

EIGENVECTOR LOCALIZATION FOR RANDOM BAND MATRICES WITH POWER LAW BAND WIDTH

JEFFREY SCHENKER

ABSTRACT. It is shown that certain ensembles of random matrices with entries that vanish outside a band around the diagonal satisfy a localization condition on the resolvent which guarantees that eigenvectors have strong overlap with a vanishing fraction of standard basis vectors, provided the band width W raised to a power μ remains smaller than the matrix size N . For a Gaussian band ensemble, with matrix elements given by i.i.d. centered Gaussians within a band of width W , the estimate $\mu \leq 8$ holds.

1. INTRODUCTION

Random band matrices, with entries that vanish outside a band of width W around the diagonal, have been suggested [7, 8] as a model to study the crossover between a strongly disordered “insulating” regime, with localized eigenfunctions and weak eigenvalue correlations, and a weakly disordered “metallic” regime, with extend eigenfunctions and strong eigenvalue repulsion. Such a crossover is believed to occur in the spectra of certain random partial differential (or difference) operators as the spectral parameter (energy) is changed.

In this paper, the strong disorder side of the band matrix crossover is analyzed. It is shown here that certain ensembles of random matrices whose entries vanish in a band of width W around the diagonal satisfy a localization condition in the limit that the size of the matrix N tends to infinity provided $W^8/N \rightarrow 0$. This result requires certain assumptions on the distribution of the entries of the matrix, and the proof given here has technical requirements that may not be necessary. Nonetheless, the conditions imposed below (see §3) allow for a large family of interesting examples. In particular, one may consider a Gaussian distributed band matrix, with distribution

$$(1.1) \quad e^{-2W \operatorname{tr} X_{W;N}^2} dX_{W;N}$$

where $dX_{W;N}$ the Lebesgue measure on the vector space of $N \times N$ matrices of band width W . That is

$$(1.2) \quad X_{W;N} = \frac{1}{\sqrt{W}} \underbrace{\begin{pmatrix} d_{1,1} & a_{1,2} \cdots & a_{1,W} \\ a_{2,1}^* & d_{2,2} & \ddots \\ \vdots & & \ddots & \ddots \\ a_{W,1} & & \ddots & \ddots & \ddots \end{pmatrix}}_{N \times N},$$

with d_i and $a_{i,j}$ independent families of i.i.d. real and complex unit Gaussian variables, respectively.

The main result obtained here is a localization estimate for the eigenvectors of the matrices $X_{W;N}$. This localization result is most conveniently stated in terms of the resolvent $(X_{W;N} - \lambda)^{-1}$, a well defined random matrix for $\lambda \in \mathbb{R}$. (We will see that λ is an eigenvalue of $X_{W;N}$ with probability zero.) Let \mathbf{e}_i denote the standard basis vectors $\mathbf{e}_i(j) = \delta_{i,j}$. Then

Theorem 1. *If $X_{W;N}$ has distribution (1.1), or more generally a distribution satisfying assumptions 1, 2, and 3 in §3 below, then there exists $\mu > 0$ and $\sigma < \infty$ such that given $r > 0$ and $s \in (0, 1)$ there are $A_s < \infty$ and $\alpha_s > 0$ such that*

$$(1.3) \quad \mathbb{E} \left(|\langle \mathbf{e}_i, (X_{W;N} - \lambda)^{-1} \mathbf{e}_j \rangle|^s \right) \leq A_s W^{s\sigma} e^{-\alpha_s \frac{|i-j|}{W^\mu}}$$

for all $\lambda \in [-r, r]$ and all $i, j = 1, \dots, N$. For the Gaussian band ensemble (1.1) $\sigma \leq \frac{1}{2}$ and $\mu \leq 8$.

Remarks: For the Gaussian Band Ensemble (1.1), the density of states, in the regime $W, N \rightarrow \infty, W/N \rightarrow 0$, is known to be the Wigner semi-circle law (see §5 below). For λ outside the support of the semi-circle law, one could obtain (1.3) with $\mu = 1$ using Lifschitz tail type estimates. This will be dealt with in a separate paper.

Theorem 1 estimates the decay of matrix elements of the resolvent away from the diagonal. Using techniques developed in the context of discrete random Schrödinger operators one may obtain from (1.3) estimates on eigenvectors.

Theorem 2 (Eigenvector localization). *Let $X_{W;N}$ have distribution (1.1), or more generally a distribution satisfying assumptions 4 and 5 in §5.*

- (1) *With probability one all eigenvalues of $X_{W;N}$ are simple.*
- (2) *If (1.3) holds for all λ in an interval $[-r, r]$ and if $\lambda_k, k = 1, \dots, N$, are the eigenvalues of $X_{W;N}$ with corresponding eigenvectors $\mathbf{v}_k, k = 1, \dots, N$, then there are $B < \infty, \tau \geq 0$, and $\beta > 0$*

$$(1.4) \quad \mathbb{E} \left(\sup_{\lambda_k \in [-r, r]} |\mathbf{v}_k(i) \mathbf{v}_k(j)| \right) \leq B W^\tau e^{-\beta \frac{|i-j|}{W^\mu}}$$

for all $i, j = 1, \dots, N$.

Remark. For the proof of this theorem, the reader is directed to the corresponding result in the context of random Schroedinger operators. See for example [15] for the non-degeneracy of the eigenvalues and [2, Theorem A.1] for a derivation of (1.4) from Green's function decay (1.3). In both cases, the proof involves only averaging over the coupling of a rank one perturbation and can be applied in the present context.

1.1. Sketch of the Proof. The proof of Theorem 1 is based on two observations, which may be summarized as follows.¹ Let $G_{W;N}(i, j) = \langle \mathbf{e}_i, (X_{W;N} - \lambda)^{-1} \mathbf{e}_j \rangle$. Then

- (1) *The random variable $G_{W;N}(i, j)$ is rarely large.* This may be expressed through a bound (uniform in N) on the tails of the distribution of $G_{W;N}(i, j)$

¹The idea to study localization via these two complementary estimates was suggested in the context of random Schroedinger operators by Michael Aizenman, and is inspired by the Dobrushin-Shlosman proof [10] of the Mermin-Wagner Theorem [17] on the absence of continuous symmetry breaking in classical statistical mechanics of dimension 2.

Lemma W. *If $X_{W;N}$ has distribution (1.1), or more generally a distribution with the properties outlined in §3 below, then there exist $\kappa > 0$ and $\sigma < \infty$ such that*

$$(1.5) \quad \text{Prob}(|G_{W;N}(i, j)| > t) \leq \kappa \frac{W^\sigma}{t}.$$

(2) *The fluctuations of $\ln |G_{W;N}(i, j)|$ grow at least linearly with $|i - j|$. One would typically express the growth of fluctuations by an inequality like*

$$\text{Var}(\ln |G_{W;N}(i, j)|) \geq \text{const. } |i - j|,$$

where $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ is the variance of a random variable X . However for present purposes a more convenient quantitative expression of this idea is the following

Lemma F. *If $X_{W;N}$ has distribution (1.1), or more generally a distribution with the properties outlined in §3 below, then there is $\nu > 0$ such that if $0 < r < s < 1$ and $|i - j| > 3W$ then*

$$(1.6) \quad \mathbb{E}(|G_{W;N}(i, j)|^r) \leq \exp(-C_{r,s}W^{-\mu}|i - j|)\mathbb{E}(|G_{W;N}(i, j)|^s)^{r/s}$$

with $C_{r,s} > 0$. For the Gaussian band ensemble (1.1) $\mu \leq 8$.

Lemmas W and F together easily imply Theorem 1. Indeed, it suffices to show that the second factor on the right hand side of (1.6) is uniformly bounded. But it follows from Lemma W that

$$(1.7) \quad \mathbb{E}(|G_{W;N}(i, j)|^s) \leq \frac{\kappa^s}{1-s}W^{s\sigma}.$$

This observation, which is the basis of the fractional moment analysis of random Schrödinger operators [3, 1, 2], follows easily from (1.5) since

$$(1.8) \quad \mathbb{E}(|G_{W;N}(i, j)|^s) = s \int_0^\infty \text{Prob}(|G_{W;N}(i, j)| > t) t^{s-1} dt,$$

and probabilities are bounded by one.

It may not be immediately clear what Lemma F has to do with large fluctuations. Towards understanding this, let $X = \ln |G_{W;N}(i, j)|$. By the Hölder inequality,

$$(1.9) \quad \mathbb{E}(e^{rX}) \leq \mathbb{E}(e^{sX})^{\frac{r}{s}}.$$

Furthermore, equality holds only if X is *non random* — if there is $x_0 \in \mathbb{R}$ so that $X = x_0$ almost surely. In other words

$$(1.10) \quad \mathbb{E}(e^{rX}) = e^{-h(r,s)}\mathbb{E}(e^{sX})^{\frac{r}{s}}$$

with $h(r, s) > 0$ unless X is non random.

If X were Gaussian with variance σ^2 (and arbitrary mean), then $h(r, s)$ would be proportional to the variance

$$(1.11) \quad h(r, s) = \frac{r(s-r)}{2}\sigma^2.$$

For a general random variable X , the associated quantity $h(r, s)$ may be taken as a measure of the fluctuations of X . In place of (1.11), we have the following identity for h in terms of the variance of X in weighted ensembles:

Proposition 3. *Let X be a random variable with $\mathbb{E}(e^{sX}) < \infty$ for some $s > 0$. If $r \in (0, s)$, then $\mathbb{E}(e^{rX}) < \infty$ and*

$$(1.12) \quad h(r, s) = \frac{1}{s} \int_0^s \min(r, q) (s - \max(r, q)) \text{Var}_q(X) dq,$$

where $h(r, s)$ is defined by (1.10) and

$$(1.13) \quad \text{Var}_q(X) = \frac{\mathbb{E}(X^2 e^{qX})}{\mathbb{E}(e^{qX})} - \left(\frac{\mathbb{E}(X e^{qX})}{\mathbb{E}(e^{qX})} \right)^2$$

is the variance of X with respect to the weighted probability measure $\text{Prob}_q(A) = \mathbb{E}(\chi_A e^{qX}) / \mathbb{E}(e^{qX})$.

Proof. Hölder's inequality is the statement that the function $\Phi(r) = \ln \mathbb{E}(e^{rX})$ is convex. In particular, if $s > 0$ then

$$(1.14) \quad \Phi(r) \leq \frac{r}{s} \Phi(s)$$

for $r \in (0, s)$, since $\Phi(1) = \ln \mathbb{E}(1) = 0$. If $\mathbb{E}(e^{sX}) < \infty$, it follows that Φ is bounded on $[0, s]$.

The identity (1.12) follows from Taylor's formula with remainder. Indeed, the second derivative of Φ at r is equal to the weighted variance $\text{Var}_r(X)$. Thus,

$$(1.15) \quad \Phi(s) = \Phi(r) + \Phi'(r)(s - r) + \int_r^s (s - q) \text{Var}_q(X) dq, \quad \text{and}$$

$$(1.16) \quad 0 = \Phi(0) = \Phi(r) - \Phi'(r)r + \int_0^r q \text{Var}_q(X) dq.$$

Taking a convex combination of these identities, chosen so the first order terms cancel, gives

$$(1.17) \quad \begin{aligned} \frac{r}{s} \Phi(s) &= \Phi(r) + \int_0^r \frac{(s - r)q}{s} \text{Var}_q(X) dq + \int_r^s \frac{(s - q)r}{s} \text{Var}_q(X) dq \\ &= \Phi(r) + \frac{1}{s} \int_0^s \min(r, q)(s - \max(r, q)) \text{Var}_q(X) dq, \end{aligned}$$

which is equivalent to (1.12). \square

Thus Lemma F may be understood as giving a lower bound on the fluctuations of $X = \ln |G_{W;N}(i, j)|$, as measured by the improvement to Hölder's inequality. The proof of this result will be accomplished using a product formula for $G_{W;N}(i, j)$ that expresses this quantity as a matrix element of a product of $O(|i - j|/W)$ matrices of size $W \times W$. Prop. (3) will be applied to factors in this product, with each factor contributing a term of size $1/W^7$ to $h(r, s)$. Since there are $O(|i - j|/W)$ terms, this produces the claimed decay.

The strategy taken below in proving Lemmas W and F has two parts. First we identify certain axioms for the distribution of $X_{W;N}$ which lead naturally to the lemmas. Second, we verify that the Gaussian band ensemble (1.1) satisfies these axioms. To motivate the form of the axioms for the distribution of $X_{W;N}$, we begin in §2 with a self contained sketch of the argument in the tri-diagonal case $W = 2$. In §3 we state the assumptions needed to adapt the proof to $W > 2$, state the associated general results and prove Lemma W. In §4 we get to the heart of the matter and prove Lemma F. In §5, we discuss examples of ensembles, including the Gaussian band ensemble (1.1), satisfying the axioms of §3. In an appendix, an elementary probability lemma used below is stated and proved.

