, \tau)P_1^{(\Lambda)}(\tilde{I}), \tag{4-42}$$

with a constant $C(E_0, u, V_{\mathrm{per}}, \tau) > 0$. Since

$$\tilde{V}_1 \le \tilde{V}_{0^\perp} := \sum_{j\in\mathbb{Z}^d\setminus\{0\}} u_j, \tag{4-43}$$

it follows that

$$P_1^{(\Lambda)}(\tilde{I})\tilde{V}_{0^\perp}^{(\Lambda)}P_1^{(\Lambda)}(\tilde{I}) \ge C(E_0, u, V_{\mathrm{per}}, \tau)P_1^{(\Lambda)}(\tilde{I}). \tag{4-44}$$

As a consequence, we get (2.21) with $\tilde{V}_\Lambda$ replaced by $\tilde{V}_{0^\perp}^{(\Lambda)}$, and hence we obtain the analogue of (2.31):

$$\mathbb{E}_{\omega^{(0)}}\big\{\operatorname{tr}\big\{P_1^{(\Lambda)}(\tilde{I})\tilde{V}_{0^\perp}^{(\Lambda)}P^{(\Lambda)}_{(\omega^{(0)},\tau)}(I)\tilde{V}_{0^\perp}^{(\Lambda)}P_1^{(\Lambda)}(\tilde{I})\big\}\big\} \le K_2\rho_+|I||\Lambda|, \tag{4-45}$$

for an appropriate constant K_2.

The desired bound (4-36) now follows as the analogue of (2.32).

If $\delta_- \ge 2$, we have

$$\sum_{j\in((j_0+\mathbb{Z}^d)\setminus\{0\})\cap\Lambda} u_j \ge u_-\chi_\Lambda, \tag{4-46}$$

so we can apply the proof of Lemma 4.1(ii) to the random operator $H_{\omega^{(0)},\tau}$ getting (4-36) with (4-37). □

5. The Minami estimate

Theorem 2.2 follows by combining Lemma 4.1(i) and the following lemma:

Lemma 5.1. *Let H_ω be an Anderson Hamiltonian with a uniform-like distribution μ. Let $E_0 > 0$ and suppose the Wegner estimate* (2-13) *holds for all intervals $I \subset [0, E_0]$ with a constant K_W such that*

$$2K_W U_+\frac{\rho_+}{\rho_-} \le 1. \tag{5-1}$$

Then there exists a constant $K_M = K_M(u, \rho_\pm, M_\rho, E_0, d)$ such that the Minami estimate (2-19) *holds for all intervals $I \subset [0, E_0]$.*

[1] by [Combes et al. 2003, Proposition 1.3]; see also [Combes et al. 2007a, Theorem 2.1].

If $\delta_- \geq 2$, *we have the estimate*

$$K_M \leq C_{d,V_{\mathrm{per}},M_\rho}(1+E_0)^{4[\![\frac{d}{4}]\!]}. \tag{5-2}$$

Proof. Let Λ be a finite box. It follows from (2-13) that $\mathbb{E}_{\boldsymbol{\omega}}\{\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(\{c\})\} = 0$ for any $c \in \mathbb{R}$. Thus we may take all bounded intervals to be of the form $]a, b]$. For such an interval we modify Lemma 3.1 as follows: Given $\delta > 0$ small, we pick a nonincreasing function $h \in C^\infty(\mathbb{R})$, such that $h(t) = 1$ for $t \leq 0$ and $h(t) = 0$ for $t \geq \delta$. Note that $0 \leq h \leq 1$, $h' \leq 0$, $\operatorname{supp} h' \subset [0, \delta]$, $\int_{\mathbb{R}} \mathrm{d}t\, h'(t) = -1$, and we can choose h so $|h'| \leq \frac{2}{\delta}$. Given $c \in \mathbb{R}$, we set $h_c(t) = h(t-c)$, and note that $h_{c-\delta} \leq \chi_{]-\infty,c]} \leq h_c$. We let $I =]a, b]$, $I_\delta =]a-\delta, b+\delta]$. Using h, we rework (3-1) in the following way. Given $j \in \widetilde{\Lambda}$ and $\tau \geq M_\rho$, we have

$$\begin{aligned}\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) &\leq \operatorname{tr} h_b(H_{\boldsymbol{\omega}}^{(\Lambda)}) - \operatorname{tr} h_{a-\delta}(H_{\boldsymbol{\omega}}^{(\Lambda)}) \\ &\leq \bigl\{\operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=0)}\bigr) - \operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}\bigr)\bigr\} + \bigl\{\operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}\bigr) - \operatorname{tr} h_{a-\delta}\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}\bigr)\bigr\} \\ &\leq \bigl\{\operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=0)}\bigr) - \operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}\bigr)\bigr\} + \operatorname{tr} P^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}(I_\delta).\end{aligned} \tag{5-3}$$

We now fix $\tau = M_\rho$ and use the Birman–Solomyak formula [Simon 1998] as in [Combes et al. 2007b, (7)–(8)], plus the hypothesis (2-8), obtaining

$$\begin{aligned}\xi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp) &:= \operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=0)}\bigr) - \operatorname{tr} h_b\bigl(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}\bigr) \\ &= -\int_0^\tau \mathrm{d}s\, \operatorname{tr}\bigl\{\sqrt{u_j}\, h_b'(H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=s)})\sqrt{u_j}\bigr\} \\ &\leq \frac{2}{\delta}\int_0^\tau \mathrm{d}s\, \operatorname{tr}\bigl\{\sqrt{u_j}\, P^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=s)}(]b, b+\delta])\sqrt{u_j}\bigr\} \\ &\leq \frac{2}{\delta\rho_-}\int \mathrm{d}s\rho(s)\, \operatorname{tr}\bigl\{\sqrt{u_j}\, P^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=s)}(]b, b+\delta])\sqrt{u_j}\bigr\}.\end{aligned} \tag{5-4}$$

Note that $\xi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp)$ is closely related to the spectral shift function associated to the pair $H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=0)}$ and $H^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\omega_j=\tau)}$.

Now fix $E_0 > 0$, let $I =]a, b] \subset [0, E_0[$, and consider $\delta > 0$ such that $b + \delta \leq E_0$, so $I_\delta \subset [0, E_0]$. If $\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) \geq 1$, it follows from (4-4) that

$$(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) - 1) \leq Q_1 \sum_{j\in\widetilde{\Lambda}} \operatorname{tr}\bigl\{\sqrt{u_j}\, P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\sqrt{u_j}\, S_j^{(\Lambda)}\bigr\}(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) - 1), \tag{5-5}$$

so, using (5-3) and (5-4), we get

$$(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) - 1) \leq Q_1 \sum_{j\in\widetilde{\Lambda}} \bigl\{\bigl(\operatorname{tr}\bigl\{\sqrt{u_j}\, P_{\boldsymbol{\omega}}^{(\Lambda)}(I)\sqrt{u_j}\, S_j^{(\Lambda)}\bigr\}\bigr)\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp)\bigr\}, \tag{5-6}$$

where for each $j \in \widetilde{\Lambda}$

$$\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp) := \bigl(\xi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp) - 1\bigr) + \operatorname{tr} P^{(\Lambda)}_{(\boldsymbol{\omega}_j^\perp,\tau)}(I_\delta) \tag{5-7}$$

is independent of the random variable ω_j. If $\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I) < 1$, we have $P_{\boldsymbol{\omega}}^{(\Lambda)}(I) = 0$, and hence we also have (5-6).

Thus, if we now take the expectation in (5-6), use (3-5) and (4-5), we get

$$\begin{aligned}\mathbb{E}\big\{(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\big\} &\le Q_1 Q_2 \rho_+ |I| \sum_{j\in\widetilde{\Lambda}} \mathbb{E}_{\boldsymbol{\omega}_j^\perp}\big\{\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_j^\perp)\big\}\\ &= Q_1 Q_2 \rho_+ |I|\,|\Lambda| \mathbb{E}_{\boldsymbol{\omega}_k^\perp}\big\{\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_k^\perp)\big\} \qquad (5\text{-}8)\end{aligned}$$

for any $k \in \widetilde{\Lambda}$.

We will now estimate $\mathbb{E}_{\boldsymbol{\omega}_k^\perp}\big\{\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_k^\perp)\big\}$. It follows from (5-4) and (2-13) that, if we have (5-1),

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\omega}_k^\perp}\big\{\xi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_k^\perp)\big\} &\le \frac{2}{\delta\rho_-}\mathbb{E}_{\boldsymbol{\omega}}\big\{\operatorname{tr}\big\{\sqrt{u_k}\,P_{\boldsymbol{\omega}}^{(\Lambda)}(]b,b+\delta])\sqrt{u_k}\big\}\big\}\\ &= \frac{2}{\delta\rho_-|\Lambda|}\mathbb{E}_{\boldsymbol{\omega}}\Big\{\sum_{j\in\widetilde{\Lambda}}\operatorname{tr}\big\{\sqrt{u_j}\,P_{(\boldsymbol{\omega})}^{(\Lambda)}(]b,b+\delta])\sqrt{u_j}\big\}\Big\}\\ &\le \frac{2U_+}{\delta\rho_-|\Lambda|}\mathbb{E}_{\boldsymbol{\omega}}\big\{\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(]b,b+\delta])\big\} \le 2K_W U_+\frac{r}{r}\rho_+\rho_- \le 1. \qquad (5\text{-}9)\end{aligned}$$

In this case, we have

$$\mathbb{E}_{\boldsymbol{\omega}_k^\perp}\big\{\Phi_{b,\tau}^{(\Lambda)}(\boldsymbol{\omega}_k^\perp)\big\} \le \mathbb{E}_{\boldsymbol{\omega}_k^\perp}\big\{\operatorname{tr} P_{(\boldsymbol{\omega}_k^\perp,\tau)}^{(\Lambda)}(I_\delta)\big\} \le \widetilde{K}_W\rho_+(|I|+2\delta)|\Lambda|, \qquad (5\text{-}10)$$

where we used Lemma 4.2, where $\widetilde{K}_W = \widetilde{K}_W(d, u, V_{\text{per}}, E_0, M_\rho)$.

Combining (5-8) and (5-10) we get

$$\mathbb{E}\big\{(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I))(\operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda)}(I)-1)\big\} \le Q_1 Q_2 \widetilde{K}_W |I|(|I|+2\delta)(\rho_+|\Lambda|)^2. \qquad (5\text{-}11)$$

Letting $\delta \to 0$ we get (2-19) with $K_M = Q_1 Q_2 \widetilde{K}_W$.

If $\delta_- \ge 2$, the estimate (5-2) follows from (4-7) and (4-37). ☐

6. Poisson statistics

In this section we prove Theorem 2.3(a).

Let $H_{\boldsymbol{\omega}}$ be an Anderson Hamiltonian, and suppose $\mathcal{I}$ is an open interval such that for all large boxes Λ the estimate (2-19) holds for any interval $I \subset \mathcal{I}$ with $|I| \le \delta_0$, for some $\delta_0 > 0$, with some constant K_M. (We will assume that a given Λ is large enough.) Recall we have (2-13) for these intervals with some constant K_W.

Let $\mathcal{E} \in \mathcal{I} \cap \Xi^{\text{CL}}$ be such that the IDS $N(E)$ is differentiable at $\mathcal{E}$ with $n(\mathcal{E}) := N'(\mathcal{E}) > 0$. It follows from (2-13) that we then have

$$0 < n(\mathcal{E}) \le K_W\rho_+. \qquad (6\text{-}1)$$

We fix an open interval $\mathcal{I}_1$ such that $\mathcal{E} \in \mathcal{I}_1 \subset \overline{\mathcal{I}}_1 \subset \mathcal{I} \cap \Xi^{\text{CL}}$. Note that for each bounded Borel set $B \subset \mathbb{R}$ there exists a finite $c_B = c_{B,\mathcal{E},\mathcal{I}_1}$ such that $\mathcal{E} + |\Lambda|^{-1}B \subset \mathcal{I}_1$ and $|\mathcal{E} + |\Lambda|^{-1}B| \le \delta_0$ if $|\Lambda| \ge c_B$. The point process $\xi_{\boldsymbol{\omega}}^{(\Lambda)} = \xi_{\mathcal{E},\boldsymbol{\omega}}^{(\Lambda)}$ of (2-17) has an intensity measure given by $\nu^{(\Lambda)}(B) := \mathbb{E}\,\xi_{\boldsymbol{\omega}}^{(\Lambda)}(B)$ for a Borel set $B \subset \mathbb{R}$; it follows from (2-13) that,

$$\nu^{(\Lambda)}(B) \le K_W\rho_+|B| \quad \text{for all } \Lambda \text{ with } |\Lambda| \ge c_B. \qquad (6\text{-}2)$$

We start with the same general strategy used in [Molchanov 1980/81; Minami 1996]. We fix $a \in]0, 1[$, and divide $\Lambda = \Lambda_L(0)$ into M_L boxes $\Lambda^{(m)} = \Lambda_\ell(k_m)$ of side $\ell \approx L^a$, $\ell \in 2\mathbb{N}$, centered at $k_m \in \Lambda \cap (2\mathbb{Z}^d)$; note $M_L = \frac{|\Lambda_L|}{|\Lambda_\ell|} \approx L^{(1-a)d}$. For each $m = 1, 2, \ldots, M_L$ we define point processes

$$\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(B) := \operatorname{tr} P_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}(\mathscr{E} + |\Lambda|^{-1} B) \quad \text{for a Borel set } B \subset \mathbb{R}. \tag{6-3}$$

Note that $\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}\}_{m=1,2,\ldots,M_L}$ are independent, identically distributed point processes, each with intensity measure (using (2-13))

$$\nu^{(\Lambda,m)}(B) := \mathbb{E}\, \xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(B) \le K_W \rho_+ |B| M_L^{-1} \quad \text{for all } \Lambda \text{ with } |\Lambda| \ge c_B. \tag{6-4}$$

We consider their superposition, the point process

$$\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)} := \sum_{m=1}^{M_L} \xi_{\boldsymbol{\omega}}^{(\Lambda,m)}, \tag{6-5}$$

with intensity measure

$$\widetilde{\nu}^{(\Lambda)}(B) := \mathbb{E}\, \widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(B) \le K_W \rho_+ |B| \quad \text{for all } \Lambda \text{ with } |\Lambda| \ge c_B. \tag{6-6}$$

We will prove that $\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)} \approx \xi_{\boldsymbol{\omega}}^{(\Lambda)}$ as $L \to \infty$, and that $\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}$ converges weakly, as $L \to \infty$, to the Poisson point process ξ with intensity measure $\nu(B) := \mathbb{E}\, \xi(B) = n(\mathscr{E})|B|$. But here we must use different methods from [Molchanov 1980/81; Minami 1996].

So let $\theta_{\boldsymbol{\omega}}^{(\Lambda)} = \theta_{\mathscr{E},\boldsymbol{\omega}}^{(\Lambda)}$ be the random measure defined in (2-24); its intensity measure is

$$\eta^{(\Lambda)}(B) := \mathbb{E}\, \theta_{\boldsymbol{\omega}}^{(\Lambda)}(B) = |\Lambda| \eta(\mathscr{E} + |\Lambda|^{-1} B), \tag{6-7}$$

where η is the density of states measure, given in (2-16). It again follows from (2-13) that

$$\eta^{(\Lambda)}(B) \le K_W \rho_+ |B| \quad \text{for all } \Lambda \text{ with } |\Lambda| \ge c_B. \tag{6-8}$$

We start with a lemma. Given a measure η on $\mathbb{R}$, we write $\eta(f) := \int_{\mathbb{R}} f \,\mathrm{d}\eta$ for suitable functions f, say, $f \in \mathscr{F}_{b,K}$, the collection of bounded Borel functions on $\mathbb{R}$ vanishing outside a compact interval. It follows from (2-17) that for all $f \in \mathscr{F}_{b,K}$ we have

$$\xi_{\boldsymbol{\omega}}^{(\Lambda)}(f) = \operatorname{tr} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}), \quad \text{where} \quad f_\Lambda(E) := f(|\Lambda|(E - \mathscr{E})), \tag{6-9}$$

with similar expressions for $\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(f)$, $\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(f)$, and $\theta_{\boldsymbol{\omega}}^{(\Lambda)}(f)$.

Lemma 6.1. *For all $f \in \mathscr{F}_{b,K}$ we have*

$$\lim_{L\to\infty} \mathbb{E}\left|\xi_{\boldsymbol{\omega}}^{(\Lambda)}(f) - \widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(f)\right| = 0 \tag{6-10}$$

and

$$\lim_{L\to\infty} \mathbb{E}\left|\xi_{\boldsymbol{\omega}}^{(\Lambda)}(f) - \theta_{\boldsymbol{\omega}}^{(\Lambda)}(f)\right| = 0. \tag{6-11}$$

Proof. In view of (6-2), (6-6), and (6-8), it suffices to prove (6-10) and (6-11) for $f \in C_K^\infty(\mathbb{R})$, since $\{f \in C_K^\infty(\mathbb{R}) : \operatorname{supp} f \subset J\}$ is dense in $\mathrm{L}^1(J, \mathrm{d}E)$ for any interval J.

So let $f \in C_K^\infty(\mathbb{R})$. To prove (6-10), we set $\ell' \approx \ell - \sqrt{\ell}$, $\Lambda^{(m,\prime)} = \Lambda_{\ell'}(k_m)$, and $\Lambda^{(m,\prime\prime)} = \Lambda_\ell(k_m) \setminus \Lambda_{\ell'}(k_m)$. Using $\chi_\Lambda = \sum_{m=1}^{M_L} \chi_{\Lambda^{(m)}}$, we get

$$\begin{aligned}
\xi_{\boldsymbol{\omega}}^{(\Lambda)}(f) - \widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(f) &= \sum_{m=1}^{M_L} \Bigl(\operatorname{tr}\bigl\{\chi_{\Lambda^{(m)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m)}}\bigr\} - \operatorname{tr} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})})\Bigr) \\
&= \sum_{m=1}^{M_L} \Bigl(\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m,\prime)}}\bigr\} - \operatorname{tr}\{\chi_{\Lambda^{(m,\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}) \chi_{\Lambda^{(m,\prime)}}\}\Bigr) \\
&\qquad + \sum_{m=1}^{M_L} \Bigl(\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\} - \operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\}\Bigr). \qquad (6\text{-}12)
\end{aligned}$$

We now use the fact that the expectation is invariant under translations in the torus to get, for any m,

$$\begin{aligned}
\mathbb{E}\bigl|\xi_{\boldsymbol{\omega}}^{(\Lambda)}(f) - \widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(f)\bigr| &\le M_L \mathbb{E}\bigl|\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m,\prime)}}\bigr\} - \operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}) \chi_{\Lambda^{(m,\prime)}}\bigr\}\bigr| \qquad (6\text{-}13) \\
&\quad + M_L \mathbb{E}\bigl|\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\} - \operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\}\bigr|. \qquad (6\text{-}14)
\end{aligned}$$

It follows from the Wegner estimate (2-13) that

$$\begin{aligned}
M_L \mathbb{E}\bigl|\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\}\bigr| &\le M_L \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda|} \mathbb{E}\operatorname{tr}\bigl\{|f_\Lambda|(H_{\boldsymbol{\omega}}^{(\Lambda)})\bigr\} \\
&\le M_L \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda|} K_W \rho_+ |\Lambda| \int_{\mathbb{R}} |f_\Lambda|(E)\,\mathrm{d}E \\
&= \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda^{(m)}|} K_W \rho_+ \|f\|_1. \qquad (6\text{-}15)
\end{aligned}$$

Similarly,

$$\begin{aligned}
M_L \mathbb{E}\bigl|\operatorname{tr}\bigl\{\chi_{\Lambda^{(m,\prime\prime)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})}) \chi_{\Lambda^{(m,\prime\prime)}}\bigr\}\bigr| &\le M_L \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda^{(m)}|} \mathbb{E}\operatorname{tr}\bigl\{|f_\Lambda|(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})})\bigr\} \\
&\le M_L \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda^{(m)}|} K_W \rho_+ |\Lambda^{(m)}| \int_{\mathbb{R}} |f_\Lambda|(E)\,\mathrm{d}E \\
&= \frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda^{(m)}|} K_W \rho_+ \|f\|_1. \qquad (6\text{-}16)
\end{aligned}$$

Since

$$\frac{|\Lambda^{(m,\prime\prime)}|}{|\Lambda^{(m)}|} \approx \frac{\ell^{d-1}\sqrt{\ell}}{\ell^d} = \frac{1}{\sqrt{\ell}} \approx \frac{1}{L^{\frac{a}{2}}} \to 0 \quad \text{as} \quad L \to \infty, \qquad (6\text{-}17)$$

the term in (6-14) goes to 0 as $L \to \infty$.

To finish the proof of (6-10) we need to show that the term in (6-13) also goes to 0 as $L \to \infty$. To do that we will use that $\overline{\mathcal{I}}_1 \subset \Xi^{\mathrm{CL}}$, the Helffer–Sjöstrand formula for smooth functions of self-adjoint operators, and estimates on Schrödinger operators.

