Some harmonic analysis questions suggested by Anderson-Bernoulli models

Carol Shubin and Ramin Vakilian
Department of Mathematics
California State University
Northridge, Ca 91330

Thomas Wolff
Department of Mathematics
253-37 Caltech
Pasadena, Ca 91125

The purpose of this paper is to prove some new variants on the harmonic analyst's uncertainty principle, i.e. interplay between the support of a function and its Fourier transform, and to apply them to some questions in spectral theory. These results were suggested by work on the one dimensional Anderson model due to Campanino-Klein and others.

The paper consists of two parts, sections 2-3 and 4-5. Both parts depend on some preliminary lemmas about characteristic functions of probability distributions, which are proved in section 1.

The results in sections 2-3 are related to the approach to the Anderson model originating in [2] and [17], based on analyzing the formulas arising from the supersymmetric replica method, and they lead to improvements of the estimates in these papers. In particular, this makes the method from [2] and [17] applicable also to questions about Bernoulli type models such as Holder continuity of the density of states.

We now describe these results from the analysis point of view. Let

$$\rho(x) = \min(1, \frac{1}{|x|}) \tag{1}$$

and make the following definition: a set $E \subset \mathbb{R}^n$ is <u>ϵ-thin</u> if

$$|E \cap D(x, \rho(x))| \leq \epsilon |D(x, \rho(x))|$$

for all $x \in \mathbb{R}^n$, where $D(x, r)$ is the disc centered at x with radius r and $|\cdot|$ is Lebesgue measure. We also let E^c be the complement of the set E.

<u>Theorem 2.1</u> There are $\epsilon > 0$ and $C < \infty$ such that if E and F are ϵ-thin sets in $\mathbb{R}^n$ then for any $f \in L^2$

$$\|f\|_2 \leq C(\|f\|_{L^2(E^c)} + \|\hat{f}\|_{L^2(F^c)}) \tag{2}$$

Here $\hat{f}(\xi) \stackrel{def}{=} \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}dx$ is the Fourier transform of f. Theorem 2.1 is evidently a version of the uncertainty principle - the fact that f and $\hat{f}$ cannot both be concentrated on small sets. There are numerous related results in the literature, e.g. Logvinenko and Sereda [20] gave sharp conditions on F under which (2) holds if $E = D(0,1)$, and Amrein and Berthier [1] showed that (2) holds if both E and F have finite volume. Further results may be found in the survey article [9] and the book [15]. The form of Theorem 2.1 is dictated by the application we want to make. However, Theorem 2.1 is also sharp in a certain sense - see the remark after the proof.

Let G be a function on $\mathbb{R}^n$ with $\|G\|_\infty = 1$ and define an operator $T_G : L^2 \to L^2$ via

$$T_G f = \widehat{Gf}$$

Clearly $\|T_G\| = 1$ and in practice, one sometimes needs to bound the spectral radius of T_G away from 1. We will use Theorem 2.1 to prove the following:

Theorem 2.3 Suppose that μ and ν are probability measures on the line, neither of which is a unit point mass and that Q_1, Q_2 are nondegenerate quadratic forms[1] in $\mathbb{R}^n$. Let G and H be functions on $\mathbb{R}^n$ such that

$$|G(x)| \le |\hat{\mu}(Q_1(x))|$$

$$|H(x)| \le |\hat{\nu}(Q_2(x))|$$

Then

$$\|T_H T_G\|_{L^2\to L^2} \le \rho$$

where $\rho < 1$ depends only on μ, ν, Q_1 and Q_2. Indeed, we can take $\rho = 1 - C^{-1}\lambda^2\gamma$ where $C > 0$ depends only on Q_1 and Q_2; here γ and λ are any numbers in $(0,1)$ such that

$$\mu \times \mu(\{(x,y) \in \mathbb{R}\times\mathbb{R} : |x-y| \ge \lambda\}) \ge \gamma \tag{3}$$

$$\nu \times \nu(\{(x,y) \in \mathbb{R}\times\mathbb{R} : |x-y| \ge \lambda\}) \ge \gamma$$

Of course, any measure other than a unit point mass will satisfy (3) for some $\lambda \in (0,1)$ and $\gamma \in (0,1)$.

We prove Theorems 2.1 and 2.3 in section 2 and in section 3 we apply Theorem 2.3 to the Anderson model. We give an alternate proof of Le Page's theorem on Holder continuity of the density of states using ideas from [2] and extend the proof of localization in [17] to the case of Bernoulli models. See Propositions 3.3 and 3.6 below.

In the second part of the paper (sections 4 and 5) we prove quantitative results on nonexistence of almost invariant vectors. First suppose that ν is a probability measure on $\mathbb{R}^n$ satisfying the condition analogous to (3):

$$\nu \times \nu(\{(x,y) \in \mathbb{R}^n\times\mathbb{R}^n : |x-y| \ge \lambda\}) \ge \gamma \tag{4}$$

[1] By this we mean that $Q(x) = < Ax, x >$ with A an invertible real symmetric matrix.

Let $\mathcal{I}$ be the unitary operator on $L^2(\mathbb{R}^n)$ associated with reflection in the unit sphere, i.e.

$$\mathcal{I}f(x) = |x|^{-n} f(x^*) \tag{5}$$

where $x^* = \frac{x}{|x|^2}$. The following result is a special case of Theorem 4.1 below.

Theorem 4.1.0 Let $Tf = \nu * \mathcal{I}f$. Then $\|T^2\|_{L^2 \to L^2} \leq 1 - C^{-1}\lambda^2$ where C depends on n and γ.

In addition to Theorem 4.1 we prove a similar result for general representations of $SL(2, \mathbb{R})$, Theorem 4.5. We then use this (via the equation for Furstenburg's invariant measure) to refine some of the known results on the Anderson model, e.g. we prove a quantitative result for the Liapunov exponent (Corollary 4.7) and a refinement of LePage's theorem (Proposition 4.8) asserting that inside the spectrum of the Laplacian the bounds are independent of the disorder as the disorder goes to zero.

In section 5 we prove an analogous result (Theorem 5.1) for the action of suitable symplectic matrices on the lagrangian grassmannian instead of projective space. Analytically this means that we prove an analogue of Theorem 4.1.0 on the space of symmetric matrices, with inversion of matrices replacing the euclidean inversion $x \to x^*$.

We now explain briefly why the results in the two parts of the paper are related. Namely, the operators $f \to \mu * \mathcal{I}f$ and the operators of type $T_{\hat{\mu} \circ Q}$ which arise from the supersymmetric replica method are related by the metaplectic representation. For example, if one takes $n = 1$ in Theorem 4.1.0 and makes some minor changes in the definitions (defines $\mathcal{I}f(x)$ to be $x^{-1}f(\frac{-1}{x})$ instead of $|x|^{-1}f(\frac{1}{x})$) then the resulting statement is an immediate corollary of Theorem 2.3, since the operator $f \to \mu * \mathcal{I}f$ has a realization where it is of the form $T_{\hat{\mu} \circ Q}$. See e.g. [12], or the proof of Theorem 4.5 for the discrete series given below. This type of argument is also implicit in papers on the Anderson model such as [2] and [18], where similar formulas are used without the group theoretic framework.

Added 10/18/97: in an appendix, we also give a different (perhaps more straightforward) approach to the results in section 4 about almost invariant vectors and prove a more general statement, Theorem A.1.

Acknowledgments The starting point for this paper was a seminar at California State University, Northridge on references [2],[16],[17] during the summer and fall of 1995. The third author is grateful to the CSUN mathematics department and Professors David Klein and William Watkins for office space. We thank Abel Klein for pointing out reference [3] (see section 4) . The first and third authors were partly supported by NSF.

1. Some lemmas on characteristic functions

The purpose of this section is to obtain a certain thinness property of the set where the characteristic function of a probability distribution is close to its maximal possible value 1, which will be used in all the subsequent arguments.

Lemma 1.1 Let μ be a probability measure on $\mathbb{R}^n$ which is not a single Dirac mass and let $E_\delta = \{\xi \in \mathbb{R}^n : |\hat{\mu}(\xi)| > 1-\delta\}$. Then for every $\epsilon > 0$ there is $\delta > 0$ such that $\sup_{a\in\mathbb{R}^n} |D(a,1)\cap E_\delta| < \epsilon$. In fact, if (4) holds (with $\lambda \in (0,1)$) we can take $\delta = C^{-1}\gamma\lambda^2\epsilon^2$ where C depends on n only.

Proof We use the following fact about the cosine function, which holds (uniformly over $k \in \mathbb{R}^n$ and α) since the image of the set in question under the map $\xi \to \xi \cdot \frac{k}{|k|}$ is a union of intervals of length $\approx |k|^{-1}\sqrt{\alpha}$ centered at integer multiples of $|k|^{-1}$.

$$\sup_{a\in\mathbb{R}^n} |\{\xi \in D(a,1) : \cos(2\pi k\cdot\xi) > 1-\alpha\}| \le C\alpha^{1/2}\max(1,|k|^{-1}) \tag{6}$$

Fix a and let $E_\delta^a = E_\delta \cap D(a,1)$. We clearly have

$$(1-2\delta)|E_\delta^a| \le \int_{E_\delta^a} |\hat{\mu}(\xi)|^2 d\xi \tag{7}$$

On the other hand

$$\int_{E_\delta^a} |\hat{\mu}(\xi)|^2 d\xi = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n}\int_{E_\delta^a} \cos(2\pi\xi\cdot(x-y))d\xi d\mu(x)d\mu(y)$$

Let α satisfy $C\alpha^{\frac{1}{2}} = \frac{1}{2}|E_\delta^a|\lambda$, with C as in (6). Then

$$\begin{aligned}\int_{E_\delta^a} \cos 2\pi\xi\cdot(x-y)d\xi &\le |E_\delta^a \cap \{\xi : \cos(2\pi\xi\cdot(x-y)) > 1-\alpha\}| \\ &\quad +(1-\alpha)|E_\delta^a \cap \{\xi : \cos(2\pi\xi\cdot(x-y)) \le 1-\alpha\}| \end{aligned} \tag{8}$$

If $|x-y| \ge \lambda$, then (6) implies

$$\begin{aligned}|E_\delta^a \cap \{\xi : \cos(2\pi\xi\cdot(x-y)) > 1-\alpha\}| &\le \frac{1}{2}|E_\delta^a|\lambda\max(1,|x-y|^{-1}) \\ &\le \frac{1}{2}|E_\delta^a|\end{aligned}$$

since $\lambda \le 1$. By (8),

$$\int_{E_\delta^a} \cos(2\pi\xi\cdot(x-y))d\xi \le (1-\frac{\alpha}{2})|E_\delta^a|$$

when $|x-y| > \lambda$. Hence,

$$\begin{aligned}\int_{E_\delta^a} |\hat{\mu}(\xi)|^2 d\xi &\le (1-\frac{\alpha}{2})|E_\delta^a|\gamma + |E_\delta^a|(1-\gamma) \\ &= (1-\frac{\alpha\gamma}{2})|E_\delta^a|\end{aligned}$$

Combining this with (7) we conclude that $\delta \geq \frac{\alpha\gamma}{4}$. By choice of α this means that $\delta \geq C^{-1}\gamma\lambda^2|E_\delta^a|^2$. This is equivalent to the lemma. $\square$

The next lemma and corollary are variants on Lemma 1.1 which will only be needed in section 5.

Lemma 1.2 Let μ be a probability measure on $\mathbb{R}^n$ and assume that for certain constants γ and A with $\gamma > 0$, and for a certain $\lambda \leq 1$,

$$\mu \times \mu(\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : |x-y| \geq A\lambda\}) \leq \frac{\gamma}{2} \tag{9}$$

If e is a unit vector in $\mathbb{R}^n$ then

$$\mu \times \mu(\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : |(x-y)\cdot e| \geq \lambda\}) \geq \gamma \tag{10}$$

Then, for a suitable constant C depending on γ, A and n the set

$$E \stackrel{def}{=} \{\xi \in \mathbb{R}^n : |\hat{\mu}(\xi)| \geq 1 - C^{-1}\epsilon^2\lambda^2\}$$

satisfies the estimate

$$|\{t \in \mathbb{R} : t = x \cdot e \text{ for some } x \in E \cap D(a,1)\}| < \epsilon$$

for all $a \in \mathbb{R}^n$ and all unit vectors $e \in \mathbb{R}^n$.

Remarks (1) Condition (10) is the natural analogue here for the assumption in Lemma 1.1. However, (10) by itself does not imply the conclusion of the lemma. This may be seen (say when $n = 2$) by taking μ to be the direct product of two copies of the measure $\frac{1}{2}(\delta_{-N} + \delta_N)$, e with irrational slope and using Dirichlet's theorem on approximation of irrationals by rationals. The conclusion will not hold uniformly in N as $N \to \infty$.

(2) Any fixed probability measure on $\mathbb{R}^n$ which is not supported on a hyperplane will satisfy the hypotheses of Lemma 1.2 for some choice of γ, any large A and some λ. In fact, the hypotheses hold uniformly over any set of probability measures which is weak$*$ compact and contains no measure supported on a hyperplane. This may be proved by a suitable compactness argument. Furthemore, the assumptions are dilation invariant in the following sense. If μ is a measure and $t \in \mathbb{R}^+$ then we let μ_t be the t-dilation of μ, i.e.

$$\mu_t(E) \stackrel{def}{=} \mu(t^{-1}E) \tag{11}$$

If μ satisfies the hypotheses with $\lambda = \lambda_0$, then μ_t will satisfy the hypotheses with the same values of A and γ and with $\lambda = t\lambda_0$.

Proof of Lemma 1.2 Define

$$\Delta = \{x_1, \ldots, x_n, y_1, \ldots, y_n \in (\mathbb{R}^n)^{2n} : |y_j - x_j| \leq A\lambda \ \forall j \text{ and } |\det(\{y_j - x_j\}_{j=1}^n)| \geq \lambda^n\}$$

Claim 1

$$\overbrace{\mu \times \ldots \times \mu}^{2n \text{ times}}(\Delta) \geq C^{-1}$$

Proof We consider the set of all $2n$-tuples x_j, y_j ($j \leq n$) which can be obtained by the following recursive process: x_1 and y_1 are arbitrary points of $\mathbb{R}^n$ such that $\lambda \leq |y_1 - x_1| \leq A\lambda$. If $2 \leq k \leq n$ and if x_j and y_j have been chosen for $j < k$, then denote the orthogonal projection of $\mathbb{R}^n$ on the orthogonal complement of $\mathrm{sp}(\{x_j - y_j\}_{j<k})$ by P, and let x_k and y_k be any points of $\mathbb{R}^n$ such that $|P(x_k - y_k)| \geq \lambda$ and $|x_k - y_k| \leq A\lambda$. The assumptions imply that the $\mu \times \mu$ measure of the allowable pairs (x_k, y_k) is bounded below by $\frac{\gamma}{2}$ at each stage. Accordingly, the $\mu \times \ldots \times \mu$-measure of the set of $2n$- tuples obtained this way is bounded below by $(\frac{\gamma}{2})^n$. On the other hand, some simple linear algebra implies that $|\det(y_j - x_j)| \geq \lambda^n$ for any such $2n$-tuple, so Claim 1 is proved.

Claim 2 below will play the same role as did (6) in the proof of Lemma 1.1.

Claim 2 Let $a \in \mathbb{R}^n$ with radius 1, assume $(x_1, \cdots, x_n, y_1, \cdots, y_n) \in \Delta$ and define

$$\Sigma = \{\xi \in D(a,1) : \Pi_{j=1}^n \cos(2\pi(x_j - y_j) \cdot \xi) \geq 1 - \rho^2\lambda^2\}$$

Then Σ is contained in the union of C discs of radius ρ. In particular, if e is any unit vector in $\mathbb{R}^n$ then

$$|\{t \in \mathbb{R} : t = x \cdot e \text{ for some } x \in \Sigma\}| \leq C\rho$$

where C depends on n and γ.

Proof First consider the case where $x_j - y_j = \lambda e_j$, where e_j is the jth standard basis vector. Then the set in question is contained in a $C\rho$-neighborhood of $(2\lambda)^{-1}\mathbb{Z}^n$, so the claim is obvious (recall that $\lambda \leq 1$). The general case then follows since by definition of Δ, there is a linear map $T \in GL(n)$ with $T(\lambda e_j) = x_j - y_j$ and such that $\|T\|$ and $\|T^{-1}\|$ are bounded by constants.

In proving the lemma we may of course assume that $e = e_n$. We will denote variables in $\mathbb{R}^n$ by $\xi = \overline{\xi} + te$ with $\overline{\xi} \in \mathbb{R}^{n-1} \times \{0\}$ and $t \in \mathbb{R}$. We fix $a \in \mathbb{R}^n$ and, for $t \in \mathbb{R}$, we let $V(t) = \sup(|\hat{\mu}(\xi)| : \xi \in D(a,1), \xi_n = t)$. Fix a number δ and consider the set $Y_\delta \overset{def}{=} \{t \in \mathbb{R} : V(t) \geq 1 - \delta\}$ and the quantity

$$\int_{Y_\delta} V(t)^{2n} dt \tag{12}$$

which is evidently $\geq (1-\delta)^{2n}|Y_\delta|$. On the other hand, (12) is

$$\begin{aligned}
&\leq \int_{Y_\delta} \sup_{|\overline{\xi}|\leq 1} |\hat{\mu}(a+te_n+\overline{\xi})|^{2n} dt \\
&= \int_{Y_\delta} \left(\sup_{|\overline{\xi}|\leq 1} \int_{\mathbb{R}^{2n}} \Pi_{j=1}^n \cos(2\pi(a+te_n+\overline{\xi})\cdot(x_j-y_j)) d\mu(x_1)d\mu(y_1)\dots d\mu(x_n)d\mu(y_n) \right) dt \\
&\leq \int_{Y_\delta} \int_{\mathbb{R}^{2n}} \sup_{|\overline{\xi}|\leq 1} \Pi_{j=1}^n \cos(2\pi(a+te_n+\overline{\xi})\cdot(x_j-y_j)) d\mu(x_1)d\mu(y_1)\dots d\mu(x_n)d\mu(y_n) dt \quad (13)
\end{aligned}$$

We will denote the measure

$$\overbrace{\mu\times\cdots\times\mu}^{2n \text{ times}}$$

by m. We subdivide the integral with respect to m in (13) into the contributions from Δ and Δ^c. Of course

$$\int_{Y_\delta} \int_{\Delta^c} \sup_{|\overline{\xi}|\leq 1} \Pi_{j=1}^n \cos(2\pi(a+te_n+\overline{\xi})\cdot(x_j-y_j)) d\mu(x_1)d\mu(y_1)\dots d\mu(x_n)d\mu(y_n) dt \leq |Y_\delta| m(\Delta^c)$$

On the other hand, if $(x_1,\dots,x_n,y_1,\dots,y_n)\in\Delta$ then for any ϵ, Claim 2 implies

$$\sup_{|\overline{\xi}|\leq 1} \Pi_{j=1}^n \cos(2\pi(a+te_n+\overline{\xi})\cdot(x_j-y_j)) \leq 1 - C^{-1}\epsilon^2\lambda^2$$

for all $t\in I$ except a set of measure ϵ. Accordingly, if $|Y_\delta|>\epsilon$ then

$$\begin{aligned}
&\int_{Y_\delta} \int_{\mathbb{R}^{2n}} \sup_{|\overline{\xi}|\leq 1} \Pi_j \cos((k+te_n+\overline{\xi})\cdot(x_j-y_j)) d\mu(x_1)d\mu(y_1)\dots d\mu(x_n)d\mu(y_n) dt \\
&\qquad\qquad \leq m(\Delta)[(|Y_\delta|-\epsilon)(1-C^{-1}\epsilon^2\lambda^2)+\epsilon]
\end{aligned}$$

So $|Y_\delta|>\epsilon$ implies

$$\begin{aligned}
(1-\delta)^{2n}|Y_\delta| &\leq m(\Delta^c)|Y_\delta| + m(\Delta)[(|Y_\delta|-\epsilon)(1-C^{-1}\epsilon^2\lambda^2)+\epsilon] \\
&= |Y_\delta| - C^{-1}m(\Delta)(|Y_\delta|-\epsilon)\epsilon^2\lambda^2
\end{aligned}$$

If $|Y_\delta|\geq 2\epsilon$ then it follows that $(1-\delta)^{2n}\leq 1-C^{-1}m(\Delta)\epsilon^2\lambda^2$. Since $m(\Delta)\geq C^{-1}$ by Claim 1 this means that $\delta\gtrsim\epsilon^2\lambda^2$, finishing the proof of the lemma. ☐

The following corollary is what will actually be used below. If μ is a measure in $\mathbb{R}^n$ then define

$$\Sigma(\mu) = \text{ span}(\{x-y : x,y\in \text{ supp}(\mu)\}) \quad (14)$$

In other words, we take the smallest affine plane containing the support of μ and translate it to the origin.