1.2. Remarks on the literature and open problems. In [7, 8] it was observed, based on numerical evidence, that the localization of eigenfunctions and eigenvalue statistics of the Gaussian band ensemble (1.1) are essentially determined by the parameter W^2/N . When $W^2/N \ll 1$ the eigenfunctions are strongly localized and the eigenvalue process is close to a Poisson process. When $W^2/N \gg 1$ the eigenfunctions are extended and the eigenvalue statistics are well described by the Gaussian unitary ensemble (GUE). A theoretical physics explanation of these numerical results was given by Fyodorov and Mirlin [13]. They considered a slightly different ensemble in which a full GUE matrix is modified by multiplying each element by a factor which decays exponentially in the distance from the diagonal. For this model, on the basis of super-symmetric functional integrals, they obtain an effective σ -model approximation which, at the level of saddle point analysis, shows a localization/delocalization transition at $W \approx \sqrt{N}$.

Theorem 1 is consistent with the above picture. However, [7, 8, 13] suggest that proper exponent on the r.h.s. of (1.3) would be $\mu = 2$.

Problem 1. *What is the optimal value of μ in (1.3)? In particular, does this equation hold with $\mu = 2$?*

In the physics literature, the nature of eigenvalue processes in the large N limit is generally expected to be related to localization properties of the eigenfunctions, with Poisson statistics corresponding to localized eigenfunctions and Wigner-Dyson statistics corresponding to extended eigenfunctions. Let us call this idea the “statistics/localization diagnostic.” (In the context of band random matrices, a vector \mathbf{v} is a function on the index set $\{1, \dots, N\}$, namely $\mathbf{v}(i) = i^{\text{th}}$ coordinate of \mathbf{v} . The statistics/localization diagnostic suggests that the eigenvalues of a random matrix should be approximately uncorrelated if a typical eigenvector is essentially supported on a vanishing fraction of $\{1, \dots, N\}$, and should show strong correlations if it is typically spread over more or less the entire index set.)

The extreme cases $W = 1$ and $W = N$ of the Gaussian band ensemble (1.1) are consistent the statistics/localization diagnostic. Indeed, with $W = 1$, the matrix is diagonal and the eigenvalues, which are just the diagonal entries $d_{j,j}$, are independent. After suitable rescaling the eigenvalue process converges to a Poisson process in the large N limit. (This is essentially the definition of a Poisson process.) Likewise the eigenfunctions are the elementary basis vectors $\mathbf{e}_i(j) = \delta_i(j)$, which are localized on single sites. On the other hand, with $W = N$ the matrix $X_{W;N}$ is sampled from the GUE. In this case, the eigenfunctions together form a uniformly distributed orthonormal frame, so they are completely extended, and a suitable rescaling of the eigenvalue process converges in distribution to an explicit determinantal point process as calculated by Dyson [12, 11].

Based on the statistics/localization diagnostic, it is reasonable to conjecture that Poisson statistics hold for local fluctuations of the eigenvalues of $X_{W;N}$ in a limit $N \rightarrow \infty$ with $W = W(N) \rightarrow \infty$ provided $W(N)^\mu/N \rightarrow 0$. (One must be a little careful with the diagnostic, as it is easy to concoct random matrices with totally extended eigenfunctions and arbitrary statistics: put N random numbers with any given joint distribution on the diagonal of a matrix and conjugate the result with a random unitary! Of course, in that ensemble the matrix elements will most likely be highly correlated. Thus, it remains plausible that the statistics/localization diagnostic is correct, at least, for matrices with independent matrix elements.)

For random Schrödinger operators, Minami has derived Poisson statistics for the local correlations of the eigenvalue process from exponential decay of the resolvent [18]. Some

aspects of Minami's proof translate to the present context. Most notably, the so-called *Minami estimate* which bounds the probability of having two eigenvalues in a small interval,

$$(1.18) \quad \frac{1}{N^2} \text{Prob} [\#\{\lambda_j \in I\} \geq 2] \leq C_W |I|^2,$$

where $|I|$ is the length of the interval and $\lambda_1 \leq \dots \leq \lambda_N$ are the eigenvalues of $X_{W;N}$, holds with

$$(1.19) \quad C_W \propto W^{2\sigma}.$$

Here σ is as in Thm. 1. (The proof of this fact may be accomplished by following Minami's argument or by one of the various alternatives that have appeared recently in the literature [14, 5, 9].)

However, one crucial ingredient is missing: *we lack sufficient control on the convergence of the density of states*. The *density of states* of $X_{W;N}$ is the measure $\kappa_{W;N}(\lambda)d\lambda$ on the real line giving the density of the eigenvalue process:

$$(1.20) \quad \int_I \kappa_{W;N}(\lambda)d\lambda = \frac{1}{N} \mathbb{E}(\#\{\lambda_j \in I\}).$$

As indicated, $\kappa_{W;N}(\lambda)d\lambda$ is absolutely continuous. In fact, it follows from the Wegner estimate — (3.8) below — that

$$(1.21) \quad |\kappa_{W;N}(\lambda)| \lesssim W^\sigma,$$

so analogous to (1.18) we have

$$(1.22) \quad \frac{1}{N} \text{Prob} [\#\{\lambda_j \in I\} \geq 1] \leq \frac{1}{N} \mathbb{E}(\#\{\lambda_j \in I\}) = \int_I \kappa_{W;N}(\lambda)d\lambda \leq W^\sigma |I|.$$

(In fact, the Minami estimate is proved in a similar way, by showing that the *expected number of eigenvalue pairs* in I is bounded by the r.h.s. of (1.18).)

To study local fluctuations of the eigenvalue processes near $\lambda_0 \in \mathbb{R}$, it is natural to consider the re-centered and re-scaled process

$$(1.23) \quad \tilde{\lambda}_j = N(\lambda_j - \lambda_0),$$

which has mean spacing $O(1)$. We say that *the eigenvalue process has Poisson statistics near λ_0* , in some limit $W = W(N)$ and $N \rightarrow \infty$, if the point process $\{\tilde{\lambda}_j\}$ converges to a Poisson process. The density of this Poisson process would then be given by the limit $\lim_{N \rightarrow \infty} \kappa_{W(N);N}(\lambda_0)$. The difficulty is *we do not know that this limit exists*.

Now, for a fairly general class of matrix ensembles with independent centered entries, e.g., for the Gaussian ensemble (1.1), it is known that the density of states $\kappa_{W;N}$ converges *weakly* to the semi-circle law, provided $W(N)/N \rightarrow 0$ or 1 (see [19]). That is,

$$(1.24) \quad \frac{1}{N} \mathbb{E}(\text{tr } f(X_{W(N);N})) = \int_{\mathbb{R}} f(\lambda) \kappa_{W;N}(\lambda) d\lambda \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \int_{-2}^2 f(t) \sqrt{4-t^2} dt.$$

However, as indicated this is a weak convergence result, and it does not follow that

$$(1.25) \quad \kappa_{W(N);N}(\lambda) \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \sqrt{4-\lambda^2} I[|\lambda| \leq 2],$$

or even that

$$(1.26) \quad \int_{(\lambda_0 - \frac{a}{N}, \lambda_0 + \frac{b}{N})} \kappa_{W;N}(\lambda) d\lambda \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \sqrt{4-\lambda^2} I[|\lambda| \leq 2],$$

which would in fact be sufficient to control the density of the putative limit process.

In this regard, let us state a couple of open problems.

Problem 2. *Improve the estimate (1.21). In particular, does this bound hold with $\sigma = 0$? (The interpretation of $\kappa_{W;N}(\lambda)/N$ as the mean eigenvalue spacing and the convergence (1.24) suggests that κ should be bounded.)*

Problem 3. *Verify either (1.25) or (1.26).*

Acknowledgments. I would like to thank Tom Spencer and Michael Aizenman for many interesting discussions related to this and other works, and to express my gratitude for the hospitality extended me by the Institute for Advanced study, where I was member when this project started, and more recently by the Isaac Newton Institute during my stay associated with the program Mathematics and Physics of Anderson Localization: 50 Years After.

2. TRIDIAGONAL MATRICES

The aim of this section is to motivate the assumptions on the distribution of $X_{W;N}$, spelled out below in §3, by examining separately the somewhat simpler case $W = 2$. Thus, consider for each $N \in \mathbb{N}$, a random tridiagonal matrix

$$(2.1) \quad X_{2;N} = \begin{pmatrix} v_1 & t_1 & & & & & \\ t_1^* & v_2 & t_2 & & & & \\ & t_2^* & v_3 & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & v_{N-1} & t_{N-1} & \\ & & & & t_{N-1}^* & v_N & \end{pmatrix},$$

with v_1, v_2, \dots and t_1, t_2, \dots two given mutually independent sequences of independent random variables, real and complex valued respectively. For such matrices, exponential decay of the Green's function and localization of eigenfunctions can be obtained by the transfer matrix approach, see [6]. Here we use a different method, which is closely related to the technique of Kunz and Souillard [16].

As discussed above, the central technical estimate is a bound on $\mathbb{E}(|\langle \mathbf{e}_i, (X_{2;N} - \lambda)^{-1} \mathbf{e}_j \rangle|^s)$, decaying exponentially in the distance $|i - j|$. To obtain this bound, it is convenient to assume that (v_k) are identically distributed and likewise (t_k) . (This assumption could be replaced by uniformity in k of certain bounds assumed below. Likewise, strict independence of (v_k) is not really the issue. The argument could easily be adapted to the situation in which (v_k) are generated by a distribution with finite range coupling, such as $\prod_k \rho(v_k - v_{k-1}) dv_k$.) The distribution of (t_k) can be arbitrary — these variables may even be deterministic as in the case of random Jacobi matrices.

To facilitate the fluctuation argument proposed above we will suppose the common distribution of v_k has a density ρ with the following property:

Definition 1. We say that a probability density ρ on \mathbb{R} is *fluctuation regular* if there are constants $\epsilon, \delta > 0$ and measurable set $\Omega \subset \mathbb{R}$ with $\int_{\Omega} \rho(v) dv > 0$ such that

$$(2.2) \quad v \in \Omega \implies \frac{\rho(v_1)}{\rho(v_2)} \geq \delta \quad \text{for all } v_1, v_2 \in (v - \epsilon, v + \epsilon)$$

Remark. A sufficient condition for fluctuation regularity is that $\ln \rho$ is Lipschitz on some open interval. For example a uniform distribution $\rho(x) \propto \chi_{[a,b]}(x)$ is fluctuation regular. So are the Gaussian and Cauchy distributions. However, fluctuation regularity is quite a bit stronger than absolute continuity of the measure ρdx , since it implies the existence of an *open* set on which ρ is strictly positive.

Our goal in this section is to prove the following result:

Theorem 4. *Let $(v_k)_{k=1}^\infty$ and $(t_k)_{k=1}^\infty$ be two sequences of i.i.d. random variables, real and complex valued respectively. Suppose that the common distribution of v_k has a density ρ which is bounded and fluctuation regular. Then for $0 < s < 1$ and $\Lambda > 0$ there are $A_s < \infty$ and $\mu_{s,\Lambda} > 0$ such that for all $\lambda \in [-\Lambda, \Lambda]$,*

$$(2.3) \quad \mathbb{E}(|\langle \mathbf{e}_i, (X_{2;N} - \lambda)^{-1} \mathbf{e}_j \rangle|^s) \leq A_s e^{-\mu_{s,\Lambda}|i-j|}.$$

Remark. We restrict λ to a compact set to facilitate the fluctuation argument below. In fact, for large $|\lambda|$ the rate of exponential decay will improve, although the mechanism will be somewhat different. One could construct a proof in this context along the lines of [3]. Thus the Λ dependence of the mass of decay $\mu_{s,\Lambda}$ may be dropped.

Let $g_N(i, j; \lambda) = \langle \mathbf{e}_i, (X_{2;N} - \lambda)^{-1} \mathbf{e}_j \rangle$. Recall that the decay of $\mathbb{E}(|g_N(i, j; \lambda)|^s)$ was to be established in two steps, the first of those being Lemma W which gives finiteness of the fractional moments. A preliminary observation is that Lemma W holds for these tridiagonal matrices:

Lemma 2.1 (Lemma W for $X_{2;N}$). *Suppose that the distribution of v_k , $k = 1, \dots, N$ satisfies*

$$(2.4) \quad \text{Prob}(v_k \in [a, b]) \leq \frac{\kappa}{2\pi} |b - a|,$$

for any interval $[a, b]$, with κ a finite constant. Then

$$(2.5) \quad \text{Prob}(|g_N(i, j; \lambda)| > t | (v_k)_{k \neq i, j}, (t_k)) \leq \frac{\kappa}{t},$$

so, in particular,

$$(2.6) \quad \text{Prob}(|g_N(i, j; \lambda)| > t) \leq \frac{\kappa}{t}$$

and

$$(2.7) \quad \mathbb{E}(|g_N(i, j; \lambda)|^s) \leq \frac{\kappa^s}{1-s}$$

for $0 < s < 1$.

Remark. The l.h.s. of (2.5) is the *conditional probability* of the event $\{|g_N(i, j; \lambda)| > t\}$ at specified values of $(v_k)_{k \neq i, j}$ and (t_k) — that is the probability conditioned on the Σ algebra generated by these variables. Eq. (2.5) is a standard estimate from the fractional moment analysis of discrete random Schrödinger operators, see [1]. The main point of this result is that to bound $\mathbb{E}(|g_N(i, j; \lambda)|^s)$, it is sufficient to average over v_i and v_j .