Given a box Λ, we identify $\mathrm{L}^2(\Lambda)$ with the subspace of $\mathrm{L}^2(\mathbb{R}^d)$ consisting of functions vanishing outside Λ. Given a function $\phi \in C_K^\infty(\mathbb{R})$, we let $W(\phi)$ to be the closure of the local first order differential operator $[\Delta, \phi]$ on $C_K^\infty(\mathbb{R})$. We set

$$\chi_\phi := \chi_{\operatorname{supp}\phi}, \qquad \chi_{\nabla\phi} := \chi_{\operatorname{supp}\nabla\phi},$$

and note that $W(\phi) = \chi_{\nabla\phi} W(\phi) = W(\phi)\chi_{\nabla\phi} = \chi_{\nabla\phi} W(\phi)\chi_{\nabla\phi}$. We recall that if $\operatorname{supp}\phi \subset \Lambda^\circ$, the interior of Λ, which here may be either a finite box or $\mathbb{R}^d$, we have

$$\big\|(H_{\boldsymbol{\omega}}^{(\Lambda)}+1)^{-1/2}W(\phi)\big\| = \big\|W(\phi)(H_{\boldsymbol{\omega}}^{(\Lambda)}+1)^{-1/2}\big\| \le C_\phi := C_1\big(\|\Delta\phi\|_\infty + \|\nabla\phi\|_\infty\big), \tag{6-18}$$

where C_1 depends only on d. We also recall that for all $x \in \Lambda$ we have

$$\big\|\chi_{\Lambda_1(x)}(H_{\boldsymbol{\omega}}^{(\Lambda)}+1)^{-1}\big\|_{p_d} \le C_2 < \infty \quad \text{with } p_d = [\tfrac{d}{2}]+1, \tag{6-19}$$

the constant C_2 being independent of x and Λ for $L \ge 2$ [Klein et al. 2002, (130)–(136)].

We now recall the Helffer–Sjöstrand formula; refer to [Hunziker and Sigal 2000, Appendix B] for details. Given $g \in C^\infty(\mathbb{R})$ and $m \in \mathbb{N}$, we set

$$\{\!\{g\}\!\}_m := \sum_{r=0}^{m} \int_{\mathbb{R}} \mathrm{d}u\, |g^{(r)}(u)|\,(1+|u|^2)^{(r-1)/2}. \tag{6-20}$$

If $\{\!\{g\}\!\}_m < \infty$ with $m \ge 2$, then for any self-adjoint operator K we have

$$f(K) = \int_{\mathbb{R}^2} \mathrm{d}\tilde{g}(z)\,(K-z)^{-1}, \tag{6-21}$$

where the integral converges absolutely in operator norm. Here $z = x+iy$, $\tilde{g}(z)$ is an *almost analytic extension* of g to the complex plane, $\mathrm{d}\tilde{g}(z) := \frac{1}{2\pi}\partial_{\bar{z}}\tilde{g}(z)\,\mathrm{d}x\,\mathrm{d}y$ with $\partial_{\bar{z}} = \partial_x + i\partial_y$, and $|\mathrm{d}\tilde{g}(z)| := (2\pi)^{-1}|\partial_{\bar{z}}\tilde{g}(z)|\,\mathrm{d}x\,\mathrm{d}y$. Moreover, for all $p \ge 0$ we have

$$\int_{\mathbb{R}^2} |\mathrm{d}\tilde{g}(z)| \frac{1}{|\Im z|^p} \le c_p\ \{\!\{g\}\!\}_m < \infty \quad \text{for} \quad m \ge p+1 \tag{6-22}$$

with a constant c_p.

Since $f \in C_K^\infty(\mathbb{R})$, we have, using the Helffer–Sjöstrand formula with $\Lambda = \Lambda_L$, $R_{\boldsymbol{\omega}}^{(\Lambda)}(z) = (H_{\boldsymbol{\omega}}^{(\Lambda)} - z)^{-1}$ and $R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) = (H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})} - z)^{-1}$, and taking $\phi_0 \in C_K^\infty(\Lambda_{\ell-10d}(k_m))$ such that $\phi_0\chi_{\Lambda_{\ell-20d}(k_m)} = \chi_{\Lambda_{\ell-20d}(k_m)}$ and $0 \le \phi_0 \le 1$, that

$$\begin{aligned} T_{\boldsymbol{\omega}}^{(\Lambda)} &:= \chi_{\Lambda^{(m,\ell)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda)})\chi_{\Lambda^{(m,\ell)}} - \chi_{\Lambda^{(m,\ell)}} f_\Lambda(H_{\boldsymbol{\omega}}^{(\Lambda^{(m)})})\chi_{\Lambda^{(m,\ell)}} \qquad (6\text{-}23)\\ &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z)\big\{\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\chi_{\Lambda^{(m,\ell)}} - \chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z)\chi_{\Lambda^{(m,\ell)}}\big\}\\ &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z)\big\{\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\phi_0\chi_{\Lambda^{(m,\ell)}} - \chi_{\Lambda^{(m,\ell)}}\phi_0 R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z)\chi_{\Lambda^{(m,\ell)}}\big\}\\ &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z)\big\{\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)W(\phi_0)R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z)\chi_{\Lambda^{(m,\ell)}}\big\}, \qquad (6\text{-}24) \end{aligned}$$

where we used the geometric resolvent identity.

Now let us pick functions $\phi_i \in C_K^\infty(\mathbb{R})$, $i = 1, 2, \dots, 2p-1$, such that $0 \le \phi_i \le 1$, $\phi_i \chi_{\nabla\phi_{i-1}} = \chi_{\nabla\phi_{i-1}}$, and $\chi_{\phi_i}\chi_{\Lambda_{\ell-30d}(k_m)} = 0$ for $i = 1, 2, \dots, 2p-1$. Using the resolvent identity $2p-1$ times we get

$$\begin{aligned}
&\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_0) \\
&\quad = \chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_{2p-1}) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_{2p-2}) \dots R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_1) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_0) \\
&\quad = \bigl\{\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\bigr\}\bigl\{W(\phi_{2p-1}) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_{2p-2})\bigr\}\bigl\{\chi_{\nabla\phi_{2p-2}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\bigr\} \\
&\qquad\qquad \times \bigl\{W(\phi_{2p-3}) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_{2p-4})\bigr\} \dots \bigl\{\chi_{\nabla\phi_2} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\bigr\}\bigl\{W(\phi_1) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_0)\bigr\}.
\end{aligned} \tag{6-25}$$

We now use that the integral in (6-24) is performed over a compact domain in $\mathbb{R}^2$, which depends only on the function f, so there is constant C_f such that for z in the region of integration we have

$$\bigl\|(H_{\boldsymbol{\omega}}^{(\Lambda)} + 1) R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\bigr\| \le \frac{C_f}{|\Im z|}, \tag{6-26}$$

and hence, using (6-18) and (6-19), we have

$$\bigl\|W(\phi_i) R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_{i-1})\bigr\| \le \frac{C_f C_{\phi_i} C_{\phi_{i-1}}}{|\Im z|} \tag{6-27}$$

and, for $B \subset \Lambda_{L'} \subset \Lambda$,

$$\bigl\|\chi_B R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\bigr\|_{p_d} \le \frac{C_f C_2}{|\Im z|} |\Lambda_{L'}|. \tag{6-28}$$

We now choose $p = p_d$ as in (6-19), and note that we can choose the functions $\phi_i \in C_K^\infty(\mathbb{R})$, $i = 1, 2, \dots, 2p_d - 1$ so that the constants C_{ϕ_i} are independent of Λ, say all $C_{\phi_i} \le C_3$ From (6-25), (6-27), and (6-28), we get

$$\begin{aligned}
\bigl\|\chi_{\Lambda^{(m,\ell)}} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) W(\phi_0) R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) \chi_{\Lambda^{(m,\ell)}}\bigr\|_1 &\le \left(\frac{C_f C_2}{|\Im z|}|\Lambda^{(m)}|\right)^{p_d} \left(\frac{C_f C_3^2}{|\Im z|}\right)^{p_d} \bigl\|\chi_{\nabla\phi_0} R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) \chi_{\Lambda^{(m,\ell)}}\bigr\| \\
&\le C_4 C_f' \ell^{p_d} |\Im z|^{-2p_d} \bigl\|\chi_{\nabla\phi_0} R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) \chi_{\Lambda^{(m,\ell)}}\bigr\|.
\end{aligned} \tag{6-29}$$

We now use that $\overline{\mathcal{I}}_1 \subset \Xi^{\mathrm{CL}}$, the region of complete localization for $H_{\boldsymbol{\omega}}$. The term in (6-13) is $M_L \mathbb{E}\{T_{\boldsymbol{\omega}}^{(\Lambda)}\}$, with $T_{\boldsymbol{\omega}}^{(\Lambda)}$ as in (6-23). It follows from (6-24), (6-25), and (6-29) that for large L,

$$\begin{aligned}
M_L \mathbb{E}\{T_{\boldsymbol{\omega}}^{(\Lambda)}\} &\le M_L C_4 C_f' \ell^{p_d} \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\, |\Im z|^{-2p_d} \mathbb{E}\bigl\{\bigl\|\chi_{\nabla\phi_0} R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) \chi_{\Lambda^{(m,\ell)}}\bigr\|\bigr\} \\
&\le M_L C_4 C_f' \ell^{p_d} \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\, |\Im z|^{-2p_d - \frac{4}{5}} \mathbb{E}\bigl\{\bigl\|\chi_{\nabla\phi_0} R_{\boldsymbol{\omega}}^{(\Lambda,m)}(z) \chi_{\Lambda^{(m,\ell)}}\bigr\|^{1/5}\bigr\} \\
&\le M_L C_4 C_f' \ell^{p_d + 2d} (\rho_+ + \sqrt{\rho_+}) \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\, |\Im z|^{-2p_d - \frac{4}{5}} \mathrm{e}^{-\ell^{1/4}} \\
&\le L^d \ell^{p_d + d} \mathrm{e}^{-\ell^{1/4}} c_{2p_d + \frac{4}{5}} C_4 C_f' (\rho_+ + \sqrt{\rho_+})\, \{\!\{f_\Lambda\}\!\}_{2p_d + 2}\,.
\end{aligned} \tag{6-30}$$

where we used (A-4) and (6-22). Note that $2p_d \le d + 1$ and

$$\{\!\{f_\Lambda\}\!\}_m \le C_{E_0, f, m} |\Lambda|^{m-1} \quad \text{for all} \quad m = 2, 3, \dots. \tag{6-31}$$

It follows that

$$M_L \mathbb{E}\{T_\omega^{(\Lambda)}\} \le L^{d^2+3d} \ell^{3d/2+1} \mathrm{e}^{-\ell^{1/4}} c_{2p_d+\frac{4}{5}} C_{f,E_0,d}(\rho_+ + \sqrt{\rho_+}) \to 0 \quad \text{as} \quad L \to \infty. \tag{6-32}$$

Thus (6-10) is proven.

The proof of (6-11) is similar. With $\Lambda = \Lambda_L(0)$, we set $L' \approx L - \sqrt{L}$, $\Lambda' = \Lambda_{L'}(0)$, and $\Lambda'' = \Lambda \setminus \Lambda'$. We have

$$\begin{aligned} \theta_\omega^{(\Lambda)}(f) - \xi_\omega^{(\Lambda)}(f) &= \operatorname{tr}\{\chi_\Lambda f_\Lambda(H_\omega)\chi_\Lambda\} - \operatorname{tr} f_\Lambda(H_\omega^{(\Lambda)}) \\ &= \big(\operatorname{tr}\{\chi_{\Lambda'} f_\Lambda(H_\omega)\chi_{\Lambda'}\} - \operatorname{tr}\{\chi_{\Lambda'} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda'}\}\big) \\ &\quad + \big(\operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega)\chi_{\Lambda''}\} - \operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda''}\}\big), \end{aligned} \tag{6-33}$$

and hence

$$\mathbb{E}\big|\theta_\omega^{(\Lambda)}(f) - \xi_\omega^{(\Lambda)}(f)\big| \le \mathbb{E}\big|\operatorname{tr}\{\chi_{\Lambda'} f_\Lambda(H_\omega)\chi_{\Lambda'}\} - \operatorname{tr}\{\chi_{\Lambda'} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda'}\}\big| \tag{6-34}$$

$$+ \mathbb{E}\big|\operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega)\chi_{\Lambda''}\} - \operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda''}\}\big|. \tag{6-35}$$

We now use the Wegner estimate (2-13) to obtain

$$\begin{aligned} \mathbb{E}\big|\operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda''}\}\big| &\le \frac{|\Lambda''|}{|\Lambda|} \mathbb{E} \operatorname{tr}\big\{|f_\Lambda|(H_\omega^{(\Lambda)})\big\} \\ &\le \frac{|\Lambda''|}{|\Lambda|} K_W \rho_+ |\Lambda| \int_{\mathbb{R}} |f_\Lambda|(E)\, \mathrm{d}E = \frac{|\Lambda''|}{|\Lambda|} K_W \rho_+ \|f\|_1, \end{aligned} \tag{6-36}$$

and

$$\begin{aligned} \mathbb{E}\big|\operatorname{tr}\{\chi_{\Lambda''} f_\Lambda(H_\omega)\chi_{\Lambda''}\}\big| &\le |\Lambda''| \mathbb{E} \operatorname{tr}\big\{\chi_0 |f_\Lambda|(H_\omega^{(\Lambda)})\chi_0\big\} = |\Lambda''| N(|f_\Lambda|) \\ &\le |\Lambda''| K_W \rho_+ \int_{\mathbb{R}} |f_\Lambda|(E)\, \mathrm{d}E = \frac{|\Lambda''|}{|\Lambda|} K_W \rho_+ \|f\|_1. \end{aligned} \tag{6-37}$$

Since $|\Lambda''|/|\Lambda| \approx 1/\sqrt{L}$, the term in (6-35) goes to 0 as $L \to \infty$.

To finish the proof of (6-11) , we need to show that the term in (6-34) also goes to 0 as $L \to \infty$. As before, we use the Helffer–Sjöstrand formula. We have, taking $\phi_0 \in C_K^\infty(\Lambda_{L-10d}(0))$ such that $0 \le \phi_0 \le 1$ and $\phi_0 \chi_{\Lambda_{L-20d}(0)} = \chi_{\Lambda_{L-20d}(0)}$, that

$$S_\omega^{(\Lambda)} := \chi_{\Lambda'} f_\Lambda(H_\omega)\chi_{\Lambda'} - \chi_{\Lambda'} f_\Lambda(H_\omega^{(\Lambda)})\chi_{\Lambda'} \tag{6-38}$$

$$\begin{aligned} &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z) \big\{\chi_{\Lambda'} R_\omega(z)\chi_{\Lambda'} - \chi_{\Lambda'} R_\omega^{(\Lambda)}(z)\chi_{\Lambda'}\big\} \\ &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z) \big\{\chi_{\Lambda'} R_\omega(z)\phi_0\chi_{\Lambda'} - \chi_{\Lambda'}\phi_0 R_\omega^{(\Lambda)}(z)\chi_{\Lambda'}\big\} \\ &= \int_{\mathbb{R}^2} \mathrm{d}\tilde{f}_\Lambda(z) \big\{\chi_{\Lambda'} R_\omega(z) W(\phi_0) R_\omega^{(\Lambda)}(z)\chi_{\Lambda'}\big\}. \end{aligned} \tag{6-39}$$

Proceeding as in (6-25)–(6-29), we get

$$\big\|\chi_{\Lambda'} R_\omega(z) W(\phi_0) R_\omega^{(\Lambda)}(z)\chi_{\Lambda'}\big\|_1 \le C_4 C_f' L^{p_d} |\Im z|^{-2p_d} \|\chi_{\nabla\phi_0} R_\omega^{(\Lambda)}(z)\chi_{\Lambda'}\|. \tag{6-40}$$

Recall that $\overline{\mathcal{F}}_1 \subset \Xi^{\mathrm{CL}}$. The term in (6-34) is $\mathbb{E}\{S_{\omega}^{(\Lambda)}\}$, with $S_{\omega}^{(\Lambda)}$ as in (6-38). It follows from (6-39) and (6-40) that for large L,

$$\begin{aligned}
\mathbb{E}\{S_{\omega}^{(\Lambda)}\} &\le C_4 C'_f L^{p_d} \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\,|\Im z|^{-2p_d}\,\mathbb{E}\big\{\|\chi_{\nabla\phi_0} R_{\omega}^{(\Lambda)}(z)\chi_{\Lambda'}\|\big\}\\
&\le M_L C_4 C'_f L^{p_d} \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\,|\Im z|^{-2p_d-\frac{4}{5}}\,\mathbb{E}\big\{\|\chi_{\nabla\phi_0} R_{\omega}^{(\Lambda)}(z)\chi_{\Lambda'}\|^{1/5}\big\}\\
&\le C_4 C'_f L^{p_d+2d}(\rho_+ + \sqrt{\rho_+}) \int_{\mathbb{R}^2} |\mathrm{d}\tilde{f}_\Lambda(z)|\,|\Im z|^{-2p_d-\frac{4}{5}}\,\mathrm{e}^{-L^{1/4}}\\
&\le L^{p_d+2d}\mathrm{e}^{-L^{1/4}} c_{2p_d+\frac{4}{5}} C_4 C'_f (\rho_+ + \sqrt{\rho_+})\, \{\!\{ f_\Lambda \}\!\}_{2p_d+2}\\
&\le L^{d^2+5d}\mathrm{e}^{-L^{1/4}} c_{2p_d+\frac{4}{5}} C_{f,E_0,d}(\rho_+ + \sqrt{\rho_+}) \to 0 \quad \text{as} \quad L\to\infty,
\end{aligned} \tag{6-41}$$

where we used (A-4) and (6-22).

Thus (6-11) is proven, and with it the lemma. □

Given point processes $\{\zeta_n\}_{n\in\mathbb{N}}$ and ζ on $\mathbb{R}$, we let $\zeta_n \Rightarrow \zeta$ denote the weak convergence of ζ_n to ζ as $n\to\infty$. We recall [Daley and Vere-Jones 1988, Proposition 9.1.VII] that $\zeta_n \Rightarrow \zeta$ if and only if

$$\lim_{n\to\infty} \mathbb{E}\,\mathrm{e}^{-\zeta_n(f)} = \mathbb{E}\,\mathrm{e}^{-\zeta(f)} \quad \text{for all} \quad f\in C_{K,+}(\mathbb{R}). \tag{6-42}$$

The following lemma shows that it suffices to prove that $\widetilde{\xi}_{\omega}^{(\Lambda)} \Rightarrow \xi$ to prove Theorem 2.3(b).

Lemma 6.2. *$\xi_{\omega}^{(\Lambda)} \Rightarrow \xi$ if and only if $\widetilde{\xi}_{\omega}^{(\Lambda)} \Rightarrow \xi$.*

Proof. If ζ_i, $i=1,2$, are point processes on $\mathbb{R}$, defined on the same probability space, we have, for all $f\in C_{K,+}(\mathbb{R})$,

$$\big|\mathbb{E}\mathrm{e}^{-\zeta_1(f)} - \mathbb{E}\mathrm{e}^{-\zeta_2(f)}\big| \le \mathbb{E}\big|\zeta_1(f)-\zeta_2(f)\big|. \tag{6-43}$$

The lemma follows immediately from (6-42), (6-43), and Lemma 6.1. □

We are now ready to prove Theorem 2.3(a). In view of Lemma 6.2, it suffices to prove that $\widetilde{\xi}_{\omega}^{(\Lambda)} \Rightarrow \xi$. By standard results from the theory of point processes (cf. [Daley and Vere-Jones 1988, Theorem 9.2.V and subsequent remark]; see also [Kritchevski 2008, Theorem 2.3]), this is equivalent to verifying the following three conditions for all bounded intervals I (recall $\Lambda = \Lambda_L(0)$):

$$\lim_{L\to\infty} \max_{m=1,2,\dots,M_L} \mathbb{P}\{\xi_{\omega}^{(\Lambda,m)}(I)\ge 1\} = 0, \tag{6-44}$$

$$\lim_{L\to\infty} \sum_{m=1}^{M_L} \mathbb{P}\{\xi_{\omega}^{(\Lambda,m)}(I)\ge 1\} = n(\mathscr{E})|I|, \tag{6-45}$$

$$\lim_{L\to\infty} \sum_{m=1}^{M_L} \mathbb{P}\{\xi_{\omega}^{(\Lambda,m)}(I)\ge 2\} = 0. \tag{6-46}$$

Since $\mathbb{P}\{\xi_{\omega}^{(\Lambda,m)}(I)\ge 1\} \le \mathbb{E}\{\xi_{\omega}^{(\Lambda,m)}(I)\}$, (6-44) follows immediately from (6-4). In addition, it follows from the definition (6-3) and the estimate (2-19), that for all Λ with $|\Lambda|\ge c_I$ we have

$$\mathbb{P}\{\xi_{\omega}^{(\Lambda,m)}(I)\ge 2\} \le \tfrac{1}{2}\mathbb{E}\big\{(\xi_{\omega}^{(\Lambda,m)}(I))(\xi_{\omega}^{(\Lambda,m)}(I)-1)\big\} \le \tfrac{1}{2}K_M(\rho_+|I|M_L^{-1})^2, \tag{6-47}$$

so (6-46) follows.