Corollary 1.3 For any given probability measure μ on $\mathbb{R}^n$ there is a constant $C = C_\mu$ such that the following holds for all $\lambda \in [0,1]$, all $\epsilon > 0$, all unit vectors $e \in \Sigma(\mu)$ and all $a \in \mathbb{R}^n$. Define

$$E_\lambda = \{\xi \in \mathbb{R}^n : |\hat{\mu}_\lambda(\xi)| \geq 1 - C^{-1}\epsilon^2\lambda^2\}$$

Then

$$|\{t \in \mathbb{R} : t = x \cdot e \text{ for some } x \in E_\lambda \cap D(a,1)\}| < \epsilon$$

The constant C may be taken uniform over any compact set of probability measures with a given $\Sigma(\mu)$.

Proof If $\Sigma(\mu) = \mathbb{R}^n$ this is the content of Remark (2) after the statement of Lemma 1.2. The general case follows since $|\hat{\mu}|$ is constant in directions perpendicular to $\Sigma(\mu)$. $\square$

2. Proof of Theorems 2.1 and 2.3

We fix a radial, real-valued Schwartz function $\phi : \mathbb{R}^n \to \mathbb{R}$ with $0 \leq \hat{\phi} \leq 1$, $\mathrm{supp}\hat{\phi} \subset D(0,2)$ and $\hat{\phi} = 1$ on $D(0,1)$. For $j \in \mathbb{Z}^+ \cup \{0\}$ we let $\phi_j(x) = 2^{jn}\phi(2^j x)$. Thus $\|\phi_j\|_1 = \|\phi\|_1$, and $\hat{\phi}_j(\xi) = \hat{\phi}(2^{-j}\xi)$ so that $\mathrm{supp}\hat{\phi}_j \subset D(0,2^{j+1})$, $\hat{\phi}_j = 1$ on $D(0,2^j)$.

Let $\psi_0 = \hat{\phi}$, and when $j \in \mathbb{Z}^+$ let $\psi_j = \hat{\phi}_j - \hat{\phi}_{j-1}$. Thus $\mathrm{supp}\psi_j \subset D(0,2^{j+1}) \backslash D(0,2^{j-1})$ when $j \geq 1$, and $\sum_{j=0}^{\infty} \psi_j = 1$. Define now

$$S_N f = \sum_{j=0}^{N} \psi_j \cdot (\phi_j * f)$$

$$T_N f = \sum_{j=0}^{N} \psi_j \cdot (f - \phi_j * f)$$

Since the $\{\phi_j\}$ form an approximate identity and $\sum \psi_j = 1$ with no more than three terms being simultaneously nonzero, we may conclude that if $f \in \mathcal{S}$ then $S_N f$ and $T_N f$ converge in the topology of $\mathcal{S}$. We denote the limit operators by S and T. Clearly $S + T$ is the identity operator. One can think of Sf and Tf as f microlocalized to the regions $|\xi| \leq |x|$ and $|\xi| \geq |x|$ respectively.

The proof of Theorem 2.1 will be based on making appropriate L^2 estimates for the operators S and T. The estimates will follow by applying Schur's test to the integral kernels of the operators. The following lemma contains the necessary calculations.

Lemma 2.2 Define

$$A_N(x) = \sum_{j=0}^{N} \psi_j(x)\phi_j(x-y)$$

$$B_N(\xi,\eta) \;=\; \sum_{j=0}^{N} \hat{\psi}_j(\xi-\eta)(1-\hat{\phi}_j(\eta))$$

Then, for a suitable constant C which is independent of N,

(i) $\int |A_N(x,y)|dy \leq C$ for all x.
(ii) $\int |A_N(x,y)|dx \leq C$ for all y.
(iii) $\int |B_N(\xi,\eta)|d\eta \leq C$ for all ξ.
(iv) $\int |B_N(\xi,\eta)d\xi \leq C$ for all η.

Furthermore, if E and F are ϵ-thin then

(v) $\int_E |A_N(x,y)|dy \leq C\epsilon$ for all x.
(vi) $\int_F |B_N(\xi,\eta)|d\xi \leq C\epsilon$ for all η.

We will first complete the proof of Theorem 2.1 assuming Lemma 2..2 and will then prove Lemma 2.2.

Proof of Theorem 2.1 Note that

$$S_N f(x) = \int A_N(x,y)f(y)dy$$

and

$$\widehat{T_N f}(\xi) = \int B_N(\xi,\eta)\hat{f}(\eta)d\eta$$

Consequently, by (i)-(iv) of Lemma 2.2 and Schur's test, the operators S and T extend to bounded operators on L^2 satisfying $S+T=$identity. Furthermore, if we let χ_E denote the indicator function of the set E, then using (ii) and (v) (respectively (iii) and (vi)) and Schur's test, we have

$$\|S(\chi_E f)\|_2 \leq C\epsilon^{\frac{1}{2}}\|f\|_2 \tag{15}$$

$$\|\chi_F \widehat{Tf}\|_2 \leq C\epsilon^{\frac{1}{2}}\|f\|_2 \tag{16}$$

Suppose now that f is given, with $\|f\|_2 = 1$. Then

$$\hat{f} = \widehat{S(\chi_{E^c} f)} + \widehat{S(\chi_E f)} + \chi_{F^c}\widehat{Tf} + \chi_F \widehat{Tf}$$

and therefore, using (15),(16),

$$\|\hat{f} - \widehat{S(\chi_{E^c} f)} - \chi_{F^c}\widehat{Tf}\|_2 \leq C\epsilon^{\frac{1}{2}}$$

If $\|f\|_{L^2(E^c)} \leq \alpha$, say, then we conclude that

$$\begin{aligned} \|\hat{f}\|_{L^2(F)} &\leq \|\hat{f} - \chi_{F^c}\widehat{Tf}\|_2 \\ &\leq C(\alpha + \epsilon^{\frac{1}{2}}) \\ &\leq \frac{1}{\sqrt{2}} \end{aligned}$$

provided ϵ and α have been chosen small. So $\|\hat{f}\|_{L^2(F^c)} \geq \frac{1}{\sqrt{2}}$ and the proof is complete.

Proof of Lemma 2.2

(i) For fixed x there are at most three values of j for which $\psi_j(x) \neq 0$. Furthermore $\|\psi_j\|_\infty \leq 1$ for any j, and $\|\phi_j\|_1 = \|\phi\|_1$ for any j. We conclude that the integral in (i) is $\leq 3\|\phi\|_1$.

(ii) Fix y and let $\sum_*$ denote the sum over all $j \in \{0, \ldots, N\}$ such that $\mathrm{dist}(y, \mathrm{supp}\ \psi_j) \geq 1$. There are at most three values of j with $\mathrm{dist}(y, \mathrm{supp}\ \psi_j) < 1$, and since $\phi \in \mathcal{S}$ we have $|\phi_j(x-y)| \leq C2^{jn}(1+2^j|x-y|)^{-3n}$, say. Hence

$$\begin{aligned}
\int |A_N(x,y)|dx &\leq 3\|\phi\|_1 + \int \sum_* |\psi_j(x)\phi_j(x-y)|dx \\
&\leq 3\|\phi\|_1 + C\int \sum_* |\psi_j(x)| 2^{jn}(1+2^j|x-y|)^{-3n}dx \\
&\leq 3\|\phi\|_1 + C\sum_* 2^{-2jn}\|\psi_j\|_1 \\
&= 3\|\phi\|_1 + C\sum_* 2^{-jn}\|\psi\|_1 \\
&\leq C
\end{aligned}$$

as claimed.

(iii) and (iv). We rewrite the definition of B_N as follows:

$$\begin{aligned}
B_N(\xi,\eta) &= \sum_{j=0}^{N} \hat{\psi}_j(\xi-\eta)\sum_{i>j}\psi_i(\eta) \\
&= \sum_{i=1}^{\infty}\psi_i(\eta)\sum_{j=0}^{\min(i-1,N)}\hat{\psi}_j(\xi-\eta) \\
&= \sum_{i=1}^{\infty}\psi_i(\eta)\phi_{i_*}(\xi-\eta) \qquad (17)
\end{aligned}$$

where we have set $i_* = \min(i-1, N)$. Note the similarity between (17) and the definition of $A_N(\eta,\xi)$. We may therefore prove (iv) by repeating the proof of (i). For (iii), we further rewrite (17) as

$$\begin{aligned}
B_N(\xi,\eta) &= \sum_{i=1}^{N}\psi_i(\eta)\phi_{i-1}(\xi-\eta) + \sum_{i>N}\psi_i(\eta)\phi_N(\xi-\eta) \\
&= \sum_{i=1}^{N}\psi_i(\eta)\phi_{i-1}(\xi-\eta) + (1-\widehat{\phi_N}(\eta))\phi_N(\xi-\eta)
\end{aligned}$$

We have $\int |(1-\widehat{\phi_N}(\eta))\phi_N(\xi-\eta)|d\eta \le \int |\phi_N(\xi-\eta)|d\eta = \|\phi_N\|_1 = \|\phi\|_1$, and the estimate $\int |\sum_{i=1}^N \psi_i(\eta)\phi_{i-1}(\xi-\eta)|d\eta \le C$ follows by repeating the proof of (ii). This proves (iii).

(v) and (vi). We will only prove (v), since (vi) once again follows by the same argument in view of (17).

Fix x and let j be such that $\psi_j(x) \neq 0$. We claim that

$$\int_E |\phi_j(x-y)|dy \le C\epsilon \tag{18}$$

If we can prove this we are done as in the proof of (i), since there are only three possible values for j and the ψ_j's are uniformly bounded.

To prove the claim, we use the following simple geometric property of the discs $D(x,\rho(x))$: if $t > \rho(x)$, then $D(x,t)$ can be covered by discs of the form $D(x_k,\rho(x_k))$ in such a way that

$$\sum_k |D(x_k,\rho(x_k))| \le C|D(x,t)|$$

This is true since

$$|y-x| \le \rho(x) \Rightarrow C^{-1}\rho(x) \le \rho(y) \le C\rho(x) \tag{19}$$

See for example [29]. The argument is as follows. Property (19) is easily verified. Now choose a maximal set of points $\{x_k\} \subset D(x,t)$ such that $|x_k - x_j| \ge \min(\rho(x_j),\rho(x_k))$ for all j and k. By maximality, the discs $D(x_k,\rho(x_k))$ cover $D(x,t)$. On the other hand, (19) implies that for a suitable constant C_0 the discs $D(x_k, C_0^{-1}\rho(x_k))$ are disjoint and contained in $D(x,C_0t)$,and therefore $\sum_k |D(x_k,\rho(x_k))| \lesssim \sum_k |D(x_k,C_0^{-1}\rho(x_k))| \le |D(x,C_0t)| \lesssim |D(x,t)|$.

It follows that

$$|D(x,t)\cap E| \le \sum_k |D(x_k,\rho(x_k))\cap E| \le \epsilon \sum_k |D(x_k,\rho(x_k))| \le C\epsilon|D(x,t)| \tag{20}$$

for any x and $t \ge \rho(x)$. Next, if $\psi_j(x) \neq 0$ then 2^j is comparable to $\rho(x)^{-1}$. Since $\phi \in \mathcal{S}$, we therefore have $|\phi_j(x-y)| \le C\rho(x)^{-n}(1+\frac{|x-y|}{\rho(x)})^{-2n}$, so we may estimate the integral in (18) by

$$\begin{aligned}\int_E |\phi_j(x-y)|dy &\lesssim \sum_{k\ge 0} 2^{-2nk}\rho(x)^{-n}|D(x,2^k\rho(x))\cap E| \\ &\lesssim \sum_{k\ge 0} 2^{-2nk}\rho(x)^{-n}\epsilon(2^k\rho(x))^n \\ &\lesssim \epsilon\end{aligned}$$

where we used (20). □

Remark Theorem 2.1 is sharp in the sense that the rate function $\rho(x)$ cannot be replaced by one which decays more slowly at ∞. We give the counterexamples in the case $n = 1$. Let ϕ be a fixed C_0^∞ function and consider the functions

$$\Phi_N(x) \stackrel{def}{=} \sum_{j=-N}^{N} \phi(N(x-j))$$

Then

$$\hat{\Phi}_N(\xi) = D_N(\xi) N^{-1} \hat{\phi}(\frac{\xi}{N})$$

where $D_N(\xi) = \frac{\sin(2\pi(N+\frac{1}{2})\xi)}{\sin(\pi\xi)}$ is the Dirichlet kernel. Let $E_N^A = \cup_{j=-AN}^{AN}(j - \frac{A}{N}, j + \frac{A}{N})$ and let F_N^A be the complement of E_N^A. Then it is not hard to see the following: for any $\eta > 0$ there is $A < \infty$ such that for any large N, we have

$$\|\Phi_N\|_{L^2(F_N^A)} + \|\widehat{\Phi_N}\|_{L^2(F_N^A)} < \eta \|\Phi_N\|_2$$

Namely, Φ_N will vanish on F_N^A due to the support property of ϕ, and $|\widehat{\Phi_N}|^2$ will have most of its mass on E_N^A since $|D_N|^2$ is concentrated near integers and $\hat{\phi}(\frac{\xi}{N})$ dies out rapidly when $|\xi|$ is large compared with N.

On the other hand, if ρ is positive and continuous and $|x|\rho(x) \to \infty$ as $|x| \to \infty$, then for any ϵ and A, we will have $|D(x, \rho(x)) \cap E_N^A| < \epsilon |D(x, \rho(x))|$ for all x provided N is sufficiently large. This shows that the rate function in Theorem 2.1 is the optimal one, as claimed.

However, we do not know whether Theorem 2.1 remains true if "There are $\epsilon > 0$ and $C < \infty \ldots$" is replaced by "For all $\epsilon < 1$ there is $C < \infty \ldots$".

Proof of Theorem 2.3 Basically, Theorem 2.3 follows by combining Theorem 2.1 and Lemma 1.1.

Any nondegenerate quadratic form Q has the following property: Q maps each disc $D(x, \rho(x))$ onto an interval I_x with $|I_x| \in [C^{-1}, C]$. Furthermore, if $E \subset I_x$ then

$$\frac{|Q^{-1}E \cap D(x, \rho(x))|}{|D(x, \rho(x))|} \le C|E|^{1/2}$$

This holds essentially because when $|x|$ is large, $|\nabla Q(x)|$ is comparable to $\rho(x)^{-1}$. We leave details to the reader.

We conclude by Lemma 1.1 that for any given ϵ if $\delta = C^{-1}\gamma\lambda^2\epsilon^2$ then the sets

$$\{x \in \mathbb{R}^n : |\hat{\mu}(Q_1(x))| > 1 - \delta\}$$

$$\{\xi \in \mathbb{R}^n : |\hat{\nu}(Q_2(\xi))| > 1 - \delta\}$$

are $\sqrt{\epsilon}-$thin. Here C depends on Q_1 and Q_2 only. Consequently by Theorem 2.1, we can choose $\beta > 0$ and $\eta > 0$ depending on Q_1 and Q_2 so that, with $\delta = \beta\gamma\lambda^2$ and

$$E = \{x : |G(x)| > 1 - \delta\}$$

$$F = \{\xi : |H(\xi)| > 1-\delta\}$$

every function f must satisfy either

$$\|f\|_{L^2(E^c)} \geq \eta \|f\|_2$$

or

$$\|\hat{f}\|_{L^2(F^c)} \geq \eta \|f\|_2$$

To finish the proof of the lemma, fix f and consider two cases:

(i) $\|f\|_{L^2(E^c)} \geq \eta \|f\|_2$.
(ii) $\|f\|_{L^2(E^c)} < \eta \|f\|_2$.

In case (i) we have

$$\begin{aligned} \|Gf\|_2^2 &= \|Gf\|^2_{L^2(E^c)} + \|Gf\|^2_{L^2(E)} \\ &\leq (1-\delta)^2 \|f\|^2_{L^2(E^c)} + \|f\|^2_{L^2(E)} \\ &\leq ((1-\delta)^2\eta^2 + 1 - \eta^2)\|f\|_2^2 \\ &= \|f\|_2^2 - (2\delta - \delta^2)\eta^2 \|f\|_2^2 \end{aligned}$$

so that $\|H\widehat{Gf}\|_2^2 \leq \|Gf\|_2^2 \leq (1-\alpha)\|f\|_2^2$ with $\alpha = (2\delta - \delta^2)\eta^2$.

In case (ii) we have also

$$\|Gf\|_{L^2(E^c)} < \eta \|Gf\|_2$$

since $\sup_{E^c} |G| \leq \inf_E |G|$. Therefore

$$\|\widehat{Gf}\|_{L^2(F^c)} \geq \eta \|\widehat{Gf}\|_2$$

Then the argument for case (i) shows that

$$\|H\widehat{Gf}\|_2^2 \leq (1-\alpha)||\widehat{Gf}\|_2^2$$

with α as above. So $\|H\widehat{Gf}\|_2^2 \leq (1-\alpha)\|f\|_2^2$ and the proof is complete, since $\alpha \approx \gamma\lambda^2$. $\square$

Remarks 1. We will need a slight variant on Theorem 2.3 where the definition of T_G is modified as follows

$$T_G f(x) \stackrel{def}{=} \widehat{Gf}(Ux)$$

with U being a fixed orthogonal map of $\mathbb{R}^n$. This version can be obtained by applying Theorem 2.3 as stated with f and Q_j replaced by $f \circ U$ and $Q_j \circ U$.