The second part of the argument is to establish large fluctuations for $g_N(i, j; \lambda)$ — this is Lemma F above. In the present context we have

Lemma 2.2 (Lemma F for $X_{2;N}$). *Under the hypotheses of Theorem 4, for each $0 < r < s < 1$ and $\Lambda \in \mathbb{R}$ there is a constant $C_{r,s;\Lambda} < \infty$ such that*

$$(2.8) \quad \mathbb{E} (|g_N(i, j; \lambda)|^r) \leq \exp(-C_{r,s;\Lambda}|i - j|) \mathbb{E} (|g_N(i, j; \lambda)|^s)^{r/s},$$

for $\lambda \in [-\Lambda, \Lambda]$.

Remark. Together Lemmas 2.1 and 2.2 prove Theorem 4.

Proof. Let us fix λ for the moment and drop it from the notation: $g_N(i, j) = g_N(i, j; \lambda)$. Suppose without loss of generality that $i < j$. A preliminary observation is that

$$(2.9) \quad g_N(i, j) = -g_{j-1}(i, j-1)t_{j-1}g_N(j, j),$$

which may be established using the resolvent identity, writing $X_{2;N}$ as a perturbation of the corresponding matrix with t_{j-1} set equal to zero (which decouples into two distinct blocks). Iteration of this identity gives

$$(2.10) \quad g_N(i, j) = (-1)^{j-i} \left[\prod_{k=i}^{j-1} g_k(k, k)t_k \right] g_N(j, j).$$

Thus

$$(2.11) \quad \ln |g_N(i, j)| = \sum_{k=i}^{j-1} \ln |t_k| + \sum_{k=i}^{j-1} \ln |g_k(k, k)| + \ln |g_N(j, j)|,$$

suggesting that if either $\ln |t_k|$ or $\ln |g_k(k, k)|$ were to exhibit fluctuations of order one, then the variance of $\ln |g_N(i, j)|$ would be of order $|i - j|$ and Lemma 2.2 would follow. However, there are substantial correlations between the various terms, making it difficult to proceed directly along this line of argument.

To make a precise analysis, let us consider the random variables

$$(2.12) \quad \gamma_k = \frac{1}{g_k(k, k)},$$

which are related by a recursion relation

$$(2.13) \quad \gamma_k = v_k - \lambda - \frac{|t_{k-1}|^2}{\gamma_{k-1}}, \quad 2 \leq k \leq N,$$

with

$$(2.14) \quad \gamma_1 = v_1.$$

These identities may be established using the Schur-complement formula. In a similar way, the Schur-complement formula may be used to show that

$$(2.15) \quad \frac{1}{g_N(j, j)} = v_j - \lambda - \frac{|t_{j-1}|^2}{\gamma_{j-1}} - |t_j|^2 \widehat{G}_{j+1} = \gamma_j - |t_j|^2 \widehat{G}_{j+1}.$$

independent after conditioning. They are, however, no longer identically distributed. Instead,

$$(2.21) \quad \text{distribution of } \alpha_k \text{ given } (t_\ell)_{\ell=1}^N \text{ and } (f_\ell)_{\ell=1}^N \\ = \frac{\rho(e^{-\alpha_k} f_k + \lambda + \frac{|t_{k-1}|^2}{f_{k-1}}) \rho(f_{k+1} + \lambda + e^{\alpha_k} \frac{|t_k|^2}{f_k}) H(\alpha_k) e^{-\alpha_k}}{Z_k} d\alpha_k$$

with

$$(2.22) \quad Z_k = \int \rho(e^{-\alpha} f_k + \lambda + \frac{|t_{k-1}|^2}{f_{k-1}}) \rho(f_{k+1} + \lambda + e^{\alpha} \frac{|t_k|^2}{f_k}) H(\alpha) e^{-\alpha} d\alpha.$$

We now express $g_N(i, j)$ in terms of the variables $(t_\ell, f_\ell, \alpha_\ell)_{\ell=1}^N$,

$$(2.23) \quad g_N(i, j) = (-1)^{j-i} \left[\prod_{\substack{k \equiv 2 \pmod{3} \\ i \leq k \leq j-1}} e^{\alpha_k} \right] \left[\prod_{k=i}^{j-1} \frac{t_k}{f_k} \right] \widehat{H}_{j+1},$$

where

$$(2.24) \quad \widehat{H}_{j+1} = \frac{1}{e^{-\alpha_j} f_j - |t_j|^2 \widehat{G}_{j+1}}$$

is a function of $(t_\ell, f_\ell, \alpha_\ell)_{\ell=j}^N$. By the conditional independence of (α_k) we find that

$$(2.25) \quad \mathbb{E} (|g_N(i, j)|^r \mid (t_\ell, f_\ell)_{\ell=1}^N, (\alpha_\ell)_{\ell=j}^N) \\ = \left(\prod_{k=i}^{j-1} \frac{|t_k|^r}{|f_k|^r} \right) \widehat{H}_{j+1} \prod_{\substack{k \equiv 2 \pmod{3} \\ i \leq k \leq j-1}} \mathbb{E} (e^{r\alpha_k} \mid (t_\ell, f_\ell)_{\ell=1}^N).$$

Applying proposition 3 to each factor $\mathbb{E} (e^{r\alpha_k} \mid (t_\ell, f_\ell)_{\ell=1}^N)$ on the right hand side, we find that

$$(2.26) \quad \mathbb{E} (|g_N(i, j)|^r \mid (t_\ell, f_\ell)_{\ell=1}^N, (\alpha_\ell)_{\ell=j}^N) \\ = \left(\prod_{k=i}^{j-1} \frac{|t_k|^r}{|f_k|^r} \right) \widehat{H}_{j+1} \prod_{\substack{k \equiv 2 \pmod{3} \\ i \leq k \leq j-1}} e^{-h_k(r, s)} \mathbb{E} (e^{s\alpha_k} \mid (t_\ell, f_\ell)_{\ell=1}^N)^{r/s},$$

with

$$(2.27) \quad h_k(r, s) = \frac{1}{s} \int_0^s \min(r, q) (s - \max(r, q)) \text{Var}_q(\alpha_k \mid (t_\ell, f_\ell)_{\ell=1}^N) dq,$$

and

$$(2.28) \quad \text{Var}_q(\alpha_k \mid (t_\ell, f_\ell)_{\ell=1}^N) = \inf_{m \in \mathbb{R}} \frac{\mathbb{E} ((\alpha_k - m)^2 e^{q\alpha_k} \mid (t_\ell, f_\ell)_{\ell=1}^N)}{\mathbb{E} (e^{q\alpha_k} \mid (t_\ell, f_\ell)_{\ell=1}^N)}.$$

Using the conditional independence of (α_k) once again to reassemble $g_N(i, j)$ inside the expectation on the r.h.s., we find that

$$(2.29) \quad \mathbb{E} (|g_N(i, j)|^r \mid (t_\ell, f_\ell)_{\ell=1}^N, (\alpha_\ell)_{\ell=j}^N) \\ = e^{-\sum_{k=i}^{j-1} h_k(r, s)} \mathbb{E} (|g_N(i, j)|^s \mid (t_\ell, f_\ell)_{\ell=1}^N, (\alpha_\ell)_{\ell=j}^N)^{r/s},$$

where we have set $h_k(r, s) = 0$ for $k \not\equiv 2 \pmod{3}$.

After averaging and applying the Hölder inequality, we conclude that

$$(2.30) \quad \mathbb{E}(|g_N(i, j)|^r) \leq \mathbb{E}\left(e^{-\frac{s}{s-r} \sum_{k=i}^{j-1} h_k(r, s)}\right)^{\frac{s-r}{s}} \mathbb{E}(|g_N(i, j)|^s)^{r/s}.$$

Eq. (2.30) is the key result. The exponent in the first factor is a sum of $O(N)$ non-negative terms, each presumably $O(1)$ and positive with positive probability. It will not be so surprising to find that the term itself is $O(N)$ with good probability. The rest is estimates.

To proceed with the estimates, let us take the *a priori* distribution of α_k , before coupling and conditioning, to be uniform in an interval $[-\eta, \eta]$ centered at the origin:

$$(2.31) \quad H(\alpha) = \frac{1}{2\eta} I[|\alpha| < \eta],$$

with η to be chosen below. Although $\text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N)$ is defined as a function of $(t_\ell, f_\ell)_{\ell=1}^N$, it is useful to express it in terms of the variables $(t_\ell, \gamma_\ell, \alpha_\ell)_{\ell=1}^N$:

$$(2.32) \quad \text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N) = \inf_{m \in \mathbb{R}} \frac{\int_{-\eta}^{\eta} (\alpha - m)^2 e^{(q-1)\alpha} \nu_k(\alpha) d\alpha}{\int_{-\eta}^{\eta} e^{(q-1)\alpha} \nu_k(\alpha) d\alpha}$$

with

$$(2.33) \quad \nu_k(\alpha) = \rho(e^{\alpha_k - \alpha} \gamma_k + \lambda + \frac{|t_{k-1}|^2}{\gamma_{k-1}}) \rho(\gamma_{k+1} + \lambda + e^{\alpha - \alpha_k} \frac{|t_k|^2}{\gamma_k}).$$

A lower bound for $\text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N)$, sufficient for our purposes, is

$$(2.34) \quad \text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N) \geq \frac{1}{3} e^{-2|q-1|\eta} \eta^2 \frac{\inf_{-\eta < \alpha < \eta} \nu_k(\alpha)}{\sup_{-\eta < \alpha < \eta} \nu_k(\alpha)}.$$

The r.h.s. still carries some dependence on α_k , through the density ν_k . We may eliminate the dependence on α_k entirely by bounding the right hand side from below:

$$(2.35) \quad \begin{aligned} \text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N) &\geq \frac{1}{3} \eta^2 e^{-2|q-1|\eta} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(e^{-\alpha} \gamma_k + \lambda + \frac{|t_{k-1}|^2}{\gamma_{k-1}}) \rho(\gamma_{k+1} + \lambda + e^{\alpha} \frac{|t_k|^2}{\gamma_k})}{\rho(e^{-\beta} \gamma_k + \lambda + \frac{|t_{k-1}|^2}{\gamma_{k-1}}) \rho(\gamma_{k+1} + \lambda + e^{\beta} \frac{|t_k|^2}{\gamma_k})}. \end{aligned}$$

It is useful to write

$$e^{-\alpha} \gamma_k + \lambda + \frac{|t_{k-1}|^2}{\gamma_{k-1}} = (e^{-\alpha} - 1) \gamma_k + v_k,$$

and similarly for the term in the denominator and the term with index $k+1$. Finally, the r.h.s. is no larger if we factor the infimum on the right hand side,

$$(2.36) \quad \begin{aligned} \text{Var}_q(\alpha_k | (t_\ell, f_\ell)_{\ell=1}^N) &\geq \frac{1}{3} \eta^2 e^{-2|q-1|\eta} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(v_k + (e^{-\alpha} - 1) \gamma_k)}{\rho(v_k + (e^{-\beta} - 1) \gamma_k)} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(v_{k+1} + (e^{\alpha} - 1) \frac{|t_k|^2}{\gamma_k})}{\rho(v_{k+1} + (e^{\beta} - 1) \frac{|t_k|^2}{\gamma_k})}. \end{aligned}$$

On the r.h.s., the only dependence on q is in the exponential term. In the integral (2.27), there is not much loss in replacing this exponential by the (smaller) $e^{-2|s-1|\eta}$, so that

$$(2.37) \quad \frac{s}{s-r} h_k(r, s) \geq \frac{rs}{6} \eta^2 e^{-2|s-1|\eta} U_k(\eta),$$

with

$$(2.38) \quad U_k(\eta) = \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(v_k + (e^{-\alpha} - 1)\gamma_k)}{\rho(v_k + (e^{-\beta} - 1)\gamma_k)} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(v_{k+1} + (e^\alpha - 1)\frac{|t_k|^2}{\gamma_k})}{\rho(v_{k+1} + (e^\beta - 1)\frac{|t_k|^2}{\gamma_k})}.$$

Plugging this estimate into eq. (2.30), we obtain

$$(2.39) \quad \mathbb{E}(|g_N(i, j)|^s) \leq \mathbb{E}\left(e^{-\frac{rs}{6}\eta^2 e^{-2|s-1|\eta} \sum_{k=i}^{j-1} U_k(\eta)}\right)^{\frac{s-r}{s}} \mathbb{E}(|g_N(i, j)|^s)^{r/s}.$$

Since ρ is fluctuation regular, there are $\delta, \epsilon > 0$ and a set $\Omega \subset \mathbb{R}$ with

$$\int_{\Omega} \rho dx = q_0 > 0$$

such that $U_k(\eta) \geq \delta^2 I[A_k]$ where $I[A_k]$ is the indicator function of the event:

$$(2.40) \quad A_k = \left\{ v_k, v_{k+1} \in \Omega, \quad |\gamma_k| \leq \frac{\epsilon}{e^{2\eta} - 1}, \quad \text{and} \quad \frac{|t_k|^2}{|\gamma_k|} \leq \frac{\epsilon}{e^{2\eta} - 1} \right\}.$$

In turn, since $\gamma_k = v_k + \lambda + |t_{k-1}|^2/\gamma_{k-1}$ and $|\lambda| \leq \Lambda$ (by assumption), we see that

$$(2.41) \quad A_k \supset \{v_k \in \Omega, |v_k| \leq L\} \cap \{v_{k+1} \in \Omega\} \cap \{|t_{k-1}|, |t_k| \leq \tau\} \\ \cap \left\{ \frac{1}{|\gamma_{k-1}|} \leq \frac{1}{\tau^2} \left(\frac{\epsilon}{e^{2\eta} - 1} - L - \Lambda \right) \right\} \cap \left\{ \frac{1}{|\gamma_k|} \leq \frac{1}{\tau^2} \frac{\epsilon}{e^{2\eta} - 1} \right\},$$

with τ and L any positive numbers.