Thus Theorem 2.3(a) is proved if we verify condition (6-45). To do so, we first notice that

$$\mathbb{E}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I)\} = \sum_{k=1}^{\infty} \mathbb{P}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I) \geq k\}, \tag{6-48}$$

and, as in [Kritchevski 2008],

$$\begin{aligned}\sum_{k=2}^{\infty} \mathbb{P}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I) \geq k\} &= \sum_{k=2}^{\infty} (k-1)\mathbb{P}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I) = k\} \\ &\leq \sum_{k=2}^{\infty} k(k-1)\mathbb{P}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I) = k\} = \mathbb{E}\big\{(\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I))(\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I)-1)\big\}.\end{aligned} \tag{6-49}$$

It thus follows, as in (6-47), that

$$0 \leq \mathbb{E}\{\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(I)\} - \sum_{m=1}^{M_L} \mathbb{P}\{\xi_{\boldsymbol{\omega}}^{(\Lambda,m)}(I) \geq 1\} \leq M_L K_M (\rho_+ |I| M_L^{-1})^2 \to 0 \quad \text{as} \quad L \to \infty. \tag{6-50}$$

We conclude that (6-45) is equivalent to

$$\lim_{L\to\infty} \mathbb{E}\{\widetilde{\xi}_{\boldsymbol{\omega}}^{(\Lambda)}(I)\} = n(\mathscr{E})|I|, \tag{6-51}$$

and hence, by Lemma 6.1, equivalent to

$$\lim_{L\to\infty} \mathbb{E}\{\theta_{\boldsymbol{\omega}}^{(\Lambda)}(I)\} = n(\mathscr{E})|I|. \tag{6-52}$$

But it follows from (6-7) that, for all Λ such that $|\Lambda| \geq c_I$

$$\mathbb{E}\{\theta_{\boldsymbol{\omega}}^{(\Lambda)}(I)\} = |\Lambda|\eta(\mathscr{E} + |\Lambda|^{-1} I) = |\Lambda| \int_{\mathscr{E}+|\Lambda|^{-1}I} n(E)\, \mathrm{d}E. \tag{6-53}$$

Since by our hypothesis $\mathscr{E}$ is a Lebesgue point of the locally integrable function $n(E)$ (cf. [Yeh 2006, Definition 25.13]), and the sets $\mathscr{E}+|\Lambda|^{-1}I$ shrink nicely to $\mathscr{E}$ as $L \to \infty$ (cf. [Yeh 2006, Definition 25.16]), we can use the Lebesgue Differentiation Theorem (cf. [Yeh 2006, Theorem 25.17]) to conclude that

$$\lim_{L\to\infty} |\Lambda| \int_{\mathscr{E}+|\Lambda|^{-1}I} n(E)\mathrm{d}E = n(\mathscr{E})|I|. \tag{6-54}$$

Thus (6-52), and hence (6-45), is proven, completing the proof of Theorem 2.3(a).

7. Simplicity of eigenvalues

We prove Theorem 2.3(b) proceeding as in [Klein and Molchanov 2006]. Let $H_{\boldsymbol{\omega}}$ be an Anderson Hamiltonian, and let $\mathscr{I}$ be an open interval such that for large boxes Λ the estimate (2-19) holds for any interval $I \subset \mathscr{I}$ with $|I| \leq \delta_0$, for some $\delta_0 > 0$, with some constant K_M. We call $\varphi \in \mathrm{L}^2(\mathbb{R}^d)$ fast decaying if it has β-decay for some $\beta > \frac{5}{2}d$, which in the continuum means that $\|\chi_x^{(1)}\varphi\| \leq C_\varphi \langle x\rangle^{-\beta}$ for some constant C_φ, where $\langle x\rangle := \sqrt{1+|x|^2}$. We will show that, with probability one, $H_{\boldsymbol{\omega}}$ cannot have an eigenvalue in $\mathscr{I}$ with 2 linearly independent fast decaying eigenfunctions.

Let $I \subset \mathcal{I}$ be a closed interval, $q > 2d$, $L \in 2\mathbb{N}$ large, $\Lambda_L = \Lambda_L(0)$. We cover the interval I by $2\big([L^q/2|I|]+1\big) \le L^q|I|+2$ intervals of length $2L^{-q}$, in such a way that any subinterval $J \subset I$ with length $|J| \le L^{-q}$ will be contained in one of these intervals. ($[x]$ denotes the largest integer $\le x$.) Let $\mathcal{B}_{L,I,q}$ denote the complement to the event that $\operatorname{tr} P_{\omega}^{(\Lambda_L)}(J) \le 1$ for all subintervals $J \subset I$ with length $|J| \le L^{-q}$. The probability of $\mathcal{B}_{L,I,q}$ can be estimated, using (2-19) and

$$\mathbb{P}\{\operatorname{tr} P_{\omega}^{(\Lambda)}(I) \ge 2\} \le \tfrac{1}{2}\mathbb{E}\big\{(\operatorname{tr} P_{\omega}^{(\Lambda)}(I))(\operatorname{tr} P_{\omega}^{(\Lambda)}(I)-1)\big\}, \tag{7-1}$$

by

$$\mathbb{P}\{\mathcal{B}_{L,I,q}\} \le \tfrac{1}{2}K_M\rho_+^2(L^q|I|+2)(2L^{-q})^2L^{2d} \le 2K_M\rho_+^2(|I|+1)L^{-q+2d}. \tag{7-2}$$

Thus, taking scales $L_k = 2^k$, $k = 1, 2, \ldots$, it follows from the Borel–Cantelli Lemma that, with probability one, the event $\mathcal{B}_{L_k,I,q}$ eventually does not occur.

Let ω be in the set of probability one for which we have pure point spectrum with exponentially decaying eigenfunctions in the region of complete localization Ξ^{CL}. Suppose there exists $E \in \mathcal{I} \cap \Xi^{\mathrm{CL}}$ which is an eigenvalue of H_ω with 2 linearly independent eigenfunctions. In particular these eigenfunctions decay exponentially, so, if we fix $\beta > \frac{5}{2}d$, they both have β-decay. Pick an open interval $I \ni E$, such that $\bar{I} \subset \mathcal{I} \cap \Xi^{\mathrm{CL}}$. [Klein and Molchanov 2006, Lemma 1] can be adapted to the continuum by using smooth functions to localize the eigenfunctions in finite boxes. It then follows that for L large enough the finite volume operator $H_\omega^{(\Lambda_L)}$ has at least 2 eigenvalues in the interval $J_{E,L} = [E-\varepsilon_L, E+\varepsilon_L]$, where $\varepsilon_L = CL^{-\beta+\frac{d}{2}}$ for an appropriate constant C independent of L. Since $\beta > \frac{5d}{2}$ there exists $q > 2d$ such that $\beta - \frac{d}{2} > q$, and hence $\varepsilon_L < L^{-q}$ for all large L. But with probability one this is impossible since the event $\mathcal{B}_{L_k,\bar{I},q}$ does not occur for large L_k.

Theorem 2.3(b) is proven.

Appendix A. The region of complete localization

In this appendix we discuss localization for an Anderson Hamiltonian H_ω. Localization is most commonly taken to be *Anderson localization*: pure point spectrum with exponentially decaying eigenstates with probability one. It is also natural to consider *dynamical localization*, where the moments of a wave packet, initially localized both in space and in energy, should remain uniformly bounded under time evolution. For the multidimensional continuum Anderson Hamiltonian, localization has been proved by a multiscale analysis [Martinelli and Holden 1984; Combes and Hislop 1994; Klopp 1995; Kirsch et al. 1998; Germinet and De Bièvre 1998; Damanik and Stollmann 2001; Germinet and Klein 2001; 2003a], and, in the case when we have the covering condition $\delta_- \ge 1$, also by the fractional moment method [Aizenman et al. 2006]. These methods give more than just Anderson or dynamical localization, although they imply both. In the case when both methods are available, that is, $\delta_- \ge 1$, they have the same region of applicability [Germinet and Klein 2006; Klein 2008].

Thus, following [Germinet and Klein 2006], we consider *the region of complete localization* Ξ^{CL} for an Anderson Hamiltonian H_ω, defined as the set of energies $E \in \mathbb{R}$ where we have the conclusions of the bootstrap multiscale analysis of [Germinet and Klein 2001], that is, as the set of $E \in \mathbb{R}$ for which there exists some open interval $I \ni E$, such that given any ζ, $0 < \zeta < 1$, and α, $1 < \alpha < \zeta^{-1}$, there is a length scale $L_0 \in 2\mathbb{N}$ and a mass $m > 0$, so if we take $L_{k+1} \approx L_k^\alpha$ with $L_{k+1} \in 2\mathbb{N}$, $k = 0, 1, \ldots$, we have

$$\mathbb{P}\{R(m, L_k, I, x, y)\} \geq 1 - \mathrm{e}^{-L_k^\zeta} \tag{A-1}$$

for all $k = 0, 1, \ldots$, and $x, y \in \mathbb{Z}^d$ with $|x - y| > L_k + \varrho$, where $\varrho > 0$ is a constant depending only on $\operatorname{supp} u$, and

$$R(m, L, I, x, y) = \omega\{; \text{for every } E' \in I \text{ either } \Lambda_L(x) \text{ or } \Lambda_L(y) \text{ is } (\omega, m, E')\text{-regular}\}. \tag{A-2}$$

Given $E \in \mathbb{R}$, $x \in \mathbb{Z}^d$ and $L \in 6\mathbb{N}$, we say that the box $\Lambda_L(x)$ is (ω, m, E)-regular for a given $m > 0$ if $E \notin \sigma(H_{\boldsymbol{\omega}}^{(\Lambda_L(x))})$ and

$$\left\| \Gamma_x^{(L)} R_{\boldsymbol{\omega}}^{(\Lambda_L(x))}(E + i\delta)\chi_{\Lambda_{\frac{L}{3}}(x)} \right\| \leq \exp\Big(-m\frac{L}{2}\Big) \quad \text{for all } \delta \in \mathbb{R}, \tag{A-3}$$

where $R_{\boldsymbol{\omega}}^{(\Lambda_L(x))}(E + i\delta) = (H_{\boldsymbol{\omega}}^{(\Lambda_L(x))} - (E + i\delta))^{-1}$ and $\Gamma_x^{(L)}$ denotes the charateristic function of the belt $\bar{\Lambda}_{L-1}(x) \backslash \Lambda_{L-3}(x)$. (See [Germinet and Klein 2001; 2004; 2006; Klein 2008]; note that all the proofs work with the definition (A-3), that is, with the insertion of "for all $\delta \in \mathbb{R}$" They also work with the finite volume operators with periodic boundary condition used in this article.)

By construction Ξ^{CL} is an open set. It can be characterized in many different ways [Germinet and Klein 2004; 2006]. For convenience, our definition includes the complement of the spectrum of $H_{\boldsymbol{\omega}}$ in the region of complete localization, that is, $\mathbb{R} \setminus \Sigma \subset \Xi^{\mathrm{CL}}$. The spectral region of complete localization, $\Xi^{\mathrm{CL}} \cap \Sigma$, is called the "strong insulator region" in [Germinet and Klein 2004].) If the conditions for the fractional moment method are satisfied, Ξ^{CL} coincides with the set of energies where the fractional moment method can be performed. (Minami [1996] proved Poisson statistics for the Anderson model in the region of validity of the fractional moment method, in other words, in the region of complete localization for the Anderson model.)

Proposition A.1. *Consider a closed bounded interval $I \subset \Xi^{\mathrm{CL}}$. Then for all $z \in \mathbb{C}$ with $\Re z \in I$, and boxes $\Lambda = \Lambda_L$, we have, for $s \in]0, \frac{1}{4}[$ and $\xi \in]0, 1[$, and $x, y \in \Lambda$ with $|x - y| \geq (\log L)^{(1/\xi)+}$,*

$$\mathbb{E}\big\{\|\chi_x^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\chi_y^{(1)}\|^s\big\} \leq C_{s,I,\zeta}(\rho_+ + \sqrt{\rho_+})\mathrm{e}^{-|x-y|^\xi} \tag{A-4}$$

for $L \geq L_1(\xi, I, s)$.

We will need the following consequence of the Wegner estimate (2-13).

Lemma A.2. *Let $I = [c, d]$ be such that* (2-13) *holds for any subinterval of* $[c-1, d+1]$ *with a constant K_W. Then for any $s \in \big]0, \frac{1}{2}\big[$, box Λ, and $z \in \mathbb{C}$ with $\Re z \in I$, we have*

$$\mathbb{E}\big\{\|R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\|^s\big\} \leq C_s K_W \rho_+ |\Lambda|. \tag{A-5}$$

Proof. Let $\Re z \in I$. It follows from (2-13) that for all $t \geq 1$

$$\mathbb{P}\big\{\|R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\| \geq t\big\} \leq \frac{2}{t} K_W \rho_+ |\Lambda| \tag{A-6}$$

Thus

$$\begin{aligned} \mathbb{E}\{\|R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\|^s\} = \int_0^\infty t\,\mathbb{P}\big\{\|R_{\boldsymbol{\omega}}^{(\Lambda)}(z)\|^s \geq t\big\}\,\mathrm{d}t &\leq 1 + \int_1^\infty t\big(2t^{-1/s} K_W \rho_+ |\Lambda|\big)\,\mathrm{d}t \\ &\leq 1 + C_s' K_W \rho_+ |\Lambda|. \end{aligned}$$

□

If we have the covering condition $\delta_- \geq 1$, (A-5) holds without the volume factor in the right hand side [Aizenman et al. 2006].

Proof of Proposition A.1. Given $0 < \xi < 1$, we pick ζ such that $\zeta^2 < \xi < \zeta < 1$ (always possible) and set $\alpha = \zeta/\xi$, note $\alpha < \zeta^{-1}$. Since $I \subset \Xi^{\mathrm{CL}}$, there is a scale $L_0 \in 2\mathbb{N}$ and a mass $m_\zeta > 0$, such that, if we set $L_{k+1} \approx L_k^\alpha$, with $L_{k+1} \in 2\mathbb{N}$, $k = 0, 1, \ldots$, we have the estimate (A-1) for $x, y \in \mathbb{Z}^d$ such that $|x - y| > L_k + \varrho$.

Let us now fix $\Lambda = \Lambda_L$, $x, y \in \Lambda_L \cap \mathbb{Z}^d$ and pick k such that $L_{k+1} + \varrho \geq |x - y| > L_k + \varrho$. In this case, if $\boldsymbol{\omega} \in R\left(m_\zeta, L_k, I, x, y\right)$, then for $\Re z \in I$ either $\Lambda_{L_k}(x)$ or $\Lambda_{L_k}(y)$ is $(\omega, m, \Re z)$-regular; say $\Lambda_{L_k}(x)$ is $(\omega, m, \Re z)$-regular. (Note that we take the boxes of size L_k in the torus Λ.) Then, using (A-3) and [Germinet and Klein 2001, (2.9)], we reach

$$\begin{aligned}\left\| \chi_y^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \chi_x^{(1)} \right\| &\leq \gamma_I \left\| \Gamma_x^{(L_k)} R_{\boldsymbol{\omega}}^{(\Lambda_{L_k}(x))}(z) \chi_x^{(1)} \right\| \left\| \chi_y^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \Gamma_x^{(L_k)} \right\| \\ &\leq \gamma_I \exp\Bigl(-m_\zeta \frac{L_k}{2}\Bigr) \left\| R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \right\|. \end{aligned} \tag{A-7}$$

Thus, with $s \in]0, \frac{1}{4}[$, using Lemma A.2,

$$\begin{aligned}\mathbb{E}\Bigl\{ \left\| \chi_y^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \chi_x^{(1)} \right\|^s : \boldsymbol{\omega} \in R(m_\zeta, L_k, I, x, y) \Bigr\} &\\ \leq \gamma_I^s \exp\Bigl(-s m_\zeta \frac{L_k}{2}\Bigr) \mathbb{E}\bigl\{ \| R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \|^s \bigr\} &\\ \leq C_s K_W \rho_+ |\Lambda| \gamma_I^s \exp\Bigl(-s m_\zeta \frac{L_k}{2}\Bigr) \leq C_{s,I} \rho_+ |\Lambda| \exp\Bigl(-s m_\zeta \frac{L_k}{2}\Bigr), & \end{aligned} \tag{A-8}$$

and

$$\begin{aligned}\mathbb{E}\Bigl\{ \left\| \chi_y^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \chi_x^{(1)} \right\|^s : \boldsymbol{\omega} \notin R(m_\zeta, L_k, I, x, y) \Bigr\} &\leq \bigl(\mathbb{E}\{ \| R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \|^{2s} \} \bigr)^{1/2} \bigl(\mathbb{P}\{ \boldsymbol{\omega} \notin R(m_\zeta, L_k, I, x, y) \} \bigr)^{1/2} \\ &\leq (C_{2s} K_W \rho_+ |\Lambda|)^{1/2} \exp\bigl(-\tfrac{1}{2} L_k^\zeta\bigr) \\ &\leq C'_{s,I} (\rho_+ |\Lambda|)^{1/2} \exp\bigl(-\tfrac{1}{2} L_k^\zeta\bigr). \end{aligned} \tag{A-9}$$

It follows that for L_k sufficiently large, that is, $|x - y|$ large, we have

$$\begin{aligned}\mathbb{E}\Bigl\{ \left\| \chi_y^{(1)} R_{\boldsymbol{\omega}}^{(\Lambda)}(z) \chi_x^{(1)} \right\|^s \Bigr\} &\leq C_{s,I,\zeta} (\rho_+ + \sqrt{\rho_+}) |\Lambda| \exp\bigl(-\tfrac{1}{2} L_k^\zeta\bigr) \\ &\leq C_{s,I,\zeta} (\rho_+ + \sqrt{\rho_+}) |\Lambda| \exp\bigl(-\tfrac{1}{2} L_{k+1}^\xi\bigr) \\ &\leq C'_{s,I,\zeta} (\rho_+ + \sqrt{\rho_+}) |\Lambda| \exp\bigl(-\tfrac{1}{2} |x - y|^\xi\bigr), \end{aligned} \tag{A-10}$$

so (A-4) follows for $|x - y| \geq (\log L)^{(1/\xi)+}$ (with a slightly smaller ξ). □

Appendix B. A convexity inequality for traces

The following inequality was used in [Combes and Hislop 1994, Proof of Proposition 4.5] and also in the derivation of (4-12) above.

Lemma B.1. *Let H_1 and H_2 be two self-adjoint operators on a Hilbert space $\mathcal{H}$, such that H_1 is diagonalizable and $H_1 \geq H_2$. Let f and g be bounded Borel functions on some open interval $I \supset \sigma(H_1)$, such that g is real-valued, nonincreasing, and convex on I. Then*

$$\operatorname{tr}\bigl\{ \bar{f}(H_1) g(H_1) f(H_1) \bigr\} \leq \operatorname{tr}\bigl\{ \bar{f}(H_1) g(H_2) f(H_1) \bigr\}. \tag{B-1}$$

Proof. Let $\varphi \in \mathcal{H}$ be an eigenvector of H_1 with eigenvalue λ and satisfying $\|\varphi\| = 1$. Then

$$\begin{aligned}\langle\varphi, \bar{f}(H_1)g(H_1)f(H_1)\varphi\rangle &= \bar{f}(\lambda)g(\lambda)f(\lambda) = \bar{f}(\lambda)g(\langle\varphi, H_1\varphi\rangle)f(\lambda) \le \bar{f}(\lambda)g(\langle\varphi, H_2\varphi\rangle)f(\lambda) \\ &\le \bar{f}(\lambda)\langle\varphi, g(H_2)\varphi\rangle f(\lambda) = \langle\varphi, \bar{f}(H_1)g(H_2)f(H_1)\varphi\rangle, \end{aligned} \tag{B-2}$$

where the first inequality follows from g nonincreasing and $H_1 \ge H_2$, and the second inequality used the convexity of the function g, Jensen's inequality (compare [Yeh 2006, Theorem 14.16]), and the spectral theorem.