2. A generalization of Theorem 2.3 is possible where one of the two measures μ or ν is allowed to be a unit point mass at a point $a \in \mathbb{R}$ provided $a \neq 0$. Namely, if we assume that μ is not a unit point mass and that ν is not a unit point mass at the origin, then $\|T_{\hat{\mu}\circ Q_1} T_{\hat{\nu}\circ Q_2} T_{\hat{\mu}\circ Q_1}\|_{L^2 \to L^2} < 1$. We sketch the proof. Assume that ν is a unit point mass at

a point $a \neq 0$ - this is no loss of generality by Theorem 2.3. Then $\hat{\nu} \circ Q_2$ is an imaginary Gaussian, which implies that the operator $f \to \widehat{T_{\hat{\nu}\circ Q_2}\hat{f}}$ may be rewritten in the following way:

$$\widehat{T_{\hat{\nu}\circ Q_2}\hat{f}} = c(\Gamma_2)(\widehat{(\Gamma_1 f)} \circ L)$$

where Γ_1 and Γ_2 are imaginary Gaussians, L is a linear change of variable and c is a constant with $|c| = |\det L|^{\frac{1}{2}}$. Let $\tilde{L} = (L^{-1})^t$. We will also let $\mathcal{F}f = \hat{f}$ and will use a lot of parentheses in order to clarify the order of operations, e.g. in the preceding formula $c(\Gamma_2)(\widehat{(\Gamma_1 f)} \circ L)$ means c times Γ_2 times the composition of $\widehat{(\Gamma_1 f)}$ with L. Then

$$\begin{aligned} T_{\hat{\mu}\circ Q_1}T_{\hat{\nu}\circ Q_2}T_{\hat{\mu}\circ Q_1}f &= c\mathcal{F}\left((\hat{\mu}\circ Q_1)(\Gamma_2)(\mathcal{F}(\Gamma_1(\hat{\mu}\circ Q_1)f)\circ L)\right) \\ &= \overline{c}^{-1}\mathcal{F}\Big((\hat{\mu}\circ Q_1)(\Gamma_2)(\mathcal{F}((\Gamma_1\circ\tilde{L})(\hat{\mu}\circ Q_1\circ\tilde{L})(f\circ\tilde{L}))\Big) \\ &= \overline{c}^{-1}T_G T_H(f\circ\tilde{L}) \end{aligned}$$

where $G = (\Gamma_2)(\hat{\mu} \circ Q_1)$ and $H = (\Gamma_1 \circ \tilde{L})(\hat{\mu} \circ Q_1 \circ \tilde{L})$. We now apply Theorem 2.3 with both measures equal to μ and with $Q_2 = Q_1 \circ \tilde{L}$. It follows that for an appropriate constant $\rho < 1$,

$$\|T_{\hat{\mu}\circ Q_1}T_{\hat{\nu}\circ Q_2}T_{\hat{\mu}\circ Q_1}f\| \le \rho|c|^{-1}\|f\circ\tilde{L}\| = \rho\|f\|$$

as claimed.

3. Consequences of the supersymmetric formalism

We will first summarize some things to be found for example in [2], with some slight modifications since we are using a different normalization of the Fourier transform and also want to work directly with two dimensional Euclidean Fourier transforms instead of using Hankel transforms as is done there. We refer to [2], [17] for further details.

If f is a (nice enough) function on the half line $\mathbb{R}^+$ then let us define

$$Df = -\frac{1}{\pi}f'$$

We may consider $f(|x|^2)$, a radial function on $\mathbb{R}^2$. Its two dimensional (Euclidean) Fourier transform is another radial function, hence of the form $g(|\xi|^2)$. We denote g by Tf. Furthermore, we denote the operator TD (i.e. D followed by T) by S. Then we have the following fact. (cf. [16] for example)

Lemma 3.1 $DS = T$.

Now consider an interval $\Lambda = \{-l, \ldots, l\} \subset \mathbb{Z}$ and a Schrodinger operator on Λ, i.e. an operator $H = \Delta + V$ where $\Delta u(n) = \frac{1}{2}(u(n+1) + u(n-1))$ and V is multiplication by a function $V(n)$, with Dirichlet boundary conditions. If $\mathrm{im}(z) > 0$, and $m, n \in \Lambda$, then we let $G_\Lambda(m,n,z)$ be the Green's function, i.e., the matrix $G_\Lambda(m,n,z)_{\substack{m\in\Lambda\\ n\in\Lambda}}$ which is inverse to $H - z\cdot$ identity. The basic formula (cf. [2]) is as follows.

Let $\beta_j(r) = e^{-2\pi i(V_j - z)r}$ and also let β_j be the operator on functions defined by multiplication by the function β_j. Then, when $m \leq n$ (and im$z > 0$), $G_\Lambda(m,n,z)$ is equal to

$$2i^{m-n+1}\int ((\Pi_{j=-l}^{m-1} S\beta_j)1)(|x|^2))((\Pi_{j=0}^{n-m-1} T\beta_{n-j})(\Pi_{j=0}^{l-n-1} S\beta_{l-j})1)(|x|^2)\beta_m(|x|^2)dx \quad (21)$$

Here, if O_j are (noncommuting) operators, then we use $\Pi_{j=1}^n O_j$ to denote the operator $O_n O_{n-1}\ldots O_1$, etc. Thus for example the expression $((\Pi_{j=-l}^{m-1} S\beta_j)1)(|x|^2)$ means start with the constant function 1, apply the operator β_{-l}, then S, then β_{-l+1} and so forth and evaluate the resulting expression at $|x|^2$. In the case of the Anderson model with single site distribution ν, taking expectations in the above formula for $G(0,0,z)$ leads [2] to

$$\mathcal{E}(G(0,0,z)) = 2i\int_{\mathbb{R}^2} ((S\Gamma)^l 1(|x|^2))^2 \Gamma(|x|^2)dx \quad (22)$$

where we have let $\Gamma(r) = \hat{\nu}(r)e^{2\pi i z r}$ and have also used Γ to denote the operator of multiplication by Γ. An analogous formula for the expectation of $|G(m,n,z)|^2$ may be obtained in the same way, cf. [17] and (34) below.

Consider functions on $\mathbb{R}^+$ of the form

$$y_\zeta(r) = e^{2\pi i\zeta r},\ \text{where im}\zeta > 0 \quad (23)$$

It is easy to check that

$$Sy_\zeta = y_{\frac{-1}{4\zeta}} \quad (24)$$

In particular, the class of such functions is closed under S. Note that it is also closed under forming products. An important consequence is that $(S\Gamma)^n 1$ in (22), and other similar expressions to be considered below, have L^∞ norm ≤ 1, since they are averages of functions of the form y_ζ with respect to a probability distribution.

Now let $g_l(z) = \mathcal{E}(G(0,0,z))$, where G is the Green's function for the Anderson model on $\{-l,\ldots,l\}$ with single site distribution ν. We assume that

$$\int |x|d\nu(x) < \infty \quad (25)$$

We denote $z = E + i\eta$, $E,\eta \in \mathbb{R}$ and will always assume that $\eta > 0$. We will prove the following:

Proposition 3.3 For any $\epsilon_0 > 0$ there are $\epsilon_1 > 0$ and $C < \infty$, depending on ν, ϵ_0 and a bound for $|E|$, so that if $\eta \leq \frac{1}{2}$ and $l > \epsilon_0 \log\frac{1}{\eta}$ then $|g_l(z)| \leq C\eta^{-(1-\epsilon_1)}$.

An immediate corollary is Holder continuity of the integrated density of states for the Anderson model on $\mathbb{Z}$ with single site distribution satisfying (25). Namely, by letting $l \to \infty$ in Proposition 3.3 we obtain the bound

$$|g(z)| \leq C\eta^{-(1-\epsilon_1)}$$

for all $z = E + i\eta$ with $0 < \eta < \frac{1}{2}$. Since $g(z)$ is the $\overline{z}$-derivative of the harmonic extension of the integrated density of states, this bound is equivalent to Holder continuity (cf. [27], ch. 5). One can also obtain an estimate of the finite volume density of states directly from Proposition 3.3 using that the expected number of states in the interval $(E - \eta, E + \eta)$ is $\leq Cl\eta \ \mathrm{im} g_l(E + i\eta)$.

Remarks The assumption on l in Proposition 3.3 is easily seen to be best possible. This argument is in [26]. Namely, if ν is Bernoulli, then there are only 2^{2l+1} possible choices for H on $\{-l, \dots, l\}$, hence (being generous) at most $(2l + 1)2^{2l+1}$ possible eigenvalues. So g_l is the Borel transform of a measure with $\leq (2l + 1)2^{2l+1}$ mass points and then it follows that $\sup_E \ \mathrm{im} g_l(E + i\eta) \gtrsim ((2l + 1)2^{2l+1}\eta)^{-1}$, which is large compared with $\eta^{-(1-\epsilon)}$ for fixed ϵ if l is small compared with $\log\frac{1}{\eta}$.

LePage's theorem is adapted to finite volume in [4], Theorem 4.1, where a result closely related to Proposition 3.3 is proved. However, it may be of interest that one can argue directly in finite volume.

We note also that the proof of Proposition 3.3 does not really use stationarity. With minor changes (mainly in the form of (22)) it works for independent, non-identically distributed V_j's provided (say) that one assumes a uniform bound in (26) and uniform lower bound on the disorder, i.e. on λ and γ in (4).

We now prove Proposition 3.3. This argument is related to the proof of Lipschitz continuity of the integrated density of states given in [2], except that by using Theorem 2.3 we can make the relevant estimates without assuming decay of $\hat{\nu}$.

Following [2] we work from formula (22) and define Hilbert spaces H_0 and H_1 on the half line as follows

$$\|f\|_{H^0}^2 = \int_{\mathbb{R}^2} (\frac{|f(|x|^2)|}{|x|})^2 dx$$

$$\|f\|_{H^1}^2 = \int_{\mathbb{R}^2} |f(|x|^2)|^2 + |Df(|x|^2)|^2 dx$$

Furthermore we define an operator R as follows: given a (nice) function f on the half line, consider the function on $\mathbb{R}^2$,

$$f(|x|^2)\frac{x}{|x|^2}$$

Its Fourier transform is again of the form

$$g(|x|^2)\frac{x}{|x|^2}$$

and we define $Rf = ig$. The following lemma is from [2].

Lemma 3.4 $Sf = f(0) + Rf$. In particular, if f vanishes at 0 then so does Sf, and Rf and Sf coincide.

Iterating the lemma, it follows that $(S\Gamma)^l 1 = \sum_{j=0}^{l} (R\Gamma)^j 1$. Furthermore, by the Plancherel theorem, R is an isometry on H^0 and (using Lemma 3.1) S is an isometry on H^1. By Theorem 2.3 applied to the functions $\frac{x}{|x|^2} f(|x|^2)$ with the quadratic form $Q(x) = |x|^2$, we also have

$$\|(R\Gamma)^j f\|_{H^0} \leq \rho^j \|f\|_{H^0} \tag{26}$$

for $j \geq 2$, where $\rho < 1$ depends on ν only. Furthermore, since (25) implies that $\hat{\nu}'$ is bounded, it is easily checked that

$$\|S\Gamma f\|_{H^1} \leq A\|f\|_{H^1} \tag{27}$$

where A depends on a bound for $|E|$.

In order to exploit (26) and (27), we let ϕ be a C_0^∞ function which is equal to 1 in a neighborhood of the origin. We define $\phi_t(x) = \phi(t^{-1}x)$. We also fix an index $k \geq 2$ and may then express $(S\Gamma)^l 1$ for $l > k$ in the following form:

$$(S\Gamma)^l 1 = \sum_{j=k}^{l} (R\Gamma)^j 1 + (S\Gamma)^k \phi + (R\Gamma)^k (1 - \phi) \tag{28}$$

In view of (27), the second term satisfies the following estimate:

$$\|(S\Gamma)^k \phi\|_{H^1} \leq CA^k \tag{29}$$

For the remaining estimates we use the fact that

$$|\Gamma(r^2)| \leq e^{-2\pi\eta r^2} \tag{30}$$

and therefore also

$$\|(1 - \phi_t)\Gamma\|_{H^0} \leq C(\log \frac{1}{\eta})^{\frac{1}{2}} \tag{31}$$

provided $t \geq \eta$, say. Taking $t = 1$ and using (26), we get the following bound for the third term in (28):

$$\|(R\Gamma)^k (1 - \phi)\|_{H^0} \leq \rho^k (\log \frac{1}{\eta})^{\frac{1}{2}} \tag{32}$$

Next, we have the following lemma:

Lemma 3.5 $\|(R\Gamma)^2 1\|_{H^0} \leq C(\log \frac{1}{\eta})^{\frac{1}{2}}$

Proof We have $\|(R\Gamma)^2(1-\phi_\eta)\Gamma\|_{H^0} \leq \|1-\phi_\eta\|_{H^0} \leq C(\log\frac{1}{\eta})^{\frac{1}{2}}$. On the other hand, define $g : \mathbb{R}^2 \to \mathbb{R}$ via $g(x) = \frac{x}{|x|^2}(\phi_\eta\Gamma)(|x|^2)$. Then $\|g\|_{L^1(\mathbb{R}^2)} \leq C\eta^{\frac{1}{2}}$. Consequently, $\|\hat{g}\|_\infty \leq C\eta^{\frac{1}{2}}$ or equivalently

$$|(R\Gamma\phi_\eta)(r^2)| \leq C\eta^{\frac{1}{2}}r$$

It follows by (30) that

$$|(\Gamma R\Gamma\phi_\eta)(r^2)| \leq C\eta^{\frac{1}{2}}re^{-2\pi\eta r^2}$$

and therefore that

$$\|\Gamma R\Gamma\phi_\eta\|_{H^0} \leq C(\int_0^\infty \eta e^{-4\pi\eta r^2}rdr)^{\frac{1}{2}}$$

which is bounded independently of η. The lemma follows, since R is an isometry. □

We conclude that the summands in the first term of (28) satisfy

$$\|(R\Gamma)^j 1\|_{H^0} \leq C\rho^j(\log\frac{1}{\eta})^{\frac{1}{2}} \tag{33}$$

Combining (28),(29),(32),(33) we conclude that for any $k \geq 2$ and $l > k$, $(S\Gamma)^l 1 = f + g$ with

$$\|f\|_{H^1} \leq CA^k$$

$$\|g\|_{H^0} \leq C\rho^k(\log\frac{1}{\eta})^{\frac{1}{2}}$$

We now set k equal to $\epsilon_0 \log\frac{1}{\eta}$ with ϵ_0 a small positive constant. Then $(S\Gamma)^l 1 = f + g$, where $\|f\|_{H^1}^2 \leq C\eta^{-\frac{1}{2}}$ and $\|g\|_{H^0}^2 \leq C\eta^{\epsilon_1}$. From (22) and (30), we have

$$\begin{aligned} |g_l(z)| &\leq C(\|f\|_{H^1}^2 + \|g\|_{H^0}^2 \sup_r(r^2 e^{-2\pi\eta r^2})) \\ &\leq C\eta^{\epsilon_1 - 1} \end{aligned}$$

completing the proof of Proposition 3.3. □

Further remarks 1. Theorem 2.3 can also be used to refine the main estimate in [17] and extend it to Bernoulli distributions. Given Theorem 2.3, this argument is identical to the corresponding argument in [17], so we will omit details. If we start from formula (21), multiply it by its complex conjugate and take expectations, then we obtain the following, where $G_l(m,n,z)$ is the Green's function on $-l,\ldots,l$ (cf. [17]):

$$\mathbb{E}(|G_l(m,n,z)|^2) = 4\int_{\mathbb{R}^2\times\mathbb{R}^2}((T\Gamma)^{m-n}\Phi_n)(x,y)\Phi_m(x,y)\Gamma(x,y)dxdy \tag{34}$$

Here T is the $\mathbb{R}^2 \times \mathbb{R}^2$ Fourier transform defined via

$$Tf(\xi,\eta) = \int_{\mathbb{R}^2\times\mathbb{R}^2} e^{-2\pi i(\xi\cdot x - \eta\cdot y)}dxdy$$

$$\Gamma(x,y) = \hat{\nu}(|x|^2 - |y|^2)e^{2\pi i E(|x|^2-|y|^2)}e^{-2\pi\eta(|x|^2+|y|^2)}$$

and we also use Γ to denote multiplication by the function Γ. Also Φ_m and Φ_n are functions with L^∞ norm ≤ 1 (see the discussion in the paragraph before (25)), and therefore also $\|T\Gamma\Phi_n\|_{L^2(\mathbb{R}^2\times\mathbb{R}^2)} \leq \|e^{-2\pi\eta(|x|^2+|y|^2)}\Phi_n\|_{L^2(\mathbb{R}^2\times\mathbb{R}^2)} \leq C\eta^{-1}$ and similarly $\|\Gamma\Phi_m\|_{L^2(\mathbb{R}^2\times\mathbb{R}^2)} \leq C\eta^{-1}$. Theorem 2.3 with the quadratic form $Q(x,y) = |x|^2 - |y|^2$ implies the estimate $\|(T\Gamma)^2\|_{L^2\to L^2} \leq \rho$, where $\rho < 1$ depends on ν only.[2] Applying this estimate $\approx |m-n|$ times to the function $T\Gamma\Phi_n(x,y)$, we readily obtain

Proposition 3.6 If ν satisfies (3) then there is an estimate

$$\mathbb{E}(|G(m,n,z)|^2) \leq Ce^{-C^{-1}\lambda^2 l}\eta^{-2}$$

for the Green's function on $\{-l,\dots,l\}$, uniformly in l and $m,n \in \{-l,\dots,l\}$ with $|m| \leq \frac{l}{4}$ and $|n| \geq \frac{l}{2}$ and $z = E + i\eta$ with $0 < \eta \leq 1$. Here C is a positive constant depending on γ only.

The usual "multiscale" arguments lead from this and Proposition 3.3 to a proof of localization, see the end of [17] and [8], Theorem 2.3.

2. More generally, the considerations in remark 1 can be applied on any tree with subexponential growth. One can show for example that the spectrum of the Anderson model is pure point provided the single site distribution is absolutely continuous with bounded density, since the analogue of Proposition 3.6 can be proved as before using a suitable generalization of (21) (essentially derived in [16]), and localization then follows using the Wegner estimate and the theory of rank one perturbations. We omit the proofs. Molchanov has previously obtained some results of this nature.

4. Inversion operators

If $x \in \mathbb{R}^n$ then in the introduction we defined

$$x^* = \frac{x}{|x|^2}$$

and we defined a unitary operator $\mathcal{I} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ via $\mathcal{I}f(x) = |x|^{-n}f(x^*)$.