We estimate the probability of A_k from below by integrating eq. (2.41) over v_{k+1} , v_k , v_{k-1} , t_k , and t_{k-1} in that order. (The need to integrate over three consecutive v variables is the reason we introduced α_k only for $k \equiv 2 \pmod{3}$.) To begin,

$$(2.42) \quad \text{Prob}(v_{k+1} \in \Omega | (v_l)_{l \neq k}, (t_l)) = \int_{\Omega} \rho(v) dv = q_0.$$

Looking now at v_k , since $\gamma_k = v_k + \lambda + |t_{k-1}|^2/\gamma_{k-1}$, we see that

$$(2.43) \quad \{v_k \in \Omega, |v_k| \leq L\} \cap \left\{ \frac{1}{|\gamma_k|} \leq \frac{1}{\tau^2} \frac{\epsilon}{e^{2\eta} - 1} \right\} \\ = \{v_k \in \Omega\} \cap \{|v_k| \leq L\} \cap \left\{ v_k \notin \left[a - \tau^2 \frac{e^{2\eta} - 1}{\epsilon}, a + \tau^2 \frac{e^{2\eta} - 1}{\epsilon} \right] \right\}$$

with $a = \lambda + |t_{k-1}|^2/\gamma_{k-1}$. Since the density ρ is bounded, it follows that

$$(2.44) \quad \text{Prob} \left(v_k \in \Omega, |v_k| \leq L, \frac{1}{|\gamma_k|} \leq \frac{1}{\tau^2} \frac{\epsilon}{e^{2\eta} - 1} \middle| (v_l)_{l \neq k, k+1}, (t_l) \right) \\ \geq q_0 - \text{Prob}(|v_k| > L) - 2 \|\rho\|_{\infty} \tau^2 \frac{e^{2\eta} - 1}{\epsilon}.$$

Similarly

$$(2.45) \quad \text{Prob} \left(\frac{1}{|\gamma_{k-1}|} \leq \frac{1}{\tau^2} \left(\frac{\epsilon}{e^{2\eta} - 1} - L - \Lambda \right) \middle| (v_l)_{l \neq k-1, k, k+1}, (t_l) \right) \\ \geq 1 - 2 \|\rho\|_{\infty} \tau^2 \frac{1}{\frac{\epsilon}{e^{2\eta} - 1} - L - \Lambda}.$$

Combining these estimates with eq. (2.41) and integrating over the identically distributed variables t_k and t_{k-1} , we find

(2.46)

$$\begin{aligned} \text{Prob}(A_k | (v_l)_{l \neq k-1, k, k+1}, (t_l)_{l \neq k, k-1}) &\geq q_0 \left(q_0 - \text{Prob}(|v_k| > L) - 2 \|\rho\|_\infty \tau^2 \frac{e^{2\eta} - 1}{\epsilon} \right) \\ &\quad \times \left(1 - 2\rho_\infty \tau^2 \frac{1}{\frac{\epsilon}{e^{2\eta} - 1} - L - \Lambda} \right) \text{Prob}(|t_k| \leq \tau)^2. \end{aligned}$$

The key things to observe is that the r.h.s. of eq. (4.28) is independent of k and can be made arbitrarily close to q_0^2 by suitable choice of large L , τ and small η .

So, for sufficiently small η we have $\text{Prob}(A_k | (v_l)_{l \neq k-1, k, k+1}, (t_l)_{l \neq k, k-1}) \geq \frac{1}{2} q_0^2$, say. Since $U_k(\eta) \geq \delta^2 I[A_k]$, we find that

$$(2.47) \quad \mathbb{E} \left(e^{-\frac{rs}{6} \eta^2 e^{-|s-1|\eta} \sum_{k=i}^{j-1} U_k(\eta)} \right)^{\frac{s-r}{s}} \leq \exp \left(-\frac{s-r}{2s} q_0^2 \left(1 - e^{-\delta^2 \eta^2 \frac{rs}{6} e^{-2|s-1|\eta}} \right) \left\lfloor \frac{|i-j|}{3} \right\rfloor \right),$$

by integrating successively over v_k, t_k from $k = i, \dots, j-1$ (see Lemma A.1 below). Combined with (2.39) this completes the proof of Lemma 2.2. \square

3. BAND MATRICES

To translate the argument of the previous section to the context of band matrices, we replace each of the variables v_j and t_j by $W \times W$ matrices. Given $W \in \mathbb{N}$, consider a sequence, $V_j, n = 1, \dots$, of independent identically distributed hermitian $W \times W$ matrices together with a sequence, $T_j, n = 1, \dots$, of independent identically distributed $W \times W$ matrices (not necessarily hermitian). With these matrix variables, we form an infinite random hermitian band matrix

$$(3.1) \quad X_W = \begin{pmatrix} V_1 & T_1 & 0 & & \\ T_1^\dagger & V_2 & T_2 & \ddots & \\ 0 & T_2^\dagger & V_3 & & \ddots \\ & \ddots & & \ddots & \\ & & \ddots & & \ddots \end{pmatrix},$$

a random operator on $\ell^2(\mathbb{N})$, and for each N the random matrix

$$(3.2) \quad X_{W;N} = Q_N X_W Q_N$$

with Q_N the projection onto $\ell^2(\{1, \dots, N\})$. For simplicity, let us consider only N a multiple of W : $N = nW$. Thus,

$$(3.3) \quad X_{W;N} = X_{W;nW} = \begin{pmatrix} V_1 & T_1 & 0 & & & \\ T_1^\dagger & V_2 & T_2 & \ddots & & \\ 0 & T_2^\dagger & V_3 & & \ddots & \\ & \ddots & & \ddots & & 0 \\ & & \ddots & & \ddots & T_n \\ & & & 0 & T_{n-1}^\dagger & V_n \end{pmatrix},$$

Let P_j denote the projection onto the j^{th} block, $\ell^2(\{(j-1)W+1, \dots, jW\})$, so

$$V_j = P_j X_W P_j \quad \text{and} \quad T_j = P_j X_W P_{j+1}.$$

Band matrix ensembles such as the Gaussian band ensemble (1.1) are of this form, with T_j lower triangular matrices. However, for the argument presented below it is not necessary that T_j be lower triangular. (Also, neither strict independence nor identicality of distribution are needed. Nonetheless, to keep things simple, let us stick to the i.i.d. case.)

In adapting the arguments from the scalar case to the matrix variables V_j, T_j , we must account for the non-commutativity of the matrix product. The basis of the argument is a change of variables $V_j \mapsto e^{\alpha_j} \Gamma_j$ with α_j a scalar random variable and Γ_j a $W \times W$ matrix obtained from the resolvent of $X_{W;jW}$. In the end we will need to estimate the ratio

$$\frac{\rho(V_j + (e^{-\alpha} - 1)\Gamma_j)}{\rho(V_j + (e^{-\beta} - 1)\Gamma_j)}$$

for small α, β , where ρ is the density of the distribution of V_j (assumed to be absolutely continuous with respect to Lebesgue measure on some vector space of matrices). In the scalar case, this change of variables was useful for all fluctuation regular densities. In the matrix case, an additional complication arises. Unless Γ_j falls in the vector space supporting the distribution of V_j there will be constraints on the matrix elements of Γ_j which manifest themselves as δ functions after the change of variables. However, Γ_j is formed from $\{V_k\}$ and $\{T_k\}$ via *non-linear* operations, so *there is no reason to expect it to fall in this vector space*. (For example when V_j are diagonal, Γ_j will in general have off-diagonal components.) To guarantee closure under non-linear operations we suppose that the vector space supporting the distribution of V_j is a *matrix algebra*:

Definition 2. A \star algebra over \mathbb{R} of $W \times W$ matrices is a set \mathcal{A} of $W \times W$ matrices that is a vector space over \mathbb{R} , under the usual addition and scalar multiplication, and such that

$$V_1, V_2 \in \mathcal{A} \implies V_1 V_2 \in \mathcal{A} \quad \text{and} \quad V_1^\dagger \in \mathcal{A}.$$

We will use

Proposition 5. *If \mathcal{A} is a matrix \star algebra over \mathbb{R} and $V \in \mathcal{A}$ is invertible then $V^{-1} \in \mathcal{A}$.*

Proof. This is a standard result for C^\star algebras. In that context, the algebra is usually assumed to be a vector space over \mathbb{C} , but that is not necessary. Here is the proof. If $V \in \mathcal{A}$ is self-adjoint and invertible, by the Weierstrass theorem we can approximate V^{-1} (in the operator norm, say) by polynomials in V with real coefficients. That is, we can approximate V^{-1} by elements of \mathcal{A} . Since a finite dimensional vector space is complete, $V^{-1} \in \mathcal{A}$. For general invertible $V \in \mathcal{A}$, we have $V^{-1} = (V^\dagger V)^{-1} V^\dagger \in \mathcal{A}$, since $V^\dagger V \in \mathcal{A}$ is self adjoint. \square

Assumption 1. Let \mathcal{S} be an increasing sequence of integers and fix, for each $W \in \mathcal{S}$, a \star algebra over \mathbb{R} of $W \times W$ matrices \mathcal{A}_W , and the set \mathcal{T}_W of matrices which preserve \mathcal{A}_W under conjugations

$$(3.4) \quad \mathcal{T}_W = \{T : T^\dagger \mathcal{A}_W T \subset \mathcal{A}_W\}.$$

Let $\mathcal{A}_W^H = \{V \in \mathcal{A}_W : V = V^\dagger\}$, the set of hermitian elements of \mathcal{A}_W . We require that $T_j \in \mathcal{T}_W$ and $V_j \in \mathcal{A}_W^H$, $j = 1, \dots$

Remark. Note that \mathcal{T}_W is closed under conjugation: $T \in \mathcal{T}_W \implies T^\dagger \in \mathcal{T}_W$.

There is a good deal of flexibility in the choice of algebras. Of course, we may take $\mathcal{A}_W = \mathcal{T}_W =$ all $n \times n$ complex matrices, so $X_{W;N}$ is complex Hermitian. On the other hand, we could restrict \mathcal{A}_W to be the set of matrices with real entries, so $X_{W;N}$ is real symmetric. In this case \mathcal{A}_W is *not* a complex vector space. Similarly we could take \mathcal{A}_W to be the set of matrices with quaternion entries, where the quaternions units are represented by 2×2 matrices, so $X_{W;N}$ would be Hermitian but anti-symmetric under transposition $X_{W;N}^T = -X_{W;N}$. In this last case, \mathcal{S} would be the set of even integers.

An important consequence of assuming that $T_j \in \mathcal{T}_W$ and $V_j \in \mathcal{A}_W^H$, is that we have some *a priori* information on the block matrices making up the resolvent of $X_{W;nW}$.

Lemma 3.1. *Suppose Y is an $nW \times nW$ matrix that is block tri-diagonal,*

$$P_i Y P_j = 0 \quad \text{if } |i - j| \geq 2,$$

and satisfies

$$V_j = P_j Y P_j \in \mathcal{A}_W, \quad j = 1, \dots, n$$

and

$$T_j = P_j Y P_{j+1} = (P_{j+1} Y P_j)^\dagger \in \mathcal{T}_W, \quad j = 1, \dots, n-1.$$

If Y is invertible then

$$(3.5) \quad P_j Y^{-1} P_j \in \mathcal{A}_W, \quad j = 1, \dots, n$$

and

$$(3.6) \quad P_i Y^{-1} P_j \in \mathcal{AT}_W, \quad i, j = 1, \dots, n,$$

where \mathcal{AT}_W is the algebra generated by \mathcal{A}_W and \mathcal{T}_W .

Remark. The off diagonal blocks of Y^{-1} need not be in \mathcal{A}_W . This is apparent already for $n = 2$, where, by the Schur complement formula,

$$\begin{aligned} P_1 Y^{-1} P_2 &= (V_1 - T_1 V_2^{-1} T_1^\dagger)^{-1} T_1 V_2^{-1} \\ &= V_1^{-1} T_1 (V_2 - T_1^\dagger V_1^{-1} T_1)^{-1}. \end{aligned}$$

In each expression on the right, the first and last factors are in \mathcal{A}_W but the middle factor, T_1 , is not.

Proof. The proof is by induction on n . The result is clear for $n = 1$. So, suppose we know that it holds if Y is a tridiagonal block matrix of size no larger than $(n-1)W \times (n-1)W$.