Since H_1 is diagonalizable, (B-1) follows by expanding the trace on an orthonormal basis of eigenvalues for H_1 and using (B-2) for each term. □

References

[Aizenman et al. 2006] M. Aizenman, A. Elgart, S. Naboko, J. H. Schenker, and G. Stolz, "Moment analysis for localization in random Schrödinger operators", *Invent. Math.* **163**:2 (2006), 343–413. MR 2006m:81097 Zbl 1090.81026

[Bellissard 2004] J. Bellissard, "Random matrix theory and the Anderson model", *J. Statist. Phys.* **116**:1-4 (2004), 739–754. MR 2005i:82033 Zbl 1142.82342

[Bellissard et al. 2007] J. V. Bellissard, P. D. Hislop, and G. Stolz, "Correlation estimates in the Anderson model", *J. Stat. Phys.* **129**:4 (2007), 649–662. MR 2009b:82048 Zbl 1131.82016

[Carmona and Lacroix 1990] R. Carmona and J. Lacroix, *Spectral theory of random Schrödinger operators*, Birkhäuser, Boston, MA, 1990. MR 92k:47143 Zbl 0717.60074

[Carmona et al. 1987] R. Carmona, A. Klein, and F. Martinelli, "Anderson localization for Bernoulli and other singular potentials", *Comm. Math. Phys.* **108**:1 (1987), 41–66. MR 88f:82027 Zbl 0615.60098

[Combes and Hislop 1994] J.-M. Combes and P. D. Hislop, "Localization for some continuous, random Hamiltonians in d-dimensions", *J. Funct. Anal.* **124**:1 (1994), 149–180. MR 95g:82047 Zbl 0801.60054

[Combes et al. 2003] J.-M. Combes, P. D. Hislop, and F. Klopp, "Hölder continuity of the integrated density of states for some random operators at all energies", *Int. Math. Res. Not.* **4** (2003), 179–209. MR 2003k:82046 Zbl 1022.47028

[Combes et al. 2007a] J.-M. Combes, P. D. Hislop, and F. Klopp, "An optimal Wegner estimate and its application to the global continuity of the integrated density of states for random Schrödinger operators", *Duke Math. J.* **140**:3 (2007), 469–498. MR 2009b:82050 Zbl 1134.81022

[Combes et al. 2007b] J.-M. Combes, P. D. Hislop, and F. Klopp, "Some new estimates on the spectral shift function associated with random Schrödinger operators", pp. 85–95 in *Probability and mathematical physics*, edited by D. A. Dawson et al., CRM Proc. Lecture Notes **42**, Amer. Math. Soc., Providence, RI, 2007. MR 2009b:47063 Zbl 1137.35048

[Combes et al. 2009] J.-M. Combes, F. Germinet, and A. Klein, "Generalized eigenvalue-counting estimates for the Anderson model", *J. Stat. Phys.* **135**:2 (2009), 201–216. MR 2505733 Zbl 1168.82016

[Daley and Vere-Jones 1988] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer, New York, 1988. MR 90e:60060 Zbl 0657.60069

[Damanik and Stollmann 2001] D. Damanik and P. Stollmann, "Multi-scale analysis implies strong dynamical localization", *Geom. Funct. Anal.* **11**:1 (2001), 11–29. MR 2002e:82035 Zbl 0976.60064

[Disertori et al. 2002] M. Disertori, H. Pinson, and T. Spencer, "Density of states for random band matrices", *Comm. Math. Phys.* **232**:1 (2002), 83–124. MR 2004c:82052 Zbl 1019.15014

[Erdős et al. 2009a] L. Erdős, B. Schlein, and H.-T. Yau, "Local semicircle law and complete delocalization for Wigner random matrices", *Comm. Math. Phys.* **287**:2 (2009), 641–655. MR 2481753 Zbl 05610582

[Erdős et al. 2009b] L. Erdős, B. Schlein, and H.-T. Yau, "Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices", *Ann. Probab.* **37**:3 (2009), 815–852. MR 2537522 Zbl 1175.15028

[Figotin and Klein 1996] A. Figotin and A. Klein, "Localization of classical waves. I. Acoustic waves", *Comm. Math. Phys.* **180**:2 (1996), 439–482. MR 2000d:35240a Zbl 0878.35109

[Fröhlich and Spencer 1983] J. Fröhlich and T. Spencer, "Absence of diffusion in the Anderson tight binding model for large disorder or low energy", *Comm. Math. Phys.* **88**:2 (1983), 151–184. MR 85c:82004 Zbl 0519.60066

[Germinet and De Bièvre 1998] F. Germinet and S. De Bièvre, "Dynamical localization for discrete and continuous random Schrödinger operators", *Comm. Math. Phys.* **194**:2 (1998), 323–341. MR 99d:82033 Zbl 0911.60099

[Germinet and Klein 2001] F. Germinet and A. Klein, "Bootstrap multiscale analysis and localization in random media", *Comm. Math. Phys.* **222**:2 (2001), 415–448. MR 2002m:82035 Zbl 0982.82030

[Germinet and Klein 2003a] F. Germinet and A. Klein, "Explicit finite volume criteria for localization in continuous random media and applications", *Geom. Funct. Anal.* **13**:6 (2003), 1201–1238. MR 2004m:82063 Zbl 1086.82008

[Germinet and Klein 2003b] F. Germinet and A. Klein, "Operator kernel estimates for functions of generalized Schrödinger operators", *Proc. Amer. Math. Soc.* **131**:3 (2003), 911–920. MR 2003k:47067 Zbl 1013.81009

[Germinet and Klein 2004] F. Germinet and A. Klein, "A characterization of the Anderson metal-insulator transport transition", *Duke Math. J.* **124**:2 (2004), 309–350. MR 2005e:82051 Zbl 1062.82020

[Germinet and Klein 2006] F. Germinet and A. Klein, "New characterizations of the region of complete localization for random Schrödinger operators", *J. Stat. Phys.* **122**:1 (2006), 73–94. MR 2007c:82041 Zbl 1127.82031

[Gol'dsheĭd et al. 1977] I. J. Gol'dsheĭd, S. A. Molchanov, and L. A. Pastur, "A random homogeneous Schrödinger operator has a pure point spectrum", *Funktsional. Anal. i Prilozhen.* **11**:1 (1977), 1–10, 96. In Russian; translated in *Funct. Anal. Appl.* **11**, (1977), 1–8. MR 57 #10266

[Graf and Vaghi 2007] G. M. Graf and A. Vaghi, "A remark on the estimate of a determinant by Minami", *Lett. Math. Phys.* **79**:1 (2007), 17–22. MR 2008b:81077 Zbl 1104.82008

[Hunziker and Sigal 2000] W. Hunziker and I. M. Sigal, "Time-dependent scattering theory of N-body quantum systems", *Rev. Math. Phys.* **12**:8 (2000), 1033–1084. MR 2001k:81347 Zbl 0978.47008

[Johansson 1998] K. Johansson, "On fluctuations of eigenvalues of random Hermitian matrices", *Duke Math. J.* **91**:1 (1998), 151–204. MR 2000m:82026 Zbl 1039.82504

[Johansson 2001] K. Johansson, "Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices", *Comm. Math. Phys.* **215**:3 (2001), 683–705. MR 2002j:15024 Zbl 0978.15020

[Killip and Stoiciu 2009] R. Killip and M. Stoiciu, "Eigenvalue statistics for CMV matrices: from Poisson to clock via random matrix ensembles", *Duke Math. J.* **146**:3 (2009), 361–399. MR 2009k:81087 Zbl 1155.81020

[Kirsch 2008] W. Kirsch, "An invitation to random Schrödinger operators", pp. 1–119 in *Random Schrödinger operators*, edited by M. Disertori et al., Panor. Synthèses **25**, Soc. Math. France, Paris, 2008. MR 2509110 Zbl 1162.82004

[Kirsch and Martinelli 1982] W. Kirsch and F. Martinelli, "On the ergodic properties of the spectrum of general random operators", *J. Reine Angew. Math.* **334** (1982), 141–156. MR 84g:60102a Zbl 0476.60058

[Kirsch et al. 1998] W. Kirsch, P. Stollmann, and G. Stolz, "Localization for random perturbations of periodic Schrödinger operators", *Random Oper. Stochastic Equations* **6**:3 (1998), 241–268. MR 99c:82038 Zbl 0927.60067

[Klein 2008] A. Klein, "Multiscale analysis and localization of random operators", pp. 121–159 in *Random Schrödinger operators*, edited by M. Disertori et al., Panor. Synthèses **25**, Soc. Math. France, Paris, 2008. MR 2509111 Zbl 05533269

[Klein and Koines 2001] A. Klein and A. Koines, "A general framework for localization of classical waves. I. Inhomogeneous media and defect eigenmodes", *Math. Phys. Anal. Geom.* **4**:2 (2001), 97–130. MR 2002i:35187 Zbl 0987.35154

[Klein and Molchanov 2006] A. Klein and S. Molchanov, "Simplicity of eigenvalues in the Anderson model", *J. Stat. Phys.* **122**:1 (2006), 95–99. MR 2007c:82042 Zbl 1152.82010

[Klein et al. 2002] A. Klein, A. Koines, and M. Seifert, "Generalized eigenfunctions for waves in inhomogeneous media", *J. Funct. Anal.* **190**:1 (2002), 255–291. MR 2004a:35165 Zbl 1043.35097

[Klopp 1995] F. Klopp, "Localization for some continuous random Schrödinger operators", *Comm. Math. Phys.* **167**:3 (1995), 553–569. MR 95m:82080 Zbl 0820.60044

[Kritchevski 2008] E. Kritchevski, "Poisson statistics of eigenvalues in the hierarchical Anderson model", *Ann. Henri Poincaré* **9**:4 (2008), 685–709. MR 2009b:82056 Zbl 1149.82017

[Martinelli and Holden 1984] F. Martinelli and H. Holden, "On absence of diffusion near the bottom of the spectrum for a random Schrödinger operator on $L^2(\mathbf{R}^\nu)$", *Comm. Math. Phys.* **93**:2 (1984), 197–217. MR 85m:82103 Zbl 0546.60063

[Minami 1996] N. Minami, "Local fluctuation of the spectrum of a multidimensional Anderson tight binding model", *Comm. Math. Phys.* **177**:3 (1996), 709–725. MR 97d:82046 Zbl 0851.60100

[Molchanov 1978] S. A. Molchanov, "Structure of the eigenfunctions of one-dimensional unordered structures", *Izv. Akad. Nauk SSSR Ser. Mat.* **42**:1 (1978), 70–103, 214. In Russian; translated in *Math. USSR-Izv.* **12** (1978) 69–101. MR 58 #6662 Zbl 0401.34023

[Molchanov 1980/81] S. A. Molchanov, "The local structure of the spectrum of the one-dimensional Schrödinger operator", *Comm. Math. Phys.* **78**:3 (1980/81), 429–446. MR 82d:35076 Zbl 0584.60072

[Pastur and Figotin 1992] L. Pastur and A. Figotin, *Spectra of random and almost-periodic operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **297**, Springer, Berlin, 1992. MR 94h:47068 Zbl 0752.47002

[Schenker and Schulz-Baldes 2007] J. Schenker and H. Schulz-Baldes, "Gaussian fluctuations for random matrices with correlated entries", *Int. Math. Res. Not. IMRN* **15** (2007), Art. ID rnm047, 36. MR 2009b:60120 Zbl 1128.60026

[Simon 1994] B. Simon, "Cyclic vectors in the Anderson model", *Rev. Math. Phys.* **6**:5A (1994), 1183–1185. MR 95i:82058 Zbl 0841.60081

[Simon 1998] B. Simon, "Spectral averaging and the Krein spectral shift", *Proc. Amer. Math. Soc.* **126**:5 (1998), 1409–1413. MR 98j:47030 Zbl 0892.47021

[Stoiciu 2006] M. Stoiciu, "The statistical distribution of the zeros of random paraorthogonal polynomials on the unit circle", *J. Approx. Theory* **139**:1-2 (2006), 29–64. MR 2007d:60034 Zbl 1088.42017

[Stoiciu 2007] M. Stoiciu, "Poisson statistics for eigenvalues: from random Schrödinger operators to random CMV matrices", pp. 465–475 in *Probability and mathematical physics*, edited by D. A. Dawson et al., CRM Proc. Lecture Notes **42**, Amer. Math. Soc., Providence, RI, 2007. MR 2009c:82029 Zbl 1132.15022

[Wegner 1981] F. Wegner, "Bounds on the density of states in disordered systems", *Z. Phys. B* **44**:1-2 (1981), 9–15. MR 83b:82060

[Yeh 2006] J. Yeh, *Real analysis*, 2nd. ed., World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006. Theory of measure and integration. MR 2007i:28001 Zbl 1098.28002

Received 9 Jul 2009. Accepted 6 Aug 2009.

JEAN-MICHEL COMBES: combes@cpt.univ-mrs.fr
Département de Mathématiques, Université du Sud: Toulon-Var, 83130 La Garde, France

and

Centre de Physique Théorique, CNRS Luminy, Case 907, 13288 Marseille, France

FRANÇOIS GERMINET: germinet@math.u-cergy.fr
Département de Mathématiques, Université de Cergy-Pontoise, 95000 Cergy-Pontoise, France
http://www.u-cergy.fr/rech/pages/germinet/

ABEL KLEIN: aklein@uci.edu
Department of Mathematics, University of California, Irvine, CA 92697-3875, United States

ANALYSIS AND PDE
Vol. 3, No. 1, 2010

BULK UNIVERSALITY AND CLOCK SPACING OF ZEROS FOR ERGODIC JACOBI MATRICES WITH ABSOLUTELY CONTINUOUS SPECTRUM

ARTUR AVILA, YORAM LAST AND BARRY SIMON

By combining ideas of Lubinsky with some soft analysis, we prove that universality and clock behavior of zeros for orthogonal polynomials on the real line in the absolutely continuous spectral region is implied by convergence of $\frac{1}{n}K_n(x,x)$ for the diagonal CD kernel and boundedness of the analog associated to second kind polynomials. We then show that these hypotheses are always valid for ergodic Jacobi matrices with absolutely continuous spectrum and prove that the limit of $\frac{1}{n}K_n(x,x)$ is $\rho_\infty(x)/w(x)$, where ρ_∞ is the density of zeros and w is the absolutely continuous weight of the spectral measure.

1. Introduction

Given a finite measure, $d\mu$, of compact and not finite support on $\mathbb{R}$, one defines the orthonormal polynomials $p_n(x)$ (or $p_n(x,d\mu)$ if the μ-dependence is important) by applying Gram–Schmidt to $1, x, x^2, \dots$. Thus, p_n is a polynomial of degree exactly n with leading positive coefficient so that

$$\int p_n(x)p_m(x)\,d\mu(x) = \delta_{nm}. \tag{1-1}$$

For background on these orthogonal polynomials on the real line (OPRL), see [Szegő 1939; Freud 1971; Simon 2010].

Associated to μ is a family of Jacobi parameters $\{a_n, b_n\}_{n=1}^{\infty}$, $a_n > 0$, b_n real, determined by the recursion relation ($p_{-1}(x) \equiv 0$):

$$xp_n(x) = a_{n+1}p_{n+1}(x) + b_{n+1}p_n(x) + a_n p_{n-1}(x). \tag{1-2}$$

The $\{p_n(x)\}_{n=0}^{\infty}$ are an orthonormal basis of $L^2(\mathbb{R}, d\mu)$ (since $\operatorname{supp} d\mu$ is compact) and (1-2) says that multiplication by x is given in this basis by the tridiagonal Jacobi matrix

$$J = \begin{pmatrix} b_1 & a_1 & 0 & \cdots \\ a_1 & b_2 & a_2 & \cdots \\ 0 & a_2 & b_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{1-3}$$

MSC2000: 26C10, 42C05, 47B36.
Keywords: orthogonal polynomials, clock behavior, almost Mathieu equation.
Y. Last was supported in part by grant 1169/06 from the Israel Science Foundation; B. Simon by grant DMS-0652919 from the NSF; and both by grant 2006483 from the United States–Israel Binational Science Foundation (BSF), Jerusalem.

If we restrict (as we normally will) to μ normalized by $\mu(\mathbb{R}) = 1$, then μ can be recovered from J as the spectral measure for the vector $(1, 0, 0, \dots)^t$. Favard's theorem says there is a one-to-one correspondence between sets of bounded Jacobi parameters, that is,

$$\sup_n |a_n| = \alpha_+ < \infty, \qquad \sup_n |b_n| = \beta < \infty, \tag{1-4}$$

and probability measures with compact and not finite support under this $\mu \to J \to \mu$ correspondence.

We will use this to justify spectral theory notation for things like supp $d\mu$ which we will denote $\sigma(d\mu)$ since it is the spectrum of J, $\sigma(J)$. We will use $\sigma_{\text{ess}}(d\mu)$ for the essential spectrum, and if

$$d\mu(x) = w(x)\,dx + d\mu_{\text{s}}(x), \tag{1-5}$$

where $d\mu_{\text{s}}$ is Lebesgue singular, then we define

$$\Sigma_{\text{ac}}(d\mu) = \{x \mid w(x) > 0\}, \tag{1-6}$$

determined up to sets of Lebesgue measure 0, so $\Sigma_{\text{ac}} \neq \varnothing$ means $d\mu$ has a nonvanishing a.c. part.

We will also suppose

$$\inf_n a_n = \alpha_- > 0, \tag{1-7}$$

which is no loss since it is known [Dombrowski 1978] that if the inf is 0, then $\Sigma_{\text{ac}} = \varnothing$, and we will only be interested in cases where $\Sigma_{\text{ac}} \neq \varnothing$.

One of our concerns in this paper is the zeros of $p_n(x, d\mu)$. These are not only of intrinsic interest; they enter in Gaussian quadrature and also as the eigenvalues of $J_{n;F}$, the upper left $n \times n$ corner of J, and so are relevant to statistics of eigenvalues in large boxes, a subject on which there is an enormous amount of discussion in both the mathematics and the physics literature.

These zeros are all simple and real. The measure $d\nu_n$ is the normalized counting measure for the zeros:

$$\nu_n(S) = \frac{1}{n}\,\#\text{ of zeros of } p_n \text{ in } S. \tag{1-8}$$

In many cases, $d\nu_n$ converges to a weak limit $d\nu_\infty$ called the density of zeros or density of states (DOS). If this weak limit exists, we say that the DOS exists. It often happens that $d\nu_\infty$ is $d\rho_{\mathfrak{e}}$, the equilibrium measure for $\mathfrak{e} = \sigma_{\text{ess}}(d\mu)$. This is true, for example, if $\rho_{\mathfrak{e}}$ is equivalent to $dx \restriction \mathfrak{e}$ and $\Sigma_{\text{ac}} = \mathfrak{e}$, a theorem of Widom [1967] and Van Assche [1986] (see also [Stahl and Totik 1992; Simon 2007]). If $d\nu_\infty$ has an a.c. part, we use $\rho_\infty(x)$ for $d\nu_\infty/dx$ and we use $\rho_{\mathfrak{e}}(x)$ for $d\rho_{\mathfrak{e}}/dx$. More properly, $d\nu_\infty$ is the *density of states measure* (so $\int_{-\infty}^x d\nu_\infty$ is the *integrated density of states*) and $\rho_\infty(x)$ the *density of states*.

We are especially interested in the fine structure of the zeros near some point $x_0 \in \sigma(d\mu)$. We define $x_j^{(n)}(x_0)$ by

$$x_{-2}^{(n)}(x_0) < x_{-1}^{(n)}(x_0) < x_0 \le x_0^{(n)}(x_0) < x_1^{(n)}(x_0) < \cdots, \tag{1-9}$$

requiring these to be all of the zeros near x_0. It is known that if x_0 is not isolated from $\sigma(d\mu)$ on either side, that is, if for all $\delta > 0$,

$$(x_0 - \delta, x_0) \cap \sigma(d\mu) \neq \varnothing \neq (x_0, x_0 + \delta) \cap \sigma(d\mu), \tag{1-10}$$

then for each fixed j,

$$\lim_{n\to\infty} x_j^{(n)}(x_0) = x_0. \tag{1-11}$$

We are interested in clock behavior, named after the spacing of numerals on a clock — meaning equal spacing of the zeros nearby to x_0:

Definition. We say that there is *quasiclock behavior* at $x_0 \in \sigma(d\mu)$ if and only if for each fixed $j \in \mathbb{Z}$,

$$\lim_{n\to\infty} \frac{x_{j+1}^{(n)}(x_0) - x_j^{(n)}(x_0)}{x_1^{(n)}(x_0) - x_0^{(n)}(x_0)} = 1. \tag{1-12}$$

We say there is *strong clock behavior* at x_0 if and only if the DOS exists and for each fixed $j \in \mathbb{Z}$,

$$\lim_{n\to\infty} n(x_{j+1}^{(n)}(x_0) - x_j(x_0)) = \frac{1}{\rho_\infty(x_0)}. \tag{1-13}$$

Obviously, strong clock behavior implies quasiclock behavior. Thus far, the only cases where it is proven there is quasiclock behavior, one has strong clock behavior but, as we will explain in Section 7, we think there are examples where one has quasiclock behavior at x_0 but not strong clock behavior. Before this paper, all examples known with strong clock behavior have $\rho_\infty = \rho_{\mathfrak{e}}$, but we will find several examples where there is strong clock behavior with $\rho_\infty \neq \rho_{\mathfrak{e}}$ in Section 7. In that section, we will say more about:

Conjecture. For any μ, quasiclock behavior holds at a.e. $x_0 \in \Sigma_{\text{ac}}(d\mu)$.