Let $O(n)$ be the orthogonal group. If $V \in O(n)$ then we let $\mathcal{I}_V f = \mathcal{I}(f \circ V^{-1})$. We note that $\mathcal{I}_V$ is unitary and interchanges functions supported on the unit disc and its complement.

Theorem 4.1 Assume that μ_1 and μ_2 are probability measures on $\mathbb{R}^n$ satisfying (4). Define $T_\mu : L^2 \to L^2$ via $T_\mu f = \mu * \mathcal{I}_V f$. Then $\|T_{\mu_1}T_{\mu_2}\|_{L^2\to L^2} \leq 1 - C^{-1}\lambda^2$ where C depends on γ and n.

[2] The $\mathbb{R}^2 \times \mathbb{R}^2$ Fourier transform differs from the $\mathbb{R}^4$ Fourier transform by an orthogonal change of variable, so Theorem 2.3 is applicable by remark 1 at the end of section 2.

The proof is quite simple. In particular, it is based on a standard version of the uncertainty principle, namely the $s = 0$ case of the following lemma. Let L^2_w be the weighted L^2 space with weight w, i.e. $\|f\|^2_{L^2_w(E)} = \int_E |f(x)|^2 w(x)dx$, and $L^2_w \stackrel{def}{=} L^2_w(\mathbb{R}^n)$. Also let $[\alpha]$ be the greatest integer $\leq \alpha$ and $\{\alpha\} = \alpha - [\alpha]$.

Lemma 4.2 Assume

$$\sup_{a\in\mathbb{R}^n} |E \cap D(a,1)| < \epsilon \tag{35}$$

Fix $s \in \mathbb{R}^+ \cup \{0\}$. Then for a certain $\kappa = \kappa_s > 0$ and $C_s < \infty$, supp$\hat{f} \subset D(0,1)$ implies

$$\|f\|_{L^2_{|x|^{-s}}(E)} \leq C_s \epsilon^\kappa \|f\|_{L^2_{|x|^{-s}}} \tag{36}$$

If $s < n$ then we can take $\kappa_s = \frac{1}{2}(1 - \frac{s}{n})$, and the constant C_s may also be taken uniform over $s \in [0,a]$ for any fixed $a < n$. If $s \geq n$ we can take $\kappa_s = \frac{1}{n}(1 - \{\frac{s-n}{2}\})$.

Proof Fix a Schwarz function ϕ with $\hat{\phi} = 1$ on $D(0,1)$. Then $f = \phi * f$. Consider convolution with ϕ as an operator from $L^2_{(1+|x|)^{-s}}$ to $L^2_{(1+|x|)^{-s}}(E)$. By Schur's test its operator norm is bounded by $\sqrt{A_1 A_2}$ where $A_1 = \sup_{x\in E} \int |\phi(x-y)| \frac{(1+|x|)^{\frac{s}{2}}}{(1+|y|)^{\frac{s}{2}}} dy$ and $A_2 = \sup_{y\in\mathbb{R}^n} \int_E |\phi(x-y)| \frac{(1+|x|)^{\frac{s}{2}}}{(1+|y|)^{\frac{s}{2}}} dx$. Using the rapid decay of ϕ it is not hard to show that $A_1 \lesssim 1$ and $A_2 \lesssim \epsilon$. We conclude that (36) holds with $\kappa = \frac{1}{2}$ if $s = 0$, and also for any s provided the weight $|x|^{-s}$ is replaced by $(1+|x|)^{-s}$. Hence

$$\|f\|_{L^2_{|x|^{-s}}(E\cap D(0,1)^c)} \lesssim \sqrt{\epsilon} \|f\|_{L^2_{|x|^{-s}}}$$

On the other hand, if supp$\hat{f} \subset D(0,1)$ and $\|f\|_{L^2_{|x|^{-s}}} = 1$ (say) then f must vanish to order k at 0, where $k = [\frac{s-n}{2}] + 1$ if $s \geq n$ and $k = 0$ if $s < n$. Furthermore, when $|x| \leq 1$ the kth derivatives of f are bounded by constants, since

$$\begin{aligned} D^\alpha f(x) &= \int f(y) D^\alpha \phi(x-y) dy \\ &= \int |y|^{-\frac{s}{2}} f(y) |y|^{\frac{s}{2}} D^\alpha \phi(x-y) dy \\ &\leq \|f\|_{L^2_{|x|^{-s}}} \|(1+|y|)^{\frac{s}{2}} D^\alpha \phi\|_2 \end{aligned}$$

Accordingly $|x| \leq 1$ implies $|f(x)| \lesssim |x|^k$, and then $\int_{E\cap D(0,1)} |x|^{-s}|f|^2 \lesssim \int_{E\cap D(0,1)} |x|^{2k-s} \lesssim \epsilon^{\frac{2k-s+n}{n}}$, and one can check that κ_s in the statement of the lemma is equal to $\frac{1}{2}(\frac{2k-s+n}{n})$. □

The next two lemmas are from elementary functional analysis; we include proofs for the reader's convenience. (Lemmas like 4.4 are often used in related contexts - see [6],[23]) Lemma 4.3 Let X, μ be a measure space, H a closed subspace of $L^2(X,\mu)$, P_H the orthogonal projection on H and U a unitary operator on L^2. Assume that for a certain $A \geq 1$

$$\|f\|^2 \leq A^2(\|P_H f\|^2 + \|P_H U f\|^2) \tag{37}$$

for all $f \in L^2$. Let G_1, G_2 be functions on X with $\|G_j\|_\infty \leq 1$, and define $T : L^2 \to L^2$ via $Tf = G_2 U(G_1 f)$. Fix $\rho > 0$ and let $E_j = \{x \in X : |G_j(x)| \geq 1 - \rho\}$. Assume the following condition: if $h \in H$, then $\|h\|_{L^2(E_j)} \leq \frac{1}{4A}\|h\|_2$, $j = 1, 2$.

Then $\|T\|_{L^2 \to L^2} \leq 1 - C^{-1}\rho$, where C depends on A only.

Proof Let $\alpha = \frac{1}{4A}$. we note for future reference that $2\sqrt{2}A\alpha = \frac{\sqrt{2}}{2}$ and $1-(2A^2+2)\alpha^2 \geq \frac{3}{4} > \frac{\sqrt{2}}{2}$. Now fix a function f. If

$$\|f\|_{L^2(E_1^c)} \geq \alpha\|f\|_2 \tag{38}$$

then $\|G_1 f\|_2^2 \leq (1-\alpha^2\rho)\|f\|_2^2$, as in the proof of Theorem 2.3. Consequently it suffices to show that either f satisfies (38) or $\tilde{f} \overset{def}{=} U(G_1 f)$ satisfies

$$\|\tilde{f}\|_{L^2(E_2^c)} \geq \alpha\|\tilde{f}\|_2 \tag{39}$$

We first prove the following:

Claim Assume that g and h are orthogonal in L^2 and $\|g+h\|_{L^2(E^c)} \leq \alpha\|g+h\|_2$, $\|h\|_{L^2(E)} \leq \alpha\|h\|_2$. Then (unless $g = h = 0$) $\|g\|_2^2 > 2A^2\|h\|_2^2$.

Proof Assume the opposite. Then $h \neq 0$ and

$$\begin{aligned}
2 < g, h >_{L^2(E^c)} &\leq \|g+h\|_{L^2(E^c)}^2 - \|h\|_{L^2(E^c)}^2 \\
&\leq \alpha^2\|g+h\|_2^2 - (1-\alpha^2)\|h\|_2^2 \\
&= \alpha^2\|g\|_2^2 - (1-2\alpha^2)\|h\|_2^2 \\
&\leq -(1-(2A^2+2)\alpha^2)\|h\|_2^2
\end{aligned}$$

whence

$$2 < g, h >_{L^2(E)} \geq (1-(2A^2+2)\alpha^2)\|h\|_2^2$$

and then by the Schwartz inequality

$$\|h\|_2^2 \leq \frac{2}{1-(2A^2+2)\alpha^2}\|g\|_{L^2}\|h\|_{L^2(E)} \leq \frac{2\sqrt{2}A\alpha}{1-(2A^2+2)\alpha^2}\|h\|_2^2 < \|h\|_2^2$$

which is a contradiction.

Suppose now that f does not satisfy (38). Then, as in the proof of Theorem 2.3, $G_1 f$ also does not satisfy (38). Accordingly, the claim with $g = P_{H^\perp}(G_1 f), h = P_H(G_1 f)$

implies that $\|P_H G_1 f\|^2 < \frac{1}{2A^2}\|G_1 f\|^2$. Likewise, if $\tilde{f}$ does not satisfy (39), then the claim implies $\|P_H \tilde{f}\|^2 < \frac{1}{2A^2}\|\tilde{f}\|^2$ which is equivalent to $\|P_H U(G_1 f)\|^2 < \frac{1}{2A^2}\|G_1 f\|^2$. Thus

$$\|G_1 f\|^2 \leq A^2(\|P_H G_1 f\|^2 + \|P_H U(G_1 f)\|^2) < \|G_1 f\|^2$$

and we have a contradiction. □

Remarks 1. Note that (37) will hold with $A = 1$ if U maps H onto its orthogonal complement.

2. Lemma 4.3 remains valid for Hilbert space valued functions. Namely, replace $L^2(X, \mu)$ in the statement by $L^2(X, \mu, V)$, the L^2 functions on X taking values in the Hilbert space V. The lemma is still valid, with exactly the same proof.

Lemma 4.4 Assume that ν is a probability measure on a space X and that $\{U_E\}_{E \in X}$ is a (measurable) family of unitary operators on a Hilbert space H. Let $T = \int U_E d\nu(E)$. Then for $f \in H$

$$\int \|U_E f - f\|_2^2 d\nu(E) \geq \frac{1}{2}(1 - \|T^2\|)$$

Proof Fix $f \in H$ with $\|f\| = 1$. Then

$$\begin{aligned} 1 - \|T^2\| &\leq 1 - \mathrm{re}\langle T^2 f, f\rangle \\ &= \int 1 - \mathrm{re}\langle U_E U_F f, f\rangle d\nu(E) d\nu(F) \\ &= \frac{1}{2}\int \|U_E U_F f - f\|^2 d\nu(E) d\nu(F) \\ &\leq \int \|U_E f - f\|^2 d\nu(E) d\nu(F) + \int \|U_E(U_F f - f)\|^2 d\nu(E) d\nu(F) \\ &= 2\int \|U_E f - f\|^2 d\nu(E) \end{aligned}$$

□

Proof of Theorem 4.1 Let H be the functions whose Fourier transforms are supported in $D(0,1)$, let $G_j = \hat{\mu}_j$ and $\rho = C^{-1}\lambda^2$, where C is a large constant. Define $E_j = \{x \in \mathbb{R}^n : |G_j(x)| \geq 1 - \rho$, as in Lemma 4.3. If $\epsilon > 0$ is given, then by choosing $C = C_\epsilon$ large enough, we can guarantee that $|E \cap D(a,1)| < \epsilon$ for all $a \in \mathbb{R}^n$, uniformly in λ - this follows from Lemma 1.1. If ϵ is small enough then we conclude by Lemma 4.2 that $\|f\|_{L^2(E)} \leq \frac{1}{4}\|f\|_2$ for all $f \in H$. The conclusion now follows from Lemma 4.3, since $\mathcal{I}_V$ is unitary and interchanges functions supported on $D(0,1)$ and $D(0,1)^c$. □

When $n = 1, Vx = -x$ the operator $\mathcal{I}_V$ is an element of the principal series representation $\mathcal{P}^{+,0}$ (cf. [19], p.36) of $SL(2,\mathbb{R})$. We now prove the analogous result for the other irreducible unitary representations. We let

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and for $E \in \mathbb{R}$ we let

$$s_E = \begin{pmatrix} 1 & 0 \\ E & 1 \end{pmatrix}$$

Theorem 4.5 Let ρ be any irreducible unitary representation of $SL(2, \mathbb{R})$ except for the trivial one dimensional representation. If μ and ν are probability measures on $\mathbb{R}$ satisfying (3) and if $\rho(\mu) = \int \rho(s_E \mathcal{I}) d\mu(E)$, then $\|\rho(\mu)\rho(\nu)\| \leq 1 - C^{-1}\lambda^2$. Here C depends on γ and on the representation ρ. The dependence on ρ is as follows: if $\epsilon > 0$ is given then C may be taken independent of ρ provided (in the notation of [19], section 2.5) ρ is not a complementary series representation $\mathcal{C}^u$ with $u > 1 - \epsilon$.

Remarks 1. The estimate in Theorem 4.5 generalizes in an obvious way to representations ρ such that for some ϵ the direct integral decomposition of ρ does not contain the trivial representation or the representations $\mathcal{C}^u$, $u > 1 - \epsilon$.

2. The assumption that ρ not contain $\mathcal{C}^u$ for $u > 1 - \epsilon$ is needed. Essentially this follows from the fact that $SL(2, \mathbb{R})$ does not have property T. For example, [6] (ch. 3, proofs of Theorem 4 and Proposition 6(i)) gives examples of representations of $SL(2, \mathbb{R})$ whose restrictions to the subgroup generated by J and τ_2 have almost invariant vectors but no invariant vectors.

3. An analogous result is valid for $SL(2, \mathbb{C})$ and in fact is a bit easier, since there are no discrete series representations. In the next section, we also prove a result like Theorem 4.1 for the real symplectic group, although at present we don't have a result for general representations in that case.

Proof of Theorem 4.5 We will follow the terminology in [19]. We have to discuss the principal, discrete and complementary series. We let $\mathcal{I} = \rho(J)$ and $\tau_E = \rho(s_E)$.

As has already been mentioned the case of the principal series representation $\mathcal{P}^{+,0}$ is simply the one dimensional case of Theorem 4.1. The other principal series representations differ from this one only in that the definition of $\mathcal{I}$ is multiplied by a unimodular factor. The operators in Theorem 4.1 are positive (in the pointwise sense), hence multiplying $\mathcal{I}$ by a unimodular factor can only decrease the norm of $\rho(\mu)\rho(\nu)$, so in the case of the principal series nothing needs to be done.

The complementary series representation $\mathcal{C}^u$ acts on the Sobolev spaces $W^{-\frac{u}{2}}$,

$$\|f\|_{W^{-\frac{u}{2}}} \stackrel{def}{=} \left(\int_{\mathbb{R}} \frac{|\hat{f}(\xi)|^2}{|\xi|^u} d\xi \right)^{\frac{1}{2}}$$

where in this case $0 < u < 1$. The operators τ_E are translations, $\tau_E f(x) = f(x - E)$. The operator $\mathcal{I}$, which we will now denote by $\mathcal{I}_u$, is given by

$$\mathcal{I}_u f(x) = |x|^{-(u+1)} f(\frac{-1}{x})$$

and $\rho(\mu) : f \to \mu * \mathcal{I}_u f$. Consider the weighted L^2 space $L^2_u \overset{def}{=} L^2_{|x|^{-u}}(\mathbb{R})$. Let $H = \{f \in L^2_u :\ \text{supp}\hat{f} \subset [-2,2]\}$. Define U via $\widehat{Uf} = \mathcal{I}_u \hat{f}$ and note that $UH = \{f \in L^2_u :\ \text{supp}\hat{f} \subset \mathbb{R}\backslash(-\frac{1}{2}, \frac{1}{2})$.

If $f \in W^{-\frac{u}{2}}$, then $f = g + h$ with supp$g \subset [-2,2]$, supp$h \subset \mathbb{R}\backslash(-\frac{1}{2}, \frac{1}{2})$ and $\|g\|_{W^{-\frac{u}{2}}} + \|h\|_{W^{-\frac{u}{2}}} \leq A_0 \|f\|_{W^{-\frac{u}{2}}}$; A_0 is independent of $u \in [0,a]$ for $a < 1$. It follows by basic Hilbert space theory that (37) is valid for an appropriate constant A depending on A_0 only. Now define $E_j = \{x \in \mathbb{R} : |\hat{\mu}_j(x)| \geq 1 - C^{-1}\lambda^2\}$ for an appropriate C. Lemma 1.1 and the $n = 1, s < 1$ case of Lemma 4.2 imply that if C is large enough (depending on a lower bound for $1 - u$ and on γ) and if $h \in H$, then $\|h\|_{L^2_{|x|^{-u}}(E_j)} \leq \frac{1}{4A} \|h\|_{L^2_{|x|^{-u}}}$. The result now follows from Lemma 4.3.

The preceding argument could also be used in the case of the discrete series, but would not show that the bounds are independent of the representation. Accordingly we will give a different proof[3] based on the metaplectic representation and Theorem 2.3. Consider the tensor product of two copies of the metaplectic representation of $SL(2,\mathbb{R})$; this gives a representation on $L^2(\mathbb{R}^2)$. The discrete series representations are subrepresentations of this one or its conjugate (one reference for this is [12] where an analogous result is proved for the holomorphic discrete series representations of the symplectic group; other references are in [10], p.216), and Theorem 2.3 can be applied to the resulting realizations of the operators $\rho(\mu)$ on $L^2(\mathbb{R}^2)$.

Details are as follows. We consider the holomorphic $m \in \mathbb{Z}^+$; the antiholomorphic ones are entirely analogous. The argument is elementary so we will avoid quoting results from the representation theory literature; however the following unitary equivalence is essentially the one in [12]. Let $\mathcal{H}_m$ be the Hilbert space for $\mathcal{D}^+_{m+1}$, i.e., the elements $f \in W^{-\frac{m}{2}}$ such that $\hat{f}(\xi)$ vanishes when $\xi < 0$. Also let $\mathcal{F}$ be the $\mathbb{R}^2$ Fourier transform. Define $V_m : \mathcal{H}^m \to L^2(\mathbb{R}^2)$ via

$$V_m f(z) = \frac{z^m}{|z|^{2m}} \hat{f}(\frac{|z|^2}{2}) \tag{40}$$

where we use complex notation for points of $\mathbb{R}^2$. Then V_m is a scalar multiple of an isometry and has the following intertwining properties

$$V_m \mathcal{I}_m f = -i\mathcal{F} V_m f \tag{41}$$

$$V_m \tau_E f = e^{-\pi i E|z|^2} V_m f \tag{42}$$

(Proofs: the isometry property follows from the definition of the norm in $W^{-\frac{m}{2}}$ using calculus. The formula (42) is obvious. One elementary way to prove (41) is to verify it

[3] Actually, it would also be possible to reduce the discrete series case to the case of $\mathcal{P}^{+,0}$ using the "principe de majoration".

by hand for functions of the form $f_a(x) = (x+ia)^{-(m+1)}$, $a > 0$; linear combinations of such functions are dense in $\mathcal{H}_m$. The verification may be reduced to the case $a = 1$ using dilations, and on the other hand $\mathcal{I}_m f_1 = i^{-(m+1)} f_1$, and also $V_m f_1$ is a scalar multiple of the function $z^m e^{-\pi|z|^2}$ whence $\mathcal{F} V_m f_1 = i^{-m} V_m f_1$ by the Hecke identity) Hence V_m intertwines the operator $\rho(\mu) : f \to \mu * \mathcal{I}_m f$ with the operator on $L^2(\mathbb{R}^2)$, $f \to -i(\hat{\mu} \circ Q)\mathcal{F} f$, where $Q(z) = \frac{1}{2}|z|^2$. The product of two operators of the latter type has norm $\leq 1 - C^{-1}\lambda^2$ by Theorem 2.3 and the result follows. □

Remark Theorems 4.1 and 4.5 assert the existence of a "spectral gap" and have the usual corollaries associated to such results. See [23] and references there.