First consider (3.5). By the Schur complement formula,

$$P_j Y^{-1} P_j = (V_j - T_j P_{j+1} Y_+^{-1} P_{j+1} T_j^\dagger - T_{j-1}^\dagger P_{j-1} Y_-^{-1} P_{j-1} T_{j-1})^{-1},$$

where

$$Y_+ = \begin{pmatrix} V_{j+1} & T_{j+1} & & \\ T_{j+1}^\dagger & \ddots & \ddots & \\ & \ddots & \ddots & T_{n-1} \\ & & T_{n-1}^\dagger & V_n \end{pmatrix}, \quad Y_- = \begin{pmatrix} V_1 & T_1 & & \\ T_1^\dagger & \ddots & \ddots & \\ & \ddots & \ddots & T_{j-1} \\ & & T_{j-1}^\dagger & V_j \end{pmatrix}.$$

As Y_+ and Y_- are of size no larger than $(n-1)W \times (n-1)W$ and $T_j, T_{j+1} \in \mathcal{T}_W$, it follows that

$$T_j P_{j+1} Y_+^{-1} P_{j+1} T_j^\dagger, \quad T_{j-1}^\dagger P_{j-1} Y_-^{-1} P_{j-1} T_{j-1} \in \mathcal{A}_W.$$

By Prop. 5 $P_j Y^{-1} P_j \in \mathcal{A}_W$.

Now consider (3.6). Suppose $i < j$ (the other case is similar). Let

$$\widehat{Y} = \begin{pmatrix} V_1 & T_1 & & & \\ T_1^\dagger & \ddots & \ddots & & \\ & \ddots & \ddots & T_{j-2} & \\ & & T_{j-2}^\dagger & & V_{j-1} \end{pmatrix}.$$

By the resolvent identity, one has

$$P_i Y^{-1} P_j = -P_i \widehat{Y}^{-1} P_{j-1} T_j P_j Y^{-1} P_j.$$

But $P_i \widehat{Y}^{-1} P_{j-1} \in \mathcal{AT}_W$ by the induction hypothesis and $P_j Y^{-1} P_j \in \mathcal{A}_W$ as we have just shown. It follows that the r.h.s. is in \mathcal{AT}_W . \square

We now consider the properties required of the distribution of V_j , denoted \mathbb{P}_W . Let $\|\cdot\|$ denote the operator norm of a matrix

$$(3.7) \quad \|A\| = \sup_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

and let $\sigma(A)$ denote the set of eigenvalues of a matrix. Recall, if A is self-adjoint, that

$$\|A\| = \max |\sigma(A)| \quad \|A^{-1}\| = \frac{1}{\min |\sigma(A)|}.$$

Assumption 2. Let $(\mathbb{P}_W)_{W \in \mathcal{S}}$ be a family of probability measures such that

- (*Absolute continuity*): Each measure \mathbb{P}_W is supported on \mathcal{A}_W^H and absolutely continuous with respect to Lebesgue measure on that space. Let $\rho_W(V)$ denote the density of \mathbb{P}_W with respect to Lebesgue measure.
- (*Wegner-type estimates*): There are $\kappa > 0$ and $\sigma \geq 0$ such that for all $A \in \mathcal{A}_W^H$, $W \in \mathcal{S}$,

$$(3.8) \quad \mathbb{P}_W \{V : \|(V - A)^{-1}\| > tW^{1+\sigma}\} \leq \kappa \frac{1}{t};$$

and for all $A, B \in \mathcal{A}_W^H$ and $C \in \mathcal{AT}_W$, $W \in \mathcal{S}$,

$$(3.9) \quad \mathbb{P}_W \otimes \mathbb{P}_W \left\{ (V_1, V_2) : \left\| \begin{pmatrix} V_1 - A & C \\ C^\dagger & V_2 - B \end{pmatrix}^{-1} \right\| > tW^{1+\sigma} \right\} \leq 2\kappa \frac{1}{t}.$$

- (*Fluctuation regularity with bounded tails*): There are constants $p_0, \delta, \epsilon > 0$, $L, a, \zeta \geq 0$ such that, for each $W \in \mathcal{S}$, there is $\Omega_W \subset \mathcal{A}_W^H$ with $\mathbb{P}_W(\Omega_W) \geq p_0$ and if $V \in \Omega_W$, then

$$(3.10) \quad \|V\| \leq LW^a$$

and

$$(3.11) \quad \frac{\rho_W(V_1)}{\rho_W(V_2)} \geq \delta \text{ for all } V_1, V_2 \in \mathcal{A}_W^h \text{ with } \|V_j - V\| \leq \epsilon W^{-\zeta}, \quad j = 1, 2.$$

Remarks.

(1) Since

$$\|(V - \lambda I)^{-1}\| = \frac{1}{\text{dist}(\lambda, \sigma(V))},$$

the Wegner-type estimate (3.8) implies

$$(3.12) \quad \mathbb{P}_W \left\{ V : \text{dist}(\lambda, \sigma(V)) \leq \frac{\epsilon}{W^{1+\sigma}} \right\} \leq \kappa \epsilon.$$

If V is suitably scaled so as to have mean eigenvalue spacing of order $1/W$, this suggests that we should be able to take $\sigma = 0$. That has not been proved, however, for the random matrix ensembles studied here. For Wigner type matrices, in particular for the Gaussian band ensemble (1.1), we will obtain the estimates (3.8, 3.9) with $\sigma = \frac{1}{2}$ in §5.

- (2) The parameters σ and a are not independent. If we rescale via $V \mapsto W^\gamma V$ this results in a shift $\sigma \mapsto \sigma - \gamma$ and $a \mapsto a + \gamma$. Nonetheless it is convenient to keep both parameters since the *natural* scaling of V is to choose the eigenvalue spacing to be of order $1/W$. This typically leads to $a = 0$, but if the entries of V have heavy tails then one may have $a > 0$.

We require very little from the distribution of T_j , denoted \mathbb{Q}_W , essentially just a uniform (in W) bound on the tails:

Assumption 3. Let $(\mathbb{Q}_W)_{W=2}^\infty$ be a family of probability measures, with \mathbb{Q}_W supported on \mathcal{T}_W . Suppose that there are $q_0, \tau > 0$ and $b \geq 0$ such that

$$(3.13) \quad \mathbb{Q}_W \{ T : \|T\| \leq \tau W^b \} \geq q_0.$$

Remark. \mathbb{Q}_W could be supported on a single point, in which case T_j would be a constant sequence. For instance, we could take $T_j = I$.

Lemma W for $X_{W;nW}$ follows easily from part (2) of assumption 1.

Lemma 3.2 (Lemma W for $X_{W;nW}$). *Let $V_j, T_j, j = 1, \dots$, be mutually independent sequences of independent random $W \times W$ matrices. Suppose each V_j has distribution \mathbb{P}_W and each T_j has distribution \mathbb{Q}_W . Then, for each $\lambda \in \mathbb{R}$,*

$$\text{Prob}[\lambda \text{ is an eigenvalue of } X_{W;nW}] = 0$$

and

$$(3.14) \quad \text{Prob} \left(\|P_i(X_{W;nW} - \lambda I)^{-1} P_j\| > t W^{1+\sigma} \mid \{T_k\}_{k=1}^{n-1} \text{ and } \{V_k\}_{k \neq i,j} \right) \leq 2\kappa \frac{1}{t}$$

for any $1 \leq i, j \leq n$.

Proof. Let us first consider the case $i = j$. The Schur complement formula shows that

$$P_i(X_{W;nW} - \lambda I)^{-1} P_i = (V_i - \lambda I + K)^{-1}$$

with

$$K = T_{i-1}^\dagger (X_- - \lambda I)^{-1} T_{i-1} + T_i (X_+ - \lambda I)^{-1} T_i^\dagger$$

with X_- and X_+ the restrictions of X to the blocks above and below i . By Lemma 3.1 $K \in \mathcal{A}_W^H$. (Note that it is self adjoint.) It follows from (3.8) that λ is an eigenvalue of $X_{W;nW}$ with probability 0 and that (3.14) holds for $i = j$.

The argument for $i \neq j$ is similar. In this case, we first estimate

$$\|P_i(X_{W;nW} - \lambda I)^{-1} P_j\| \leq \|(P_i + P_j)(X_{W;nW} - \lambda I)^{-1}(P_i + P_j)\|.$$

As above, we have

$$(P_i + P_j)(X_{W;nW} - \lambda I)^{-1}(P_i + P_j) = \left[\begin{pmatrix} V_i - \lambda I & 0 \\ 0 & V_j - \lambda I \end{pmatrix} + \begin{pmatrix} A & C \\ C^\dagger & B \end{pmatrix} \right],$$

where A, B and C are formed from blocks of the resolvents of restrictions of $X_{W;nW}$. One may verify that $A, B \in \mathcal{A}_W^H$ and $C \in \mathcal{AT}_W$. Thus the result follows from (3.9). \square

It follows that

$$(3.15) \quad \mathbb{E} \left(\|P_i(X_{W;nW} - \lambda I)^{-1} P_j\|^s \right) \leq \frac{2^s \kappa^s}{1-s} W^{(1+\sigma)s},$$

and so

$$(3.16) \quad \mathbb{E} \left(|\langle \mathbf{v}, P_i(X_{W;nW} - \lambda I)^{-1} P_j \mathbf{w} \rangle|^s \right) \leq \frac{2^s \kappa^s}{1-s} W^{(1+\sigma)s},$$

for any two vectors \mathbf{v}, \mathbf{w} . (See (1.8).)

Lemma F in this context is as follows:

Lemma 3.3 (Lemma F for $X_{W;nW}$). *Let $V_j, T_j, j = 1, \dots$, be mutually independent sequences of independent random $W \times W$ matrices. Suppose each V_j has distribution \mathbb{P}_W and each T_j has distribution \mathbb{Q}_W . Let D_W denote the real dimension of \mathcal{A}_W^H . Fix a positive number ν large enough that $\sup_W D_W W^{-\nu} < \infty$ and suppose also that $\nu \geq \zeta + \max(a, 1 + \sigma + 2b)$, with σ, a, ζ as in assumption 2 and b as in assumption 3. Then for each $0 < r < s < 1$ and $0 < \Lambda < \infty$ there is $C_{r,s} > 0$ such that if $|i - j| \geq 3$ then*

$$(3.17) \quad \mathbb{E} \left([\Phi(P_i(X_{W;nW} - \lambda I^{-1} P_j))]^r \right) \leq \exp(-C_{r,s} W^{-2\nu} |i - j|) \mathbb{E} \left([\Phi(P_i(X_{W;nW} - \lambda I)^{-1} P_j)]^s \right)^{r/s}$$

for any $\lambda \in [-\Lambda, \Lambda]$ and any non-negative, positive-homogeneous function $\Phi : \mathcal{AT}_W \rightarrow \mathbb{R}$ — i.e., $\Phi(Y) \geq 0$ and $\Phi(\alpha Y) = \alpha \Phi(Y)$ for $\alpha \geq 0$.

Remarks.

- (1) Below we will apply the result with $\Phi(Y)$ a semi-norm such as the the absolute value of a matrix element $|\langle \mathbf{v}, Y \mathbf{w} \rangle|$ or the norm $\|Y\|$. However the proof does not make use of the triangle inequality, so the result also applies, for example, to $\Phi(Y) =$ spectral radius (Y) or $\Phi(Y) =$ smallest singular value of Φ .
- (2) Under rescaling of the matrix elements $X_{W;nW} \mapsto W^\gamma X_{W;nW}$ the localization length $1/C_{r,s} W^{-2\nu}$ should not change. That this is indeed so follows since $\zeta \mapsto \zeta - \gamma$, $a \mapsto a + \gamma$, $\sigma \mapsto -\gamma$ and $b \mapsto b + \gamma$, so the combination $\zeta + \max(a, 1 + \sigma + 2b)$ is invariant under rescaling.

Combining Lemma 3.3 and (3.16) we have

Theorem 6. *Let $V_j, T_j, j = 1, \dots$, be mutually independent sequences of independent random $W \times W$ matrices. Suppose each V_j has distribution \mathbb{P}_W and each T_j has distribution \mathbb{Q}_W . Let D_W denote the real dimension of \mathcal{A}_W^H . Fix a positive number ν large enough that $\sup_W D_W W^{-\nu} < \infty$ and suppose also that $\nu \geq \zeta + \max(a, 1 + \sigma + 2b)$, with σ, a, ζ as in assumption 2 and b as in assumption 3. For $0 < t < 1$ let*

$$(3.18) \quad M(W, t) = \max_{1 \leq x, y \leq nW} \mathbb{E} \left(|\langle \mathbf{e}_x, (X_{W;nW} - \lambda)^{-1} \mathbf{e}_y \rangle|^t \right),$$

where \mathbf{e}_x and \mathbf{e}_y denote elementary basis vectors. Then

$$(3.19) \quad M(W, t) \leq \frac{2^t \kappa^t}{1-t} W^{(1+\sigma)t}$$

and given $0 < s < t$ there are constants C, μ such that for any $1 \leq x, y \leq nW$

$$(3.20) \quad \mathbb{E} \left(|\langle \mathbf{e}_x, (X_{W;nW} - \lambda)^{-1} \mathbf{e}_y \rangle|^s \right) \leq C M(W, t)^{s/t} e^{-\mu W^{-2\nu-1} |x-y|}.$$

Proof. This amounts to special cases of (3.16) and Lemma 3.3. The exponent $2\nu + 1$ appears in (3.20) because the difference $|i - j|$ of the blocks to which x and y belong is estimated by $|x - y|/W$. The constant C compensates for the exponential factor $e^{-\mu W^{-\nu-1}|x-y|}$ when $|x - y|$ is smaller than $3W$, in which case the estimate of Lemma 3.3 does not hold. \square

Remark. Putting (3.20) and (3.19) together we have

$$(3.21) \quad \mathbb{E}(|\langle \mathbf{e}_x, (X_{W;N} - \lambda)^{-1} \mathbf{e}_y \rangle|^s) \leq \text{const.} W^{(1+\sigma)s} e^{-\mu W^{-\nu-1}|x-y|}.$$

If the diagonal blocks V_j are Wigner matrices, as in assumption 4 in §5 below, one may obtain the estimate

$$(3.22) \quad M(W, t) \leq \frac{2^t \kappa^t}{1-t} W^{\frac{1}{2}},$$

resulting in a very slight improvement on the estimate on the r.h.s. of (3.21),

$$(3.23) \quad \mathbb{E}(|\langle \mathbf{e}_x, (X_{W;N} - \lambda)^{-1} \mathbf{e}_y \rangle|^s) \leq \text{const.} W^{\frac{s}{2}} e^{-\mu W^{-\nu-1}|x-y|}.$$

This improvement is not very significant, as the main point here is the exponential factor, which dominates any power of W as long as $|x - y| \gg W^{2\nu+1}$.