In this paper, one of our main goals is to prove this result for ergodic Jacobi matrices. A major role will be played by the Christoffel–Darboux (CD) kernel, defined for $x, y \in \mathbb{C}$ by

$$K_n(x,y) = \sum_{j=0}^{n} \overline{p_j(x)}\, p_j(y), \tag{1-14}$$

the integral kernel for the orthogonal projection onto polynomials of degree at most n in $L^2(\mathbb{R}, d\mu)$; see Simon [2008a] for a review of some important aspects of the properties and uses of this kernel. We will repeatedly make use of the CD formula:

$$K_n(x,y) = \frac{a_{n+1}[\,\overline{p_{n+1}(x)}\, p_n(y) - \overline{p_n(x)} p_{n+1}(y)]}{\bar{x} - y}; \tag{1-15}$$

the Schwarz inequality:

$$|K_n(x,y)|^2 \le K_n(x,x) K_n(y,y); \tag{1-16}$$

and the reproducing property:

$$\int K_n(x,y) K_n(y,z)\, d\mu(y) = K_n(x,z). \tag{1-17}$$

It is a theorem [Simon 2009] that if the DOS exists, then

$$\frac{1}{n+1} K_n(x,x)\, d\mu(x) \xrightarrow{\text{weak}} d\nu_\infty(x), \tag{1-18}$$

and, in general, $\dfrac{1}{n+1} K_n(x,x)\, d\mu(x)$ has the same weak limit points as $d\nu_n$. This suggests that a.c. parts converge pointwise; that is, one hopes that for a.e. $x_0 \in \Sigma_{\mathrm{ac}}$,

$$\frac{1}{n+1}\, K_n(x_0, x_0) \to \frac{\rho_\infty(x_0)}{w(x_0)}. \tag{1-19}$$

This has been proven for regular measures (in the sense of [Stahl and Totik 1992]; see also [Simon 2007]) with a local Szegő condition in a series of papers, of which the seminal ones are [Máté et al. 1991; Totik 2000]. We will prove it for ergodic Jacobi matrices.

We say *bulk universality* holds at $x_0 \in \operatorname{supp} d\mu$ if and only if uniformly for a, b in compact subsets of $\mathbb{R}$, we have

$$\frac{K_n(x_0 + a/n,\ x_0 + b/n)}{K_n(x_0, x_0)} \to \frac{\sin(\pi\rho(x_0)(b-a))}{\pi\rho(x_0)(b-a)}. \tag{1-20}$$

We use the term *bulk* here because (1-20) fails at edges of the spectrum [Lubinsky 2008a]. We also note that when (1-20) holds, typically (and in all cases below) for z, w complex, one has

$$\frac{K_n(x_0 + z/n,\ x_0 + w/n)}{K_n(x_0, x_0)} \to \frac{\sin(\rho(x_0)(w-\bar{z}))}{\rho(x_0)(w-\bar{z})}. \tag{1-21}$$

Freud [1971] proved bulk universality for measures on $[-1, 1]$ with $d\mu_{\mathrm{s}} = 0$ and strong conditions on $w(x)$. Because of related results (but with variable weights) in random matrix theory, this result was reexamined and proven in multiple interval support cases with analytic weights by Kuijlaars and Vanlessen [2002]. A significant breakthrough was made by Lubinsky [2009], whose contributions we return to shortly.

The following theorem is a basic result of Freud [1971], rediscovered by Levin.[1]

Theorem 1.1 (Freud–Levin Theorem). *Bulk universality at x_0 implies strong clock behavior at x_0.*

Remarks. 1. The proof [Freud 1971; Levin and Lubinsky 2008; Simon 2008a] relies on the CD formula (1-15), which implies that if y_0 is a zero of p_n, then the other zeros of p_n are the points y solving $K_n(y, y_0) = 0$ and the fact that the zeros of $\sin(\pi\rho(x_0)(b-a))$ are at $b - a = j/\rho(x_0)$ with $j \in \mathbb{Z}$.

2. Szegő [1939] proved strong clock behavior for Jacobi polynomials and Erdős and Turán [1940] for a more general class of measures on $[-1, 1]$. Simon has a series on the subject [2005; 2006a; 2006b; Last and Simon 2008]. The last of these papers was one motivation for [Levin and Lubinsky 2008].

It is also useful to define

$$\rho_n = \frac{1}{n}\, w(x_0) K_n(x_0, x_0), \tag{1-22}$$

so (1-19) is equivalent to

$$\rho_n \to \rho_\infty(x_0). \tag{1-23}$$

We say *weak bulk universality* holds at x_0 if and only if, uniformly for a, b on compact subsets of $\mathbb{R}$, we have

$$\frac{K_n(x_0 + a/(n\rho_n), x_0 + b/(n\rho_n))}{K_n(x_0, x_0)} \to \frac{\sin(\pi(b-a))}{\pi(b-a)}, \tag{1-24}$$

[1] See [Levin and Lubinsky 2008]. Lubinsky (private communication) has emphasized to us that this part of the paper is due to Levin alone — hence our name for the result.

the form in which universality is often written, especially in the random matrix literature. Notice that

$$\text{weak universality} + \text{(1-23)} \;\Rightarrow\; \text{universality}. \tag{1-25}$$

Notice also that (1-24) could hold in case where ρ_n does not converge as $n \to \infty$. The same proof that verifies Theorem 1.1 implies:

Theorem 1.2 (Weak Freud–Levin Theorem). *Weak bulk universality at x_0 implies quasiclock behavior at x_0.*

With this background in place, we can turn to describing the main results of this paper: five theorems, proven one per section in Sections 2–6.

The first theorem is an abstraction, extension, and simplification of Lubinsky's second approach to universality [2008b]. Lubinsky [2009] found a beautiful way of going from control of the diagonal CD kernel to the off-diagonal (i.e., to universality). It depended on the ability to control limits not only of $(1/n)K_n(x_0, x_0)$ but also $(1/n)K_n(x_0 + a/n,\ x_0 + a/n)$ — what we call the Lubinsky wiggle. We will especially care about the *Lubinsky wiggle condition*:

$$\lim_{n\to\infty} \frac{K_n(x_0 + a/n,\ x_0 + a/n)}{K_n(x_0, x_0)} = 1 \tag{1-26}$$

uniformly for $a \in [-A, A]$ for each A. In addition to this, Lubinsky [2009] needed a simple but clever inequality and, most significantly, a comparison model example where one knows universality holds. For $[-1, 1]$, he took Legendre polynomials (that is, $d\mu = (1/2)\chi_{[-1,1]}(x)\,dx$). In extending this to more general sets, one uses approximation by finite gap sets as pioneered by Totik [2001]. Simon [2008b] then used Jacobi matrices in isospectral tori for a comparison model on these finite gap sets, while Totik [$\geq$ 2010] used polynomials mappings and the results for $[-1, 1]$.

For ergodic Jacobi matrices, where $\sigma(d\mu)$ is often a Cantor set, it is hard to find comparison models, so we will rely on a second approach developed by Lubinsky [2008b] that seems to be able to handle any situation that his first approach can and which does not rely on a comparison model. Our first theorem, proven in Section 2, is a variant of this approach. We need a preliminary definition.

Definition. Let $d\mu$ be given by (1-5). A point x_0 is called a *Lebesgue point* of $d\mu$ if $w(x_0) > 0$ and

$$\lim_{\delta\downarrow 0} (2\delta)^{-1} \int_{x_0-\delta}^{x_0+\delta} |w(x) - w(x_0)|\,dx = 0, \tag{1-27}$$

$$\lim_{\delta\downarrow 0} (2\delta)^{-1} \mu_{\mathrm{s}}(x_0 - \delta, x_0 + \delta) = 0. \tag{1-28}$$

Standard maximal function methods [Rudin 1987] show that Lebesgue almost every $x_0 \in \Sigma_{\mathrm{ac}}(d\mu)$ is a Lebesgue point.

Theorem 1. *Let x_0 be a Lebesgue point of μ. Suppose that*:

(i) *The Lubinsky wiggle condition* (1-26) *holds uniformly for $a \in [-A, A]$ and any $A < \infty$.*

(ii) *We have*

$$\liminf_{n\to\infty} \frac{1}{n+1} K_n(x_0, x_0) > 0. \tag{1-29}$$

(iii) *For any* ε, *there is* $C_\varepsilon > 0$ *so that for any* $R < \infty$, *there is an* N *so that for all* $n > N$ *and all* $z \in \mathbb{C}$ *with* $|z| < R$, *we have*

$$\frac{1}{n+1}\, K_n\Big(x_0 + \frac{z}{n}\,, x_0 + \frac{\bar{z}}{n}\Big) \le C_\varepsilon \exp(\varepsilon |z|^2). \tag{1-30}$$

Then weak bulk universality, *and so*, *quasiclock behavior*, *holds at* x_0.

Remarks. 1. If one replaces the right-hand side of (1-30) by

$$C \exp(A|z|), \tag{1-31}$$

then the result can be proven by following Lubinsky's argument in [2008b]. He does not assume (1-31) directly but rather hypotheses that he shows imply it (but which are invalid when the support of $d\mu$ is a Cantor set).

2. Because our Theorem 3 below is so general, we doubt there are examples where (1-30) holds but (1-31) does not, but we feel our more general abstract result is clarifying.

3. The strategy we follow is Lubinsky's, but the tactics differ and, we feel, are more elementary and illuminating.

In [Lubinsky 2008b], the only examples where the wiggle condition can be verified are the situations where Totik [$\geq$ 2010] proves universality using Lubinsky's first method. To go beyond that, we need the following, proven in Section 3:

Theorem 2. *Let* $\Sigma \subset \Sigma_{\mathrm{ac}}$. *Suppose for a.e.* $x_0 \in \Sigma$, *condition* (iii) *of Theorem 1 holds and*

(iv) $\lim_{n\to\infty}(1/(n+1))K_n(x_0, x_0)$ *exists and is strictly positive.*

Then condition (i) *of Theorem 1 holds for a.e.* $x_0 \in \Sigma$.

Of course, (iv) implies condition (ii). So we obtain:

Corollary 1.3. *If* (iii) *and* (iv) *hold for a.e.* $x_0 \in \Sigma$, *then for a.e.* $x_0 \in \Sigma$, *we have weak universality and quasiclock behavior.*

By (1-25), we see:

Corollary 1.4. *If* (iii) *and* (iv) *hold for a.e.* $x_0 \in \Sigma$, *and if the DOS exists and the limit in* (iv) *is* $\rho_\infty(x)/w(x)$, *then for a.e.* $x \in \Sigma$, *we have universality and strong clock behavior.*

Next, we need to examine when (1-30) holds. We will not only obtain a bound of the type (1-31) but one that does not need to vary N with R and is universal in z. We will use transfer matrix techniques and notation.

Given Jacobi parameters, $\{a_n, b_n\}_{n=1}^\infty$, we define

$$A_j(z) = \begin{pmatrix} \frac{z-b_j}{a_j} & -\frac{1}{a_j} \\ a_j & 0 \end{pmatrix}, \tag{1-32}$$

so that (1-2) is equivalent to

$$\begin{pmatrix} p_n(x) \\ a_n p_{n-1}(x) \end{pmatrix} = A_n(x) \begin{pmatrix} p_{n-1}(x) \\ a_{n-1} p_{n-2}(x) \end{pmatrix}. \tag{1-33}$$

We normalize, placing a_n on the lower component, so that

$$\det(A_j(z)) = 1. \tag{1-34}$$

The transfer matrix is then defined by

$$T_n(z) = A_n(z)\dots A_1(z), \tag{1-35}$$

so

$$\begin{pmatrix} p_n(x) \\ a_n p_{n-1}(x) \end{pmatrix} = T_n(x) \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{1-36}$$

If $\tilde{p}_n$ are the OPRL associated to the once stripped Jacobi parameters $\{a_{n+1}, b_{n+1}\}_{n=1}^{\infty}$, and

$$q_n(x) = -a_1^{-1} \tilde{p}_{n-1}(x) \tag{1-37}$$

with $q_0 = 0$, then

$$T_n(z) = \begin{pmatrix} p_n(z) & q_n(z) \\ a_n p_{n-1}(z) & a_n q_{n-1}(z) \end{pmatrix}. \tag{1-38}$$

Here is how we will establish (1-30) and (1-31):

Theorem 3. *Fix $x_0 \in \mathbb{R}$. Suppose that*

$$\sup_n \frac{1}{n+1} \sum_{j=0}^{n} \|T_j(x_0)\|^2 \le C < \infty. \tag{1-39}$$

Then for all $z \in \mathbb{C}$ and all n,

$$\frac{1}{n+1} \sum_{j=0}^{n} \left\| T_j\left(x_0 + \frac{z}{n+1}\right)\right\|^2 \le C \exp(2C\alpha_-^{-1}|z|). \tag{1-40}$$

Moreover, if

$$\sup_n \|T_n(x_0)\|^2 = C < \infty, \tag{1-41}$$

then for all $z \in \mathbb{C}$ and n,

$$\left\| T_n\left(x_0 + \frac{z}{n+1}\right)\right\| \le C^{1/2} \exp(C\alpha_-^{-1}|z|). \tag{1-42}$$

Remarks. 1. Our proof is an abstraction of ideas of Avila and Krikorian [2006], who only treated the ergodic case.

2. α_- is given by (1-7).
3. There is a conjecture, called the Schrödinger conjecture [Maslov et al. 1993], that says (1-41) holds for a.e. $x_0 \in \Sigma_{\mathrm{ac}}(d\mu)$.

Our last two theorems below are special to the ergodic situation. Let Ω be a compact metric space, $d\eta$ a probability measure on Ω, and $S\colon \Omega \to \Omega$ an ergodic invertible map of Ω to itself. Let A, B be continuous real-valued functions on Ω with $\inf_\omega A(\omega) > 0$. Let

$$\alpha_+ = \|A\|_\infty, \qquad \beta = \|B\|_\infty, \qquad \alpha_- = \|A^{-1}\|_\infty^{-1}. \tag{1-43}$$

For each $\omega \in \Omega$, J_ω is the Jacobi matrix with

$$a_n(\omega) = A(S^{n-1}\omega), \qquad b_n(\omega) = B(S^{n-1}\omega). \tag{1-44}$$

Equation (1-43) is consistent with (1-4) and (1-7). Usually one only takes Ω, a measure space, and A, B bounded measurable functions, but by replacing Ω by $([\alpha_-, \alpha_+] \times [-\beta, \beta])^\infty \equiv \widetilde{\Omega}$ and mapping $\Omega \to \widetilde{\Omega}$ by $\omega \mapsto (A(S^n\omega), B(S^n\omega))_{n=-\infty}^{\infty}$, we get a compact space model equivalent to the original measure model. We use $d\mu_\omega$ for the spectral measure of J_ω and $p_n(x, \omega)$ for $p_n(x, d\mu_\omega)$.

The canonical example of the setup with a.c. spectrum is the almost Mathieu equation. Let α be a fixed irrational, λ a nonzero real, and $\Omega = \partial\mathbb{D}$ the unit circle $\{e^{i\theta} \mid \theta \in [0, 2\pi)\}$. Then take

$$a_n = 1, \qquad b_n = 2\lambda \cos(\pi\alpha n + \theta),$$

(so $S(e^{i\theta}) = e^{i\theta}e^{i\pi\alpha}$, $d\eta(\theta) = d\theta/2\pi$). If $0 \neq |\lambda| < 1$, it is known [Avila 2008; Avila and Damanik 2008; Avila and Jitomirskaya 2008; Jitomirskaya 2007] that the spectrum is purely a.c. and is a Cantor set. It is also known [Jitomirskaya 2007] that if $|\lambda| \geq 1$, there is no a.c. spectrum.

Theorem 4. *Let $\{J_\omega\}_{\omega\in n}$ be an ergodic family with Σ_{ac}, the common essential support of the a.c. spectrum of J_ω, of positive Lebesgue measure. Then for a.e. pairs $(x, \omega) \in \Sigma_{\mathrm{ac}} \times \Omega$,*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{j=0}^{n} |p_j(x, w)|^2 \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n+1} \sum_{j=0}^{n} |q_j(x, w)|^2 \tag{1-45}$$

exist.

Theorem 5. *For a.e. (x, ω) in $\Sigma_{\mathrm{ac}} \times \Omega$, the first limit in* (1-45) *is $\rho_\infty(x)/w_\omega(x)$, where ρ_∞ is the density of the a.c. part of the DOS.*

This is, of course, an analog of the celebrated results of Máté et al. [1991] (for $[-1, 1]$) and Totik [2000] (for general sets $\mathfrak{e}$ containing open intervals) for regular measures obeying a local Szegő condition.

Theorems 3–5 show the applicability of Theorem 2, and so lead to:

Corollary 1.5. *For any ergodic Jacobi matrix, we have universality and strong clock behavior for a.e. ω and a.e. $x_0 \in \Sigma_{\mathrm{ac}}$.*

In particular, the almost Mathieu equation has strong clock behavior for the zeros.

Remark. It is possible to show that for the almost Mathieu equation there is universality for a.e. $x_0 \in \Sigma_{\mathrm{ac}}$ and *every* ω. Our current approach to this uses that the Schrödinger conjecture is true for the almost Mathieu operator, a recently announced result [Avila et al. $\geq$ 2010].

For $n = 1, 2, 3, 4, 5$, Theorem n is proven in Section $n + 1$. Section 7 has some further remarks.

2. Lubinsky's second approach

In this section, we will prove Theorem 1. We begin with two overall visions relevant to the proof. First, the *sinc kernel* $\sin \pi z/\pi z$ [Lund and Bowers 1992] enters as the Fourier transform of a suitable multiple of the characteristic function of $[-\pi, \pi]$.

Second, the ultimate goal of quasiclock spacing is that on a $1/n\rho_n$ scale, zeros are a unit distance apart, so on this scale

$$\#\text{ of zeros in } [0,n] \sim n. \tag{2-1}$$

Lubinsky's realization is that the Lubinsky wiggle condition and Markov–Stieltjes inequalities (see below) imply that the difference of the two sides of (2-1) is bounded by 1. This is close enough that, together with some complex variable magic, one gets unit spacing.

The complex variable magic is encapsulated in the following result whose proof we defer until the end of the section.

Theorem 2.1. *Let f be an entire function with the following properties*:

(a) $f(0) = 1$.

(b) $\sup_{x\in\mathbb{R}}|f(x)| < \infty$.

(c) $\displaystyle\int_{-\infty}^{\infty} |f(x)|^2\,dx \le 1$.

(d) *f is real on $\mathbb{R}$.*

(e) *All the zeros of f lie on $\mathbb{R}$ and if these zeros are labeled by $\cdots \le z_{-2} \le z_{-1} < 0 < z_1 \le z_2 \le \cdots$, with $z_0 \equiv 0$, then*

$$|z_j - z_k| \ge |j-k| - 1. \tag{2-2}$$

(f) *For each $\varepsilon > 0$, there is C_ε with*

$$|f(z)| \le C_\varepsilon e^{\varepsilon |z|^2}. \tag{2-3}$$

Then

$$f(z) = \frac{\sin(\pi z)}{\pi z}. \tag{2-4}$$

Remarks. 1. Equation (2-2) allows f a priori to have double zeros but not triple or higher zeros.

2. It is easy to see there are examples where (2-3) holds for some but not all ε and where (2-4) is false, so (2-3) is sharp.

Proof of Theorem 1 given Theorem 2.1. (This part of the argument is essentially in [Lubinsky 2008b].) Fix $a \in \mathbb{R}$ and let

$$f_n(z) = \frac{K_n\bigl(x_0 + a/(n\rho_n),\, x_0 + (a+z)/(n\rho_n)\bigr)}{K_n(x_0,x_0)}. \tag{2-5}$$

By (1-29), (1-30), and (1-16), the f_n are uniformly bounded on each disk $\{z \mid |z| < R\}$, so by Montel's theorem, we have compactness that shows it suffices to prove that any limit point $f(z)$ has the form (2-4). We will show that this putative limit point obeys conditions (a)–(f) of Theorem 2.1.

The Lubinsky wiggle condition (1-26) implies (a). From the Schwarz inequality, (1-11) and the wiggle condition, we get

$$\sup_{x\in\mathbb{R}} |f(x)| = 1, \tag{2-6}$$

which is stronger than (b).

By (1-17),

$$\int_{|y-x_0-(a/n\rho_n)|\le(R/n\rho_n)} |K_n(x,y)|^2 w(y)\,dy \;\le\; K_n(x,x) \tag{2-7}$$

for each $R < \infty$. Changing variables and using the Lebesgue point condition leads to

$$\int_{-R}^{R} |f(y)|^2\, dy \le 1, \tag{2-8}$$

which yields (c) (see Lubinsky [2008b] for more details). In this, one uses (1-29) and (1-30) to see that

$$0 < \inf \rho_n < \sup \rho_n < \infty. \tag{2-9}$$

That f is real on $\mathbb{R}$ is immediate; the reality of zeros follows from Hurwitz's theorem and the fact [Simon 2008a] that $p_{n+1}(x) - cp_n(x)$ has only real zeros for c real.

The Markov–Stieltjes inequalities [Markoff 1884; Freud 1971; Simon 2008a] assert that if $x_1, x_2, \dots$ are successive zeros of $p_n(x) - cp_{n-1}(x)$ for some c, then for $j \ge k+2$,

$$\mu([x_j, x_k]) \ge \sum_{\ell=k+1}^{j-1} \frac{1}{K_n(x_\ell, x_\ell)}. \tag{2-10}$$

Using the fact that the z_j (including z_0) are, by Hurwitz's theorem, limits of x_j's scaled by $n\rho_n$ and the Lubinsky wiggle condition to control limits of $n\rho_n / K_n(x_\ell, x_\ell)$, one finds that (2-2) holds (see [Lubinsky 2008b] for more details). Here one uses that x_0 is a Lebesgue point to be sure that

$$\frac{1}{x_k - x_j} \int_{x_j}^{x_k} d\mu(y) \to w(x_0). \tag{2-11}$$

Finally, (1-30) implies (2-3). Thus, (2-4) holds. □

We now reduce the proof of Theorem 2.1 to using conditions (a)–(e) to improve the bound (2-3).