For example, let Γ be a lattice in $PSL(2,\mathbb{R})$, let $M = PSL(2,\mathbb{R})/\Gamma$ and let ρ be the representation of $PSL(2,\mathbb{R})$ on $L^2_0(M) \overset{def}{=} \{f \in L^2(M) : \int_M f = 0\}$ resulting from the left action of $PSL(2,\mathbb{R})$ on M. We will assume Γ is cocompact in order to have a convenient reference (namely [13], p. 47) for the fact that a complementary series representation $\mathcal{C}^u$ can occur in the direct integral decomposition of ρ only if $\frac{1-u^2}{4}$ is an eigenvalue of the Laplacian on the corresponding hyperbolic surface $\mathbb{H}/\Gamma$. It follows that ρ satisfies the hypothesis in remark 1 after the statement of Theorem 4.5. Fix $\alpha_1, \alpha_2 \in \mathbb{R}$ and let $g_j = s_{\alpha_j} J$, $j = 1, 2$. Let $\{\gamma_j\}_{j=1}^{2^n}$ be all words of length n in g_1 and g_2. Then we have an estimate

$$\|2^{-n} \sum_{j=1}^{2^n} \rho(\gamma_j) f\|_2 \leq e^{-C^{-1}|\alpha_1 - \alpha_2|^2 n} \|f\|_2, \quad f \in L^2_0(M)$$

uniformly in $n \geq 2$, and the resulting mixing properties. Here C depends on a lower bound for the first eigenvalue of the Laplacian on $\mathbb{H}/\Gamma$. This follows by applying Theorem 4.5 with $\mu = \frac{1}{2}(\delta_{\alpha_1} + \delta_{\alpha_2})$ and expanding out the definition of T^n_μ.

Next let G be the conformal group in $\mathbb{R}^n$, i.e. the group generated by translations, conformal linear maps and the inversion $z \to z^*$; G acts on $L^2(\mathbb{R}^n)$ via

$$U_{\gamma^{-1}} : f \to |\det(D\gamma)|^{\frac{1}{2}} f \circ \gamma, \quad \gamma \in G$$

Let Γ be a finitely generated subgroup of G with generators $\{\gamma_j\}_{j=1}^n$. Assume (i) Γ contains a parabolic element (i.e. an element which is conjugate to a translation) and (ii) there is no point of $\mathbb{R}^n \cup \{\infty\}$ which is fixed by all elements of Γ. Then there is $\epsilon_0 > 0$ such that

$$\forall f \in L^2 \exists j \in \{1, \dots, n\} : \ \|f - U_{\gamma_j} f\|_2 \geq \epsilon_0 \|f\|_2$$

Namely, by appropriate conjugations we may assume that Γ contains a translation τ_p and an element γ mapping ∞ to 0 and furthermore (conjugate γ by a dilation if necessary) that γ has the form

$$z \to (Vz - b)^*$$

with $V \in O(n)$. But then $\gamma \circ \tau_p$ is another element of the same form with the same V and a different b so the result follows in a standard way from Theorem 4.1.

We now consider applications of Theorem 4.5 to the Anderson model; this is related e.g. to considerations in [7] and [18]. We would like to consider the strip so we will now prove a result like Theorem 4.1 which is applicable to the equation for the invariant measure on projective space.

We will denote points of $\mathbb{R}^{2n}$ by (x, y), $x, y \in \mathbb{R}^n$, or by $(\overline{x}, x_n, \overline{y}, y_n)$, $\overline{x}, \overline{y} \in \mathbb{R}^{n-1}$. We define $\mathbb{P}^{2n-1}$ in the usual way, i.e. as the lines through the origin in $\mathbb{R}^{2n}$, and will use the coordinate system obtained by normalizing y_n, i.e. the line $\ell \in \mathbb{P}^{2n-1}$ has coordinates $\overline{x}, x_n, \overline{y}$ if $(\overline{x}, x_n, \overline{y}, 1) \in \ell$.

By viewing the action of $Sp(n, \mathbb{R})$ on $\mathbb{P}^{2n-1}$ in these coordinates, we get a unitary representation on $L^2(\mathbb{R}^{2n-1})$. Thus, if we denote the action of the element $g \in Sp(n, \mathbb{R})$ on $\ell \in \mathbb{P}^{2n-1}$ by $g\ell$, then the element of the representation corresponding to $g \in Sp(n, \mathbb{R})$ maps f to $f \circ g^{-1}$ times an appropriate Jacobian factor. We let $\mathcal{I}$ be the element of this representation corresponding to the matrix

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

Also, we let $M(n)$ be the $n \times n$ symmetric matrices, and if $m \in M(n)$ we let τ_m be the representation element corresponding to the matrix

$$s_m \stackrel{def}{=} \begin{pmatrix} I & m \\ 0 & I \end{pmatrix}$$

Explicitly,

$$\mathcal{I}f(\overline{x}, t, \overline{y}) = |t|^{-n} f(\frac{-\overline{y}}{t}, -\frac{1}{t}, \frac{\overline{x}}{t})$$

and if

$$m = \begin{pmatrix} A & b \\ b^t & c \end{pmatrix}$$

with A an $(n-1) \times (n-1)$ symmetric matrix, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$, then

$$\tau_m f(\overline{x}, t, \overline{y}) = f(\overline{x} - A\overline{y} - b, t - b \cdot \overline{y} - c, \overline{y})$$

If ν is a measure on $M(n)$, then we define τ_ν by integrating the representation, i.e

$$\tau_\nu f(\overline{x}, t, \overline{y}) = \int \tau_m f(\overline{x}, t, \overline{y}) d\nu(m)$$

We also let $\tilde{\nu}$ be the measure on symplectic matrices of the form

$$\begin{pmatrix} m & -I \\ I & 0 \end{pmatrix}$$

determined by the measure ν on $M(n)$. Explicitly, we set $g_m = s_m J$ and let $\tilde{\nu}$ be the push forward of ν by the map $m \to g_m$.

We will work with the following property of ν:

Condition C_λ: for some k, the conditional measures obtained by freezing all but the (k,k) entry satisfy (3) with probability at least β.

Explicitly, this means that if we identify $M(n)$ with $\mathbb{R}^{\frac{n(n+1)}{2}}$ by $m \leftrightarrow \{m_{ij}\}_{i\le j}$ then for continuous compactly supported f

$$\int f(m)d\nu(m) = \int f(\overline{m}, m_{kk})d\sigma_{\overline{m}}(m_{kk})d\rho(\overline{m}) \tag{43}$$

where $\overline{m}$ means $\{m_{ij}\}_{i\le j,(i,j)\ne(k,k)}$, and where ρ is a probability measure on $\mathbb{R}^{\frac{n(n+1)}{2}-1}$, and $\sigma_{\overline{m}}$ is a probability measure on $\mathbb{R}$ for each $\overline{m}$ and $\rho(\{\overline{m} : \sigma_{\overline{m}} \text{ satisfies } (3)\}) \ge \beta$.

Note this condition is satisfied if the distribution of M_{kk} is independent of the other entries and satisfies (3), e.g. by the Anderson model with disorder parameter λ, even if one randomizes only at one site. Note also that the condition is invariant under translation of ν so in the Anderson model case the resulting estimates are uniform in the energy E. Corollary 4.7 below is known (in stronger form) if say $n = 1$ and the single site distribution is sufficiently smooth, cf. [22] and [3].

Theorem 4.6 Assume that ν_1, ν_2 satisfy condition C_λ. Then

$$\|\tau_{\nu_1}\mathcal{I}\tau_{\nu_2}\|_{L^2(\mathbb{R}^{2n-1})\to L^2(\mathbb{R}^{2n-1})} \le 1 - C^{-1}\lambda^2$$

where C depends on γ and β.

Corollary 4.7 Assume that ν satisfies condition C_λ. Then the largest Liapunov exponent of the measure $\tilde{\nu}$ (in the sense of [5]) is $\ge C^{-1}\frac{\lambda^2}{n}$; C depends on γ and β.

Proof of Theorem 4.6 This is similar to the proof of Theorem 4.1, except that we work with vector valued functions. Note first that we can assume $k = n$, since the statement is invariant under the natural changes of coordinates on projective space. Next, we claim that it suffices to prove the result in the special case where σ_1 and σ_2 are two probability measures on $\mathbb{R}$ satisfying (3) and $\nu_j = \delta_0 \times \sigma_j$, where δ_0 is the δ-mass at the origin in $\mathbb{R}^{\frac{n(n+1)}{2}-1}$.

For if this is proved, then for any ν_1 and ν_2 we express as in (43)

$$d\nu_j(m) = d\sigma^j_{\overline{m}}(m_{nn})d\rho^j(\overline{m})$$

Using that $s_m s_{m'} = s_{m+m'}$ for any symmetric matrices m and m', we have

$$\begin{aligned}
\|\tau_{\nu_1}\mathcal{I}\tau_{\nu_2}\|_{L^2\to L^2} &= \|\int \tau_{(\overline{m_1},0)}\tau_{\delta_0\times\sigma^1_{\overline{m_1}}}\mathcal{I}\tau_{\delta_0\times\sigma^2_{\overline{m_2}}}\tau_{(\overline{m_2},0)}d\rho^1(\overline{m_1})d\rho^2(\overline{m_2})\|_{L^2\to L^2} \\
&\le \int \|\tau_{\delta_0\times\sigma^1_{\overline{m_1}}}\mathcal{I}\tau_{\delta_0\times\sigma^2_{\overline{m_2}}}\|_{L^2\to L^2}d\rho^1(\overline{m_1})d\rho^2(\overline{m_2}) \\
&\le (1 - C^{-1}\lambda^2)\beta^2 + (1-\beta^2) \\
&\le 1 - C^{-1}\lambda^2
\end{aligned}$$

Now let S be the "slab" $\{(\overline{x}, x_n, \overline{y}) : -1 \le x_n \le 1\}$ and let $H = \{f \in L^2(\mathbb{R}^{2n-1}) : \text{supp} f \subset S\}$. Then $\mathcal{I}$ interchanges H and $H^\perp$. Let E be a set in $\mathbb{R}$ with $\sup_{a\in\mathbb{R}^n} |E \cap D(a,1)| < \epsilon$. If $f \in H$ then, by the $n = 1$, $m = 0$ case of Lemma 4.2 applied to the partial Fourier transform,

$$\int_E |\hat{f}(\overline{\xi}, \xi_n, \overline{\eta})|^2 d\xi_n \lesssim \epsilon \int_{\mathbb{R}} |\hat{f}(\overline{\xi}, \xi_n, \overline{\eta})|^2 d\xi_n$$

for any $\overline{\xi}, \overline{\eta}$. Hence also

$$\int_E \|\hat{f}(\cdot, \xi_n, \cdot)\|^2_{L^2(\mathbb{R}^{n-1}\times\mathbb{R}^{n-1})} d\xi_n \lesssim \epsilon \|\hat{f}\|_2^2$$

With $\nu_j = \delta_0 \times \sigma_j$ we have

$$\widehat{\tau_{\nu_j} f}(\overline{\xi}, \xi_n, \overline{\eta}) = \hat{\sigma_j}(\xi_n) \hat{f}(\overline{\xi}, \xi_n, \overline{\eta})$$

Since σ_j satisfies (3) the result now follows using Lemma 1.1 and the vector valued version of Lemma 4.3. $\square$

Proof of Corollary 4.7 We will follow the original proof of positivity of the Liapunov exponent due to Furstenburg [11], incorporating Theorem 4.6 in order to make the result quantitative. We first reformulate Theorem 4.6 as follows: if ν satisfies C_λ then

$$\int \|(\tau_m \mathcal{I})^{-1} f - f\|^2_{L^2(\mathbb{R}^{2n-1})} d\nu(m) \ge C^{-1}\lambda^2 \tag{44}$$

This follows immediately from Theorem 4.6 and Lemma 4.4.

Because of the upper semicontinuity of the Liapunov exponent under weak convergence, we can assume that ν is absolutely continuous to Lebesgue measure with a C_0^∞ density provided we obtain estimates which are independent of this assumption.

Claim: if ν is absolutely continuous with a C_0^∞ density then $\tilde{\nu}$ has a unique invariant measure and it is absolutely continuous with respect to the Cauchy measure (i.e. rotation invariant measure) on $\mathbb{P}^{2n-1}$ with a continuous density.

Here by invariant measure, we mean of course the invariant measure in the sense of [11],[5]. Thus an invariant measure for $\tilde{\nu}$ is a measure μ on $\mathbb{P}^{2n-1}$ such that

$$\int_{M(n)} \int_{\mathbb{P}^{2n-1}} f(g_m z) d\mu(z) d\nu(m) = \int_{\mathbb{P}^{2n-1}} f d\mu \tag{45}$$

for continuous functions f on $\mathbb{P}^{2n-1}$. The claim is certainly known but we could not find a reference where it is proved (except when $n = 1$ [26]) and a very simple argument is available so we give the argument.

Namely, any invariant measure with respect to $\tilde{\nu}$ is also invariant with respect to any convolution power of $\tilde{\nu}$ on the symplectic group, so by [11], Lemma 8.5 it suffices to show that some such convolution power is absolutely continuous (to Haar measure on $Sp(n,\mathbb{R})$) with a continuous density. The matrices of the form

$$\begin{pmatrix} m & -I \\ I & 0 \end{pmatrix}$$

generate $Sp(n,\mathbb{R})$ as a Lie group. So by a general lemma of Ricci and Stein (see [28], p. 209), a suitable convolution power $\tilde{\nu}^{(k)}$ is absolutely continuous with a density ρ satisfying an L^1 Holder estimate, i.e. in local coordinates $d\tilde{\nu}^{(k)}(x) = \rho(x)dx$ with $\int |\rho(x+h) - \rho(x)|dx \lesssim |h|^{\epsilon}$. It follows that $\rho \in L^p$ for some $p > 1$. Hence, by Young's inequality on $Sp(n,\mathbb{R})$, a high convolution power of ρ will be continuous.

We also note the following fact. Suppose that f and g are positive L^2 functions on a measure space X, dx, $\int f^2 dx = \int g^2 dx = 1$ and $\int (f-g)^2 dx = \beta$. Then $f^2 \log_+ \frac{g}{f}$ is integrable and

$$\int f^2 \log \frac{g}{f} dx \leq -\frac{\beta}{2} \tag{46}$$

Namely, the integrability follows since $f^2 \log_+ \frac{g}{f} \leq f^2 \frac{g}{f} = fg \in L^1$. Also

$$\int f^2 \log \frac{g}{f} dx \leq \int (\frac{g}{f} - 1) f^2 dx = \int fg - f^2 dx = \frac{1}{2} \int g^2 - (f-g)^2 - g^2 dx = -\frac{\beta}{2}$$

as claimed.

The Liapunov exponent can be expressed as

$$\lambda_1 = -\frac{1}{2n} \int \log \frac{dg_m^{-1}\sigma}{d\sigma}(x) d\mu(x) d\nu(m)$$

where σ is the Cauchy measure. See [5], propositions IV.6.2 and IV.6.4. If $\phi : \mathbb{P}^{2n-1} \to (0,\infty)$ is continuous, and if we let $d\tilde{\sigma} = \phi d\sigma$, then we have $\log \frac{dg_m^{-1}\tilde{\sigma}}{d\tilde{\sigma}} = \log \frac{\phi \circ g_m}{\phi} + \log \frac{dg_m^{-1}\sigma}{d\sigma}$. By the invariance property of μ,

$$\lambda_1 = -\frac{1}{2n} \int \log \frac{dg_m^{-1}\tilde{\sigma}}{d\tilde{\sigma}}(x) d\mu(x) d\nu(m) \tag{47}$$

We would like to replace $\tilde{\sigma}$ by μ here. The Radon-Nikodym derivative $f = \frac{d\mu}{d\sigma}$ is a continuous nonnegative function on $\mathbb{P}^{2n-1}$ by the claim. We haven't shown that it is nonzero, but we have $\int |\log \frac{d\mu}{d\sigma}| d\mu = \int f |\log f| d\sigma < \infty$. Using this, an approximation of $\log f$ by continuous functions and the dominated convergence theorem it follows that (47) is applicable with $\tilde{\sigma} = \mu$. This can be expressed in coordinates in the following way. Let g be the density of μ with respect to Lebesgue measure dz on $\mathbb{R}^{2n-1}$. Then

$$\lambda_1 = -\frac{1}{2n} \int_{M(n)} \int_{\mathbb{R}^{2n-1}} \log \left(\frac{\mathcal{I}_1^{-1} \tau_m^{-1} g(z)}{g(z)} \right) g(z) dz d\nu(m)$$

where $\mathcal{I}_1^{-1} f(\overline{x}, t, \overline{y}) = t^{-2n} f(-\frac{\overline{y}}{t}, -\frac{1}{t}, \frac{\overline{x}}{t})$.

Note that $\mathcal{I}_1^{-1} g = (\mathcal{I}^{-1}\sqrt{g})^2$. So by (46),

$$\lambda_1 \gtrsim \frac{1}{n} \int \|\sqrt{g} - (\tau_m \mathcal{I})^{-1}\sqrt{g}\|_2^2 d\nu(m)$$

and now we are done by (44). □

It is also possible to use the $m = 1$ case of Theorem 4.5 to prove Holder continuity of the density of states. In fact we will prove a refinement asserting that inside the spectrum of the Laplacian the bounds in Le Page's theorem are independent of the disorder for small disorder. This can also be proved by an extension of the argument in section 3, in fact we originally did it that way (see [24]). The two arguments are in a sense equivalent since the operator $\mathcal{I}_1$ corresponds to the operator S of section 3 under the unitary equivalence V_1 given by (40). The supersymmetric approach has the advantage that it works directly in finite volume, but the approach via Theorem 4.5 is shorter so we will do it that way.

We do not have a version of Proposition 4.8 below for the strip and will assume $n = 1$ for the remainder of this section. The zero moment assumption on ν is just a normalization. The compact support could perhaps be weakened but we wanted to avoid technicalities as much as possible.

Proposition 4.8 Suppose $\lambda \leq 1$ and let ν be a measure which is supported on $\{x : -A\lambda \leq x \leq A\lambda\}$ and satisfies (3), and assume $\int x d\nu(x) = 0$. Then, on compact subintervals[4] $I \subset (-2, 2)$ the integrated density of states for the Anderson model with single site distribution ν satisfies $|k(x) - k(y)| \leq C|x - y|^\alpha$ where C and α depend only on A, γ and I (and not for example on λ).