4. FLUCTUATIONS

We now prove Lemma 3.3. Following the proof of Lemma 2.2, let us fix λ and set

$$G_n(i, j) = P_i(X_{W;nW} - \lambda I)^{-1} P_j.$$

Since $G_n(i, j)^\dagger = G_n(j, i)$, in estimating $\|G_n(i, j)\|$ we may assume without loss that $i \leq j$. We have, by the resolvent identity,

$$(4.1) \quad G_n(i, j) = -G_{j-1}(i, j-1) T_{j-1} G_n(j, j).$$

Iteration gives

$$(4.2) \quad G_n(i, j) = (-1)^{j-i} G_i(i, i) T_i G_{i+1}(i+1, i+1) T_{i+1} \cdots G_{j-1}(j-1, j-1) T_{j-1} G_n(j, j).$$

Let us define $W \times W$ random matrices

$$(4.3) \quad \Gamma_k = G_k(k, k)^{-1},$$

related by a recursion relation

$$(4.4) \quad \Gamma_k = V_k - \lambda I - T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}.$$

As in the $W = 2$ case, these identities may be established using the Schur-complement formula — compare with (2.9) and (2.13). Similarly,

$$(4.5) \quad G_n(j, j)^{-1} = V_j - \lambda I - T_{j-1}^\dagger \Gamma_{j-1}^{-1} T_{j-1} - T_j^\dagger \widehat{G}_{j+1} T_j = \Gamma_j - T_j \widehat{G}_{j+1} T_j^\dagger$$

where $\widehat{G}_{j+1} = P_{j+1}(\widehat{X}_{W;nW} - \lambda)^{-1} P_{j+1}$ with $\widehat{X}_{W;nW}$ the matrix obtained from $X_{W;nW}$ by setting $T_j = 0$. Thus \widehat{G}_{j+1} is a function of the matrix variables $(V_k)_{k=j+1}^N$ and $(T_k)_{k=j+1}^N$.

We now make the change of variables $V_k \mapsto \Gamma_k$ in our probability space. By Lem. 3.1 and Prop. 5, $\Gamma_k \in \mathcal{A}_W^H$. As in the tri-diagonal case, the Jacobian determinant is 1, so

$$(4.6) \quad \begin{aligned} & \text{Joint distribution of } (\Gamma_k)_{k=1}^n \text{ given } (T_k)_{k=1}^{n-1} \\ &= \rho(\Gamma_1 + \lambda I) \prod_{k=2}^n \rho(\Gamma_k + \lambda I + T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}) \prod_{k=1}^n d\Gamma_k, \end{aligned}$$

where $d\Gamma_k$ denotes Lebesgue measure on \mathcal{A}_W^H . In terms of the matrices Γ_k , we have

$$(4.7) \quad G_n(i, j) = (-1)^{|i-j|} \Gamma_i^{-1} T_i \Gamma_{i+1}^{-1} T_{i+1} \cdots \Gamma_{j-1}^{-1} T_{j-1} \cdot (\Gamma_j - T_j \widehat{G}_{j+1} T_j^\dagger)^{-1},$$

where \widehat{G}_{j+1} is a function of $(\Gamma_k)_{k=j}^n$ and $(T_k)_{k=j}^n$ (since $V_k = \Gamma_k + \lambda I - T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}$).

The matrix product in (4.7) is non-commutative, so it is not clear if the heuristic analysis that the ‘‘log of G is a sum of terms with only local correlations’’ is valid. Nonetheless, we may use the trick employed above of coupling the system to a family of independent identically distributed scalar variables $\alpha_2, \alpha_5, \dots$, each with absolutely continuous distribution

$$(4.8) \quad H(\alpha_k) d\alpha_k = \frac{1}{2\eta} I[|\alpha_k| \leq \eta] d\alpha_k,$$

with $\eta > 0$ to be chosen below. We define

$$(4.9) \quad F_k = e^{\alpha_k} \Gamma_k,$$

where we take $\alpha_k = 0$ for $k \not\equiv 2 \pmod{3}$. The Jacobian of the transformation $(\Gamma_k, \alpha_k) \mapsto (F_k, \alpha_k)$ is $\prod_{k=2,5,8,\dots}^N e^{-D_W \alpha_k}$, where $D_W = \dim \mathcal{A}_W^H$ is the dimension of \mathcal{A}_W^H . Thus

$$(4.10) \quad \text{joint distribution of } (F_k)_{k=1}^n \text{ and } (\alpha_k)_{k=1}^N, \text{ given } (T_k)_{k=1}^{n-1} = \\ \prod_{k \equiv 2 \pmod{3}}^n \rho(F_{k-1} + \lambda I + T_{k-2}^\dagger F_{k-2}^{-1} T_{k-2}) \rho(e^{-\alpha_k} F_k + \lambda I + T_{k-1}^\dagger F_{k-1}^{-1} T_{k-1}) \\ \times \rho(F_{k+1} + \lambda I + e^{\alpha_k} T_k^\dagger F_k^{-1} T_k) H(\alpha_k) e^{-D_W \alpha_k} dF_{k-1} dF_k dF_{k+1} d\alpha_k,$$

with the convention that $T_0 = 0$.

As in the tri-diagonal case, the variables α_k remain independent after conditioning on $(F_k)_{k=1}^N$. Also, the

$$(4.11) \quad \text{distribution of } \alpha_k \text{ given } (T_\ell)_{\ell=1}^{n-1} \text{ and } (F_\ell)_{\ell=1}^n = \\ \frac{\rho(e^{-\alpha_k} F_k + \lambda I + T_{k-1}^\dagger F_{k-1}^{-1} T_{k-1}) \rho(F_{k+1} + \lambda I + e^{\alpha_k} T_k^\dagger F_k^{-1} T_k) H(\alpha_k) e^{-D_W \alpha_k}}{Z_k} d\alpha_k$$

with

$$(4.12) \quad Z_k = \frac{1}{2\eta} \int_{-\eta}^{\eta} \rho(e^{-\alpha} F_k + \lambda I + T_{k-1}^\dagger F_{k-1}^{-1} T_{k-1}) \rho(F_{k+1} + \lambda I + e^{\alpha} T_k^\dagger F_k^{-1} T_k) e^{-D_W \alpha} d\alpha.$$

Now fix a non-negative positive homogeneous Φ as in the statement of the Lemma. Replacing Γ_j in (4.7) by $e^{-\alpha_j} F_j$, we find that

$$(4.13) \quad \Phi(G_n(i, j)) = \prod_{\substack{k \equiv 2 \pmod{3} \\ i \leq k \leq j-1}} e^{\alpha_k} \Phi \left((-1)^{|i-j|} \left(\prod_{k=i}^{j-1} F_k^{-1} T_k \right) \widehat{H}_{j+1} \right),$$

where

$$(4.14) \quad \widehat{H}_{j+1} = \frac{1}{\Gamma_j - T_j \widehat{G}_{j+1} T_j^\dagger}$$

is a function of $(T_k, F_k, \alpha_k)_{k=j}^N$. Since (α_k) are conditionally independent, it follows that

$$(4.15) \quad \mathbb{E}([\Phi(G_n(i, j))]^r | (T_\ell, F_\ell)_{\ell=1}^N, (\alpha_\ell)_{k=j}^N) \\ = \left[\Phi \left((-1)^{|i-j|} \left(\prod_{k=i}^{j-1} F_k^{-1} T_k \right) \widehat{H}_{j+1} \right) \right]^r \prod_{\substack{k \equiv 2 \pmod{3} \\ i \leq k \leq j-1}} \mathbb{E}(e^{r\alpha_k} | (T_k, F_k)_{k=1}^N).$$

By proposition 3 and the Hölder inequality, we conclude that (compare with (2.30)):

$$(4.16) \quad \mathbb{E}([\Phi(G_n(i, j))]^r) \leq \mathbb{E} \left(e^{-\frac{s}{s-r} \sum_{k=i}^{j-1} h_k(r, s)} \right)^{\frac{s-r}{s}} \mathbb{E}([\Phi(G_n(i, j))]^s)^{r/s},$$

where for $k \equiv 2 \pmod{3}$

$$(4.17) \quad h_k(r, s) = \frac{1}{s} \int_0^s \min(r, q)(s - \max(r, q)) \text{Var}_q(\alpha_k | (T_\ell, F_\ell)_{\ell=1}^N) dq,$$

with Var_q as in (2.28), and we have set $h_k(r, s) = 0$ for $k \not\equiv 2 \pmod{3}$.

Let us express $\text{Var}_q(\alpha_k | (T_\ell, F_\ell)_{\ell=1}^N)$ in terms of $(T_\ell, \Gamma_\ell, \alpha_\ell)_{\ell=1}^N$:

$$(4.18) \quad \text{Var}_q(\alpha_k | (T_\ell, F_\ell)_{\ell=1}^N) = \inf_{m \in \mathbb{R}} \frac{\int_{-\eta}^{\eta} (\alpha - m)^2 e^{(q-D_W)\alpha} \nu_k(\alpha) d\alpha}{\int_{-\eta}^{\eta} e^{(q-D_W)\alpha} \nu_k(\alpha) d\alpha}$$

with

$$(4.19) \quad \nu_k(\alpha) = \rho(e^{\alpha_k - \alpha} \Gamma_k + \lambda I + T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}) \rho(\Gamma_{k+1} + \lambda I + e^{\alpha - \alpha_k} T_k^\dagger \Gamma_k^{-1} T_k).$$

Thus (compare with (2.36)),

$$(4.20) \quad \text{Var}_q(\alpha_k | (T_\ell, F_\ell)_{\ell=1}^N) \geq \frac{1}{3} \eta^2 e^{-2q\eta} e^{-2D_W\eta} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(V_k + (e^{-\alpha} - 1)\Gamma_k)}{\rho(V_k + (e^{-\beta} - 1)\Gamma_k)} \\ \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(V_{k+1} + (e^\alpha - 1)T_k^\dagger \Gamma_k^{-1} T_k)}{\rho(V_{k+1} + (e^\beta - 1)T_k^\dagger \Gamma_k^{-1} T_k)}.$$

With (4.16) this implies

$$(4.21) \quad \mathbb{E}([\Phi(G_n(i, j))]^r) \leq \mathbb{E} \left(e^{-\frac{r}{6}\eta^2 e^{-2s\eta} e^{-2D_W\eta} \sum_{k=i}^{j-1} U_k(\eta)} \right)^{\frac{s-r}{s}} \mathbb{E}([\Phi(G_n(i, j))]^s)^{r/s},$$

where

$$(4.22) \quad U_k(\eta) = \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(V_k + (e^{-\alpha} - 1)\Gamma_k)}{\rho(V_k + (e^{-\beta} - 1)\Gamma_k)} \inf_{-2\eta < \alpha, \beta < 2\eta} \frac{\rho(V_{k+1} + (e^\alpha - 1)T_k^\dagger \Gamma_k^{-1} T_k)}{\rho(V_{k+1} + (e^\beta - 1)T_k^\dagger \Gamma_k^{-1} T_k)}.$$

By fluctuation regularity of \mathbb{P}_W , we have $U_k(\eta) \geq \delta^2 I[A_k]$ where $I[A_k]$ is the indicator function of the event:

$$(4.23) \quad A_k = \left\{ V_k, V_{k+1} \in \Omega_W, \quad \|\Gamma_k\| \leq \frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta}, \quad \text{and} \quad \|T_k^\dagger \Gamma_k^{-1} T_k\| \leq \frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta} \right\},$$

with $\delta, \epsilon > 0$, $\zeta \geq 0$ and Ω_W as in assumption (3). In turn, since $\Gamma_k = V_k + \lambda + T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}$, we see that

$$(4.24) \quad A_k \supset \{V_{k+1} \in \Omega_W\} \cap \{V_k \in \Omega_W\} \cap \{\|T_{k-1}\|, \|T_k\| \leq \tau W^b\} \\ \cap \left\{ \|\Gamma_{k-1}^{-1}\| \leq \frac{1}{\tau^2} \left(\frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta} - LW^a - |\lambda| \right) W^{-2b} \right\} \\ \cap \left\{ \|\Gamma_k^{-1}\| \leq \frac{1}{\tau^2} \frac{\epsilon}{e^{2\eta} - 1} W^{-2b-\zeta} \right\},$$

with $L, a \geq 0$ as in assumption 2 and $\tau, b \geq 0$ as in assumption 3. This allows us to estimate the probability of A_k from below by successively integrating over V_{k+1} , V_k , V_{k-1} , T_k , and T_{k-1} in that order. To begin, by assumption 2,

$$(4.25) \quad \text{Prob}(V_{k+1} \in \Omega_W | (V_l)_{l \neq k}, (T_l)) = \mathbb{P}_W(\Omega_W) \geq p_0 > 0.$$