Proposition 2.2. (a) *Fix $a > 0$. If f is measurable, real-valued and supported on $[-a, a]$ with*

$$\int_{-a}^{a} f(x)^2\, dx \le 2a \quad \text{and} \quad \int_{-a}^{a} f(x)\, dx = 2a, \tag{2-12}$$

then

$$f(x) = \chi_{[-a,a]}(x) \quad \text{a.e.} \tag{2-13}$$

(b) *If f is real-valued and continuous on $\mathbb{R}$ and $\hat{f}$ is supported on $[-\pi, \pi]$ with*

$$\int_{-\infty}^{\infty} f(x)^2\, dx \le 1 \quad \text{and} \quad f(0) = 1, \tag{2-14}$$

then

$$f(x) = \frac{\sin(\pi x)}{\pi x}. \tag{2-15}$$

(c) *If f is an entire function, real on $\mathbb{R}$ with* (2-14), *and for all $\delta > 0$, there is C_δ with*

$$|f(z)| \le C_\delta \exp((\pi + \delta)|\operatorname{Im} z|), \tag{2-16}$$

then (2-4) *holds.*

Proof. (a) Essentially this follows from equality in the Schwarz inequality. More precisely, (2-12) implies

$$\int_{-a}^{a} |f(x) - \chi_{[-a,a]}(x)|^2 \, dx \leq 0. \tag{2-17}$$

(b) Apply Proposition 2.2 (a) to $(2\pi)^{1/2} \widehat{f}(k)$ with $a = \pi$.

(c) By the Paley–Wiener theorem, (2-16) implies that $\widehat{f}$ is supported on $[-\pi, \pi]$. □

Thus, we are reduced to going from (2-3) to (2-16).

By $f(0) = 1$, the reality of the zeros and (2-3), we have, by the Hadamard factorization theorem [Titchmarsh 1932, Section 8.24] that

$$f(z) = e^{Az} \prod_{j \neq 0} \Big(1 - \frac{z}{z_j}\Big) e^{z/z_j}, \tag{2-18}$$

with A real. For $x \in \mathbb{R}$, define $z_j(x)$ to be a renumbering of the z_j, so

$$\ldots \leq z_{-1}(x) < x \leq z_0(x) \leq z_1(x) \leq \ldots. \tag{2-19}$$

By $|z_j - z_k| \geq |k - j| - 1$, we see that

$$z_{n+1}(x) - x \geq n, \qquad x - z_{-(n+1)}(x) \geq n. \tag{2-20}$$

In particular, $(x - 1.1, x + 1.1)$ can contain at most $z_0(x)$, $z_{\pm 1}(x)$, $z_{\pm 2}(x)$. Removing the open intervals of size $2/10$ about each of the five points $|z_\ell(x) - x|$ ($\ell = 0, \pm 1, \pm 2$) from $[0, 1]$ leaves at least one $\delta > 0$, that is, we can pick $\delta(x)$ in $[0, 1]$ so for all j,

$$|z_j(n) - (x \pm \delta)| \geq \tfrac{1}{10}. \tag{2-21}$$

Moreover, by (2-20), for $n = 1, 2, \ldots,$

$$|z_{\pm(n+2)}(x) - (x \pm \delta)| \geq n. \tag{2-22}$$

Since

$$\frac{|1 - (x + iy)/z_j|^2}{|(1 - (x + \delta/z_j)(1 - x - \delta)/z_j)|} \leq 1 + \frac{(y^2 + \delta^2)}{|z_j - (x + \delta)||z_j - (x - \delta)|}, \tag{2-23}$$

we conclude from (2-18) that

$$\frac{|f(x + iy)|^2}{|f(x - \delta)||f(x + \delta)|} \leq \Big[1 + \frac{y^2 + 1}{(1/100)}\Big]^5 \prod_{n=1}^{\infty} \Big(1 + \frac{1 + y^2}{n^2}\Big)^2 \leq C(1 + y^{10}) \Big(\frac{\sinh \pi \sqrt{y^2 + 1}}{\pi \sqrt{y^2 + 1}}\Big)^2. \tag{2-24}$$

Thus, for any ε, there is a C_ε with

$$|f(x + iy)| \leq C_\varepsilon \exp((\pi + \varepsilon)|y|), \tag{2-25}$$

for every $x + iy \in \mathbb{C}$, which is (2-16). This concludes the proof of Theorem 2.1.

Remark. It is possible to show, using the Phragmén–Lindelöf principle [Titchmarsh 1932], that if one assumes, instead of (2-3), the stronger $|f(z)| \leq Ce^{|z|^\delta}$, then it is possible to weaken (2-2) to

$$|z_j| \geq |j| - 1, \tag{2-26}$$

for if (2-26) holds, then (2-18) implies that

$$|f(iy)| \leq C(1+|y|)e^{\pi|y|}. \tag{2-27}$$

Applying Phragmén–Lindelöf to $(1-iz)^{-1} f(z)e^{i\pi z}$ on the sectors $\arg z \in [0, \pi/2]$ and $[\pi/2, \pi]$ proves that

$$|f(x+iy)| \leq C(1+|z|)e^{\pi|y|}. \tag{2-28}$$

3. Doing the Lubinsky wiggle

Our goal in this section is to prove Theorem 2.

Proof of Theorem 2. By Egorov's theorem [Rudin 1987, p. 73], for every ε, there exists a compact set $\mathcal{L} \subset \Sigma$ with $|\Sigma \setminus \mathcal{L}| < \varepsilon$ (with $|\ | =$ Lebesgue measure) so that on $\mathcal{L}$, the sequence $\frac{1}{n+1} K_n(x,x) \equiv \tilde{q}_n(x)$ converges uniformly to a limit, which we call $\tilde{q}(x)$. If we prove that (1-26) holds for a.e. $x_0 \in \mathcal{L}$, then by taking a sequence of ε's going to 0, we get that (1-26) holds for a.e. $x_0 \in \Sigma$

By Lebesgue's theorem on differentiability of integrals of L^1-functions [Rudin 1987, Theorem 7.7] applied to the characteristic function of $\mathcal{L}$, for a.e. $x_0 \in \mathcal{L}$, we get

$$\lim_{\delta \downarrow 0} (2\delta)^{-1} |(x_0 - \delta, x_0 + \delta) \cap \mathcal{L}| = 1. \tag{3-1}$$

We will prove that (1-26) holds for all x_0 with (3-1) and with condition (iv) of Theorem 2.

The expression $\frac{1}{n+1} K_n\bigl(x + \frac{a}{n} + \frac{\bar{z}}{n}, x + \frac{a}{n} + \frac{z}{n}\bigr)$ is analytic in z, so by a Cauchy estimate and a real,

$$\left|\frac{d}{da}\tilde{q}_n\Bigl(x+\frac{a}{n}\Bigr)\right| \leq \sup_{|z|\leq 1} \frac{1}{n+1}\left|K_n\Bigl(x+\frac{a}{n}+\frac{\bar{z}}{n}, x+\frac{a}{n}+\frac{z}{n}\Bigr)\right| = \sup_{|z|\leq 1}\left|\tilde{q}_n\Bigl(x+\frac{a}{n}+\frac{z}{n}\Bigr)\right|. \tag{3-2}$$

By a Schwarz inequality, for $x, y \in \mathbb{C}$,

$$\frac{1}{n+1}\,|K_n(x,y)| \leq (\tilde{q}_n(x)\tilde{q}_n(y))^{1/2}. \tag{3-3}$$

Thus, using the assumed (1-30), for any x_0 for which (1-30) holds and any $A < \infty$, there are N_0 and C so for $n \geq N_0$,

$$\left|\tilde{q}_n\Bigl(x_0+\frac{a}{n}\Bigr) - \tilde{q}_n\Bigl(x_0+\frac{b}{n}\Bigr)\right| \leq C|a-b|, \tag{3-4}$$

for all a, b with $|a| \leq A$, $|b| \leq A$.

Since each $\tilde{q}_n$ is continuous and the convergence is uniform on $\mathcal{L}$, $\tilde{q}$ is continuous on $\mathcal{L}$. Thus, we have for each $A < \infty$,

$$\sup\left\{\left|\tilde{q}\Bigl(x_0+\frac{a}{n}\Bigr) - \tilde{q}(x_0)\right| \;\middle|\; |a| < A,\ x_0+\frac{a}{n} \in \mathcal{L}\right\} \to 0, \tag{3-5}$$

as $n \to \infty$. By the uniform convergence theorem,

$$\sup\left\{\left|\tilde{q}_n\Bigl(x_0+\frac{a}{n}\Bigr) - \tilde{q}_n(x_0)\right| \;\middle|\; |a| < A,\ x_0+\frac{a}{n} \in \mathcal{L}\right\} \to 0. \tag{3-6}$$

We next note that (3-1) implies

$$\sup_{|b|\le A} n \operatorname{dist}\Bigl(x_0 + \frac{b}{n}, \mathcal{L}\Bigr) \to 0; \tag{3-7}$$

equivalently, for any ε, there is an N_1 so for $n \ge N_1$ and $|b| < A$, there exists $|a| < A$ (a will be n-dependent) so that $|a-b| < \varepsilon$ and $x_0 + a/n \in \mathcal{L}$. We have

$$\Bigl|\tilde{q}_n\Bigl(x_0+\frac{b}{n}\Bigr)-\tilde{q}_n(x_0)\Bigr| \le \Bigl|\tilde{q}_n\Bigl(x_0+\frac{b}{n}\Bigr)-\tilde{q}_n\Bigl(x_0+\frac{a}{n}\Bigr)\Bigr| + \Bigl|\tilde{q}_n\Bigl(x_0+\frac{a}{n}\Bigr)-\tilde{q}_n(x_0)\Bigr|, \tag{3-8}$$

where $|b-a| < \varepsilon$ and $x_0 + a/n \in \mathcal{L}$. By (3-4), if $n \ge \max(N_0, N_1)$, the first term is bounded by $C\varepsilon$ and, by (3-7), the second term goes to zero, that is,

$$\sup_{|b|<A} \Bigl|\tilde{q}_n\Bigl(x_0+\frac{b}{n}\Bigr)-\tilde{q}_n(x_0)\Bigr| \to 0. \tag{3-9}$$

Since $\tilde{q}_n(x_0) \to \tilde{q}(x_0) \ne 0$, we have

$$\sup_{|b|<A} \Bigl|\frac{\tilde{q}_n(x_0+b/n)}{\tilde{q}_n(x_0)} - 1\Bigr| \to 0, \tag{3-10}$$

as $n \to \infty$, which is (1-26). □

4. Exponential bounds for perturbed transfer matrices

In this section, our goal is to prove Theorem 3. As noted in the Introduction, our approach is an extension of a theorem of Avila and Krikorian [2006, Lemma 3.1] exploiting that one can avoid using cocycles and so go beyond the apparent limitation to ergodic situations. The argument here is related to but somewhat different from variation of parameters techniques [Jitomirskaya and Last 1999; Killip et al. 2003] and should have wide applicability.

Proof of Theorem 3. Fix n and define, for $j = 1, 2, \dots, n$,

$$\tilde{A}_j = A_j\Bigl(x_0 + \frac{z}{n+1}\Bigr), \tag{4-1}$$

$$A_j = A_j(x_0), \tag{4-2}$$

$$T_j = A_j \dots A_1, \qquad \tilde{T}_j = \tilde{A}_j \dots \tilde{A}_1. \tag{4-3}$$

(Note that $\tilde{A}_j$ and $\tilde{T}_j$ depend on n as well as j.)

Note that, by (1-32),

$$\tilde{A}_j - A_j = a_j^{-1}\begin{pmatrix} z/(n+1) & 0 \\ 0 & 0 \end{pmatrix}, \tag{4-4}$$

so that

$$\|\tilde{A}_j - A_j\| \le \alpha_-^{-1} \quad \frac{|z|}{n+1}. \tag{4-5}$$

Write

$$T_j^{-1}\tilde{T}_j = (T_j^{-1}\tilde{A}_j T_{j-1})(T_{j-1}^{-1}\tilde{A}_{j-1}T_{j-2})\dots(T_1^{-1}\tilde{A}_1 T_0) \\ = (1+B_j)(1+B_{j-1})\dots(1+B_1), \tag{4-6}$$

where

$$B_k = T_k^{-1}(\tilde{A}_k - A_k)T_{k-1}. \tag{4-7}$$

Here we used

$$A_k T_{k-1} = T_k. \tag{4-8}$$

Since T_k has determinant 1 (see (1-34)), we have

$$\|T_k^{-1}\| = \|T_k\|. \tag{4-9}$$

So, by (4-5),

$$\|B_k\| \le \|T_k\|\,\|T_{k-1}\|\alpha_-^{-1}\,\frac{|z|}{n+1}. \tag{4-10}$$

Thus, since

$$\|1+B_j\| \le 1 + \|B_j\| \le \exp(\|B_j\|), \tag{4-11}$$

Equation (4-6) implies that

$$\|\tilde{T}_j\| \le \|T_j\|\exp\Bigl(\alpha_-^{-1}|z|\Bigl[\frac{1}{n+1}\sum_{k=1}^{j}\|T_k\|\,\|T_{k-1}\|\Bigr]\Bigr). \tag{4-12}$$

By the Schwarz inequality, for $j = 1, 2, \dots, n$,

$$\frac{1}{n+1}\sum_{k=1}^{j}\|T_k\|\,\|T_{k-1}\| \le \frac{1}{n+1}\sum_{k=0}^{j}\|T_k\|^2 \le \frac{1}{n+1}\sum_{k=0}^{n}\|T_k\|^2. \tag{4-13}$$

Using (1-39) and (4-12), we find

$$\|\tilde{T}_j\| \le \|T_j\|\exp(C\alpha_-^{-1}|z|). \tag{4-14}$$

This clearly holds for $j = 0$ also. Squaring and summing,

$$\frac{1}{n+1}\sum_{j=0}^{n}\|\tilde{T}_j\|^2 \le \Bigl(\frac{1}{n+1}\sum_{j=0}^{n}\|T_j\|^2\Bigr)\exp(2C\alpha_-^{-1}|z|), \tag{4-15}$$

which is (1-40).

Note that (1-41) implies (1-39) so that (1-42) is just (4-14). □

We note that the argument above can also be used for more general perturbative bounds. For example, suppose that

$$C_1 \equiv \sup_n \|T_n(x_0)\| < \infty, \tag{4-16}$$

for a given set of Jacobi parameters. Let $a_n' = a_n + \delta a_n$ and $b_n' = b_n + \delta b_n$ with

$$C_2 \equiv \sum_{n=1}^{\infty} |\delta a_n| + |\delta b_n| < \infty \tag{4-17}$$

and

$$\alpha'_- = \inf a_n' > 0. \tag{4-18}$$

Defining $\tilde{A}_n, \tilde{T}_n$ at energy x_0 but with $\{a_n', b_n'\}_{n=1}^{\infty}$ Jacobi parameters, one gets

$$\|\tilde{A}_k - A_k\| \le C_3[\alpha_-^{-1} + (\alpha'_-)^{-1}](|\delta a_k| + |\delta b_k|) \tag{4-19}$$

for some universal constant C_3. Thus

$$\|B_k\| \le C_3 C_1^2 [\alpha_-^{-1} + (\alpha_n')^{-1}](|\delta a_k| + |\delta b_k|) \tag{4-20}$$

and

$$\|\tilde{T}_n\| \le C_1 \exp(C_1^2 C_2 C_3 [\alpha_-^{-1} + (\alpha'_-)^{-1}]), \tag{4-21}$$

providing another proof of a standard ℓ^1 perturbation result.

5. Ergodic Jacobi matrices and Cesàro summability

In this section, our goal is to prove Theorem 4. We fix an ergodic Jacobi matrix setup. We will need to use certain special solutions:

Theorem 5.1 [Deift and Simon 1983]. *For any Jacobi matrix with* $\Sigma_{\mathrm{ac}}(d\mu_\omega)$ (*which is a.e.* ω*-independent*) *of positive measure*, *for a.e. pair* $(x, \omega) \in \Sigma_{\mathrm{ac}} \times \Omega$ (*a.e. with respect to* $dx \otimes d\eta(\omega)$), *there exist sequences* $\{u_n^{\pm}(x,\omega)\}_{n=-\infty}^{\infty}$ *such that*

$$T_n(x,\omega)\begin{pmatrix} u_1^{\pm}(x,\omega) \\ a_0 u_0^{\pm}(x,\omega)\end{pmatrix} = \begin{pmatrix} u_{n+1}^{\pm}(x,\omega) \\ a_n u_n^{\pm}(x,\omega)\end{pmatrix}, \tag{5-1}$$

with the following properties:

(i) $u_n^-(x,\omega) = \overline{u_n^+(x,\omega)}$;

(ii) $a_n(u_{n+1}^+ u_n^- - u_{n+1}^- u_n^+) = -2i$;

(iii) $|u_n^+(x,\omega)| = |u_0^+(x, S^n\omega)|$;

(iv) $\int |u_n^+(x,\omega)|^2 \, d\eta(\omega) < \infty$;

(v) $u_0^{\pm}$ *is real.*

Of course, by (iii), the integral in (iv) is n-independent. For later purposes (see Section 6), we will need an explicit formula for this integral. In fact, we will need explicit formulae for u_0, u_{-1} in terms of the m-function.

For $\operatorname{Im} z > 0$, one defines $\tilde{u}_n^+(z,\omega)$ so as to solve the following equation equivalent to (5-1):

$$a_n \tilde{u}_{n+1}^+ + (b_n - z)\tilde{u}_n^+ + a_{n-1}\tilde{u}_{n-1}^+ = 0, \tag{5-2}$$

with $\sum_{n=1}^{\infty}|\tilde{u}_n^+|^2<\infty$. This determines $\tilde{u}_n^+$ up to a constant, and so

$$m(z,\omega)=-\frac{\tilde{u}_1^+(z,\omega)}{a_0\tilde{u}_0^+(z,\omega)} \tag{5-3}$$

is normalization-independent and, by (5-2), obeys

$$m(z,\omega)=\frac{1}{-z+b_1-a_1^2m(z,S\omega)}. \tag{5-4}$$

(Note: We have suppressed the ω-dependence of a_n, b_n.)

As usual with solutions of (5-4),

$$m(z,\omega)=\int\frac{d\mu_\omega^+(x)}{x-z}, \tag{5-5}$$

where $d\mu_\omega^+$ is the measure associated to the half-line Jacobi matrix J_ω.

For a.e. $x\in\Sigma_{\mathrm{ac}}$ and a.e. ω, $m(x+i0,\omega)$ exists and has

$$\operatorname{Im} m(x+i0,\omega)>0 \qquad (\text{a.e. } x\in\Sigma_{\mathrm{ac}}), \tag{5-6}$$

We normalize the solution u^+ obeying Theorem 5.1 by defining:

$$u_0^+(x,\omega)=\frac{1}{a_0[\operatorname{Im} m(x+i0,\omega)]^{1/2}}, \tag{5-7}$$

$$u_1^+(x,\omega)=-\frac{m(x+i0,\omega)}{[\operatorname{Im} m(x+i0,\omega)]^{1/2}}. \tag{5-8}$$

(We have listed all the formulae because [Deift and Simon 1983] only considers the case $a_n\equiv 1$.) The u_n^+ are then determined by the difference equation, and the u_n^- by condition (i).

Of course, we have

$$p_n=\frac{u_{n+1}^+-u_{n+1}^-}{u_1^+-u_1^-}, \tag{5-9}$$

since both sides obey the same difference equations with $p_{-1}=0$ (since $u_0^+=u_0^-$) and $p_0=1$.

By (5-9), to prove Theorem 4 we need to show that

$$\frac{1}{n}\sum_{j=0}^{n-1}(u_{j+1}^+-u_{j+1}^-)^2 \tag{5-10}$$

exists. This follows from the existence of

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}|u_j^+|^2 \tag{5-11}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}(u_j^+)^2. \tag{5-12}$$

From condition (iii) and the ergodic theorem (plus (iv)), the a.e. ω existence of the limit in (5-11) is immediate. In cases like the almost Mathieu equation with Diophantine frequencies where u_n^+ is almost periodic, one also gets the existence of the limit in (5-12) directly, but there are examples, like the almost Mathieu equation with frequencies whose dual has singular continuous spectrum, where the phase of u_n^+ is not almost periodic. So this argument does not work in general. In fact, we will eventually prove that for a.e. (x,ω) in $\Sigma_{\mathrm{ac}} \times \Omega$ (see Theorem 6.3):

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (u_j^+)^2 = 0. \tag{5-13}$$

It would be interesting to have a direct proof of this (for the periodic case, see [Simon 2010]) rather than the indirect path we will take.

Define the 2×2 matrix

$$U_n(x,\omega) = \frac{1}{(-2i)^{1/2}} \begin{pmatrix} u_{n+1}^+(x,\omega) & u_{n+1}^-(x,\omega) \\ a_n u_n^+(x,\omega) & a_n u_n^-(x,\omega) \end{pmatrix}, \tag{5-14}$$

(where we fix once and for all a choice of $\sqrt{-2i}$). By condition (ii),

$$\det(U_n(x,\omega)) = 1 \tag{5-15}$$

and, by (5-1),

$$T_n(x,\omega) U_0(x,\omega) = U_n(x,\omega) \tag{5-16}$$

or

$$T_n(x,\omega) = U_n(x,\omega) U_0(x,\omega)^{-1}. \tag{5-17}$$

For now, we fix $x \in \Sigma_{\mathrm{ac}}$ with

$$E([a_0(\omega)^2 \operatorname{Im} m(x+i0,\omega)]^{-1}) < \infty, \tag{5-18}$$

(known Lebesgue a.e. by Kotani theory; see [Simon 1983; Deift and Simon 1983]), so U_n can be defined and is in L^2.