Proof We will actually consider the Liapunov exponent instead of the density of states which of course is equivalent. We let τ_E be translation by E, i.e. $\tau_E f(x) = f(x - E)$ for distributions f, and $d\nu_E = \tau_E d\nu$. We let μ_E be the invariant measure at energy E, i.e. the measure μ_E corresponding to $\nu = \nu_E$ in (45), and we let $\lambda(E)$ be the corresponding Liapunov exponent. We must prove that for $E_0, E \in I$ there is an estimate $|\lambda(E) - \lambda(E_0)| \lesssim |E - E_0|^\alpha$.

Let $Y_0 = W^{-\frac{1}{2}}$, let Y_1 be the subspace of $L^2_{1+x^2}(\mathbb{R})$ consisting of functions with $\int_{\mathbb{R}} f dx = 0$, and let Y_θ be the complex interpolation space $[Y_0, Y_1]_\theta$. Let μ_E be the invariant measure corresponding to energy E. We first prove the following

Claim There are constants $\alpha > 0$ and $C < \infty$ such that for sufficiently small θ, the estimate $\|\mu_E - \mu_{E_0}\|_{Y_\theta} \leq C|E - E_0|^\alpha \lambda^{-2}$ is valid.

[4] We are normalizing the free laplacian via $\Delta f(k) = f(k+1) + f(k-1)$ here, rather than including a factor of $\frac{1}{2}$ as in section 3.

Since the invariant measure behaves correctly under weak convergence we can assume that ν has a smooth density provided we get bounds which are independent of this assumption. If ν has a smooth density then the invariant measures μ_E will have (say) continuous densities, $d\mu_E(x) = f_E(x)dx$ with $f_E(x) \leq \min(1, x^{-2})$. This is the one dimensional case of the claim in the proof of Corollary 4.7. We also let $P_E(x) = \frac{1}{\pi}\frac{\mathrm{im}\zeta}{|x-\zeta|^2}$, where

$$\zeta = \frac{E}{2} + i\sqrt{1 - \frac{E^2}{4}} \tag{48}$$

Thus ζ is a point in the upper half plane with absolute value 1. It is a fixed point of the map $z \to \frac{-1}{z-E}$, and P_E is the corresponding Poisson kernel. It follows that $P_E dx$ is an "invarian measure at zero disorder" i.e.

$$\tau_E \mathcal{I}_1 P_E = P_E \tag{49}$$

a well-known fact (cf. [3]) which can easily be verified by direct computation.

We note the following properties of the spaces Y_θ:

(i) If $\theta < \mu$ then $Y_\mu \subset Y_\theta$ and $\|f\|_{Y_\theta} \lesssim \|f\|_{Y_\mu}$. Furthermore $f \in Y_\mu$ implies translation of f is Holder continuous into Y_θ,

$$\|\tau_v f - f\|_{Y_\theta} \lesssim |v|^{\frac{1}{2}(\mu-\theta)} \|f\|_{Y_\mu} \tag{50}$$

Since the Y_θ are defined by interpolation it suffices to give the proof when $\mu = 1,\ \theta = 0$. If $f \in Y_1$ then $\hat{f}(0) = 0$ and $\int |\hat{f}|^2 + |\hat{f}'|^2 < \infty$ and it follows that $\int \frac{|\hat{f}(\xi)|^2}{|\xi|} < \infty$, i.e. $f \in Y_0$. (50) is a standard fact about Sobolev spaces: $\|\tau_t f - f\|^2_{Y_0} = \int \frac{|e^{2\pi i t\xi} - 1|^2}{|\xi|} |\hat{f}(\xi)|^2 d\xi \lesssim \int \min(t^2|\xi|, |\xi|^{-1}) |\hat{f}(\xi)|^2 d\xi \lesssim |t| \|f\|^2_{L^2} \leq |t| \|f\|^2_{Y_1}$.

(ii) If $\int_{\mathbb{R}} f dx = 0$ and $|f(x)| \leq B\min(1, x^{-2})$ then $\|f\|_{Y_1} \lesssim B$. Consequently$\|f\|_{Y_\theta} \lesssim B$ for all θ.

This is obvious. It follows that for any θ:

$$\|P_E - \tau_v P_E\|_{Y_\theta} \lesssim |v| \tag{51}$$

$$\|P_E - \nu_E * \mathcal{I}_1 P_E\|_{Y_\theta} \lesssim \lambda^2 \tag{52}$$

$$f_E - P_E \in Y_\theta \tag{53}$$

(51) just follows from (ii) using that $|\frac{dP_E}{dx}| \lesssim \min(1, x^{-2})$. As for (52), by (i) and (49) it is equivalent to prove that $\|P_E - \nu * P_E\|_{Y_1} \lesssim \lambda^2$. The latter is an easy calculation using that ν has zero first moment. And (53) follows from our a priori hypothesis and the obvious fact that $\int f_E - P_E dx = 0$.

(iii) Define $T_E f = \nu_E * \mathcal{I}_1 f$. Then for a certain $\theta_0 > 0$ and any $\theta \in [0, \theta_0]$, the operator id$-T_E$ is invertible on Y_θ with $\|(\mathrm{id} - T_E)^{-1}\|_{Y_\theta \to Y_\theta} \lesssim \lambda^{-2}$.

We use id for the identity operator. To prove (iii), we introduce an equivalent norm $\| \cdot \|_{\theta,E}$ on Y_θ as follows: the new norm on Y_1 is the $L^2_{|x-\zeta|^2}$ norm, the norm on Y_0 remains the same and the norm on Y_θ when $0 < \theta < 1$ is obtained by interpolation. (this trick is suggested by an idea in [3]). We let $Y_{\theta,E}$ be Y_θ with the $\| \ \|_{\theta,E}$ norm. The point is that the operator $f \to \tau_E \mathcal{I}_1 f$ may easily be seen to be an isometry on $Y_{1,E}$. Furthermore, using that ν has zero first moment, it is not hard to show that convolution with ν has norm $\leq 1 + \mathcal{O}(\lambda^2)$ on $Y_{1,E}$. It follows that T_E^2 has norm $\leq 1 + \mathcal{O}(\lambda^2)$ on $Y_{1,E}$. On the other hand T_E^2 has norm $\leq 1 - C^{-1}\lambda^2$ on Y_0 by Theorem 4.5. It follows by interpolation that T_E^2 has norm $\leq 1 - C^{-1}\lambda^2$ on $Y_{\theta,E}$ for $\theta \leq \theta_0$, hence $\|(\mathrm{id} - T_E)^{-1}\|_{Y_{\theta,E} \to Y_{\theta,E}} \leq \|\mathrm{id} + T_E\|_{Y_{\theta,E} \to Y_{\theta,E}} \ \|(\mathrm{id} - T_E^2)^{-1}\|_{Y_{\theta,E} \to Y_{\theta,E}} \lesssim \lambda^{-2}$. This implies (iii) since $\| \ \|_\theta$ and $\| \ \|_{\theta,E}$ are equivalent norms.

The function f_E satisfies $\nu_E * \mathcal{I}_1 f_E = f_E$ and therefore $\nu_E * \mathcal{I}_1(f_E - P_E) = f_E - P_E + P_E - \nu_E * \mathcal{I}_1 P_E$. By (53), this means that $f_E - P_E$ is obtained from $\nu_E * \mathcal{I}_1 P_E - P_E$ by applying the operator $(\mathrm{id} - T_E)^{-1} : Y_\theta \to Y_\theta$. Hence by (52) and (iii),

$$\|f_E - P_E\|_{Y_\theta} \lesssim \lambda^{-2} \cdot \lambda^2 = 1 \tag{54}$$

We now finish with the proof of the claim. We have $f_E - f_{E_0} \in Y_\theta$ by (53) and the obvious (from (ii)) fact that $P_E - P_{E_0} \in Y_\theta$, and

$$\begin{aligned} \nu_{E_0} * \mathcal{I}_1(f_E - f_{E_0}) - (f_E - f_{E_0}) &= (\nu_{E_0} - \nu_E) * \mathcal{I}_1 f_E \\ &= (\tau_{E_0 - E} - \mathrm{id}) f_E \\ &= (\tau_{E_0 - E} - \mathrm{id})(f_E - P_E) + (\tau_{E_0 - E} - \mathrm{id}) P_E \end{aligned}$$

If θ is small, then the first term on the right side in the bottom formula has Y_θ norm $\lesssim |E - E_0|^\alpha$ for suitable $\alpha > 0$, by (54) and (50). The second term has norm $\lesssim |E - E_0|$ by (51). The claim now follows from (iii).

To prove Proposition 4.8, we apply (47) with $d\tilde{\sigma} = P_{E_0}(x)dx$. This gives the formula for the difference of Liapunov exponents

$$\lambda(E) - \lambda(E_0) = \int \Gamma(x)(f_E(x) - f_{E_0}(x))dx \tag{55}$$

where

$$\Gamma(x) = \int \log\left(\frac{x^2 P_{E_0}(x)}{P_{E_0}(E_0 + v - x^{-1})}\right) d\nu(v)$$

Letting ζ correspond to E_0 as in (48) we may rewrite (55) as follows:

$$\begin{aligned} \frac{x^2 P_{E_0}(x)}{P_{E_0}(E_0 + v - x^{-1})} &= x^2 \frac{|E_0 + v - x^{-1} - \zeta|^2}{|x - \zeta|^2} \\ &= |(1 + v\zeta)\frac{z - \frac{\zeta}{1+v\zeta}}{z - \zeta}|^2 \end{aligned}$$

hence

$$\lambda(E)-\lambda(E_0)=\int\tilde{\Gamma}(x)(f_E(x)-f_{E_0}(x))dx$$

where

$$\tilde{\Gamma}(z)=2\int\log\left|\frac{z-\frac{\zeta}{1+v\zeta}}{z-\zeta}\right|d\nu(v) \tag{56}$$

Because of the claim (with $\theta=0$) it now suffices to show that $\|\tilde{\Gamma}\|_{W^{\frac{1}{2}}}\lesssim\lambda^2$. We consider (56) with z in the lower half space $\mathbb{R}^2_-$, and $v\in$suppν. The functions

$$\frac{z-\frac{\zeta}{1+v\zeta}}{z-\zeta}$$

are bounded and bounded away from zero when $z\in\mathbb{R}^2_-$, uniformly over v for fixed ζ. Thus (56) defines a bounded harmonic function of z on $\mathbb{R}^2_-$. Differentiating under the integral sign and using the zero moment condition, we obtain

$$|\frac{\partial\tilde{\Gamma}}{\partial z}|\lesssim\lambda^2|z-\zeta|^{-2}$$

which implies

$$(\int_{\mathbb{R}^2_-}|\frac{\partial\tilde{\Gamma}}{\partial z}|^2)^{\frac{1}{2}}\lesssim\lambda^2$$

The $W^{\frac{1}{2}}$ norm coincides with the Dirichlet integral of the harmonic extension, so the proof is complete. □

Remarks 1) If one compares Proposition 4.8 and Proposition 3.6 then one sees that it should be possible to use multiscale analysis as in [8] to prove the following estimate for the "localization length" in the one dimensional Anderson model, under the same assumptions as in Proposition 4.8:

Let I be a compact interval contained in $(-2,2)$. Then with probability 1, all eigenfunctions ϕ with eigenvalue in I satisfy $|\phi(m)|\leq C_\phi e^{-C^{-1}\lambda^2|m|}$, with $C=C_{A\gamma I}$ being a fixed constant and C_ϕ a constant depending on ϕ.

We omit the proof. It may be found in [24] together with analogous results when the randomization decays at ∞.

2) We have been unable to answer the following question. Fix a compact subinterval $I\subset(-2,2)$, and let ν be as in Proposition 4.8. Is the density of states Lipschitz (or better) on I provided λ is sufficiently small?

5. The lagrangian grassmannian

The purpose of this section is to prove a result like Theorem 4.6 for the action of $Sp(n)$ on the Lagrangian (maximal isotropic) subspaces instead of projective space.

We will work in the "noncompact picture", i.e. with the action on $M(n)$, the $n \times n$ symmetric matrices. We define translation operators τ_A ($A \in M(n)$) acting on functions on $M(n)$ via $\tau_A f(T) = f(T-A)$. The inner product on $M(n)$ is given by $\langle A, B\rangle =$tr(AB), and other quantities defined in terms of the inner product such as norm and Fourier transform will be interpreted accordingly. We let $\mathcal{D} = \{A \in M(n) : |\det A| \leq 1\}$ and define an inversion operator on functions via

$$\mathcal{I}f(A) = |\det A|^{-\frac{n+1}{2}} f(A^{-1})$$

Then $\mathcal{I}$ is unitary on L^2 and evidently interchanges functions supported on $\mathcal{D}$ and on $\mathcal{D}^c$.

Theorem 5.1 Let μ be a probability measure on $M(n)$ and assume that the plane $\Sigma(\mu) \subset M(n)$ (cf. (14)) contains a positive definite matrix. Define $T_\lambda : L^2(M(n)) \to L^2(M(n))$ via

$$T_\lambda f = \mathcal{I}f * \mu_\lambda$$

where μ_λ is the dilation of μ by λ, cf. (11). Then, for $\lambda \in (0, 1]$,

$$\|T_\lambda^2\|_{L^2 \to L^2} \leq 1 - C^{-1}\lambda^2$$

where C depends on μ.

Remarks 1. More precisely, let $\lambda_n(P)$ and $\lambda_1(P)$ be the largest and smallest eigenvalues of the matrix P. Then C depends on n, on a bound for the quantity

$$\inf\left(\frac{\lambda_n(P)}{\lambda_1(P)} : \ P \in \Sigma(\mu),\ P \text{ positive definite}\right)$$

and on constants A and γ for which (10) and (9) hold for all $e \in \Sigma(\mu)$. It follows for example that the bound is uniform over the family of all translates of a given measure μ and over any weak$*$ compact family with a given $\Sigma(\mu)$.

2. As with our earlier results, one can also consider the product of operators corresponding to two different measures. Thus if μ^1 and μ^2 are two measures satisfying the hypotheses and if $T_{j,\lambda} f = If * \mu^j_\lambda$, then $\|T_{1,\lambda} T_{2,\lambda}\| \leq 1 - C^{-1}\lambda^2$. This follows by making appropriate (purely notational) changes in the last part of the proof below.

Theorem 5.1 differs from Theorem 4.1 in that the set $\mathcal{D}$ is not compact, which means that the usual "elliptic" uncertainty principle (e.g. Lemma 4.2) is not applicable. We will base the proof on the fact that the relevant differential operator (the Cayley operator) is hyperbolic in positive definite directions together with Corollary 1.3 and an uncertainty principle type lemma for hyperbolic operators, Lemma 5.2 below. This approach is likely not the shortest possible but it is natural from a certain point of view and Lemma 5.2

may be of some independent interest. We recall that the Cayley operator (for symmetric matrices) is the order n constant coefficient differential operator acting on functions on $M(n)$,

$$\Omega = \det \begin{pmatrix} \frac{\partial}{\partial t_{11}} & \cdots & \frac{\partial}{\partial t_{1n}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial}{\partial t_{1n}} & \cdots & \frac{\partial}{\partial t_{nn}} \end{pmatrix}$$

where $\{t_{ij}\}$ are the entries of the matrix in question.

Let p be a homogenous polynomial in $\mathbb{R}^d$ with real coefficients, $P(D) = p(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_j})$ the corresponding constant coefficient differential operator, $e \in \mathbb{R}^d$ a unit vector, and assume
(i) p is hyperbolic in the e direction, i.e., if $\xi \in \mathbb{R}$ then the equation $p(\xi - te) = 0$ has only real roots.
(ii) $|p(\xi + ie)| \geq C^{-1}(|\xi| + 1)$ for any $\xi \in \mathbb{R}^d$.
(iii) If $\{\lambda_j(\xi)\}$ are the roots of $p(\xi + te) = 0$, ordered so that $\lambda_j(\xi) \leq \lambda_{j+1}(\xi)$, then $|\lambda_j(\xi) - \lambda_j(\eta)| \leq C|\xi - \eta|$.
(iv) the set $\{\xi \in \mathbb{R}^d : p(\xi + te) = 0 \text{ has a multiple root}\}$ has measure zero.

Also let $E \subset \mathbb{R}$ be a set with the following property:

$$\sup_{a \in \mathbb{R}^d} |\{t \in \mathbb{R} : t = x \cdot e \text{ for some } x \in E \cap D(a, 1)\}| < \epsilon \tag{57}$$

<u>Lemma 5.2</u> If p satisfies (i) - (iv) and E satisfies (57) then there is an a priori estimate on Schwarz functions

$$\|u\|_{L^2(E)} \leq C\epsilon^{1/2}(\|u\|_{L^2} + \|p(D)u\|_{L^2})$$

<u>Remark</u>: The example we care about is $d = \frac{n(n+1)}{2}$, $p(D) =$ Cayley operator on $n \times n$ symmetric matrices, e any positive definite matrix.

We show that (i) - (iv) are then satisfied. If $\xi \in M(n)$, then

$$p(\xi - te) = \det(e) \cdot \det(e^{-\frac{1}{2}}\xi e^{-\frac{1}{2}} - t)$$

So the roots $\{\lambda_j\}$ of $p(\xi + te) = 0$ are the eigenvalues of the symmetric matrix $e^{-\frac{1}{2}}\xi e^{-\frac{1}{2}}$. This gives (i), and (ii) also follows: $|p(\xi + ie)| = \det e \prod_j |\lambda_j - i|$. All the factors are ≥ 1 and at least one must be comparable to $|\xi| + 1$. Property (iv) is obvious, and (iii) is a basic fact in eigenvalue perturbation theory (see [25], Theorem 1.20(b) for example).

Note though that the Cayley operator is not strictly hyperbolic if $n \geq 3$. In order to avoid assuming strict hyperbolicity we need some estimates on trigonometric polynomials which are independent of any lower bounds on the gaps between frequencies, which we now record.

Lemma 5.3 Let $q(t) = \sum_{j=1}^{N} a_j e^{2\pi i \lambda_j t}$ where the $\{\lambda_j\}$ are real numbers. Then

(i) For any R, $||q||_{L^\infty([-R,R])} \leq C(N,R)||q||_{L^1([-1,1])}$

(ii) If $|\lambda_j| \leq 1$ for all j, then $\|q'\|_{L^\infty([-1,1])} \leq C(N)\|q\|_{L^\infty([-1,1])}$

(iii) Suppose all $\{\lambda_j\}$ belong to the set $\{|\lambda| \geq 1\} \cup \{|\lambda| \leq \rho\}$, where $\rho < 1$. Let $q_1 = \sum_{j:|\lambda_j|\leq\rho} a_j e^{2\pi i \lambda_j t}$ Then $\|q_1\|_{L^\infty([-1,1])} \leq C(N,\rho)\|q\|_{L^\infty([-1,1])}$.

Proof All of this must clearly be known, but we do not know a reference for (iii), so we will give proofs starting from a basic inequality of Turan (see [14], p. 62 or [21] - the latter reference also gives several applications of Turan's method to questions about the uncertainty principle). Turan's inequality is

$$\|q\|_{L^\infty([-R,R])} \leq C(N,R)\|q\|_{L^\infty([-1,1])} \tag{58}$$

with an explicit $C(N,R)$, whose value we do not need here.