Since $\Gamma_k = V_k + \lambda I + T_{k-1}^\dagger \Gamma_{k-1}^{-1} T_{k-1}$, we see from the Wegner estimate (3.8) that

$$(4.26) \quad \text{Prob} \left(V_k \in \Omega_W, \|\Gamma_k^{-1}\| \leq \frac{1}{\tau^2} \frac{\epsilon}{e^{2\eta} - 1} W^{-2b-\zeta} \middle| (V_l)_{l \neq k, k+1}, (T_l) \right) \\ \geq p_0 - \kappa \tau^2 \frac{e^{2\eta} - 1}{\epsilon} W^{1+\sigma+\zeta+2b}.$$

Similarly

$$(4.27) \quad \text{Prob} \left(\|\Gamma_{k-1}^{-1}\| \leq \frac{1}{\tau^2} \left(\frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta} - LW^a - |\lambda| \right) W^{-2b} \middle| (V_l)_{l \neq k-1, k, k+1}, (T_l) \right) \\ \geq 1 - \kappa \tau^2 \frac{1}{\frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta} - LW^a - |\lambda|} W^{1+\sigma+2b}.$$

Combining these estimates and using assumption 3 to integrate over T_k and T_{k-1} , we find

$$(4.28) \quad \text{Prob}(A_k | (V_l)_{l \neq k-1, k, k+1}, (T_l)_{l \neq k, k-1}) \\ \geq q_0^2 p_0 \left(p_0 - \kappa \tau^2 \frac{e^{2\eta} - 1}{\epsilon} W^{1+\sigma+2b+\zeta} \right) \left(1 - \kappa \tau^2 \frac{1}{\frac{\epsilon}{e^{2\eta} - 1} W^{-\zeta} - LW^a - |\lambda|} W^{1+\sigma+2b} \right).$$

Taking $\eta = cW^{-\nu}$ with $\nu \geq \max(a, 2b + \sigma + 1) + \zeta$, we may choose c sufficiently small to make the r.h.s. larger than $\frac{1}{2} q_0^2 p_0^2$, say. Since $U_k(\eta) \geq \delta^2 I[A_k]$ we find, integrating successively over V_k, T_k from $k = i, \dots, j-1$ (see Lemma A.1), that

$$(4.29) \quad \mathbb{E} \left(e^{-\frac{rs}{6} \eta^2 e^{-s\eta} e^{-DW\eta} \sum_{k=i}^{j-1} U_k(\eta)} \right)^{\frac{s-r}{s}} \\ \leq \exp \left(-\frac{s-r}{2s} q_0^2 p_0^2 \left(1 - e^{-c^2 \delta^2 \frac{rs}{6} W^{-2\nu} e^{-csW^{-\nu}} e^{-cDWW^{-\nu}}} \right) \left\lfloor \frac{|i-j|}{3} \right\rfloor \right).$$

Increasing ν , if necessary, so that $\sup_W DWW^{-\nu} < \infty$ completes the proof. \square

5. ENSEMBLES

In this section, we consider several examples of band matrix ensembles satisfying assumptions 1, 2, and 3 of section 3. Assumption 1 is simply the choice of an algebra \mathcal{A}_W to support the distribution of the diagonal blocks, and the corresponding set \mathcal{T}_W for the off-diagonal blocks. In this regard, we will consider two cases:

(\mathbb{R}) $\mathcal{A}_W = W \times W$ matrices with real entries,
or

(\mathbb{C}) $\mathcal{A}_W = W \times W$ matrices with complex entries.

In each case the dimension of the algebra D_W is comparable to W^2 and $\mathcal{T}_W = \mathcal{A}_W$.

5.1. Wigner-matrix blocks and the Wegner estimate. We shall suppose that the diagonal blocks V_j of $X_{W;N}$ are *Wigner matrices*:

Assumption 4. The distribution of the diagonal blocks, $d\mathbb{P}_W(V)$, written in terms of the matrix elements

$$(5.1) \quad V = \frac{1}{\sqrt{W}} \begin{pmatrix} d_1 & a_{1,2} & \cdots & \cdots & a_{1,W} \\ a_{1,2}^* & d_2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & a_{W-1,W} \\ a_{1,W}^* & \cdots & \cdots & a_{W-1,W}^* & d_W \end{pmatrix},$$

has the form

$$(5.2) \quad d\mathbb{P}_W(V) = \prod_{j=1}^W h(d_j) dd_j \prod_{1 \leq i < j \leq W} g(a_{i,j}) da_{i,j},$$

where dd_j is Lebesgue measure on the real line, $h \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$ is non-negative with $\int h = 1$, and either

(\mathbb{R}) $da_{i,j}$ is Lebesgue measure on \mathbb{R} and $g \in L^\infty(\mathbb{R}) \cap L^1(\mathbb{R})$ is non-negative with $\int_{\mathbb{R}} g = 1$,
or

(\mathbb{C}) $da_{i,j}$ is Lebesgue measure on \mathbb{C} and $g \in L^\infty(\mathbb{C}) \cap L^1(\mathbb{C})$ is non-negative with $\int_{\mathbb{C}} g = 1$.

Furthermore, we require

$$(5.3) \quad \int_{\mathbb{R}} \lambda^2 h(\lambda) d\lambda < \infty,$$

$$(5.4) \quad \int |a|^4 g(a) da < \infty, \quad \text{and} \quad \int ag(a) da = 0.$$

Clearly the measure $d\mathbb{P}_W$ is absolutely continuous with respect to Lebesgue measure on \mathcal{A}_W^H — this is part 1 of assumption 2. Regarding the Wegner estimates — part 2 of assumption 2 — we then have the following

Theorem 7 (Wegner estimate). *Under assumption 4, the Wegner estimates (3.8) and (3.9) hold with $\sigma = \frac{1}{2}$ and $\kappa = 2\pi \text{ess-sup}_\lambda h(\lambda)$.*

Proof. This result, which is obtained by averaging over the diagonal variables $\{d_j\}$ only, is a standard estimate from the theory of random Schrödinger operators, first obtained by Wegner [20]. For completeness, we sketch the proof.

Note that $\|(V - A)^{-1}\| > t$ if and only if $V - A$ has an eigenvalue in the interval $(-\frac{1}{t}, \frac{1}{t})$. It follows that

$$(5.5) \quad \begin{aligned} \text{Prob} \left\{ \|(V - A)^{-1}\| > t \right\} &\leq \frac{2}{t^2} \mathbb{E} \left(\text{tr} \left[(V - A)^2 + \frac{1}{t^2} \right]^{-1} \right) \\ &= \frac{2}{t} \mathbb{E} \left(\text{Im tr} \left[V - A - i\frac{1}{t}I \right]^{-1} \right) = \frac{2}{t} \sum_{i=1}^W \mathbb{E} \left(\text{Im} \left\langle \mathbf{e}_i, \left[V - A - i\frac{1}{t}I \right]^{-1} \mathbf{e}_i \right\rangle \right). \end{aligned}$$

By the Schur complement formula,

$$(5.6) \quad \left\langle \mathbf{e}_i, \left[V - A - i\frac{1}{t}I \right]^{-1} \mathbf{e}_i \right\rangle = \frac{1}{\frac{1}{\sqrt{W}}d_i - \frac{1}{t}i - \gamma},$$

where γ is a function of all matrix elements of V *except* d_i . Thus γ is a random variable independent of d_i , so

$$(5.7) \quad \mathbb{E} \left(\text{Im} \left\langle \mathbf{e}_i, \left[V - A - i\frac{1}{t}I \right]^{-1} \mathbf{e}_i \right\rangle \right) = \mathbb{E} \left(\frac{\frac{1}{t} + \text{Im} \gamma}{\left(\frac{1}{\sqrt{W}}d_i - \text{Re} \gamma\right)^2 + \left(\frac{1}{t} + \text{Im} \gamma\right)^2} \right) \leq \|h\|_\infty \pi \sqrt{W},$$

where the inequality follows from replacing the average $\int \bullet h(d_i) dd_i$ by the upper bound $\|h\|_\infty \int_{\mathbb{R}} \bullet dd_i$. Summing over i gives the result.

The proof of (3.9) is analogous. However in that case the trace is over a $2W$ dimensional space, resulting in the additional factor of 2 on the r.h.s. of that equation. \square

The scaling factor \sqrt{W} that appears in (5.11) is natural, as with this scaling the matrix V has a finite density of states in the large W limit [19]:

$$(5.8) \quad \lim_{W \rightarrow \infty} \frac{1}{W} \int_{\mathcal{A}_W^H} \text{tr} f(V) d\mathbb{P}_W(V) = \frac{1}{2\sigma^2\pi} \int_{-2\sigma}^{2\sigma} f(\lambda) \sqrt{4\sigma^2 - \lambda^2} d\lambda,$$

with $\sigma^2 = \int |a|^2 g(a) da$. A key fact below is the following related result

Theorem 8 (Bai and Yin [4]). *Let V be a $W \times W$ random matrix of the form (5.11), with $\{d_i\}$ and $\{a_{i,j}\}$ mutually independent sets of independent random variables. If*

$$\mathbb{E}(|d_i|) < \infty, \quad \mathbb{E}(|a_{i,j}|^4) < \infty, \quad \mathbb{E}(a_{i,j}) = 0,$$

and $\sigma^2 = \mathbb{E}(|a_{i,j}|^2)$ then

$$(5.9) \quad \lim_{W \rightarrow \infty} \text{Prob} [\|V\| > 2\sigma + \eta] = 0.$$

Remark. This follows from Theorem A of ref. [4], which gives the convergence of λ_1 , the largest eigenvalue of V , to 2σ with probability one. Symmetrizing the assumptions of Theorem A and applying the result also to show that λ_W , the smallest eigenvalue of V , converges to -2σ , this result follows. (The proof in [4] is written out in the real symmetric case, but carries over to the complex hermitian case with only very minor modifications.)

Corollary 9. *Under assumption 4, we may find $p_0, L > 0$ such that*

$$(5.10) \quad \text{Prob} [\|V\| \leq L] \leq p_0.$$

We require very little of the off diagonal blocks T_j . They need only satisfy the estimate (3.13) analogous to (3.10) and (5.10). In particular, they could be deterministic, say $T_j = I$ for all j or T_j given by a Toeplitz matrix. In this section we consider a few examples of random off-diagonal blocks modeled on the blocks for the Gaussian band ensemble (1.1). In that case, the off-diagonal blocks T_j are lower triangular matrices with Gaussian entries. More generally we may suppose

Assumption 5. The distribution of the off-diagonal blocks, $d\mathbb{Q}_W(T)$, written in terms of the matrix elements

$$(5.11) \quad T = \frac{1}{\sqrt{W}} \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ t_{2,1} & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ t_{W,1} & \cdots & \cdots & t_{W,W-1} & 0 \end{pmatrix},$$

has the form

$$(5.12) \quad d\mathbb{Q}_W(T) = \prod_{1 \leq j < i \leq W} d\mu(t_{i,j}),$$

where either

(\mathbb{R}) $\mu(t_{i,j})$ is a probability measure on \mathbb{R}
or

(\mathbb{C}) $\mu(t_{i,j})$ is a probability measure on \mathbb{C} ,

and

$$(5.13) \quad \int |t|^4 d\mu(t) < \infty, \quad \text{and} \quad \int t d\mu(t) = 0.$$

Theorem 10. *Under assumption 5, we may find $q_0, \tau > 0$ such that assumption 3 holds with $b = 0$, i.e.,*

$$(5.14) \quad \text{Prob} [\|T\| \leq \tau] \leq q_0.$$

Proof. It follows from [4, Theorem A] that, with $\sigma^2 = \int |t|^2 d\mu(t)$,

$$(5.15) \quad \lim_{W \rightarrow 0} \text{Prob} [\|T + T^\dagger\| > \sigma + \eta] = 0, \quad \lim_{W \rightarrow 0} \text{Prob} [\|i(T - T^\dagger)\| > \sigma + \eta] = 0,$$

for any $\eta > 0$. Since

$$(5.16) \quad T = \frac{1}{2}(T + T^\dagger) + \frac{1}{2i}i(T - T^\dagger),$$

it follows that

$$(5.17) \quad \lim_{W \rightarrow 0} \text{Prob} [\|T\| > \sigma + \eta] = 0.$$

Thus (5.14) holds. □

5.2. Fluctuation regularity. A particular example of distributions satisfying assumption 4 are the Gaussian Orthogonal Ensemble (GOE), corresponding to case (\mathbb{R}) , and the Gaussian Unitary Ensemble (GUE), corresponding to case (\mathbb{C}) . In these cases, the measure \mathbb{P} is of the form

$$(5.18) \quad d\mathbb{P}(V) \propto e^{-\beta W \operatorname{tr} V^2} dV, \quad V \in \mathcal{A}_W^H,$$

with $\beta = 1$ (\mathbb{R}) or $\beta = 2$ (\mathbb{C}). That is,

$$(5.19) \quad h(d) = \frac{1}{\sqrt{\pi}} e^{-\beta d^2}, \quad g(a) = \frac{1}{(2\beta\pi)^{\frac{\beta}{2}}} e^{-2\beta|a|^2}.$$