Theorem 5.2. *Fix a matrix Q. For a.e. ω, the limit of matrices*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} T_j(x,\omega)^t \, Q \, T_j(x,\omega) \tag{5-19}$$

exists.

Proof of Theorem 4 given Theorem 5.2. Pick

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then the 1,1 matrix element of $T_j(x,\omega)^t Q T_j(x,\omega)$ is $p_j(x,\omega)^2$, and the 2,2 element is $q_j(x,\omega)^2$. Since the limits in (1-45) exist, we are done. $\square$

Equation (5-17) plus condition (iv) will imply critical a priori bounds on $\|T_n(x,\,\cdot\,)\|_{L^1(d\eta)}$. It will be convenient to use the Hilbert–Schmidt norm on these 2×2 matrices.

Lemma 5.3. *We have*

$$\sup_n \int \|T_n(x,\omega)\|\,d\eta(\omega) < \infty. \tag{5-20}$$

Proof. Since $\det(U_n) = 1$,

$$\|U_n(x,\omega)^{-1}\| = \|U_n(x,\omega)\|. \tag{5-21}$$

Thus, by (5-17),

$$\|T_n(x,\omega)\| \le \|U_n(x,\omega)\|\,\|U_0(x,\omega)\|. \tag{5-22}$$

By the Schwarz inequality,

$$\sup_n \int \|T_n(x,\omega)\|\,d\eta(\omega) \le \sup_n \int \|U_n(x,\omega)\|^2\,d\eta(\omega) = \int \|U_0(x,\omega)\|^2\,d\eta(\omega) < \infty,$$

where we also have used condition (iv) and the equality

$$\|U_j(x,\omega)\| = \|U_0(x,S^j\omega)\|, \tag{5-23}$$

a consequence of condition (iii) and our use of Hilbert–Schmidt norms. □

Let $A_j(\omega)$ be the matrix (1-32) with $a_j = a_j(\omega)$, $b_j = b_j(\omega)$ and let

$$A(\omega) \equiv A_1(\omega), \tag{5-24}$$

so

$$A_j(\omega) = A(S^{j-1}\omega), \tag{5-25}$$

and the transfer matrix for J_ω is

$$T_n(\omega) = A(S^{n-1}\omega)\dots A(\omega). \tag{5-26}$$

Now form the suspension

$$\widehat{\Omega} = \Omega \times \mathbb{SL}(2,\mathbb{C}) \tag{5-27}$$

and define $\widehat{S}\colon \widehat{\Omega} \to \widehat{\Omega}$ by

$$\widehat{S}(\omega, C) = (S\omega, A(\omega)C), \tag{5-28}$$

so

$$\widehat{S}^n(\omega, C) = (S^n\omega, T_n(\omega)C). \tag{5-29}$$

Theorem 5.4. *There exists an $\widehat{S}$-invariant probability measure $d\nu$ on $\widehat{\Omega}$ whose projection onto Ω is $d\eta$ and with*

$$\int \|C\|\,d\nu(\omega, C) < \infty. \tag{5-30}$$

Proof. Pick any probability measure μ_0 on $\mathbb{SL}(2,\mathbb{C})$ with $\int \|C\|^k\,d\mu_0(C) < \infty$ for all k. For example, one could take $d\mu_0(C) = Ne^{-\|C\|^2}d\,\mathrm{Haar}(C)$ where N is a normalization constant. Let $\widehat{S}_*$ be induced on measures on $\widehat{\Omega}$ by $[\widehat{S}_*(\nu)](f) = \nu(f\circ\widehat{S})$. Let

$$\nu_n = \widehat{S}_*^n(\eta \otimes \mu_0). \tag{5-31}$$

Then the invariance of η under S_* implies the projection of ν_n is η and

$$\int \|C\|\,d\nu_n = \int \|T_n(\omega)C\|\,d\eta \otimes d\mu_0 \le \Big(\int \|T_n(\omega)\|\,d\eta\Big)\Big(\int \|C\|\,d\mu_0\Big), \tag{5-32}$$

which, by (5-20), is uniformly bounded in n.

Let $\tilde{\nu}_n$ be the Cesàro averages of ν_n, that is,

$$\tilde{\nu}_n = \frac{1}{n}\sum_{j=0}^{n-1} \nu_j. \tag{5-33}$$

So, by (5-32),

$$\sup_n \int \|C\|\,d\tilde{\nu}_n < \infty, \tag{5-34}$$

so $\{\tilde{\nu}_n\}$ are tight, that is,

$$\lim_{K\to\infty} \sup_n \tilde{\nu}_n\{C \mid \|C\| \ge K\} \to 0,$$

which implies that $\tilde{\nu}_n$ has a weak limit point in probability measures on $\widetilde{\Omega}$. This weak limit point is invariant and, by (5-34), it obeys (5-30). ☐

Lemma 5.5. *Let $L < \infty$. Let*

$$\widehat{\Omega}_L = \{(\omega, C) \mid \|U_0(\omega)\| < L, \|C\| < L\}. \tag{5-35}$$

Then for any ε, there is a K so that for a.e. $(\omega, C) \in \widehat{\Omega}_L$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\substack{j\in B(K,\omega,C)\\ 0\le j\le n-1}} \|T_j(\omega)C\|^2 \le \varepsilon, \tag{5-36}$$

where

$$B(K,\omega,C) = \{j \mid \|T_j(\omega)C\| \ge K\}. \tag{5-37}$$

Proof. Since $U_0(\omega) \in L^2(d\eta)$, we have

$$\lim_{s\to\infty} \int_{\|U_0(\omega)\|\ge s} \|U_0(\omega)\|^2 d\eta(\omega) = 0, \tag{5-38}$$

so for any $\delta > 0$, there exists $s(\delta)$ so that the integral is less than δ.

Let $\widetilde{B}(\widetilde{K},\omega)$ be defined by

$$\widetilde{B}(\widetilde{K},\omega) = \{j \mid \|U_j(\omega)\| \ge \widetilde{K}\}. \tag{5-39}$$

By the Birkhoff ergodic theorem and (5-23) for a.e. ω,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\substack{j\in \widetilde{B}(\widetilde{K},\omega)\\ 0\le j\le n-1}} \|U_j(\omega)\|^2 = \int_{\|U_0(\omega)\|\ge \widetilde{K}} \|U_0(\omega)\|^2 d\eta \le \delta, \tag{5-40}$$

if $\widetilde{K} \ge s(\delta)$.

Given ε and L, let $\delta = \varepsilon/L^2$ and $K \ge L^2 s(\delta)$. Since

$$\|T_j(\omega)C\| \le \|U_j(\omega)\| L^2 \tag{5-41}$$

if $(\omega, C) \subset \Omega_L$,

$$B(K,\omega,C) \subset \widetilde{B}\Big(\frac{K}{L^2},\omega\Big).$$

So, by (5-40) and (5-41),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\substack{j\in B(K,\omega,C)\\ 0\le j\le n-1}} \|T_j(\omega)C\|^2 \le L^2\delta = \varepsilon, \tag{5-42}$$

which is (5-35). $\square$

Proof of Theorem 5.2. Without loss, suppose $\|Q\| \le 1$. Define on $\widehat{\Omega}$

$$f_n(\omega, C) = \frac{1}{n} \sum_{j=0}^{n-1} C^t T_j(x,\omega)^t Q T_j(x,\omega) C. \tag{5-43}$$

If we prove that this has a pointwise limit for ν a.e. (ω, C), we are done: since η is the projection of ν, for η a.e. ω, there are some C for which (5-43) has a limit. But C is invertible, so $(C^t)^{-1} f_n C^{-1}$ has a limit, that is, (5-19) does.

Notice that if

$$h(\omega, C) = C^t Q C, \tag{5-44}$$

then $f_n(\omega, C)$ is a Cesàro average of $h(\widehat{S}^j(\omega, C))$, so we can almost use the ergodic theorem except we only know a priori that $\int \|h(\omega, C)\|^{1/2}\, d\nu < \infty$, not $\int \|h(\omega, C)\|\, d\nu < \infty$, so we need to use Lemma 5.5.

Fix L and consider $(\omega, C) \in \widehat{\Omega}_L$. Let

$$h_K(\omega, C) = \begin{cases} C^t Q C & \text{if } \|C\| \le K, \\ 0 & \text{if } \|C\| > K. \end{cases} \tag{5-45}$$

Then, since $\|Q\| \le 1$,

$$\|h_K(\widehat{S}^j(\omega, C)) - h(\widehat{S}^j(\omega, C))\| \le \begin{cases} 0 & \text{if } j \notin B(K,\omega,C), \\ \|T_j(\omega)C\|^2 & \text{if } j \in B(K,\omega,C). \end{cases} \tag{5-46}$$

It follows that if

$$f_n^{(K)}(\omega, C) = \frac{1}{n} \sum_{j=0}^{n-1} h_K(\widehat{S}^j(\omega, C)), \tag{5-47}$$

then

$$\|f_n^{(K)}(\omega, C) - f_n(\omega, C)\| \le \text{sum on left side of (5-36)}.$$

So, by Lemma 5.5,

$$\limsup_{n\to\infty} \|f_n^{(K)}(\omega, C) - f_n(\omega, C)\| \le \varepsilon, \tag{5-48}$$

if

$$K \ge K(\varepsilon, L) \tag{5-49}$$

given by the lemma.

For any finite K, h_K is bounded, so the Birkhoff ergodic theorem and the invariance of ν imply, for a.e. (ω, C), $\lim f_n^{(K)}(\omega, C)$ exists. Thus (5-48) and (5-49) imply that $\lim f_n^{(K)}(\omega, C)$ forms a Cauchy sequence as $K \to \infty$ (among, say, integer values), and that its limit is also $\lim f_n(\omega, C)$, for a.e. (ω, C) in $\widehat{\Omega}_L$.

Since L is arbitrary and $\nu(\widehat{\Omega} \setminus \widehat{\Omega}_L) \to 0$ on account of $\int \|U_0(\omega)\|^2 \, d\nu < \infty$, we see that f_n has a limit for a.e. ω, C. □

6. Equality of the local and microlocal DOS

Our main goal in this section is to prove Theorem 5. We know from Theorem 4 that for a.e. $\omega \in \Omega$ and $x_0 \in \Sigma_{\text{ac}}$, we have

$$\frac{1}{n+1} K_n(x_0, x_0) \to k_\omega(x_0) \tag{6-1}$$

some positive function. By Theorems 1 and 2, this implies that the spacing of zeros at a.e. Lebesgue point is

$$x_{j+1}^{(n)}(x_0) - x_j^{(n)}(x_0) \sim \frac{1}{n w_\omega(x_0) k_\omega(x_0)}. \tag{6-2}$$

Thus, for fixed K large, in an interval $(x_0 - K/n, x_0 + K/n)$, the number of zeros is $2Kw(x_0)k(x_0)$. On the other hand, if $\rho_\infty(x_0)$ is the density of states, for a.e. x_0 in the a.c. part of the support of $d\nu_\infty$, the number of zeros in $(x_0 - \delta, x_0 + \delta)$ is approximately $2\delta n \rho(x_0)$. If δ were K/n, this would tell us that

$$w_\omega(x_0) k_\omega(x_0) = \rho_\infty(x_0), \tag{6-3}$$

which is precisely (1-23).

Of course, ρ_∞ is defined by first taking $n \to \infty$ and then $\delta \downarrow 0$, so we cannot set $\delta = K/n$, but (6-3) is an equality of a local density of zeros obtained by taking intervals with $O(n)$ zeros as $n \to \infty$ and a microlocal individual spacing as in (6-2).

So define

$$\rho_L(x_0, \omega) = w_\omega(x_0) k_\omega(x_0), \tag{6-4}$$

the microlocal DOS. Notice that we have indicated an ω-dependence of ρ_L because, at this point, we have not proven ω-independence. ω-independence often comes from the ergodic theorem — we determined the existence of $k_\omega(x_0)$ using the ergodic theorem, but unlike for ρ_∞, the underlying measure was only invariant, not ergodic, and indeed, k_ω, the object we controlled is *not* ω-independent.

Of course, once we prove $\rho_L = \rho_\infty$, ρ_L will be proven ω-independent, but we will, in fact, go the other way: we first prove that ρ_L is ω-independent, use that to show that if u is the Deift–Simon wave function, then the average of u^2 (not $|u|^2$) is zero, and use that to prove that $\rho_L = \rho_\infty$.

Theorem 6.1. *Suppose that J_ω is a family of ergodic Jacobi matrices. Let $\rho_L(x, \omega)$ be determined by* (6-1) *and* (6-4) *for $x \in \Sigma_{\text{ac}}$, $\omega \in \Omega$. Then for a.e. $x \in \Sigma_{\text{ac}}$, $\rho_L(x, \omega)$ is a.e. ω-independent.*

Proof. Since $\rho_L(x, \omega)$ is jointly measurable for $(x, \omega) \in \Sigma_{\text{ac}} \times \Omega$, $\rho_L(x, \cdot)$ is measurable for a.e. x. Since S is ergodic, it suffices to prove that $\rho_L(x, S\omega) = \rho_L(x, \omega)$ for a.e. (x, ω).

Let $p_n(x, \omega)$ be the OPs for J_ω. Then the zeros of $p_{n-1}(x, S\omega)$ and $p_n(x, \omega)$ interlace. It follows, for any interval $I_{n,A}(x_0) = [x_0 - A/n, x_0 + A/n]$, that

$$\left| \# \text{ of zeros of } p_n(x, \omega) \text{ in } I_{n,A}(x_0) - \# \text{ of zeros of } p_{n-1}(x, S\omega) \text{ in } I_{n,A}(x_0) \right| \leq 2. \tag{6-5}$$

If $\rho_L(x_0, S\omega) \neq \rho_L(x_0, \omega)$ and $A = k\rho_L(x_0, \omega)^{-1}$ with k large, it is easy to get a contradiction between (6-5) and (6-2). Thus, $\rho_L(x, \omega) = \rho_L(x, S\omega)$ as claimed. $\square$

Next, we need a connection between ρ_L and u. Recall from (5-9) that

$$p_n(x, \omega) = \frac{\operatorname{Im} u_{n+1}^+(x, \omega)}{\operatorname{Im} u_1^+(x, \omega)}, \tag{6-6}$$

while (5-8) and (5-5) give, respectively,

$$\operatorname{Im} u_1^+(x, \omega) = -[\operatorname{Im} m(x + i0, \omega)]^{1/2}, \tag{6-7}$$

$$\operatorname{Im} m(x + i0, \omega) = \pi w_\omega(x) \quad \text{for a.e. } x \in \Sigma_{\text{ac}}. \tag{6-8}$$

Thus, if we define

$$\operatorname{Av}_\omega(f_j(\omega)) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f_j(\omega), \tag{6-9}$$

then

$$\rho_L(x, \omega) = \frac{1}{\pi} \operatorname{Av}_\omega([\operatorname{Im} u_j^+(x, \omega)]^2). \tag{6-10}$$

Note that $\operatorname{Im} u_j^+(x, \omega)$ is not $\operatorname{Im} u_0^+(x, S^j\omega)$, so we cannot write (6-10) as an integral. In fact, the ω-independence of the right side of (6-10) (because of ω-independence of the left side) will have important consequences.

To see where we are heading, we note the following result (see also [Damanik 2007, Theorem 5]).

Theorem 6.2 [Kotani 1997]. *For a.e. $x \in \Sigma_{\text{ac}}$,*

$$\rho_\infty(x) = \frac{1}{2\pi} \int |u_0^+(x, \omega)|^2 \, d\eta(x). \tag{6-11}$$

Remarks. 1. Kotani [1997] and Damanik [2007] treat $a_n \equiv 1$, but it is easy to accommodate general a_n.

2. Kotani's theorem is not stated in this form but rather as (see Equation (22) in [Damanik 2007]):

$$\pi\rho_\infty(x) = \int \operatorname{Im} G_\omega(0, 0; x + i0) \, d\eta(\omega), \tag{6-12}$$

where G_ω is the whole-line Green's function. Because G_ω is reflectionless, G_ω is pure imaginary and

$$\operatorname{Im}(G_\omega(0, 0; x + i0)) = [2a_0^2 \operatorname{Im} m(x + i0, \omega)]^{-1} = \tfrac{1}{2} |u_0^+(x, \omega)|^2, \tag{6-13}$$

by (5-7).

Thus, the key to proving $\rho_L = \rho_\infty$ will be to show that

$$\operatorname{Av}_\omega([\operatorname{Im} u_j^+(x, \omega)]^2) = \operatorname{Av}_\omega([\operatorname{Re} u_j^+(x, \omega)]^2). \tag{6-14}$$

Note that (6-10) includes that the $\operatorname{Av}_\omega([\operatorname{Im} u_j^+]^2)$ exists and, by the ergodic theorem, $\operatorname{Av}_\omega(|u_j^+|^2)$ exists, so we know for a.e. $(x, \omega) \in \Sigma_{\text{ac}} \times \Omega$ that $\operatorname{Av}_\omega([\operatorname{Re} u_j^+(x, \omega)]^2)$ exists. We are heading towards:

Theorem 6.3. *Suppose $x \in \Sigma_{\mathrm{ac}}$ is such that $\rho_L(x,\omega)$ exists for a.e. ω and is ω-independent, and that*

$$\nu_\infty((-\infty, x]) \neq \tfrac{1}{2}. \tag{6-15}$$

Then for a.e. ω,

$$\mathrm{Av}_\omega((u_j^+(x,\omega))^2) = 0. \tag{6-16}$$

Proof of Theorem 5 given Theorem 6.3. (6-15) fails at most a single x in Σ_{ac}, so (6-16) holds for a.e. $(x,\omega) \in \Sigma_{\mathrm{ac}} \times \Omega$. Its real part implies (6-14), and so for a.e. (x,ω),

$$\mathrm{Av}_\omega([\mathrm{Im}\, u_j^+(x,\omega)]^2) = \tfrac{1}{2}\,\mathrm{Av}_\omega(|u_j^+(x,\omega)|^2) = \tfrac{1}{2}\int |u_0^+(x,\omega)|^2\, d\eta(x), \tag{6-17}$$

by the ergodic theorem. By (6-10), (6-11), and the definition of ρ_L in (6-4) and the paragraphs preceding it, we see that the first limit in (1-45) is $\rho_\infty(x)/w_\omega(x)$. □

Proof of Theorem 6.3. Fix $x \in \Sigma_{\mathrm{ac}}$ (at each stage, we work up to sets of Lebesgue measure 0). Define $\varphi(\omega) \in (0, 2\pi)$ by

$$\mathrm{Arg}(-m(x+i0,\omega)) = -\varphi(\omega). \tag{6-18}$$

Then $\varphi(\omega) \in (0,\pi)$ by $\mathrm{Im}\, m > 0$. Let (φ and s_n also depend on x)

$$s_n(\omega) = \sum_{j=1}^{n} \varphi(S^{j-1}\omega). \tag{6-19}$$

Then, by (5-3) and condition (iii),

$$u_n^+(x,\omega) = e^{-is_n(\omega)} u_0^+(x, S^n\omega) \quad \text{and} \quad u_{n+j}^+(x,\omega) = e^{-is_n(\omega)} u_j^+(x, S^n\omega). \tag{6-20}$$

It follows that for each fixed n,

$$\mathrm{Av}_\omega(\mathrm{Im}\, u_j^+((x, S^n\omega))^2) = \mathrm{Av}_\omega((\mathrm{Im}\, e^{is_n(\omega)} u_j^+(x,\omega))^2). \tag{6-21}$$

If s, x, y are real,

$$\begin{aligned}(\mathrm{Im}(e^{is}(x+iy)))^2 &= (x\sin s + y\cos s)^2 \\ &= y^2 + (\sin^2 s)(x^2 - y^2) + xy(\sin 2s),\end{aligned} \tag{6-22}$$

and thus we can write for the left-hand side of (6-21)

$$\mathrm{Av}_\omega(\mathrm{Im}\, u_j^+((x, S^n\omega))^2) = \mathrm{Av}_\omega([\mathrm{Im}(u_j^+(x,\omega))]^2) + \sin^2 s_n(\omega) R(\omega) + \tfrac{1}{2}\sin(2s_n(\omega)) I(\omega), \tag{6-23}$$

where

$$R(\omega) = \mathrm{Av}_\omega(\mathrm{Re}((u_j^+(x,\omega))^2)), \quad I(\omega) = \mathrm{Av}_\omega(\mathrm{Im}((u_j^+(x,\omega))^2)), \tag{6-24}$$

(all such averages having been previously shown to exist).