(i) This is a direct consequence of [21], Theorem I, but we will give a proof since the same argument will be used below to prove (iii). Consider the function

$$\phi(x) = 1 - \prod_{j=1}^{N}(1 - (\frac{\sin(x-\lambda_j)}{x-\lambda_j})^2)$$

This is an entire function of exponential type $\lesssim N$ and also satisfies $\|\phi\|_{L^1(\mathbb{R})} \lesssim N$ since $\phi(x) \leq \sum_j (\frac{\sin(x-\lambda_j)}{x-\lambda_j})^2$. It follows that $\hat{\phi}$ is supported in $|x| \leq CN$ and $\|\hat{\phi}\|_\infty \leq CN$. Furthermore, $\phi(\lambda_j) = 1$ for all j. So (constants depend on N)

$$|q(0)| = |\sum_j a_j| = |\langle \hat{q}, \phi\rangle = |\langle q, \hat{\phi}\rangle| \lesssim \|q\|_{L^1([-CN,CN])}$$

where we used the distributional Fourier transform. It follows by translation invariance that $\|q\|_{L^\infty([-1,1])} \lesssim \|q\|_{L^1([-CN-1,CN+1])}$. Then (58) implies

$$\|q\|_{L^\infty([-R,R])} \lesssim \|q\|_{L^1([-CN-1,CN+1])}$$

for any fixed R and then (i) follows by rescaling.

(ii) We skip this argument since a stronger result is explicitly proved in [14].

(iii) We will construct a function ϕ such that

- ϕ has exponential type $\leq C(N,\rho)$
- $\|\phi\|_{L^1(\mathbb{R})} \leq C(N,\rho)$
- $\phi(\lambda_j) = 1$ when $|\lambda_j| \leq \rho$, $\phi(\lambda_j) = 0$ when $|\lambda_j| \geq 1$

The following function has the indicated properties: let b_{ij} be numbers such that $b_{ij}(\lambda_j - \lambda_i) = \pi (\text{mod } 2\pi\mathbb{Z})$ and $1 \leq |b_{ij}| \leq \frac{\pi}{1-\rho}$ and define

$$\phi(x) = \prod_{i:|\lambda_i|\geq 1} \left(1 - \prod_{j:|\lambda_j|\leq\rho} (1 - (\frac{\sin b_{ij}(x-\lambda_j)}{b_{ij}(x-\lambda_j)})^2) \right)$$

The same argument as for (i) now shows that $\|q_1\|_{L^\infty([-1,1])} \leq C_N \|q\|_{L^1([-CN-1,CN+1])}$ and then (iii) follows from (i). $\square$

Remark Part (iii) will be used in the following way. Suppose that

$$q(t) = \sum_{j=1}^{N} a_j e^{2\pi i \lambda_j t}, \quad \lambda_j < \lambda_{j+1}$$

is a trigonometric polynomial. Then we can put the frequencies into groups $\{g_n\}_{n=-\infty}^{\infty}$ so that

(i) $\lambda_j \in g_n \Rightarrow |\lambda_j - n| < 1$

(ii) $\lambda_j \in g_n, \lambda_k \notin g_n \Rightarrow |\lambda_j - \lambda_k| \geq \frac{1}{N}$

Namely, for each n there is an interval $(c_n, d_n) \subset [n, n+1]$ of length $\geq N^{-1}$ which contains no λ_j's. We fix such an interval for each n and define

$$g_n = \{\lambda_j : d_{n-1} \leq \lambda_j \leq c_n\}$$

If we define

$$b_n(t) = \sum_{\lambda_j \in g_n} a_j e^{2\pi i(\lambda_j - n)t}$$

so that

$$q(t) = \sum_n b_n(t) e^{2\pi i n t}$$

then Lemma 5.3(iii) implies (uniformly in n) that

$$\|b_n\|_{L^\infty([-1,1])} \lesssim \|q\|_{L^2([-1,1])}$$

Proof of Lemma 5.2 We can assume $e = e_d$ and will denote points of $\mathbb{R}^d$ by $(x,t), x \in \mathbb{R}^{d-1}, t \in \mathbb{R}$.

We subdivide $\mathbb{R}^d$ in "slabs"

$$S_n \stackrel{def}{=} \{(x,t) : n \le t < n+1\}, n \in \mathbb{Z}$$

and will also denote S_0 by S. It suffices to show that

$$\|u\|_{L^2(E\cap S)} \lesssim \sqrt{\epsilon}(\|u\|_{L^2(S)} + \|p(D)u\|_{L^2(S)}) \tag{59}$$

since the corresponding statement for S_n then follows by translation invariance and the lemma follows by taking an ℓ^2 sum over n.

Let $f = \chi_S p(D)u$ and consider the solution of $p(D)v = f$ given by

$$\widehat{e^{2\pi t}v} = p((\xi,\tau) + i\cdot(0,1))^{-1}\widehat{e^{2\pi t}f}$$

Property (ii) implies that

$$\|e^{2\pi t}v\|_{W^{12}} \lesssim \|p(D)u\|_{L^2(S)}$$

where W^{12} is the inhomogeneous Sobolev space, i.e. $\|g\|_{W^{12}}^2 \stackrel{def}{=} \|g\|_2^2 + \|\nabla g\|_2^2$. It follows in particular that

$$\|v\|_{L^2(S)} + \|\dot{v}\|_{L^2(S)} \lesssim \|p(D)u\|_{L^2(S)} \tag{60}$$

where $\dot{v} = \frac{\partial v}{\partial t}$. Applying the one dimensional inequality

$$\|v\|_{L^\infty([0,1])} \lesssim \|v\|_{L^2([0,1])} + \|\dot{v}\|_{L^2([0,1])} \tag{61}$$

in the t variable and using (57), we obtain

$$\|v\|_{L^2(S\cap E)} \lesssim \sqrt{\epsilon}\|p(D)u\|_{L^2(S)} \tag{62}$$

Set $w = u - v$. We are going to show that

$$\|w\|_{L^2(S\cap E)} \lesssim \sqrt{\epsilon}\|w\|_{L^2(S)} \tag{63}$$

This will finish the proof of (59) (hence of the lemma) since then

$$\begin{aligned} \|u\|_{L^2(S\cap E)} &\lesssim \sqrt{\epsilon}(\|w\|_{L^2(S)} + \|p(D)u\|_{L^2(S)}) \quad \text{by (61) and (63)} \\ &\lesssim \sqrt{\epsilon}(\|u\|_{L^2(S)} + \|v\|_{L^2(S)} + \|p(D)u\|_{L^2(S)}) \\ &\lesssim \sqrt{\epsilon}(\|u\|_{L^2(S)} + \|p(D)u\|_{L^2(S)}) \end{aligned}$$

by (60).

Let $\tilde{w}(\xi, t)$ be the partial Fourier transform in the x variables, i.e.

$$\tilde{w}(\xi, t) = \int w(x, t) e^{2\pi i x \cdot \xi} dx$$

Then, since $p(D)w = 0$ on S we have (here we use assumption (iv))

$$\tilde{w}(\xi, t) = \sum a_j(\xi) e^{2\pi i \lambda_j(\xi) t}$$

where $\{\lambda_j(\xi)\}$ are the roots of $p((\xi, 0) + t \cdot (0, 1)) = 0$. By the remark after Lemma 5.3 this means that

$$\tilde{w}(\xi, t) = \sum_{m \in \mathbb{Z}} b_m(\xi, t) e^{2\pi i m t}$$

where the b_m are trigonometric polynomials with frequencies in $[-1, 1]$ and

$$\int_0^1 |b_m(\xi, t)|^2 dt \lesssim \int_0^1 |\hat{w}(\xi, t)|^2 dt$$

uniformly in m and ξ. For fixed ξ there are only a bounded number of values of m with $b_m(\xi, \cdot) \neq 0$, so in fact

$$\sum_{m \in \mathbb{Z}} \int_0^1 |b_m(\xi, t)|^2 dt \lesssim \int_0^1 |\hat{w}(\xi, t)|^2 dt \tag{64}$$

uniformly in ξ.

Now let ϕ be a fixed Schwarz function in $\mathbb{R}^{d-1}$ such that $\hat{\phi}$ has compact support and $|\phi(x)| \geq 1$ when $|x| \leq \sqrt{d}$. For $k \in \mathbb{Z}^{d-1}$, let $\phi_k(x) = \phi(x - k)$. Let $\psi_k = \hat{\phi}_k$ and let $b_m^{(k)} = \psi_k * b_m$, where the convolution is on $\mathbb{R}^{d-1}$. Then

$$\sum_{k \in \mathbb{Z}^{d-1}} \|b_m^{(k)}(\cdot, t)\|^2_{L^2(d\xi)} \lesssim \|b_m(\cdot, t)\|^2_{L^2(d\xi)} \tag{65}$$

$$\sum_{k \in \mathbb{Z}^{d-1}} \|\dot{b}_m^{(k)}(\cdot, t)\|^2_{L^2(d\xi)} \lesssim \|\dot{b}_m(\cdot, t)\|^2_{L^2(d\xi)} \tag{66}$$

uniformly in $t \in [0, 1]$ and m; these estimates just follow from the Plancherel theorem in $\mathbb{R}^{d-1}$ since $\sum_{k \in \mathbb{Z}^{d-1}} |\phi(x - k)|^2$ is bounded. Furthermore, for fixed ξ we can apply parts (i) and (ii) of Lemma 5.3 to the trigonometric polynomial $b_m(\xi, \cdot)$ obtaining

$$\sup_{t \in [0,1]} (|b_m(\xi, t)|^2 + |\dot{b}_m(\xi, t)|^2) \lesssim \int_0^1 |b_m(\xi, t)|^2 dt \tag{67}$$

uniformly in ξ. If we integrate (67) dt and (65),(66) $d\xi$ we get

$$\sum_{k \in \mathbb{Z}^{d-1}} \|b_m^{(k)}\|^2_{L^2(S)} + \|\dot{b}_m^{(k)}\|^2_{L^2(S)} \lesssim \|b_m\|^2_{L^2(S)} \tag{68}$$

uniformly in m. Now, by the thinnness assumption (57),

$$\|w\|_{L^2(S\cap E)} \lesssim \sum_{k\in\mathbb{Z}^{d-1}} \int_{D(k,\sqrt{d})\times Y_k} |w(x,t)|^2 dxdt$$

where the Y_k are subsets of $[0,1]$ with measure $\lesssim \epsilon$. Consequently

$$\begin{aligned}\|w\|_{L^2(S\cap E)} &\lesssim \sum_{k\in\mathbb{Z}^{d-1}} \int_{\mathbb{R}^{d-1}\times Y_k} |\phi_k w|^2 dxdt \\ &= \int_{\mathbb{R}^{d-1}\times Y_k} |\sum_m b_m^{(k)}(\xi,t) e^{2\pi imt}|^2 d\xi dt \qquad (69)\end{aligned}$$

by Plancherel. Assumption (iii) implies (since $\hat{\phi}$ has compact support) that for fixed ξ and k there are only a bounded number of values of m for which $b_m^{(k)}(\xi,t) \neq 0$. So by Cauchy-Schwarz,

$$(69) \lesssim \int \sum_{k,m} \int_{Y_k} |b_m^{(k)}(\xi,t)|^2 dtd\xi$$

Using (61) we obtain

$$\int_{Y_k} |b_m^{(k)}(\xi,t)|^2 dt \lesssim \epsilon(\int_0^1 |b_m^{(k)}(\xi,t)|^2 dt + \int_0^1 |\dot{b}_m^{(k)}(\xi,t)|^2 dt)$$

so (68) implies

$$\begin{aligned}\|w\|^2_{L^2(S\cap E)} &\lesssim \int \sum_m \int |b_m(\xi,t)|^2 dtd\xi \\ &\lesssim \epsilon \|w\|^2_{L^2(S)}\end{aligned}$$

where the last line follows from (64). □

Corollary 5.4 Assume that $E \subset M(n)$ satisfies (57), with e being positive definite. Let $u \in L^2(M(n))$ be such that $\text{supp}\hat{u} \subset \mathcal{D}$. Then

$$\|u\|_{L^2(E)} \leq C\epsilon^{\frac{1}{2}} \|u\|_2 \qquad (70)$$

Proof We can assume that u is a Schwarz function, and then (70) follows by applying Lemma 5.2 to the Cayley operator, since $\text{supp}\hat{u} \subset \mathcal{D}$ implies $\|\Omega u\|_2 \leq \|u\|_2$. □

Proof of Theorem 5.1 Define $\mathcal{J}$ via $\widehat{\mathcal{J}f} = \mathcal{I}\hat{f}$. Taking Fourier transforms, we see it suffices to prove that $\|\hat{\mu}_\lambda \mathcal{J} \hat{\mu}_\lambda\|_{L^2\to L^2} \leq 1 - C^{-1}\lambda^2$.

Let $H = \{u : \text{supp}\hat{u} \subset \mathcal{D}\}$. Let e be a positive definite matrix in $\Sigma(\mu)$ and for suitable constants C and ϵ let $E = \{m \in M(n) : |\hat{\mu}_\lambda(m)| \geq 1 - C^{-1}\epsilon^2\lambda^2\}$. Corollary 1.3 implies that E satisfies (57). Hence if ϵ is small then by Corollary 5.4

$$\|u\|_{L^2(E)} \leq \frac{1}{4}\|u\|_2$$

for all $u \in H$. Since $\mathcal{J}$ interchanges H and $H^\perp$ the result now follows from Lemma 4.3.□

References

[1] W. O. Amrein, A.-M. Berthier, On support properties of L^p functions and their Fourier transforms, J. Funct. Anal. 24, 258-267 (1977).

[2] M. Campanino, A. Klein, A supersymmetric transfer matrix and differentiability of the density of states in the one-dimensional Anderson model, Commun. Math. Phys. 108, 41-66 (1987).

[3] M. Campanino, A. Klein, Anomalies in the one dimensional Anderson model at weak disorder, Comm. Math. Phys. 130, 441-456 (1990).

[4] R. Carmona, A. Klein, F. Martinelli, Anderson localization for Bernoulli and other singular potentials, Commun. Math. Phys. 108, 41-66 (1987).

[5] R. Carmona, J. Lacroix, *Spectral Theory of Random Schrodinger Operators*, Birkhauser, 1990.

[6] P. de la Harpe, A. Vallette, La propriété T de Kazhdan pour les groupes localement compacts, Astérisque 175, Société Mathématique de France, 1989.

[7] F. Delyon, H. Kunz, B. Souillard, One dimensional wave equations in random media, J. Phys. A 16, 25-42 (1983).

[8] H. von Dreifus, A. Klein, A new proof of localization in the Anderson tight binding model, Comm. Math. Phys. 124, 285-299 (1989).

[9] C. Fefferman, The uncertainty principle, Bull. Amer. Math. Soc. 9, 129-206 (1983).

[10] G. Folland, *Harmonic Analysis in Phase Space*, Princeton University Press, 1989.

[11] H. Furstenburg, Noncommuting random products, Trans. Amer. Math. Soc. 108, 277-428 (1963).

[12] S. Gelbart, Holomorphic discrete series for the real symplectic group, Inv. Math. 19, 49-58 (1973).

[13] I. M. Gelfand, M. I. Graev, I. I. Pyatetskii- Shapiro, *Representation Theory and Automorphic Functions*, translated by K. A. Hirsch, W. B. Saunders, 1969.

[14] G. Halasz, H. Montgomery, Bernstein's inequality for finite intervals, pp. 60-65 in *Conference in Harmonic Analysis in Honor of Antoni Zygmund, Chicago, 1981*, ed. W. Beckner et al, Wadsworth, 1983.

[15] V. Havin, B. Joricke, *The Uncertainty Principle in Harmonic Analysis*, Springer-Verlag, 1994.

[16] A. Klein, Extended states in the Anderson model on the Bethe lattice, preprint 1994.

[17] A. Klein, F. Martinelli, J. F. Perez, A rigorous replica trick approach to Anderson localization in one dimension, Commun. Math. Phys. 106, 623-633 (1986).

[18] A. Klein, A. Speis, Regularity of the density of states in the one dimensional Anderson model, J. Funct. Anal. 898, 211-227 (1989).

[19] A. Knapp, *Representation Theory of Semisimple Lie Groups: an Overview Based on Examples*, Princeton University Press, 1986.

[20] V. N. Logvinenko, Yu. F. Sereda, Equivalent norms in spaces of entire functions of exponential type, Teor. Funtsii, Functional Anal. i. Prilozhen 19, 234-246 (1973).

[21] F. L. Nazarov, Local estimates of exponential polynomials and their application to inequalities of uncertainty principle type, St. Petersburg Math. J. 5, 663-717 (1994).

[22] L. Pastur, A. Figotin, *Spectral Theory of Random and Almost Periodic Operators*, Springer-Verlag, 1992.

[23] P. Sarnak, *Some Applications of Modular Forms*, Cambridge University Press, 1988.

[24] C. Shubin, R. Vakilian, T. Wolff, On some operators connected with the Anderson model, unpublished manuscript, 1996, available electronically at http://www.math.caltech.edu/wolff.htm

[25] B. Simon, *Trace Ideals and their Applications*, Cambridge University Press, 1979.

[26] B. Simon, M. Taylor, Harmonic analysis on $SL(2,\mathbb{R})$ and smoothness of the density of states in the one dimensional Anderson model, Comm. Math. Phys. 101, 1-19 (1985).

[27] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[28] E. M. Stein, Problems in harmonic analysis related to curvature and oscillatory integrals, *ICM Proceedings, Berkeley, 1986*, p. 196-221.

[29] E. M. Stein, *Harmonic Analysis*, Princeton University Press, 1993.

Appendix: A general contraction property in $PSL(2,\mathbb{R})$

After submitting this paper we realized that some of the results in the later sections can be proved without using Fourier transforms. In particular this means that there is no need to restrict to situations involving parabolic elements. Here we will prove a general form of Theorem 4.5. This can be read independently of the rest of the paper.

Before stating it we introduce some notation. $\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$ will be the upper half space and $\mathbb{R}^* = \mathbb{R} \cup \{\infty\}$ its boundary. The closure of a set in $\mathbb{H}$ will be its closure relative to $\mathbb{H} \cup \mathbb{R}^*$. An interval in $\mathbb{R}^*$ will be allowed to contain ∞, i.e. will be any connected subset of $\mathbb{R}^*$. We regard $PSL(2,\mathbb{R})$ as acting on $\mathbb{H}$ and on $\mathbb{R}^*$ via Mobius transformations, normalizing as in [19]. Thus, if

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL(2,\mathbb{R})$$

and $x \in \mathbb{H} \cup \mathbb{R}^*$ then $gx = \frac{dx+c}{bx+a}$. We will use the subgroups stabilizing a one or two point set:

If $z \in \mathbb{H}$, then $\mathcal{F}_z \stackrel{def}{=} \{g \in PSL(2,\mathbb{R}) : gz = z\}$.
If $x \in \mathbb{R}^*$, then $\mathcal{F}_x \stackrel{def}{=} \{g \in PSL(2,\mathbb{R}) : gx = x\}$.
If $x, y \in \mathbb{R}^*$, $x \neq y$, then $\mathcal{F}_{x,y} \stackrel{def}{=} \{g \in PSL(2,\mathbb{R}) : g$ maps the set $\{x, y\}$ into itself$\}$.