Theorem 11. *If V is a GUE or GOE matrix of size W then assumption 2 of section 3 holds with $\sigma = \frac{1}{2}$, $\zeta = 2$ and $a = 0$.*

Corollary 12. *Assumptions 1, 2, and 3 hold for the Gaussian band ensemble (1.1).*

Proof. We have already derived the Wegner estimates (Thm. 7). It remains only to show the fluctuation regularity. For the Gaussian ensembles, we have

$$(5.20) \quad \frac{\rho(V_1)}{\rho(V_2)} = e^{-\beta W \operatorname{tr}(V_1^2 - V_2^2)} = e^{-\beta W \operatorname{tr}(V_1 - V_2)(V_1 + V_2)} \geq e^{-\beta W^2 \|V_1 - V_2\| \|V_1 + V_2\|}.$$

If $\|V_1 - V\|, \|V_2 - V\| \leq \epsilon W^{-2}$ we have

$$(5.21) \quad \frac{\rho(V_1)}{\rho(V_2)} \geq e^{-2\beta\epsilon(\|V\| + \epsilon W^{-2})}.$$

Letting p_0 and L be as in Cor. 9, we set $\Omega_W := \{\|V\| \leq L\}$. Then $\operatorname{Prob}(\Omega_W) \geq p_0 > 0$ and if $V \in \Omega_W$, we have

$$(5.22) \quad \frac{\rho(V_1)}{\rho(V_2)} \geq e^{-2\epsilon(L + \epsilon)} := \delta,$$

whenever $\|V_1 - V\|, \|V_2 - V\| \leq \epsilon W^{-2}$. □

To obtain fluctuation regularity for general Wigner matrices (3.11) we require additional assumptions on h and g . For instance, we have the following

Theorem 13. *If V satisfies assumption 4 with $\ln h$ and $\ln g$ uniformly Hölder continuous with exponent α , then assumption 2 of section 3 holds with $\sigma = \frac{1}{2}$, $\zeta = \frac{2}{\alpha} + \frac{1}{2}$ and $a = 0$.*

Remark. For example $h(\lambda) = g(\lambda) = c_\alpha e^{-|\lambda|^\alpha}$ with $0 < \alpha \leq 1$ satisfies the hypotheses of the theorem.

Proof. We have

$$(5.23) \quad \frac{\rho(V_1)}{\rho(V_2)} = \exp \left(\sum_i \ln h(d_{i;1}) - \ln h(d_{i;2}) + \sum_{i,j} \ln g(a_{i,j;1}) - \ln g(a_{i,j;2}) \right) \\ \geq \exp \left[-C \left(\sum_i |d_{i;1} - d_{i;2}|^\alpha + \sum_{i,j} |a_{i,j;1} - a_{i,j;2}|^\alpha \right) \right]$$

If $\|V_1 - V\|, \|V_2 - V\| \leq \epsilon W^{-\zeta}$, then

$$|d_{i;1} - d_{i;2}|, |a_{i,j;1} - a_{i,j;2}| \leq \sqrt{W} \|V_1 - V\| + \sqrt{W} \|V_2 - V\| \leq 2\epsilon W^{\frac{1}{2} - \zeta}.$$

It follows that

$$(5.24) \quad \frac{\rho(V_1)}{\rho(V_2)} \geq \exp[-C\epsilon W^2 W^{\frac{\alpha}{2}-\zeta\alpha}] = e^{-C\epsilon} =: \delta.$$

This estimate holds for *every* V , so in particular for all V in $\Omega_W = \{\|V\| \leq L\}$. \square

Theorem 13 cannot apply if h or g has compact support. Nonetheless compactly supported densities can be handled. A general result of this type would somewhat involve to state, so let us simply note that assumption 2 holds if h and g are characteristic functions of open neighborhoods of the origin.

Theorem 14. *Suppose that V satisfies assumption 4 and that*

$$h(d) = \frac{1}{2D} I[|d| < D], \quad g(a) = \frac{1}{c_\beta A^\beta} I[|a| < A],$$

with $c_1 = 2$ and $c_2 = \pi$. Then assumption 2 of section 3 holds with $\sigma = \frac{1}{2}$, $\zeta = \frac{5}{2}$ and $a = 0$.

Proof. Clearly the moment conditions of assumption 4 hold. Thus by Cor. 9 we can find p_0 and L so that (5.10) holds.

Now suppose $\|V_1 - V\|, \|V_2 - V\| \leq \epsilon W^{-\frac{5}{2}}$. Suppose also that the matrix elements of V satisfy $W^{-\frac{1}{2}}|d_i| \leq W^{-\frac{1}{2}}D - \epsilon W^{-\frac{5}{2}}$, $W^{-\frac{1}{2}}|a_{i,j}| \leq W^{-\frac{1}{2}}A - \epsilon W^{-\frac{5}{2}}$ for all i, j . Then

$$(5.25) \quad \frac{\rho(V_1)}{\rho(V_2)} = 1.$$

But

$$(5.26) \quad \begin{aligned} & \text{Prob}(W^{-\frac{1}{2}}|d_i| \leq W^{-\frac{1}{2}}D - \epsilon W^{-\frac{5}{2}}, W^{-\frac{1}{2}}|a_{i,j}| \leq W^{-\frac{1}{2}}A - \epsilon W^{-\frac{5}{2}}) \\ & \geq 1 - \sum_i \text{Prob}(|d_i| > D - \epsilon W^{-2}) - \sum_{i,j} \text{Prob}(|a_{i,j}| > A - \epsilon W^{-2}) \geq 1 - C\epsilon. \end{aligned}$$

Now let

$$(5.27) \quad \Omega_W = \left\{ \|V\| \leq L, W^{-\frac{1}{2}}|d_i| \leq W^{-\frac{1}{2}}D - \epsilon W^{-\frac{5}{2}}, \text{ and } W^{-\frac{1}{2}}|a_{i,j}| \leq W^{-\frac{1}{2}}A - \epsilon W^{-\frac{5}{2}} \right\},$$

with ϵ sufficiently small that

$$(5.28) \quad \text{Prob}(\Omega_W) \geq p_0 - C\epsilon > 0. \quad \square$$

5.3. Summary. Putting the results of this section together with Thm. 6 we have:

Theorem 15. . *Let $\mathcal{A}_W = \mathcal{T}_W =$ set of $W \times W$ matrices with real or complex entries and suppose \mathbb{P} and \mathbb{Q} satisfy assumptions 4 and 5.*

- (1) *If \mathbb{P} is either the Gaussian orthogonal or Gaussian unitary ensemble, then given $r > 0$ and $s \in (0, 1)$ there are $A_s < \infty$ and $\alpha_s > 0$ such that*

$$(5.29) \quad \mathbb{E} \left(\left| \langle \mathbf{e}_i, (X_{W;N} - \lambda)^{-1} \mathbf{e}_j \rangle \right|^s \right) \leq A_s W^{\frac{s}{2}} e^{-\alpha_s \frac{|i-j|}{W^s}}, \quad \lambda \in [-r, r].$$

In particular, (5.29) holds for the Gaussian band ensemble (1.1).

- (2) *If $\ln h$ and $\ln g$ are uniformly Hölder continuous with exponent α , then given $r > 0$ and $s \in (0, 1)$ there are $A_s < \infty$ and $\alpha_s > 0$ such that*

$$(5.30) \quad \mathbb{E} \left(\left| \langle \mathbf{e}_i, (X_{W;N} - \lambda)^{-1} \mathbf{e}_j \rangle \right|^s \right) \leq A_s W^{\frac{s}{2}} e^{-\alpha_s \frac{|i-j|}{W^\mu}}, \quad \lambda \in [-r, r],$$

with $\mu = 5 + \frac{4}{\alpha}$

- (3) If h and g are proportional to characteristic functions of open neighborhoods of the origin, then given $r > 0$ and $s \in (0, 1)$ there are $A_s < \infty$ and $\alpha_s > 0$ such that (5.30) holds with $\mu = 9$.

APPENDIX A. A LEMMA ON CONDITIONAL AVERAGES

In the proofs of the various versions of Lemma F above, a key step was to estimate averages of the form

$$(A.1) \quad \mathbb{E} \left(e^{-\sum_{j=1}^n U_j} \right)$$

in which U_j are non-negative, strictly positive with good probability, but not independent. The following Lemma gives the relevant estimate, which can be seen as a simple version of *stochastic domination*. As the proof shows, under appropriate assumptions, we can estimate (A.1) in terms of the same expression with U_j replaced by i.i.d. non-negative Bernoulli variables taking 0 with probability less than 1.

Lemma A.1. *Let Σ_j be a sequence of σ -algebras of events on a probability space and let U_j be a sequence of non-negative random variables with U_j measurable with respect to Σ_k for $k \neq j$. If for some $\delta > 0$,*

$$\text{Prob}(U_j \geq \delta | \Sigma_j) \geq p_0$$

for each j , then

$$\mathbb{E} \left(e^{-\sum_{j=1}^n U_j} \right) \leq e^{-(1-e^{-\delta})p_0 n}.$$

Proof. This follows by induction, since

$$\mathbb{E} \left(e^{-\sum_{j=1}^n U_j} | \Sigma_n \right) = e^{-\sum_{j=1}^{n-1} U_j} \mathbb{E} \left(e^{-U_n} | \Sigma_n \right) \leq [(1 - p_0) + e^{-\delta} p_0] e^{-\sum_{j=1}^{n-1} U_j}$$

and

$$(1 - p_0) + e^{-\delta} p_0 \leq e^{-(1-e^{-\delta})p_0}. \quad \square$$

REFERENCES

1. M Aizenman and S Molchanov, *Localization at large disorder and at extreme energies: An elementary derivation*, Comm. Math. Phys. **157** (1993), no. 2, 245–278.
2. M Aizenman, JH Schenker, RM Friedrich, and D Hundertmark, *Finite-volume fractional-moment criteria for Anderson localization*, Comm. Math. Phys. **224** (2001), no. 1, 219–253.
3. M Aizenman, *Localization at weak disorder: some elementary bounds*, Rev. Math. Phys. **6** (1994), no. 5A, 1163–1182, Special issue dedicated to Elliott H. Lieb.
4. Z D Bai and Y Q Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*, Ann. Probab. **16** (1988), no. 4, 1729–1741.
5. J V Bellissard, P D Hislop, and G Stolz, *Correlation estimates in the Anderson model*, J. Stat. Phys. **129** (2007), no. 4, 649–662.
6. R Carmona and J Lacroix, *Spectral theory of random Schrödinger operators*, Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1990.
7. G Casati, L Molinari, and F Izrailev, *Scaling properties of band random matrices*, Phys. Rev. Lett. **64** (1990), no. 16, 1851–1854.
8. G Chirikov, B V Guarneri, I Izrailev, and F M Casati, *Band-random-matrix model for quantum localization in conservative systems*, Phys. Rev. E **48** (1993), no. 3, R1613.
9. J-M Combes, F Germinet, and A Klein, *Generalized eigenvalue-counting estimates for the anderson model*, Preprint, 2008.
10. R L Dobrushin and S B Shlosman, *Absence of breakdown of continuous symmetry in two-dimensional models of statistical physics*, Comm. Math. Phys. **42** (1975), no. 1, 31–40.

11. F J Dyson, *Statistical theory of the energy levels of complex systems. I*, J. Math. Phys. **3** (1962), no. 1, 140.
12. ———, *Statistical theory of the energy levels of complex systems. II*, J. Math. Phys. **3** (1962), no. 1, 157.
13. Y V Fyodorov and A D Mirlin, *Scaling properties of localization in random band matrices: A σ -model approach*, Phys. Rev. Lett. **67** (1991), no. 18, 2405–2409.
14. G-M Graf and A Vaghi, *A remark on the estimate of a determinant by Minami*, Lett. Math. Phys. **79** (2007), no. 1, 17–22.
15. A Klein and S Molchanov, *Simplicity of eigenvalues in the Anderson model*, J. Stat. Phys. **122** (2006), no. 1, 95–99.
16. H Kunz and B Souillard, *Sur le spectre des opérateurs aux différences finies aléatoires*, Comm. Math. Phys. **78** (1980), 201–246.
17. N D Mermin and H Wagner, *Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models*, Phys. Rev. Lett. **17** (1966), no. 22, 1133–1136.
18. N Minami, *Local fluctuation of the spectrum of a multidimensional Anderson tight binding model*, Comm. Math. Phys. **177** (1996), no. 3, 709–725.
19. S A Molchanov, L A Pastur, and A M Khorunzhii, *Limiting eigenvalue distribution for band random matrices*, Theor. Math. Phys. **90** (1992), no. 2, 108–118.
20. F Wegner, *Bounds on the density of states in disordered systems*, Zeit. Phys. B **44** (1981), no. 1-2, 9–15.

MICHIGAN STATE UNIVERSITY, EAST LANSING, MICHIGAN 48824

E-mail address: jeffrey@math.msu.edu