We know that for a.e. (x,ω), for $n = 0, 1, 2, \ldots$, the left side of (6-21) exists and is n-independent (and equal to $\rho_L(x,\omega)$). For such (x,ω), (6-23) implies that for all n,

$$\sin s_n(\omega)[\sin s_n(\omega) R(\omega) + \cos s_n(\omega) I(\omega)] = 0. \tag{6-25}$$

We want to consider two cases:

Case 1. For a positive measure set of ω,

$$s_2(\omega) = \pi, \quad s_4(\omega) = 2\pi, \quad s_6(\omega) = 3\pi, \dots. \tag{6-26}$$

Case 2. For a.e. ω, there is an $n(\omega)$ so

$$s_{2j}(\omega) = j\pi \quad (j = 1, \dots, n-1) \qquad s_{2n}(\omega) \neq n\pi. \tag{6-27}$$

In Case 1, for such ω, we have $s_n(\omega)/(n\pi) \to \frac{1}{2}$. It follows by standard Sturm oscillation theory [Johnson and Moser 1982] that $s_n(\omega)/(n\pi) \to \nu_\infty((-\infty, x])$ for almost every ω. Thus, the hypothesis (6-15) eliminates Case 1.

For Case 2, suppose first that n is odd, so $s_{2(n-1)}(\omega)$ is a multiple of 2π and (6-19), for $2n-1$ and $2n$ imply

$$\sin(\varphi_{2n-1})[\sin(\varphi_{2n-1})R + \cos(\varphi_{2n-1})I] = 0, \tag{6-28}$$

$$\sin(\varphi_{2n-1} + \varphi_{2n})[\sin(\varphi_{2n-1} + \varphi_{2n})R + \cos(\varphi_{2n-1} + \varphi_{2n})I] = 0. \tag{6-29}$$

Since $\varphi_{2n-1} \in (0, \pi)$, $\sin(\varphi_{2n-1}) \neq 0$ and since $\varphi_{2n-1} + \varphi_{2n} \in (0, 2\pi) \setminus \{\pi\}$, (for if it equals π, then $s_{2n} = n\pi$!), $\sin(\varphi_{2n-1} + \varphi_{2n}) \neq 0$.

The determinant of equations (6-28)/(6-29) is

$$-\sin(\varphi_{2n-1}) \sin(\varphi_{2n-1} + \varphi_{2n}) \sin(\varphi_{2n}) \neq 0 \tag{6-30}$$

since

$$\sin(A)\cos(B) - \sin(B)\cos(A) = \sin(A - B). \tag{6-31}$$

Here $\neq 0$ in (6-30) comes from $\varphi_{2n} \in (0, \pi)$, so $\sin(\varphi_{2n}) \neq 0$.

The nonzero determinant means that (6-28)/(6-29) $\Rightarrow I = R = 0$, that is, $\mathrm{Av}_\omega((u_j^+)^2) = 0$ for a.e. ω. If n is even, $s_{2(n-1)}(\omega)$ is an odd multiple of π and all equations pick up minus signs, so the argument is unchanged. $\square$

7. Concluding remarks

1. We have proven for general ergodic Jacobi matrices that for a.e. $(x, \omega) \in \Sigma_{\mathrm{ac}} \times \Omega$,

$$\frac{1}{n+1} K_n(x, x; \omega) \to \frac{\rho_\infty(x)}{w_\omega(x)}. \tag{7-1}$$

Here ρ_∞ is the Radon–Nikodým derivative of the a.c. part of $d\rho_\infty$. Based on [Máté et al. 1991; Totik 2000], where results of this type are proven for regular measures, one expects

$$\rho_\infty(x) = \rho_{\mathfrak{e}}(x). \tag{7-2}$$

Here $\mathfrak{e}$ is the essential spectrum of J_ω and $\rho_{\mathfrak{e}}$ its equilibrium measure. Simon [2007, Theorem 1.15] proves

Theorem 7.1. *If Σ_{ac} is not empty, then* (7-2) *holds if and only if, for $\rho_{\mathfrak{e}}$ a.e. x, the Lyapunov exponent, $\gamma(x)$, obeys*

$$\gamma(x) = 0. \tag{7-3}$$

In particular, for examples where (7-3) fails on a set of positive Lebesgue measure in $\mathfrak{e}$ [Bjerklöv 2006; Bourgain 2002a; 2002b; Fedotov and Klopp 2005; 2006], (7-2) may not hold. On the other hand, for examples like the almost Mathieu equation where it is known that (7-3) holds on all of $\mathfrak{e}$ [Bourgain and Jitomirskaya 2002], (7-2) holds. The moral is that (7-2) holds some, but not all, of the time for ergodic Jacobi matrices.

2. Here is an interesting example that provides a deterministic problem where one has strong clock behavior but with a density of zeros, ρ_∞, which is not $\rho_{\mathfrak{e}}$. Let $d\mu$ be a measure on $[-2,2]$ of the form (N is a normalization constant)

$$d\mu(x) = \frac{1}{N}\biggl(\chi_{[-1,1]}(x)\,dx + \sum_{n=1}^{\infty} e^{-n^2}\delta_{x_n}\biggr), \tag{7-4}$$

where $\{x_n\}$ is a dense subset of $[-2,2]\setminus(-1,1)$. Then, as in [Simon 2007, Example 5.8], ρ_∞ exists and is the equilibrium measure for $[-1,1]$ (not $\mathfrak{e}=[-2,2]$). Moreover, the method of [Lubinsky 2009] shows that for $x\in(-1,1)$,

$$\frac{1}{n+1}\,K_n(x,x) \to \frac{\rho_\infty(x)}{N^{-1}}. \tag{7-5}$$

Using either the method of this paper (that is, of [Lubinsky 2008b]) or the method of [Lubinsky 2009], one proves universality with ρ_∞.

3. Simon [2007, Example 5.8] provides a measure with $\sigma_{\text{ess}}(\mu)=[-2,2]$ but $\Sigma_{\text{ac}}=[-2,0]$ and where ν_n has multiple weak limits, including the equilibrium measures for $[-2,0]$ and for $[-2,2]$. By general principles [Stahl and Totik 1992], the set of limits is connected, so uncountable. One would like to prove that quasiclock behavior nevertheless holds for the a.c. spectrum of this model as this will provide a key test for the conjecture that quasiclock behavior always holds on Σ_{ac}.

4. What has sometimes been called the Schrödinger conjecture [Maslov et al. 1993] says that for any Jacobi matrix and a.e. $x\in\Sigma_{\text{ac}}(\mu)$, we have a solution, u_n, with

$$0 < \inf_n |u_n| \le \sup_n |u_n| < \infty \tag{7-6}$$

and $u_{-1}=0$. Invariance of Σ_{ac} under rank one perturbations then proves that for a.e. $x\in\Sigma_{\text{ac}}(\mu)$, the transfer matrix is bounded. Thus, Theorem 3 in the strong form would always be applicable.

5. While (6-15) is harmless since it only eliminates at most one x, one can ask if (6-16) holds even if (6-15) fails. Using periodic problems, it is easy to construct ergodic cases where $\arg u_n^+ = -\pi n/2$, so (6-25) provides no information on $I(\omega)$. Nevertheless, in these cases, one can show $R(\omega)=I(\omega)=0$. We have not been able to find an example where for a set of positive measure ω's, $s_{2n}(\omega)=n\pi$, $s_{2n+1}(\omega)=n\pi+\varphi$ with φ some fixed point in $(0,\pi)\setminus\{\pi/2\}$. In that case, it might happen that $R(\omega)\neq 0$, $I(\omega)\neq 0$. So it remains open if we need to exclude the x with (6-15).

6. While we could use soft methods in Section 3, at one point in our research we used an explicit formula for the derivative of $(1/n)K_n(x_0+a/n,\ x_0+a/n)$ as a function of a that may be useful in other contexts, so we want to mention it. We start with a variation of parameters formula (discussed, for example, in [Jitomirskaya and Last 1999; Killip et al. 2003]) that says that, in terms of the second kind polynomials

of (1-38),

$$p_n(x) - p_n(x_0) = (x - x_0) \sum_{m=0}^{n-1} (p_n(x_0)q_m(x_0) - p_m(x_0)q_n(x_0))\, p_m(x), \tag{7-7}$$

which implies

$$p_n'(x_0) = \sum_{m=0}^{n-1} (p_n(x_0)q_m(x_0) - p_m(x_0)q_n(x_0))\, p_m(x_0). \tag{7-8}$$

Since

$$\frac{d}{da}\frac{1}{n} K_n\Big(x_0 + \frac{a}{n}, x_0 + \frac{a}{n}\Big)\Big|_{a=0} = \frac{1}{n^2}\sum_{j=0}^{n} 2p_j'(x_0)p_j(x_0), \tag{7-9}$$

this leads to

$$\frac{d}{da}\frac{1}{n} K_n\Big(x_0+\frac{a}{n}, x_0+\frac{a}{n}\Big)\Big|_{a=0} = \frac{2}{n^2}\sum_{j=0}^{n}\Big[p_j(x_0)^2\Big(\sum_{k=0}^{j} p_k(x_0)q_k(x_0)\Big) - q_j(x_0)p_j(x_0)\sum_{k=0}^{j} p_k(x_0)^2\Big]. \tag{7-10}$$

As noted in [Simon 2008a], if $(1/n)\sum_{j=0}^{n} p_j(x_0)^2$ and $(1/n)\sum_{j=0}^{n} p_j(x_0)q_j(x_0)$ have limits and $\sup_n[(1/n)\sum_{j=0}^{n} q_j(x_0)^2] < \infty$, then the right side of (7-10) goes to 0.

Acknowledgments

A. Avila thanks M. Flach and T. Tombrello for the hospitality of Caltech. B. Simon would like to thank E. de Shalit for the hospitality of Hebrew University. This research was partially conducted during the period Avila served as a Clay Research Fellow. We would like to thank H. Furstenberg and B. Weiss for useful comments.

References

[Avila 2008] A. Avila, "Absolutely continuous spectrum for the almost Mathieu operator", preprint, 2008. arXiv 0810.2965

[Avila and Damanik 2008] A. Avila and D. Damanik, "Absolute continuity of the integrated density of states for the almost Mathieu operator with non-critical coupling", *Invent. Math.* **172**:2 (2008), 439–453. MR 2009i:47066 Zbl 1149.47021

[Avila and Jitomirskaya 2008] A. Avila and S. Jitomirskaya, "Almost localization and almost reducibility", preprint, 2008. To appear in *J. Eur. Math. Soc.* arXiv 0805.1761

[Avila and Krikorian 2006] A. Avila and R. Krikorian, "Reducibility or nonuniform hyperbolicity for quasiperiodic Schrödinger cocycles", *Ann. of Math.* (2) **164**:3 (2006), 911–940. MR 2008h:81044

[Avila et al. ≥ 2010] A. Avila, B. Fayad, and R. Krikorian, "A KAM scheme for SL(2, $\mathbb{R}$) cocycles with Liouvillean frequencies", in preparation.

[Bjerklöv 2006] K. Bjerklöv, "Explicit examples of arbitrarily large analytic ergodic potentials with zero Lyapunov exponent", *Geom. Funct. Anal.* **16**:6 (2006), 1183–1200. MR 2008b:47069 Zbl 1113.82028

[Bourgain 2002a] J. Bourgain, "On the spectrum of lattice Schrödinger operators with deterministic potential", *J. Anal. Math.* **87** (2002), 37–75. MR 2003m:47062

[Bourgain 2002b] J. Bourgain, "On the spectrum of lattice Schrödinger operators with deterministic potential, II", *J. Anal. Math.* **88** (2002), 221–254. MR 2004e:47046

[Bourgain and Jitomirskaya 2002] J. Bourgain and S. Jitomirskaya, "Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential", *J. Statist. Phys.* **108**:5-6 (2002), 1203–1218. MR 2004c:47073 Zbl 1039.81019

[Damanik 2007] D. Damanik, “Lyapunov exponents and spectral analysis of ergodic Schrödinger operators: a survey of Kotani theory and its applications”, pp. 539–563 in *Spectral theory and mathematical physics*, edited by F. Gesztesy et al., Proc. Sympos. Pure Math. **76**, Amer. Math. Soc., Providence, RI, 2007. MR 2008c:82040

[Deift and Simon 1983] P. Deift and B. Simon, “Almost periodic Schrödinger operators, III: The absolutely continuous spectrum in one dimension”, *Comm. Math. Phys.* **90**:3 (1983), 389–411. MR 85i:34009b

[Dombrowski 1978] J. Dombrowski, “Quasitriangular matrices”, *Proc. Amer. Math. Soc.* **69**:1 (1978), 95–96. MR 57 #7232 Zbl 0379.47013

[Erdös and Turán 1940] P. Erdös and P. Turán, “On interpolation, III: Interpolatory theory of polynomials”, *Ann. of Math.* (2) **41** (1940), 510–553. MR 1,333e

[Fedotov and Klopp 2005] A. Fedotov and F. Klopp, “Strong resonant tunneling, level repulsion and spectral type for one-dimensional adiabatic quasi-periodic Schrödinger operators”, *Ann. Sci. École Norm. Sup.* (4) **38**:6 (2005), 889–950. MR 2008j:81037

[Fedotov and Klopp 2006] A. Fedotov and F. Klopp, *Weakly resonant tunneling interactions for adiabatic quasi-periodic Schrödinger operators*, Mém. Soc. Math. Fr. (N.S.) **104**, 2006. MR 2007f:47039

[Freud 1971] G. Freud, *Orthogonal Polynomials*, New York, 1971.

[Jitomirskaya 2007] S. Jitomirskaya, “Ergodic Schrödinger operators (on one foot)”, pp. 613–647 in *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon’s 60th birthday*, edited by F. Gesztesy et al., Proc. Sympos. Pure Math. **76**, Amer. Math. Soc., Providence, RI, 2007. MR 2008g:82055

[Jitomirskaya and Last 1999] S. Jitomirskaya and Y. Last, “Power-law subordinacy and singular spectra, I. Half-line operators”, *Acta Math.* **183**:2 (1999), 171–189. MR 2001a:47033

[Johnson and Moser 1982] R. Johnson and J. Moser, “The rotation number for almost periodic potentials”, *Comm. Math. Phys.* **84**:3 (1982), 403–438. MR 83h:34018 Zbl 0497.35026

[Killip et al. 2003] R. Killip, A. Kiselev, and Y. Last, “Dynamical upper bounds on wavepacket spreading”, *Amer. J. Math.* **125**:5 (2003), 1165–1198. MR 2005a:81255 Zbl 1053.81020

[Kotani 1997] S. Kotani, “Generalized Floquet theory for stationary Schrödinger operators in one dimension”, *Chaos Solitons Fractals* **8**:11 (1997), 1817–1854. MR 98m:34173

[Kuijlaars and Vanlessen 2002] A. B. J. Kuijlaars and M. Vanlessen, “Universality for eigenvalue correlations from the modified Jacobi unitary ensemble”, *Int. Math. Res. Not.* **2002** (2002), 1575–1600. MR 2003g:30043 Zbl 1122.30303

[Last and Simon 2008] Y. Last and B. Simon, “Fine structure of the zeros of orthogonal polynomials, IV. A priori bounds and clock behavior”, *Comm. Pure Appl. Math.* **61**:4 (2008), 486–538. MR 2009d:42070

[Levin and Lubinsky 2008] E. Levin and D. S. Lubinsky, “Applications of universality limits to zeros and reproducing kernels of orthogonal polynomials”, *J. Approx. Theory* **150**:1 (2008), 69–95. MR 2008k:42083 Zbl 1138.33006

[Lubinsky 2008a] D. S. Lubinsky, “A new approach to universality limits at the edge of the spectrum”, pp. 281–290 in *Integrable systems and random matrices*, edited by J. Baik et al., Contemp. Math. **458**, Amer. Math. Soc., Providence, RI, 2008. MR 2010a:42097 Zbl 1147.15306

[Lubinsky 2008b] D. S. Lubinsky, “Universality limits in the bulk for arbitrary measures on compact sets”, *J. Anal. Math.* **106** (2008), 373–394. MR 2009j:30085 Zbl 1156.42005

[Lubinsky 2009] D. S. Lubinsky, “A new approach to universality limits involving orthogonal polynomials”, *Ann. of Math.* (2) **170**:2 (2009), 915–939. MR 2552113 Zbl 1176.42022

[Lund and Bowers 1992] J. Lund and K. L. Bowers, *Sinc methods for quadrature and differential equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. MR 93i:65004 Zbl 0753.65081

[Markoff 1884] A. Markoff, “Démonstration de certaines inégalités de M. Tchébychef”, *Math. Ann.* **24**:2 (1884), 172–180. MR 1510283

[Maslov et al. 1993] V. P. Maslov, S. A. Molchanov, and A. Y. Gordon, “Behavior of generalized eigenfunctions at infinity and the Schrödinger conjecture”, *Russian J. Math. Phys.* **1**:1 (1993), 71–104. MR 95a:81059

[Máté et al. 1991] A. Máté, P. Nevai, and V. Totik, “Szegő’s extremum problem on the unit circle”, *Ann. of Math.* (2) **134**:2 (1991), 433–453. MR 92i:42014

[Rudin 1987] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1987. MR 88k:00002 Zbl 0925.00005

[Simon 1983] B. Simon, "Kotani theory for one-dimensional stochastic Jacobi matrices", *Comm. Math. Phys.* **89**:2 (1983), 227–234. MR 85d:60122 Zbl 0534.60057

[Simon 2005] B. Simon, "Fine structure of the zeros of orthogonal polynomials, II: OPUC with competing exponential decay", *J. Approx. Theory* **135**:1 (2005), 125–139. MR 2006g:42048

[Simon 2006a] B. Simon, "Fine structure of the zeros of orthogonal polynomials, I. A tale of two pictures", *Electron. Trans. Numer. Anal.* **25** (2006), 328–368. MR 2007k:42075

[Simon 2006b] B. Simon, "Fine structure of the zeros of orthogonal polynomials, III. Periodic recursion coefficients", *Comm. Pure Appl. Math.* **59**:7 (2006), 1042–1062. MR 2006k:42052

[Simon 2007] B. Simon, "Equilibrium measures and capacities in spectral theory", *Inverse Probl. Imaging* **1**:4 (2007), 713–772. MR 2008k:31003 Zbl 1149.31004

[Simon 2008a] B. Simon, "The Christoffel–Darboux kernel", pp. 295–335 in *Perspectives in partial differential equations, harmonic analysis and applications*, edited by D. Mitrea and D. Mitrea, Proc. Sympos. Pure Math. **79**, Amer. Math. Soc., Providence, RI, 2008. MR 2500498 Zbl 1159.42020

[Simon 2008b] B. Simon, "Two extensions of Lubinsky's universality theorem", *J. Anal. Math.* **105** (2008), 345–362. MR 2010c:42054 Zbl 1168.42304

[Simon 2009] B. Simon, "Weak convergence of CD kernels and applications", *Duke Math. J.* **146**:2 (2009), 305–330. MR 2009k:33027 Zbl 1158.33003

[Simon 2010] B. Simon, *Szegő's Theorem and its descendants: Spectral theory for L^2 perturbations of orthogonal polynomials*, Princeton Univ. Press, Princeton, 2010.

[Stahl and Totik 1992] H. Stahl and V. Totik, *General orthogonal polynomials*, Encyclopedia of Mathematics and its Applications **43**, Cambridge University Press, Cambridge, 1992. MR 93d:42029 Zbl 0791.33009

[Szegő 1939] G. Szegő, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ. **23**, American Mathematical Society, Providence, R.I., 1939.

[Titchmarsh 1932] E. C. Titchmarsh, *The Theory of Functions*, Oxford University Press, 1932.

[Totik 2000] V. Totik, "Asymptotics for Christoffel functions for general measures on the real line", *J. Anal. Math.* **81** (2000), 283–303. MR 2001j:42021 Zbl 0966.42017

[Totik 2001] V. Totik, "Polynomial inverse images and polynomial inequalities", *Acta Math.* **187**:1 (2001), 139–160. MR 2002h:41017 Zbl 0997.41005

[Totik $\geq$ 2010] V. Totik, "Universality and fine zero spacing on general sets", in preparation.

[Van Assche 1986] W. Van Assche, "Invariant zero behaviour for orthogonal polynomials on compact sets of the real line", *Bull. Soc. Math. Belg. Sér. B* **38**:1 (1986), 1–13. MR 88e:42049 Zbl 0622.33006

[Widom 1967] H. Widom, "Polynomials associated with measures in the complex plane", *J. Math. Mech.* **16** (1967), 997–1013. MR 35 #346 Zbl 0182.09201

Received 20 Oct 2009. Accepted 19 Nov 2009.

ARTUR AVILA: artur@math.sunysb.edu
CNRS UMR 7599, Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, Boîte Courrier 188, 75252 Paris Cedex 05, France

and

Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, RJ, Brazil
http://w3.impa.br/~avila/

YORAM LAST: ylast@math.huji.ac.il
Institute of Mathematics, The Hebrew University, 91904 Jerusalem, Israel

BARRY SIMON: bsimon@caltech.edu
Department of Mathematics, California Institute of Technology, MC 253-37, Pasadena, CA 91125, United States
http://www.math.caltech.edu/people/simon.html

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at pjm.math.berkeley.edu/apde.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in APDE are usually in English, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.org with details about how your graphics were generated.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ANALYSIS & PDE

Volume 3 No. 1 2010