A convenient reference for basic properties of Mobius transformations is [30]. Abusing slightly the terminology in [30], we define an elementary subgroup of $PSL(2,\mathbb{R})$ to be a subgroup which is of one of the three types $\mathcal{F}_z, \mathcal{F}_x$ or $\mathcal{F}_{x,y}$. Any left coset of an elementary subgroup is also a right coset of some elementary subgroup and vice versa.

By a representation we will always mean a unitary representation of $PSL(2,\mathbb{R})$. For $0 < \alpha < 1$, we let $\mathcal{C}^\alpha$ be the complementary series representation as defined in [19]; thus, if $u = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$\mathcal{C}^\alpha(u)f(x) = |-bx + d|^{-(1+\alpha)} f(\frac{ax - c}{-bx + d}) \tag{71}$$

We will view $\mathcal{C}^\alpha$ as acting on the L^2 Sobolev space

$$W^{-\frac{\alpha}{2}}(\mathbb{R}) = \{f : \|f\|^2 \stackrel{def}{=} \int_{\mathbb{R}} |\hat{f}(\xi)|^2 |\xi|^{-\alpha} d\xi < \infty\}$$

This norm is the same as the one in [19] as may be seen by taking Fourier transforms.

For reasons that will be clear below, we will use $\mathcal{C}^0$ to denote the principal series representation obtained by inducing the trivial representation of the upper triangular subgroup, i.e. the representation called $\mathcal{P}^{+,0}$ in [19]. Thus $\mathcal{C}^0$ acts on $L^2(\mathbb{R})$ and $\mathcal{C}^0 u$ is given by (71) with $\alpha = 0$.

Theorem A.1 Let μ be a probability measure on $PSL(2,\mathbb{R})$ and assume that $\text{supp}\mu$ is not contained in a coset of an elementary subgroup.

Let ρ be a representation of $PSL(2,\mathbb{R})$ such that for some $\delta > 0$, the direct integral decomposition of ρ does not contain the trivial representation nor the representations $\mathcal{C}^u, u > 1 - \delta$. Define $\rho(\mu) = \int \rho(g) d\mu(g)$. Then $\|\rho(\mu)\| < 1$.

Corollary A.2 Let μ be a probability measure on $PSL(2,\mathbb{R})$ such that suppμ is not contained in an elementary subgroup. Let ρ be a representation satisfying the condition in Theorem A.1. Then $\|\rho(\mu)^3\| < 1$.

Proof of the corollary If $E \subset PSL(2,\mathbb{R})$ then let $E^k = \{\Pi_{j=1}^k e_j : e_1, e_2, \ldots, e_k \in E\}$. Using that a Mobius transformation has at most two fixed points it is easy to see that if E, E^2 and E^3 are each contained in a coset of an elementary subgroup, then E is contained in an elementary subgroup. We leave details of this argument to the reader. Let μ be a probability measure, $E = \text{supp}\mu$. If $\|\rho(\mu)^3\| = 1$, then of course also $\|\rho(\mu)^2\| = \|\rho(\mu)\| = 1$. Combining the preceding observation with Theorem A.1 we get the result. $\square$

Remarks 1. Theorem A.1 can be applied to the left regular representation of $PSL(2,\mathbb{R})$, whose direct integral decomposition contains no complementary series representations. It follows that if μ is a probability measure not supported on a coset of an elementary subgroup, then left (or right) convolution with μ is a strict contraction on $L^2(PSL(2,\mathbb{R}))$. Theorem A.1 also be applied to $L^2(PSL(2,\mathbb{R})/\Gamma)$ as indicated in the remarks in section 4.

2. The hypothesis that μ not be supported on an elementary subgroup is exactly the hypothesis of Furstenburg's theorem [11] in the case of $PSL(2,\mathbb{R})$, so our assumptions are the natural ones. In fact it is easy to construct various examples showing that if μ is supported on an elementary subgroup then $\rho(\mu)$ can have spectral radius 1.

Now the proof of Theorem A.1. It suffices to consider irreducible representations, and although this could be avoided, we will simplify matters a bit further by using the "principe de majoration" of Herz, which implies that it suffices to consider the representations $\mathcal{C}^\alpha, 0 \le \alpha < 1$. This follows just as in [31] for example: the principe de majoration (e.g. [31], Lemma 5.2) shows that if ρ is induced from a unitary representation of the upper triangular subgroup then $\|\rho(\mu)\| \le \|\mathcal{C}^0(\mu)\|$. The other principal series representations as well as the regular representation are induced from the upper triangular subgroup, and the discrete series representations are direct summands of the regular representation. Thus, Theorem A.1 is equivalent to the following

Theorem A.1$'$ Let μ be a probability measure on $PSL(2,\mathbb{R})$ not supported on a coset of an elementary subgroup. Then, for any $\delta > 0$ there is $\epsilon > 0$ such that $\|\mathcal{C}^\alpha(\mu)\| < 1 - \epsilon$ for all $\alpha \in [0, 1-\delta)$.

We define an ϵ-invariant vector for an operator T in the usual way, i.e. as a vector f such that $\|Tf - f\| < \epsilon\|f\|$, and a vector is ϵ-invariant for a collection of operators $\mathcal{T}$ if it is ϵ- invariant for each $T \in \mathcal{T}$. Thus no vector is 0- invariant. It is well-known that bounding $\|\rho(\mu)\|$ away from 1 reduces to a question of nonexistence of ϵ-invariant vectors for small $\epsilon > 0$. The following lemma is a convenient formulation of this principle.

Lemma A.3 Let k be a positive integer, ρ a unitary representation, and for each $g, g_1, \cdots, g_k$, let $\epsilon(g; g_1 \dots g_k)$ be such that collection of operators $\{\rho(g_1 g^{-1}), \dots, \rho(g_k g^{-1})\}$ has no $\epsilon(g; g_1 \dots g_k)$-invariant vectors. (We allow the possibility that $\epsilon(g; g_1 \dots g_k) = 0$) Then for any probability measure μ

$$1 - \|\rho(\mu)\|^2 \geq (2k)^{-1} \int \epsilon(g; g_1 \dots g_k)^2 d\mu(g) d\mu(g_1) \dots d\mu(g_k)$$

Proof For any unit vector f,

$$\begin{aligned} 1 - \|\rho(\mu)f\|^2 &= \frac{1}{2} \int \|\rho(g)f - \rho(\overline{g})f\|^2 d\mu(g) d\mu(\overline{g}) \\ &= \frac{1}{2k} \int \sum_{i=1}^{k} \|\rho(g)f - \rho(g_i)f\|^2 d\mu(g) d\mu(g_1) \dots d\mu(g_k) \\ &\geq \frac{1}{2k} \int \epsilon(g; g_1 \dots g_k)^2 d\mu(g) d\mu(g_1) \dots d\mu(g_k) \end{aligned}$$

□

Lemma A.4 Let $k \in (0,1) \cup (1, \infty)$ and define

$$u_k = \begin{pmatrix} k^{\frac{1}{2}} & 0 \\ 0 & k^{-\frac{1}{2}} \end{pmatrix} \in PSL(2, \mathbb{R}) \tag{72}$$

Let I be an interval in $\mathbb{R}$ of the form $[a, b]$ or $[-b, -a]$, where $0 < a < \frac{b}{2}$. Let $n \in \mathbb{Z}^+$ be such that $\max(k^n, k^{-n}) \geq \frac{b}{a}$. Then, for $0 \leq \alpha \leq 1 - \delta$, any unit vector f which is $\epsilon-$invariant under $\mathcal{C}^\alpha(u_k)$ must satisfy

$$\|\chi_I f\|_{W^{-\frac{\alpha}{2}}} \leq C(n\epsilon)^{\frac{1}{3}}$$

where C depends on δ only.

Proof We note that $\mathcal{C}^\alpha(u_k)f(x) = k^{\frac{1+\alpha}{2}} f(kx)$.

We consider the interval $[a, b]$, where $0 < a \leq \frac{b}{2}$, and we may assume that $n = 1$ and $k \geq \frac{b}{a}$, since any vector ϵ- invariant for a unitary operator T is $\epsilon|n|$-invariant for T^n. We will use the following fact:

For $n \in \mathbb{Z}$, let $I_n = [k^n a, k^n b]$. Define $\Delta_n f$ via $\widehat{\Delta_n f} = \chi_{I_n} \hat{f}$, where χ_{I_n} is the indicator function of I_n. Let Sf be the Littlewood-Paley square function, $Sf = (\sum_n |\Delta_n f|^2)^{\frac{1}{2}}$. Then, for $0 \leq \alpha \leq 1 - \delta$,

$$\int_{\mathbb{R}} |Sf(x)|^2 |x|^{-\alpha} dx \leq C_\delta \int_{\mathbb{R}} |f(x)|^2 |x|^{-\alpha} dx \tag{73}$$

where C_δ depends on δ only.

When $\alpha = 0$ this is obvious, and the general case is also well-known. For example, since $|x|^{-\alpha}$ is an A_2 weight uniformly in $\alpha \leq 1-\delta$, it follows from Theorem 2 of [32] by the randomization argument used there.

Using (73) and taking Fourier transforms, we have

$$\sum_j \|\chi_{I_j} f\|^2_{W^{-\frac{\alpha}{2}}} \leq C_\delta \|f\|^2_{W^{-\frac{\alpha}{2}}} \tag{74}$$

In particular,

$$\|\chi_I f\|^2_{W^{-\frac{\alpha}{2}}} \leq C_\delta \|f\|^2_{W^{-\frac{\alpha}{2}}} \tag{75}$$

Now let f be an ϵ-invariant vector with norm 1 and define $f_m = \mathcal{C}^\alpha(u_{k^m})f$. Then $\|f - f_m\| \leq \epsilon|m|$ and therefore

$$\|\chi_I(f - f_m)\| \leq C\epsilon|m|$$

by (75). It follows that

$$\text{If } |m| \leq \frac{\|\chi_I f\|}{2C\epsilon}, \text{ then } \|\chi_I f_m\| \geq \frac{1}{2}\|\chi_I f\|$$

On the other hand, we have $\chi_I f_m = \mathcal{C}^\alpha(u_{k^m})(\chi_{I_m} f)$, and $\mathcal{C}^\alpha(u_{k^m})$ is unitary, so if $|m| \leq \frac{\|\chi_I f\|}{2C\epsilon}$, then $\|\chi_{I_m} f\| \geq \frac{1}{2}\|\chi_I f\|$. If we square this, sum over m and use (74) we obtain the following:

$$\begin{aligned} \epsilon^{-1}\|\chi_I f\|^3 &\leq 8C \sum_{|m| \leq \frac{\|\chi_I f\|}{2C\epsilon}} \|\chi_{I_m} f\|^2 \\ &\lesssim \|f\|^2 \end{aligned}$$

and therefore the lemma. □

Lemma A.5 Let h_1 and h_2 be two hyperbolic elements of $PSL(2,\mathbb{R})$ without a common fixed point. Let $\{p_i, q_i\}$ be the fixed point set of h_i and let $[p_1, q_2, p_2, q_1] \stackrel{def}{=} \frac{(p_1-p_2)(q_2-q_1)}{(p_1-q_2)(p_2-q_1)}$ be the cross ratio. Assume that

$$\text{trace}(h_i) \geq 2 + \eta^2,\ i = 1, 2 \tag{76}$$

$$\tau^{-1} \leq |[p_1, q_1, q_2, p_2]| \leq \tau \tag{77}$$

for certain numbers $\eta \in (0,1]$ and $\tau \in [2,\infty)$.

Fix $\delta > 0$ and let $\epsilon = C^{-1}\frac{\eta}{\log\tau}$ for a suitable constant $C = C_\delta$. Then the pair $\{\mathcal{C}^\alpha(h_1), \mathcal{C}^\alpha(h_2)\}$ has no ϵ-invariant vectors if $0 \leq \alpha \leq 1-\delta$.

Remark Of course any hyperbolic h_1 and h_2 without common fixed points will satisfy (76) and (77) for some $\eta \in (0,1]$ and $\tau \in [2,\infty)$. Further, η, τ and therefore also ϵ may be chosen uniformly over all pairs (h_1, h_2) in a small neighborhood of a given one. The quantitative statement in Lemma A.5 is used only through this uniformity property.

Proof The statement is clearly invariant under conjugation in $PSL(2,\mathbb{R})$, so, with u_k as in (72), we can assume that $h_1 = u_k$ and that h_2 has fixed points ω and 1, with $\tau^{-1} \leq \omega < 1$. Thus $h_2 = v u_m v^{-1}$, where v maps $\{0, \infty\}$ to $\{\omega, 1\}$. Here $k, m \in (0,1) \cup (1, \infty)$ must satisfy $k^{\frac{1}{2}} + k^{-\frac{1}{2}} \geq 2 + \eta^2$, $m^{\frac{1}{2}} + m^{-\frac{1}{2}} \geq 2 + \eta^2$.

Consider the four intervals in $\mathbb{R}^*$ defined by $I_1 = [-2, -\frac{\omega}{2}]$, $I_2 = [-\frac{\omega}{2}, \frac{\omega}{2}]$, $I_3 = [\frac{\omega}{2}, 2]$, $I_4 = [2, -2]$. I_1 and I_3 are of the form in Lemma A.4 with $\frac{b}{a} \approx \omega^{-1} \leq \tau$ while the images of I_2 and I_4 under v^{-1} are of the form in Lemma A.4 with $\frac{b}{a} \leq$constant. Let $n = C\frac{\log \tau}{\eta}$ for a suitable fixed constant C. By taking C large enough, we can make $\max(k^n, k^{-n}) \geq C_1 \tau$ (and also $\max(m^n, m^{-n}) \geq C_1$) for any fixed constant C_1. If f is ϵ-invariant for $\mathcal{C}^\alpha(h_1)$ and $\mathcal{C}^\alpha(h_2)$ then we can apply Lemma A.4 for u_k with the intervals I_1 and I_3 and the function f and for u_m with the intervals $v^{-1}I_2$ and $v^{-1}I_4$ and the function $\mathcal{C}^\alpha(v^{-1})f$. It follows that $\|\chi_{I_j} f\| \lesssim (\epsilon \frac{\log \tau}{\eta})^{\frac{1}{3}} \|f\|$ for $j = 1$ or 3 and $\|\chi_{I_j} f\| \lesssim (\frac{\epsilon}{\eta})^{\frac{1}{3}} \lesssim (\epsilon \frac{\log \tau}{\eta})^{\frac{1}{3}} \|f\|$ for $j = 2$ or 4. Since f is the sum of the four functions $\chi_{I_j} f$ we conclude that $\epsilon \frac{\log \tau}{\eta} \gtrsim 1$, as claimed. □

Proof of Theorem A.1$'$ We fix $\delta > 0$; it is understood below that constants may depend on δ. Fix $\gamma \in \text{supp}\mu$, let $\Sigma = \{g\gamma^{-1} : g \in \text{supp}\mu\}$ and let Γ be the group generated (algebraically) by Σ. By [30], Theorem 5.1.3, Γ must contain two hyperbolic elements h_1^0, h_2^0 without common fixed points. Hyperbolic elements without common fixed points are open in $PSL(2,\mathbb{R}) \times PSL(2,\mathbb{R})$; further, we can choose a neighborhood V of (h_1^0, h_2^0) in $PSL(2,\mathbb{R}) \times PSL(2,\mathbb{R})$ and numbers $\tau \in [2, \infty)$ and $\eta \in (0, 1]$ such that if $(h_1, h_2) \in V$, then h_1 and h_2 satisfy (76) and (77) with these values of τ and η. It follows by Lemma A.5 that there is $\epsilon > 0$ such that if $(h_1, h_2) \in V$ and $\alpha \leq 1 - \delta$ then the pair $\{\mathcal{C}^\alpha(h_1), \mathcal{C}^\alpha(h_2)\}$ has no ϵ-invariant vectors. Regarding h_1^0 and h_2^0 as words in the elements of Σ, we see that there are a positive integer k and open sets $S \subset PSL(2,\mathbb{R}) \times \overset{k\, times}{\cdots} \times PSL(2,\mathbb{R})$ with $\mu \times \ldots \times \mu(S) > 0$ and $U \subset PSL(2,\mathbb{R})$ with $\mu(U) > 0$ such that if $g \in U$ and $(g_1, \ldots, g_k) \in S$ then there are two words h_1 and h_2 of length $\leq k$ in $(g_1 g^{-1}, \ldots, g_k g^{-1})$ such that $(h_1, h_2) \in V$. It follows that if $(g_1, \ldots, g_k) \in S$ and $g \in U$ then $\{\mathcal{C}^\alpha(g_i g^{-1})\}_{i=1}^k$ have no $\frac{\epsilon}{k}$-invariant vectors. The theorem now follows by Lemma A.3. □

Remark The above proof does not immediately give a quantitative result, because of the appeal to Theorem 5.1.3 of [30]. It is not very difficult to modify the argument so as to obtain such a result, but we will not pursue this here.

On the other hand, the above proof does give the following uniformity statement:

Corollary A.6 Let K be a weak*-compact set of probability measures on $PSL(2,\mathbb{R})$ and ρ a representation satisfying the hypothesis of Theorem A.1. If no measure in K is supported on a coset of an elementary subgroup, then there is $\epsilon > 0$ such that $\|\rho(\mu)\| \leq 1 - \epsilon$ for all $\mu \in K$. If no measure in K is supported on an elementary subgroup, then there is $\epsilon > 0$ such that $\|\rho(\mu)^3\| \leq 1 - \epsilon$ for all $\mu \in K$.

To see this, fix $\mu \in K$ and $\gamma \in \text{supp}\mu$, and define S and U as in the proof of Theorem A.1$'$. Observe that for ν in a suitable weak* neighborhood $\mathcal{N}$ of μ, the quantities $\nu(U)$ and $\nu \times \ldots \times \nu(S)$ are bounded away from zero independently of ν. It follows that the bound in Theorem A.1$'$ is uniform in $\nu \in \mathcal{N}$, hence so is the bound in Theorem A.1. The first part of Corollary A.6 now follows using the Heine-Borel property. The second part follows from the first since taking the third convolution power preserves weak* compactness. ☐

Additional References

[30] A. F. Beardon, *The Geometry of Discrete Groups*, Springer-Verlag, 1983.

[31] M. Cowling, The Kunze-Stein phenomenon, Ann. Math. 107(1978), 209-234.

[32] D. S. Kurtz, Littlewood-Paley and multiplier theorems on weighted L^p spaces, Trans. Amer. Math. Soc. 259(1980), 235-254.